Fonction de complexité en facteurs et un théorème de PANSIOT

Auteur : Cisternino, Célia
Promoteur(s) : Rigo, Michel
Faculté : Faculté des Sciences
Diplôme : Master en sciences mathématiques, à finalité spécialisée en informatique
Année académique : 2017-2018
URI/URL : http://hdl.handle.net/2268.2/4934

Université de Liège
Faculté des Sciences
Département de Mathématique
Année académique 2017–2018

Fonction de complexité en facteurs et un théorème de Pansiot

Mémoire de fin d'études présenté en vue de l'obtention d'un Master en Sciences Mathématiques, à finalité.

Mémoire réalisé par :
Célia Cisternino
ccisternino@student.uliege.be
2e Master en Sciences Mathématiques à finalité informatique

Promoteur :
Michel Rigo
m.rigo@uliege.be

Remerciements

J'aimerais adresser mes remerciements à toutes les personnes qui, de près ou de loin, ont contribué à l'élaboration de ce mémoire de fin d'études.

Tout d'abord, je tiens à remercier mon promoteur, Michel Rigo, pour sa disponibilité, ses conseils avisés et sa générosité à partager ses connaissances et sa vision des mathématiques.

J'exprime également ma gratitude auprès de mes lecteurs, membres et président du jury, Émilie Charlier, Julien Leroy et Pierre Mathonet.

Je remercie chaleureusement mon "parrain", Laurent Loosveldt, pour avoir rempli son rôle jusqu'à aujourd'hui et plus particulièrement pour la relecture et la correction de ce mémoire.

Ensuite, j'aimerais exprimer mon amitié sincère auprès de Marie Lejeune avec qui j'ai eu le plaisir de partager cinq années d'études en binôme.

Finalement, je remercie toutes les personnes qui ont contribué à ma vie universitaire. Plus particulièrement, je pense à mes parents et à ma soeur qui m'ont soutenue et encouragée tout au long de ces années.

"Plusieurs se plaisent à penser qu'ils peuvent se débrouiller seuls,
mais le sage sait que rien ne vaut les encouragements
d'une personne qui nous aime."

Table des matières

Introduction

La combinatoire des mots est une branche des mathématiques discrètes, datant du début du XX$^{\text{e}}$ siècle, qui étudie les suites finies ou infinies écrites sur un alphabet fini. Ces suites sont en général appelées *mots* (finis ou infinis). Un *facteur* d'un mot infini u est un mot fini w apparaissant dans u, c'est-à-dire qu'il existe un mot fini v éventuellement vide et un mot infini x tels que $u = vwx$. L'ensemble des mots finis muni de l'application naturelle de concaténation est une monoïde libre. Ainsi, les mots peuvent être considérés comme des objets algébriques combinatoires discrets dans une structure non commutative. C'est pourquoi, bien que les objets soient intrinsèquement simples, de nombreuses questions non triviales se posent. Les mots sont, par exemple, les objets centraux de la théorie des automates et, par conséquent, de beaucoup de modèles informatiques et algorithmiques. Les propriétés étudiées sur les mots sont, entre autres, la présence de régularités ou de certains motifs tels que des palindromes ou encore, des puissances, c'est-à-dire la répétition d'un même facteur. Historiquement, la combinatoire des mots a commencé au début du siècle passé lorsque A. Thue a prouvé l'existence d'un mot ternaire infini sans carré. L'étude de telles propriétés est un sujet vaste ayant une multitude d'applications.

Le *langage* d'un mot u est l'ensemble des facteurs apparaissant dans u. Ainsi, la *fonction de complexité en facteurs* p_u d'un mot infini u est définie sur l'ensemble des naturels et à n associe le nombre de facteurs de longueur n appartenant au langage de u. Formellement, si l'on note $\mathcal{A}$ l'alphabet considéré et $L(u)$ le langage du mot u, on a

$$p_u : \mathbb{N} \to \mathbb{N},\ n \mapsto \#(L(u) \cap \mathcal{A}^n).$$

Intuitivement, si le mot infini considéré est "simple", on s'attend à ce que le langage de ses facteurs le soit également, pour un sens à donner précisément. En 1938, Gustav A. Hedlund et Marston Morse ont confirmé et formalisé le sens de cette intuition en prouvant [11] que la fonction de complexité est bornée par une constante si et seulement si le mot u est ultimement périodique. La fonction de complexité en facteurs, associée à un mot infini, a des applications, par exemple, en théorie des nombres.

Le but de ce mémoire est d'étudier les principales propriétés de cette fonction de complexité en facteurs et, plus particulièrement, un théorème de Pansiot [12]. Ce résultat classique est important en combinatoire des mots. Il décrit entièrement les types possibles du comportement asymptotique de la fonction de complexité en facteurs des mots purement morphiques. Ce travail s'inspire essentiellement du chapitre écrit par Cassaigne et Nicolas [6]. Dans ce mémoire, nous introduisons des notions et des outils classiques tels que les facteurs spéciaux et les graphes de Rauzy permettant de mieux appréhender la fonction de complexité d'un mot infini.

Détaillons à présent le contenu de ce travail. Le premier chapitre de ce mémoire introduit des concepts de base de la combinatoire des mots. Nous étudions ensuite les propriétés les plus classiques de la fonction de complexité en facteurs d'un mot infini telles que sa croissance ainsi que ses caractéristiques dans le cas particulier de mots périodiques.

Le deuxième chapitre étudie et démontre le théorème de Morse et Hedlund [11] énoncé précédemment. En d'autres termes, ce théorème montre que les mots périodiques sont les seuls à avoir une fonction de complexité en $\Theta(1)$. Ce théorème est tout d'abord démontré sans outil au préalable. Ensuite, les concepts de facteurs spéciaux et de première et seconde différences s et b sont introduits. Une formule de comptage est obtenue : pour tout $n \in \mathbb{N}$, la première différence $s(n)$ est égale au nombre de facteurs w spéciaux à droite de longueur n, comptés avec une multiplicité égale à $d^+(w) - 1$. Finalement, ce résultat nous permet d'apporter une preuve alternative au théorème de Morse et Hedlund.

Dans le troisième chapitre, les graphes de Rauzy sont introduits en tant qu'outil permettant la visualisation des facteurs spéciaux et de la fonction de complexité d'un mot infini. Nous allons caractériser les graphes de Rauzy des mots *Sturmiens*, qui sont les mots infinis apériodiques ayant une fonction de complexité égale à $n + 1$ pour tout $n \in \mathbb{N}$. Remarquons que, par le théorème de Morse et Hedlund, $n + 1$ est la borne inférieure pour une fonction de complexité en facteurs d'un mot apériodique. Ensuite, pour apporter une seconde illustration à ce chapitre, on définit deux nouveaux concepts. Un *mot de retour* d'un facteur w dans le mot infini u est un facteur $u_i \cdots u_{j-1}$ tel que i et j sont deux occurrences successives de w. Un *mot lacunaire* est un mot binaire $u \in \{0, 1\}^{\mathbb{N}}$ tel que 0^n est un facteur de u pour tout $n \in \mathbb{N}$ et l'ensemble des mots de retour de 0^n dans u est composé de deux facteurs différents. Ainsi, un exemple de calcul de complexité en facteurs est établi par l'étude des mots de retour et des mots lacunaires.

Le chapitre suivant définit les notions, indispensables pour le théorème de Pansiot, de morphisme et de mots purement morphiques. Un mot purement morphique est le point fixe d'un morphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ tel qu'il existe une lettre $a \in \mathcal{A}$ satisfaisant $\sigma(a) = av$ où v est un mot non vide et $\lim_{n\to+\infty} |\sigma^n(a)| = +\infty$. Ces mots, construits récursivement en itérant un morphisme, fournissent un procédé de définition simple d'une large classe de mots, riches en propriétés, permettant de répondre à de nombreux problèmes. Par exemple, le célèbre mot Thue-Morse,

$$t = abbabaabbaababba \cdots ,$$

est un mot purement morphique obtenu en itérant le morphisme $\sigma : a \mapsto ab, b \mapsto ba$, en débutant par la lettre a, et le mot de Fibonacci,

$$f = aababaaabaababaababaabaab \cdots ,$$

est le point fixe du morphisme $\sigma : a \mapsto ab, b \mapsto a$. En second lieu, nous étudions les mots automatiques étant des codages de points fixes de morphismes où les images de chaque lettre ont même longueur. Les mots automatiques forment une sous-classes importantes des mots morphiques. En effet, le fait d'avoir des images de longueur constante rend la description plus simple en recourant aux développements en base entière et aux automates. On établit, en obtenant une borne plus fine, un résultat classique de COBHAM montrant que la fonction de complexité en facteurs de mots automatiques est en $\mathcal{O}(n)$. En particulier, le mot de THUE-MORSE étant automatique et apériodique, sa fonction de complexité est en $\Theta(n)$.

Ensuite, le cinquième chapitre est consacré à l'élaboration du théorème de PANSIOT montrant que la fonction de complexité en facteurs d'un mot purement morphique ne peut avoir que l'un des cinq comportements asymptotiques suivants :

$$\Theta(1), \Theta(n), \Theta(n \log \log n), \Theta(n \log n) \text{ et } \Theta(n^2).$$

De nombreuses classifications intermédiaires sont établies afin d'obtenir une démonstration courte et simple de ce célèbre théorème initialement démontré dans [12], de manière difficile à suivre pour le lecteur non expert.

Finalement, le dernier chapitre fait l'objet de l'étude de la forme exacte de la fonction de complexité du mot de THUE-MORSE et de la démonstration du caractère *Sturmien* du mot de FIBONACCI. Ces illustrations permettent de mettre en œuvre les techniques développées dans les chapitres précédents.

Chapitre 1

Définitions et familiarisation

Dans ce chapitre, nous allons tout d'abord introduire quelques concepts de combinatoire des mots nécessaires à ce mémoire ainsi que leurs notations usuelles. Nous allons ensuite définir la fonction de complexité en facteurs d'un mot infini et étudier quelques-unes de ses propriétés. Cette fonction fera l'objet de l'étude de ce mémoire.

Débutons par quelques définitions générales de combinatoire des mots.

Définitions 1.1. Un *alphabet* est un ensemble fini, généralement noté $\mathcal{A} = \{a_1, a_2, ..., a_d\}$ tel que $\#(\mathcal{A}) = d \in \mathbb{N}_0$. Pour tout $i \in \{1, ..., d\}$, les éléments a_i d'un alphabet $\mathcal{A}$ sont des *lettres*. Un *mot fini* (resp. *infini*) est une suite finie (resp. infinie) de lettres. L'ensemble des mots finis (resp. infinis) de $\mathcal{A}$ est noté $\mathcal{A}^*$ (resp. $\mathcal{A}^{\mathbb{N}}$). Le *mot vide*, noté ε, appartient à $\mathcal{A}^*$, pour tout alphabet $\mathcal{A}$. L'ensemble $\mathcal{A}^+$ désigne $\mathcal{A}^* \setminus \{\varepsilon\}$. La *longueur d'un mot fini* $u \in \mathcal{A}^*$, notée $|u|$, est le nombre de lettres qui le composent. L'ensemble des mots de longueur $n \in \mathbb{N}$ (resp. longueur au plus $n \in \mathbb{N}$) écrits sur l'alphabet $\mathcal{A}$ est noté $\mathcal{A}^n$ (resp. $\mathcal{A}^{\leq n}$).

Soit un mot $u \in \mathcal{A}^*$ (resp. $\mathcal{A}^{\mathbb{N}}$). Un mot $v \in \mathcal{A}^*$ est un *facteur* de u s'il existe $v' \in \mathcal{A}^*$ et $u' \in \mathcal{A}^*$ (resp. $\mathcal{A}^{\mathbb{N}}$) tels que $u = v'vu'$. Dans ce cas, le naturel $|v'|$ est une *position d'occurrence* du facteur v le mot u. Si le facteur v' est le mot vide, on dit que v est un *préfixe* de u. De la même façon, si le mot u est fini, un mot v est un *suffixe* de u si le mot u' est vide. L'*ensemble des préfixes* et l'*ensemble des suffixes* de u sont respectivement notés $\mathrm{Pref}(u)$ et $\mathrm{Suff}(u)$. Le préfixe (resp. suffixe) de longueur $n \in \mathbb{N}$ de u est noté $\mathrm{Pref}_n(u)$ (resp. $\mathrm{Suff}_n(u)$).

Définition 1.2. L'opération de concaténation

$$\cdot : \mathcal{A}^* \times (\mathcal{A}^* \cup \mathcal{A}^{\mathbb{N}}) \to (\mathcal{A}^* \cup \mathcal{A}^{\mathbb{N}}), \quad (u, v) \mapsto u \cdot v$$

est définie comme suit : si $u = u_0 \cdots u_{n-1} \in \mathcal{A}^*$ et $v = \begin{cases} v_0 \cdots v_{m-1} & \in \mathcal{A}^* \\ v_0 v_1 \cdots & \in \mathcal{A}^{\mathbb{N}} \end{cases}$,

$$\text{alors } u \cdot v = \begin{cases} u_0 \cdots u_{n-1} v_0 \cdots v_{m-1} \in \mathcal{A}^* \\ u_0 \cdots u_{n-1} v_0 v_1 \cdots \in \mathcal{A}^{\mathbb{N}} \end{cases}.$$

On note généralement $u \cdot v = uv$.

Remarque 1.3. L'ensemble $\mathcal{A}^*$ muni de l'opération de concaténation est un monoïde libre non commutatif avec pour neutre ε.
En effet, la concaténation restreinte aux mots finis est une loi interne, associative, ayant pour neutre ε car, pour tout mot $u \in \mathcal{A}^*$, on a $u\varepsilon = \varepsilon u = u$. De plus, ce monoïde est libre car une base est donnée par l'alphabet $\mathcal{A}$. Cependant, il n'est pas commutatif puisque, par exemple, sur l'alphabet binaire $\mathcal{A} = \{a, b\}$, on a $a \cdot b \neq b \cdot a$.

Définitions 1.4. Soient deux mots $u, v \in \mathcal{A}^{\mathbb{N}}$. La *distance* entre u et v est définie telle que

$$d : \mathcal{A}^{\mathbb{N}} \times \mathcal{A}^{\mathbb{N}} \to \mathbb{R}_{\geq 0},\ (u, v) \mapsto \begin{cases} 2^{-i(u,v)} & \text{si } u \neq v \\ 0 & \text{sinon} \end{cases},$$

où $i(u, v)$ est la longueur du plus long préfixe commun de u et v.
Soit $(u_n)_{n\in\mathbb{N}}$ une suite de mots infinis écrits sur l'alphabet $\mathcal{A}$. Cette suite converge vers un mot $v \in \mathcal{A}^{\mathbb{N}}$ si $\lim\limits_{n\to\infty} d(u_n, v) = 0$. Le mot v est appelé la limite de la suite $(u_n)_{n\in\mathbb{N}}$.

La notion de distance peut être étendue à l'ensemble des mots finis. En effet, la distance entre deux mots $u, v \in \mathcal{A}^*$ est définie comme la distance entre les mots $ub^\omega, vb^\omega \in (\mathcal{A} \cup \{b\})^{\mathbb{N}}$ où b est une lettre n'appartenant pas à l'alphabet $\mathcal{A}$. De plus, une suite $(u_n)_{n\in\mathbb{N}}$ de $\mathcal{A}^*$ converge vers un mot infini $v \in \mathcal{A}^{\mathbb{N}}$ si la suite de mots infinis $(u_n b^\omega)_{n\in\mathbb{N}}$ converge vers v.

Définissons désormais la fonction de complexité d'un mot infini.

Définitions 1.5. Soient un alphabet $\mathcal{A}$ et un mot infini $u \in \mathcal{A}^{\mathbb{N}}$.

(i) L'ensemble des facteurs de u, couramment appelé le langage de u, est noté $L(u)$.

(ii) Pour tout $n \in \mathbb{N}$, on note $L_n(u)$ l'ensemble des facteurs de longueur n de u. Formellement, on a $L_n(u) = L(u) \cap \mathcal{A}^n$.

(iii) Pour tout $n \in \mathbb{N}$, on note $p_u(n)$ le cardinal de l'ensemble $L_n(u)$. Ainsi, la fonction

$$p_u : \mathbb{N} \to \mathbb{N},\ n \mapsto \#(L_n(u))$$

est appelée la *fonction de complexité (en facteurs) de* u. Lorsqu'aucune confusion n'est possible quant au mot u, on notera p pour parler de la fonction p_u.

Remarques 1.6. Pour tout mot $u \in \mathcal{A}^{\mathbb{N}}$, étant donné que $L_0(u) = \{\varepsilon\}$, on a $p_u(0) = 1$. De plus, si l'on note $\mathrm{alph}(u)$ l'alphabet minimal sur lequel le mot u est écrit, on a $\mathrm{alph}(u) = L_1(u)$ et donc $p_u(1) = \#(\mathrm{alph}(u))$.

Remarque 1.7. Les Définitions 1.5 pourraient être généralisées tout d'abord en définissant la fonction de complexité p_u d'un mot fini $u \in \mathcal{A}^*$. Une autre généralisation consisterait à définir pour un langage $X \subseteq \mathcal{A}^*$ où $X \subseteq \mathcal{A}^{\mathbb{N}}$, l'ensemble de ses facteurs $L(X)$ par

$$L(X) = \bigcup_{u\in X} L(u).$$

Ainsi, la fonction de complexité du langage X associe à tout nombre naturel n le nombre de mots de longueur n dans $L(X)$. Cependant, dans ce mémoire, nous allons nous concentrer uniquement sur l'étude de la fonction de complexité des mots infinis.

Remarque 1.8. Soient deux mots $u \in \mathcal{A}^{\mathbb{N}}$ et $w \in \mathcal{A}^*$. Pour tout entier naturel n, on a

$$p_u(n) \leq p_{wu}(n) \leq p_u(n) + |w|.$$

En effet, en ajoutant un préfixe w, on risque d'ajouter des facteurs n'étant pas encore présents dans u. Plus précisément, à chaque position $i \in \mathbb{N}$ dans w correspond un facteur $w^{(i)}$ de longueur n de wu. On a

$$L_n(wu) = L_n(u) \cup \{w^{(i)} \mid i = 0, \ldots, |w| - 1\}.$$

Ainsi, en fonction de l'appartenance au préalable du mot $w^{(i)}$ à l'ensemble $L(u)$, on ajoute au plus $|w|$ nouveaux facteurs.

1.1 Complexité maximale

Proposition 1.9. *Soient un mot $u \in \mathcal{A}^{\mathbb{N}}$ et un entier naturel $n \in \mathbb{N}$. On a $p_u(n) \leq d^n$, où $d = \#(\mathcal{A})$.*

Démonstration. Étant donné que $d = \#(\mathcal{A})$, il y a d^n mots dans $\mathcal{A}^n$. En effet, si l'on fixe la longueur d'un mot écrit sur l'alphabet $\mathcal{A}$ comme étant égale à n, alors pour chaque position dans ce mot, on a d lettres possibles. Dès lors, on a $p_u(n) \leq d^n$ car, au plus, tous les facteurs possibles de longueur n écrits sur l'alphabet $\mathcal{A}$ apparaissent dans le mot u.

$\square$

L'exemple suivant illustre la possibilité d'atteindre cette borne.

Exemple 1.10. *Le mot de* CHAMPERNOWNE
Le monoïde libre $\mathcal{A}^*$ est un ensemble dénombrable. En effet, cela peut être démontré en fournissant une énumération de l'ensemble $\mathcal{A}^*$. Pour ce faire, définissons un ordre radiciel $\leq_r$ sur $\mathcal{A}^*$;
soient $p, q \in \mathcal{A}^*, p \leq_r q$ si et seulement si l'une des deux affirmations suivantes est vérifiée :
- $|p| < |q|$,
- $|p| = |q|$ et $p \leq_l q$ où $\leq_l$ est l'ordre lexicographique [1].

Ainsi, considérons $\{w_0, w_1, w_2, \ldots\}$ une énumération de $\mathcal{A}^*$. En concaténant successivement les mots de $\mathcal{A}^*$ dans cet ordre, on obtient le mot

$$u = w_0 w_1 w_2 \cdots \in \mathcal{A}^{\mathbb{N}}.$$

1. L'ordre lexicographique sur $\mathcal{A}^*$ noté $\leq_l$ est défini comme suit. On a $u \leq_l v$ si et seulement si l'une des deux affirmations suivantes est vérifiée :
— u est un préfixe de v,
— $u = wu'$, $v = wv'$ avec $w \in \mathcal{A}^*$, $u' \neq \varepsilon$ et $v' \neq \varepsilon$ et la première lettre de u' précède celle de v' en suivant l'ordre (total) de $\mathcal{A}$

Sa complexité est donnée par $p(n) = d^n$ étant donné que, par construction, tous les mots de $\mathcal{A}^*$ ont une occurrence dans u. Ce mot est appelé *le mot de* CHAMPERNOWNE *sur* $\mathcal{A}$. Par exemple, le mot de CHAMPERNOWNE binaire est le mot u suivant :

$$\begin{aligned}
&\mathcal{A} = \{a, b\},\\
&\mathcal{A}^* = \{\varepsilon, a, b, aa, ab, ba, bb, aaa, aab, \ldots\},\\
&u = abaaabbabbaaaaababaabbbaababbbabbbaaaaaaabaaba \cdots .
\end{aligned}$$

1.2 Monotonie et sous-additivité

Dans la section précédente, nous avons établi une borne que nous pouvons qualifier de "grossière" pour notre fonction de complexité. Cependant, cela n'implique pas que, pour toute fonction f définie sur les naturels et vérifiant

$$1 \leq f(n) \leq (\#(\mathcal{A}))^n,$$

il existe un mot $u \in \mathcal{A}^{\mathbb{N}}$ tel que f est la fonction de complexité de u. En réalité, la caractérisation des fonctions qui sont effectivement des fonctions de complexité de mots est un problème ouvert. Toutefois, quelques conditions nécessaires peuvent être énoncées. Notons $u \in \mathcal{A}^{\mathbb{N}}$ le mot générique pour lequel la fonction de complexité $p = p_u$ sera étudiée dans cette section.

Proposition 1.11. *La fonction de complexité d'un mot est croissante : pour tout* $n \in \mathbb{N}$, $p(n) \leq p(n+1)$.

Démonstration. Pour tout facteur $w \in L_n(u)$, définissons $i(w)$ comme étant un naturel pour lequel w apparaît dans le mot u à la position $i(w)$. Étant donné que le mot u est infini, le facteur w à une continuation à droite dans u. Dès lors, notons $e(w)$ le mot de longueur $n+1$ commençant à la position $i(w)$. Par définition, tout facteur w de longueur n est un préfixe d'un facteur $e(w)$ de longueur $n+1$. Cette construction définit une fonction $e : L_n(u) \to L_{n+1}(u)$.

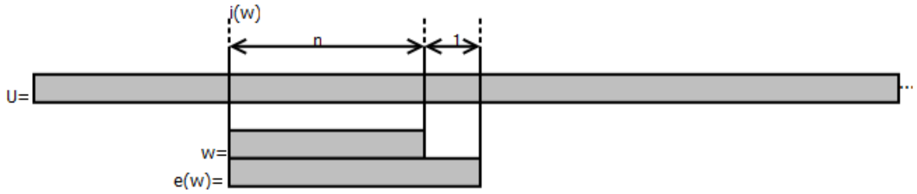


FIGURE 1.1 – Fonction $e : L_n(u) \to L_{n+1}(u)$.

Vu que deux mots différents de longueur n ne peuvent pas être préfixes d'un même mot de taille $n+1$, la fonction e est injective. Dès lors, on obtient

$$\#(L_n(u)) \leq \#(L_{n+1}(u)).$$

Par définition, c'est équivalent à dire que $p_u(n) \leq p_u(n+1)$.

□

En itérant la Proposition 1.11, on obtient le corollaire suivant :

Corollaire 1.12. *Si m, n sont deux nombres naturels tels que $m \leq n$, on a $p(m) \leq p(n)$.*

De plus, on remarque que si w est un facteur de longueur n de u, alors il y a au plus $\#(\mathcal{A})$ facteurs de longueur $n+1$ dans u ayant w comme préfixe. Ceux-ci correspondent aux mots wa avec $a \in \mathcal{A}$. Dès lors, pour tout entier positif n,

$$p(n+1) \leq \#(\mathcal{A})\ p(n).$$

De manière plus générale, on a :

Proposition 1.13. *Soit p la fonction de complexité du mot infini $u \in \mathcal{A}^{\mathbb{N}}$ considéré. Pour tous entiers positifs m et n, on a $p(m+n) \leq p(m)p(n)$. En d'autres termes, la fonction à valeurs réelles $\log(p)$ est sous-additive : pour tous m, n on a*

$$\log(p(m+n)) \leq \log(p(m)) + \log(p(n)).$$

Démonstration. Pour tout facteur $w \in L_{m+n}(u)$, notons w_1 (resp. w_2) son préfixe (resp. suffixe) de longueur m (resp. n). En particulier, les mots w_1 et w_2 sont des facteurs de u et donc $w_1 \in L_m(u)$ et $w_2 \in L_n(u)$. Cette procédure définit une fonction

$$f : L_{m+n}(u) \rightarrow L_m(u) \times L_n(u)$$

telle que $f(w) = (w_1, w_2)$. Par construction, la fonction f est injective car deux mots différents de longueur $m+n$ ne peuvent pas avoir à la fois le même préfixe de longueur m et le même suffixe de longueur n. Dès lors, on obtient

$$\#(L_{m+n}(u)) \leq \#(L_m(u) \times L_n(u)) = \#(L_m(u))\#(L_n(u)).$$

Ainsi, on en tire que $p_u(m+n) \leq p_u(m)p_u(n)$. La suite de la preuve découle des propriétés classiques du logarithme.

□

Pour énoncer un corollaire à cette Proposition 1.13, rappelons la notation de Landau communément appelée "*Grand O*".

Définition 1.14. Soient f et g deux fonctions numériques. On a

$$f(n) = \mathcal{O}(g(n))$$

s'il existe $k > 0$, $n_0 \geq 1$ tels que pour tout entier $n \geq n_0$, $|f(n)| \leq k|g(n)|$. C'est-à-dire f est bornée asymptotiquement par g.

En utilisant cette notation asymptotique, on peut exprimer le corollaire suivant :

Corollaire 1.15. *Si le langage $L(u)$ du mot u n'est pas égal à l'ensemble $\mathcal{A}^*$ tout entier, alors il existe un nombre réel α tel que $1 \leq \alpha < \#(\mathcal{A})$ et $p(n) = \mathcal{O}(\alpha^n)$.*

Démonstration. Notons $d = \#(\mathcal{A})$. Par hypothèse, il existe $m \in \mathbb{N}_0$ tel que $L_m(u) \neq \mathcal{A}^m$. Ainsi, on en tire que $p(m) = \#(L_m(u)) < \#(\mathcal{A}^m) = d^m$. Posons $\alpha = p(m)^{\frac{1}{m}}$, on a $1 \leq \alpha < d$. Soit $n \in \mathbb{N}$ et $q = \lceil \frac{n}{m} \rceil$, on a $n \leq mq$ et $q < \frac{n}{m} + 1 \Leftrightarrow mq < n + m$. Dès lors, on a $n \leq mq < m + n$. En itérant la Proposition 1.13, on obtient

$$p(mq) = p(m + m(q-1)) \leq p(m)p(m(q-1)) \leq ... \leq p(m)^q.$$

De plus, par le Corollaire 1.12, l'inégalité $p(n) \leq p(mq)$ est vérifiée. Ainsi, on en tire que

$$p(n) \leq p(mq) \leq p(m)^q = \alpha^{mq} \leq \alpha^{m+n} = \alpha^m \alpha^n = p(m)\alpha^n.$$

Par conséquent, étant donné que $p(m)$ est une constante, la fonction de complexité $p(n)$ est en $\mathcal{O}(\alpha^n)$.

□

Ce résultat peut être traduit comme suit : s'il manque des facteurs dans un mot, alors sa fonction de complexité est bornée asymptotiquement par une exponentielle en base strictement inférieure au cardinal de l'alphabet.

Définition 1.16. Considérons deux fonctions numériques f et g. La fonction f est dite *asymptotiquement équivalente* à la fonction g, ce que l'on note $f \sim g$, si pour tout $\epsilon > 0$, il existe un entier $n_0 \geq 1$, tel que pour tout $n \geq n_0$, $|f(n) - g(n)| < \epsilon|g(n)|$. De façon équivalente,

$$\lim_{n \to \infty} \frac{f(n)}{g(n)} = 1.$$

Une conséquence du corollaire précédent est que, par exemple, une fonction asymptotiquement équivalente à $\frac{2^n}{n}$ ne peut pas être la fonction de complexité d'un mot binaire infini.
En effet, considérons un alphabet $\mathcal{A}$ tel que $\#(\mathcal{A}) = 2$ et un mot $u \in \mathcal{A}^{\mathbb{N}}$. Si $L(u) = \mathcal{A}^*$, pour tout $n \in \mathbb{N}$, on a $p(n) = 2^n$. Par conséquent, la fonction p n'est pas asymptotiquement équivalente à la fonction $\frac{2^n}{n}$. Il nous reste à considérer le cas où $L(u) \neq \mathcal{A}^*$. Par le corollaire précédent, il existe un nombre réel α tel que $1 \leq \alpha < 2$ et $p(n) = \mathcal{O}(\alpha^n)$. Dans ce cas, la fonction p ne peut pas non plus être asymptotiquement équivalente à la fonction

$\frac{2^n}{n}$. En effet, dans le cas contraire, nous aurions une fonction de l'ordre de $\frac{2^n}{n}$ en $\mathcal{O}(\alpha^n)$. C'est-à-dire qu'il existe $k > 0$ tel que

$$\frac{2^n}{n} \leq k.\alpha^n$$
$$\Leftrightarrow \left(\frac{2}{\alpha}\right)^n \leq k.n \quad .$$

Cependant, le réel $\frac{2}{\alpha}$ est strictement supérieur à 1. On en tire que cette inégalité n'est pas valide lorsque n tend vers l'infini.

1.3 Étude des mots périodiques

Dans cette section, nous allons étudier les propriétés de la fonction de complexité des mots périodiques et ultimement périodiques.

Définition 1.17. Soit $z \in \mathcal{A}^+$, on dit que z est un *mot primitif* s'il n'est pas une puissance d'un mot plus court.

Définition 1.18. Le mot $u \in \mathcal{A}^{\mathbb{N}}$ est *infini périodique* s'il existe $z \in \mathcal{A}^+$ tel que $u = z^\omega = zzz\cdots$. L'entier $|z|$ est *une période* de u. Le plus petit mot z satisfaisant à la propriété est appelé le *cycle de période* de u et définit *la période* $|z|$ de u.

Remarque 1.19. Si le mot $u \in \mathcal{A}^{\mathbb{N}}$ est infini périodique de cycle de période $z \in \mathcal{A}$, c'est-à-dire tel que $|z| = 1$, alors le mot u est dit *constant*.

Remarque 1.20. Si u est un mot infini de périodes p et q, alors il est périodique de période $\mathrm{pgcd}(p, q)$.
En effet, si $p \geq q$, par le théorème de Bezout, il existe $m, n \geq 0$ tels que $mp - nq = \mathrm{pgcd}(p, q)$. Ainsi, on a

$$u_i = u_{i+mp} = u_{i+mp-nq} = u_{i+\mathrm{pgcd}(p,q)}.$$

Définitions 1.21. Soit $z \in \mathcal{A}^+$ et $t \in \mathcal{A}^+$, le mot $u = tz^\omega = tzzz\cdots$ est un *mot infini ultimement périodique*. Le mot t et l'entier naturel $|t|$ sont respectivement une *partie transitoire* et une *prépériode* de u.

Remarque 1.22. Soit $u = tz^\omega$ un mot ultimement périodique. On peut toujours écrire $u = t'z'^\omega$ de sorte que z' soit primitif et t' soit vide ou se termine par une lettre différente de la dernière lettre de z'.
Ainsi, dans la définition précédente, si l'on considère $z, t \in \mathcal{A}^+$ tels que $|z|$ est minimal et que $|t|$ est minimal pour cette longueur de z, ce critère sera rempli. Ce mot t est *la partie transitoire* de u et définit *la prépériode* $|t|$ de u.

Proposition 1.23. *Soit $u = tz^\omega$ un mot ultimement périodique. Pour tout $n \in \mathbb{N}$, on a $p_u(n) \leq |tz|$. De plus, si z est primitif et soit t est vide, soit t se termine par une lettre différente de la dernière lettre de z, alors on a $p_u(n) = |tz|$ pour tout $n \geq |tz|$.*

Démonstration. Considérons une position $i \geq |tz|$ et un entier $n \in \mathbb{N}$. Par périodicité, le facteur de longueur n apparaissant à la position i dans u a également une occurrence à la position $i - |z|$. Ainsi, en itérant le processus, tous les facteurs de longueur n ont une occurrence à une position inférieure ou égale à $|tz|$. Dès lors, vu qu'il y a au plus $|tz|$ positions ayant cette condition, il y a au plus $|tz|$ facteurs différents de longueur n dans u. On en tire que, pour tout entier positif n, $p_u(n) \leq |tz|$.
Supposons maintenant que z et t soient respectivement primitif et minimal et que $n \geq |tz|$. Procédons par l'absurde. Si $p(n) < |tz|$, alors il existe deux positions $0 \leq i < j < |tz|$ telles que le même facteur $w = w_0 \cdots w_{n-1}$ de longueur n apparaît aux positions i et j.
On a $|w| = n \geq |tz| \geq |z|$ et le suffixe de longueur $|z|$ de w a des occurrences dans u aux positions $i + n - |z|$ et $j + n - |z|$. Or, on a

$$i + n - |z| \geq i + |tz| - |z| = i + |t| \geq |t|$$

et de la même façon $j + n - |z| \geq |t|$. Le mot z étant primitif, cela est possible uniquement si $(j - i)$ est un multiple de $|z|$. En effet, dans le cas contraire, par la Remarque 1.20, nous aurions une période égale à $\mathrm{pgcd}(j - i, |z|) < |z|$ pour le mot w^ω obtenu tel que $z^\omega = u_{|t|} \cdots u_{i+n+|z|-1} w^\omega$. Et dans le mot u on aurait choisi la partie transitoire $t' = tu_{|t|} \cdots u_{i+n+|z|-1}$ correspondante. Ceci n'est pas possible par minimalité de $|z|$.
Ainsi, il existe $m \in \mathbb{N}_0$, tel que $j - i = m|z|$, c'est-à-dire

$$i \leq j - |z|.$$

Par conséquent, étant donné que $j < |tz|$, on a $j - |z| < |tz| - |z|$, c'est-à-dire $j - |z| < |t|$. On obtient donc

$$i \leq j - |z| < |t|$$

et t ne peut donc pas être le mot vide. Ainsi, on a

$$|t| - i - 1 \geq 0.$$

Par définition du facteur w de u, on en tire $w_{|t|-i-1} = u_{|t|-1}$. Or, la lettre $u_{|t|-1}$ est la dernière t. Cependant, la lettre $w_{|t|-i-1}$ est aussi égale à $u_{|t|-1+m|z|}$ par périodicité. Cependant, la lettre $u_{|t|-1+m|z|}$ est la dernière lettre de z étant donné que $m \geq 1$. Nous avons donc obtenu une contradiction.

□

Vu qu'un mot périodique est un mot ultimement périodique de prépériode vide, on obtient le corollaire suivant.

Corollaire 1.24. *La fonction de complexité en facteurs d'un mot infini périodique $u = z^\omega$ est telle que $p_u(n) = |z|$ pour tout $n \geq |z|$.*

Exemple 1.25. Considérons le mot $u = (aab)^\omega = aabaabaabaab\cdots$. On a

$$L(u) = (aab)^*\{\varepsilon, a, aa\} \cup (aba)^*\{\varepsilon, a, ab\} \cup (baa)^*\{\varepsilon, b, ba\}$$

et donc

- $p(0) = 1$ car $L_0(u) = \{\varepsilon\}$
- $p(1) = 2$ car $L_1(u) = \{a, b\}$
- $p(n) = 3$ pour $n \geq 2$ car $L(u)$ contient exactement un mot de longueur n dans chaque ensemble de l'union le définissant ci-avant. Autrement dit, dans u il y a exactement un facteur de longueur n commençant par aa, un commençant par ab et finalement un par ba.

Chapitre 2

Théorème de Morse et Hedlund

2.1 Énoncé et première démonstration

Nous avons vu dans le chapitre précédent que les mots ultimement périodiques ont une complexité en facteurs bornée. Le théorème suivant de Morse et Hedlund, initialement démontré dans l'article [11], nous affirme que non seulement la réciproque est vraie, mais également que la complexité d'un mot non-ultimement périodique est strictement croissante et donc, en aucun cas bornée.

Notation 2.1. L'ensemble des entiers compris entre a et $b \in \mathbb{R}$ est noté $[\![a, b]\!]$.

Théorème 2.2. *(Morse et Hedlund, 1938)*
Si $u \in \mathcal{A}^{\mathbb{N}}$ est un mot infini, alors soit u est ultimement périodique, soit sa fonction de complexité est strictement croissante.

Démonstration. Par la Proposition 1.11, on sait que la fonction de complexité en facteurs de u, notée p, est croissante. Supposons qu'elle ne soit pas strictement croissante et montrons que dans ce cas u est obligatoirement un mot ultimement périodique.
Étant donné que p est croissante mais pas strictement croissante, il existe $n \in \mathbb{N}$ tel que $p(n) = p(n+1)$. Notons $M = p(n)$.
Considérons la fonction e définie dans la preuve de la Proposition 1.11. Cette fonction définit une injection entre les ensembles $L_n(u)$ et $L_{n+1}(u)$ de même cardinal. C'est donc une bijection. Pour tout facteur $w \in L_n(u)$, définissons $f(w)$ comme étant le suffixe de longueur n de $e(w)$: f est une fonction de $L_n(u)$ dans lui-même. Cependant, remarquons que la fonction f n'est pas spécialement injective. En effet, soient par exemple $w_0 = w_{0,0} \cdots w_{0,n-1}$ et $w_1 = w_{1,0} \cdots w_{1,n-1}$ deux mots tels que $w_{0,0} \neq w_{1,0}$ mais $w_{0,1} \cdots w_{0,n-1} = w_{1,1} \cdots w_{1,n-1}$ et tels que aux positions $i(w_0) + n$ et $i(w_1) + n$ dans u, il y ait la même lettre $a \in \mathcal{A}$. Alors $w_0 \neq w_1$ mais $f(w_0) = w_{0,1} \cdots w_{0,n-1}a = w_{1,1} \cdots w_{1,n-1}a = f(w_1)$.
Prenons un élément quelconque $y \in L_n(u)$ de départ et considérons la suite de $M + 1$ éléments de $L_n(u)$:

$$y, f(y), f^2(y), ..., f^M(y).$$

Par hypothèse, on a $|L_n(u)| = p_u(n) = M$, dès lors il existe $k, l \in \{0, ..., M\}$ tels que $k < l$ et $f^k(y) = f^l(y)$. Ainsi, en posant $x = f^k(y) \in L_n(u)$ et $m = l - k \in [\![1, M]\!]$, on a $f^m(x) = x$.
Considérons le facteur $w \in L_n(u)$ apparaissant à la position i dans u. Par définition des fonctions e et f, le suffixe de longueur n de $e(w)$ est $f(w)$ et il apparaît à la position $i+1$ dans u. Comme illustré sur le schéma suivant, en itérant le résultat, on obtient que, pour tout $j \geq 0$, $f^j(w)$ apparaît à la position $i + j$ dans u.

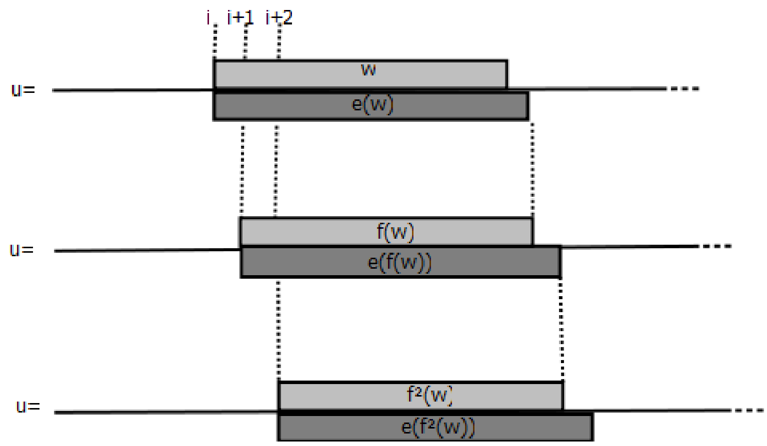


FIGURE 2.1 – Itérations de la fonction f.

En particulier, lorsque $w = x$, vu que $f^m(x) = x$, on obtient que $f^{j+m}(x) = f^j(x)$ a des occurrences aux positions $i + j + m$ et $i + j$, si x débute à la position i. Ainsi, deux à deux, les lettres aux positions correspondantes dans le mot de longueur n débutant à ces positions coïncident : $u_{i+j+m} = u_{i+j}$ pour tout $j \in \mathbb{N}$. Par définition, ceci montre que u est ultimement périodique de période m et de prépériode au plus i.

□

Corollaire 2.3. *S'il existe $n \in \mathbb{N}$ tel que $p(n) \leq n$, alors u est ultimement périodique.*

Démonstration. Procédons par contraposée. Si le mot u n'est pas ultimement périodique, alors par le Théorème 2.2 de MORSE et HEDLUND, p est strictement croissante, c'est-à-dire $p(n + 1) \geq p(n) + 1$. Puisque $p(0) = 1$, on obtient $p(n) \geq n + 1$ pour tout $n \in \mathbb{N}$.

□

Le Théorème 2.2 de MORSE et HEDLUND, nous permet donc d'affirmer que la complexité la plus basse pour un mot non-ultimement périodique est telle que $p(n) = n + 1$ pour tout $n \in \mathbb{N}$. L'ensemble des mots ayant cette caractéristique sont appelés les mots *Sturmiens*. Remarquons que ces derniers sont forcément binaires puisque $p(1) = 2$. Cette classe de mots binaires infinis peut être étudiée et caractérisée de différentes façons. Leurs propriétés combinatoires sont multiples. Ces mots sont, en particulier, étudiés dans les ouvrages [8] et [10].

Exemple 2.4. Une autre conséquence du Théorème 2.2 est que des fonctions asymptotiquement équivalentes à, par exemple, $\sqrt{n}$ ou $\frac{n}{2}$, ne peuvent être des fonctions de complexité d'aucun mot.
En effet, considérons une fonction $f : \mathbb{N} \to \mathbb{N}$ telle que $f(n) \sim \sqrt{n}$ (resp. $f(n) \sim \frac{n}{2}$) et supposons qu'il existe un mot $u \in \mathcal{A}^{\mathbb{N}}$ tel que $f(n) = p_u(n)$ pour tout $n \in \mathbb{N}$. Par le Corollaire 2.3, vu qu'il existe n tel que $p_u(n) \leq n$, alors u est ultimement périodique et la fonction $f(n) = p_u(n)$ est bornée. Ce résultat est absurde car, pour les deux fonctions considérées, on a

$$\lim_{n \to +\infty} f(n) = +\infty.$$

2.2 Facteurs spéciaux et bispéciaux

Nous pouvons remarquer que dans la preuve du Théorème 2.2, l'astuce était de regarder comment les facteurs peuvent être prolongés. Ainsi, dans cette section, nous allons tout d'abord formaliser ce concept d'extension et faire le lien avec la fonction de complexité étudiée. Ceci nous permettra en particulier, par la suite, d'avoir une preuve plus courte du Théorème 2.2 de Morse et Hedlund.
Tout au long de cette section, le mot infini générique étudié sera noté $u \in \mathcal{A}^{\mathbb{N}}$. Pour cette raison, lorsque nous allons définir une notation, par soucis de simplicité, nous n'allons pas spécialement mentionner la dépendance en u. Par exemple, si dans un autre contexte une ambiguïté entre plusieurs mots est possible, alors la notation que nous allons introduire comme étant $d^+(w)$ pourra être rendue plus précise. En effet, dans un tel cas, on la notera $d_u^+(w)$ pour spécifier qu'elle s'applique au mot u.

2.2.1 Définitions

Définition 2.5. Un mot infini u est dit *récurrent* si tous ses facteurs apparaissent infiniment souvent.

Définitions 2.6. Soit w un facteur de u. Une lettre qui apparaît dans u directement après une occurrence de w est une *extension à droite* de w dans u. L'ensemble des extensions à droite du facteur w est noté $E^+(w)$. La *valence à droite* de w, $d^+(w)$, est alors définie par

$$d^+(w) = \#(\{a \in \mathcal{A} \,|\, wa \in L(u)\}) = \#(E^+(w)).$$

De la même façon, on définit une *extension à gauche* de w dans u, l'ensemble des extensions à gauche $E^-(w)$ de w et la *valence à gauche* $d^-(w)$ de w.

Définitions 2.7. Un facteur ayant une valence à droite (resp. à gauche) plus grande ou égale à deux est appelé un *facteur spécial à droite* (resp. *à gauche*). Un facteur qui est à la fois spécial à gauche et à droite est dit *bispécial*. On note les ensembles des facteurs spéciaux à droite, spéciaux à gauche et bispéciaux de longueur n dans u respectivement [1] $RS_n(u)$, $LS_n(u)$ et $BS_n(u)$.

1. Les notations anglo-saxonnes L et R signifiant respectivement "Left" et "Right" ont été conservées.

Remarquons que dans la Proposition 1.11, on utilise le fait que la valence à droite de tout facteur est toujours plus grande ou égale à un. En effet, le mot u étant infini, tout facteur de u admet une continuation à droite. Ainsi, on peut dire que w est un facteur spécial à droite si $d^+(w) \neq 1$.

Définitions 2.8. Un facteur w est un *préfixe unioccurrent* de u si c'est un préfixe de u qui n'a qu'une seule occurrence dans le mot u. On note $LS'_n(u)$ l'ensemble contenant les facteurs spéciaux à gauche de longueur n ainsi que le préfixe unioccurent de longueur n de u, s'il existe. Formellement,

$$LS'_n(u) = \{w \in L_n(u) | d^-(w) \neq 1\}.$$

Ainsi, dans certaines circonstances $d^-(w)$ peut être égal à 0. Cela arrive lorsque w est un préfixe unioccurrent. Dans la plupart des exemples, cela ne se produit pas car le mot infini u est récurrent.

Lorsque le mot u n'est pas récurrent, un de ses préfixes va jouer un rôle particulier.

Définitions 2.9. Le plus long préfixe u qui n'est pas unioccurent est appelé le *préfixe exceptionnel* de u. On définit $BS'_n(u)$ comme étant l'ensemble contenant les facteurs bispéciaux de longueur n ainsi que le préfixe exceptionnel, s'il existe et s'il est de longueur n.

Ainsi, lorsque le mot u est récurrent, on a $LS'_n(u) = LS_n(u)$ et $BS'_n(u) = BS_n(u)$.

Définition 2.10. Le *type d'extensions* d'un facteur w de u est l'ensemble $E(w)$ des paires de lettres $(a, b) \in \mathcal{A}^2$ telles que w peut être étendu dans les deux directions dans u en tant que awb :

$$E(w) = \{(a, b) \in \mathcal{A}^2 | awb \in L(u)\}.$$

Remarquons que l'on peut avoir deux lettres a_0 et b_0 telles que $a_0 \in d^-(w)$ et $b_0 \in d^+(w)$, c'est-à-dire $a_0w \in L(u)$ et $wb_0 \in L(u)$, sans pour autant que le couple (a_0, b_0) appartienne à $E(w)$. Ainsi, on en tire

$$\# (E(w)) \leq d^-(w)d^+(w).$$

Définition 2.11. On appelle la *multiplicité (bilatérale)* d'un facteur w le nombre

$$m(w) = \# (E(w)) - d^-(w) - d^+(w) + 1.$$

Finalement, définissons deux nouvelles fonctions en lien direct avec la fonction de complexité p.

Définitions 2.12. On note respectivement s et b les *première* et *seconde différences (finies)* de la fonction de complexité p. Formellement, pour tout $n \in \mathbb{N}$, on a

$$s(n) = p(n + 1) - p(n)$$

et

$$b(n) = s(n + 1) - s(n) = p(n + 2) - 2p(n + 1) + p(n).$$

2.2.2 Propriétés et liens avec la fonction de complexité

Avec toutes ces définitions, nous pouvons maintenant énoncer quelques petits résultats utilisant ces notions de facteurs spéciaux.

Proposition 2.13. *Si w est un facteur du mot u tel que $m(w) \neq 0$, alors w est soit un facteur bispécial de u, soit son préfixe exceptionnel.*

Démonstration. Supposons que le facteur w ne soit pas bispécial et montrons que l'hypothèse $m(w) \neq 0$ ne peut se produire que lorsque w est le préfixe exceptionnel de u.
Si w n'est pas bispécial, alors on a $d^+(w) = 1$ ou $d^-(w) \leq 1$
Cas 1 : si $d^+(w) = 1$.
Il existe une lettre $b \in \mathcal{A}$ telle que $E^+(w) = \{b\}$. Ainsi, chaque occurrence de w est suivie par b et on a $E(w) = E^-(w) \times \{b\}$ donc $\#(E(w)) = d^-(w)$ et

$$\begin{aligned} m(w) &= d^-(w) - d^-(w) - d^+(w) + 1 \\ &= -d^+(w) + 1 = 0. \end{aligned}$$

Ceci est absurde par hypothèse.
Cas 2 : si $d^-(w) = 0$.
Par définition, le facteur w est un préfixe unioccurrent. Ainsi, étant donné que le facteur w n'a qu'une seule occurrence, il n'a qu'une seule extension à droite. On en tire $d^+(w) = 1$, $E(w) = \emptyset$ et obtient

$$\begin{aligned} m(w) &= \#(E(w)) - d^-(w) - d^+(w) + 1 \\ &= 0 - 0 - 1 + 1 = 0. \end{aligned}$$

Ceci est également absurde.
Cas 3 : si $d^-(w) = 1$.
Considérons tout d'abord le cas où le facteur w n'est pas un préfixe de u.
Ce cas se traite de manière similaire au cas 1. En effet, dans ce cas, il existe $a \in \mathcal{A}$ tel que $E^-(w) = \{a\}$. Ainsi, comme ci-avant, chaque occurrence de w est précédée par a et on a $E(w) = \{a\} \times E^+(w)$ donc $\#(E(w)) = d^+(w)$ et

$$\begin{aligned} m(w) &= d^+(w) - d^-(w) - d^+(w) + 1 \\ &= -d^-(w) + 1 = 0. \end{aligned}$$

Ceci est de nouveau absurde.
Le dernier cas à considérer est le cas où w est un préfixe de u.
Étant donné que le premier cas a déjà étudié la situation où on a $d^+(w) < 2$, nous pouvons donc supposer que $d^+(w) \geq 2$. Sachant que $d^-(w) = 1$, il existe $a \in \mathcal{A}$ tel que $E^-(w) = \{a\}$. De plus, le mot w est un préfixe de u, dès lors, posons $b \in \mathcal{A}$ la lettre telle que wb est un préfixe de u.

- Si le facteur wb a au moins une autre occurrence dans u, alors elle doit être précédée par a et on a $E(w) = \{a\} \times E^+(w)$. En effet, toutes les prolongations à droite de w, la lettre b y compris, sont telles que la prolongation à gauche de w correspondante est a uniquement. On obtient que $\#(E(w)) = d^+(w)$ et donc

$$\begin{aligned} m(w) &= \#(E(w)) - d^-(w) - d^+(w) + 1 \\ &= d^+(w) - 1 - d^+(w) + 1 = 0. \end{aligned}$$

  Ceci est absurde par hypothèse.

- Sinon, le mot wb est unioccurent, ce qui signifie par définition que w est le préfixe exceptionnel de u car c'est le plus long préfixe qui n'est pas unioccurent. En effet, soit v un préfixe de u tel que $|v| \geq |wb|$, alors wb est un facteur de v et donc v ne peut pas avoir une autre occurrence dans u, sinon wb en aurait une autre également. Remarquons que dans ce cas, $E(w) = \{a\} \times (E^+(w) \setminus \{b\})$, ainsi on obtient $\#(E(w)) = d^+(w) - 1$ et

$$\begin{aligned} m(w) &= \#(E(w)) - d^-(w) - d^+(w) + 1 \\ &= d^+(w) - 1 - d^-(w) - d^+(w) + 1 \\ &= -d^-(w) = -1. \end{aligned}$$

□

Exemple 2.14. Considérons le mot $u = abc(abbbbc)^\omega$. Le mot u est un mot infini ultimement périodique. Vu que le facteur abc n'a qu'une seule occurrence dans u, le mot u n'est pas récurrent. Qui plus est, les préfixes de longueur au moins 3 sont unioccurrents dans u, et le préfixe exceptionnel est ab. Les facteurs bispéciaux sont :

— ε car $E^-(\varepsilon) = \{a, b, c\}$ et $E^+(\varepsilon) = \{a, b, c\}$, donc $d^-(\varepsilon) = 3 = d^+(\varepsilon)$,

— b car $E^-(b) = \{a, b\}$ et $E^+(b) = \{b, c\}$, donc $d^-(b) = 2 = d^+(b)$,

— bb car $E^-(bb) = \{a, b\}$ et $E^+(bb) = \{b, c\}$, donc $d^-(bb) = 2 = d^+(bb)$, et

— bbb car $E^-(bbb) = \{a, b\}$ et $E^+(bbb) = \{b, c\}$, donc $d^-(bbb) = 2 = d^+(bbb)$.

Ensuite, on a $E(\varepsilon) = \{(a, b), (b, c), (c, a), (b, b)\}$, $E(b) = \{(a, c), (a, b), (b, b), (b, c)\}$, $E(bb) = \{(a, b), (b, b), (b, c)\}$ et $E(bbb) = \{(a, b), (b, c)\}$. Et donc, on obtient $m(\varepsilon) = -1$, $m(b) = 1$, $m(bb) = 0$ et $m(bbb) - 1$. De plus, vu que ab est le préfixe exceptionnel, on a $\#(E(ab)) = 0 = d^-(ab)$ et $m(ab) = -2 + 1 = -1$, car $E^+(ab) = \{c, b\}$.

Remarque 2.15. En appliquant la Proposition 2.13 à un mot u récurrent, étant donné qu'il n'y a pas de préfixe exceptionnel dans u, on en tire que seuls les facteurs bispeciaux peuvent avoir une multiplicité non nulle.

Définition 2.16. Un facteur bispécial w de u est dit *fort* si $m(w) > 0$, *faible* si $m(w) < 0$ et *neutre* si $m(w) = 0$.

Définition 2.17. Un facteur bispécial w de u, qui n'est pas son préfixe exceptionnel et pour lequel il existe $(a, b) \in E(w)$ tel que $E(w) \subseteq (\{a\} \times \mathcal{A}) \cup (\mathcal{A} \times \{b\})$, est dit *ordinaire*.

Proposition 2.18. *Tout facteur bispécial ordinaire est neutre.*

Démonstration. Soit w un facteur bispécial ordinaire de u. Par définition, il existe un couple $(a, b) \in E(w)$ tel que $E(w) \subseteq (\{a\} \times \mathcal{A}) \cup (\mathcal{A} \times \{b\})$. Notons k et $l \in \mathbb{N}$ les entiers naturels définis tels que $k = d^-(w) - 1$ et $l = d^+(w) - 1$. Par définition d'un facteur bispécial, on a $k \geq 1$ et $l \geq 1$.
Posons $D^- = \{\alpha_1^-, \ldots, \alpha_k^-\}$ et $D^+ = \{\alpha_1^+, \ldots, \alpha_l^+\}$. On obtient

$$E(w) = \{(a, b)\} \cup \underbrace{\{(a, \alpha_i^+) \mid i = 1, \ldots, l\}}_{:=E_1} \cup \underbrace{\{(\alpha_j^-, b) \mid j = 1, \ldots, k\}}_{:=E_2}.$$

En représentant les extensions de l'ensemble E_1 par des lignes en traits d'union et celles de e_2 en pointillés, les extensions à gauche et à droite du facteur w peuvent être illustrées par le schéma suivant :

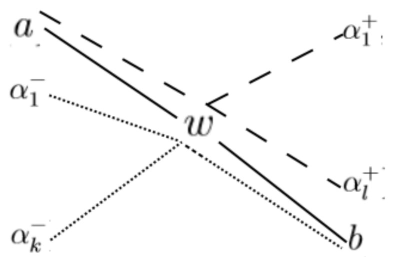


FIGURE 2.2 – Extensions d'un facteur bispécial ordinaire.

Calculons le cardinal de l'ensemble $E(w)$ en utilisant la définition des ensembles E_1 et E_2 :

$$\begin{aligned} \#\,(E(w)) &= \#(\{(a, b)\}) + \#(E_1) + \#(E_2) \\ &= 1 + \#(D^+) + \#(D^-) \\ &= 1 + (d^+(w) - 1) + (d^-(w) - 1) \\ &= d^-(w) + d^+(w) - 1 \end{aligned}$$

et

$$\begin{aligned} m(w) &= \#\,(E(w)) - d^-(w) - d^+(w) + 1 \\ &= d^-(w) + d^+(w) - 1 - d^-(w) - d^+(w) + 1 = 0. \end{aligned}$$

□

Remarques 2.19. Étude de la réciproque.

- Si $\#(\mathcal{A}) = 2$, la réciproque est vraie.
  En effet, si l'on considère l'alphabet $\mathcal{A} = \{a, b\}$, alors on a
  $$E(w) \subseteq \{(a, a), (a, b), (b, a), (b, b)\}.$$
  Dès lors, si w est un facteur bispécial neutre, on a $d^-(w) = 2 = d^+(w)$ et pour que la multiplicité bilatérale soit nulle, il faut que l'on ait $\#(E(w)) = 3$. Et pour tous les sous-ensembles de taille 3 de $E(w)$, les hypothèses de la définition d'un facteur bispécial ordinaire sont vérifiées.
- Si $\#(\mathcal{A}) \geq 3$, la réciproque est fausse.
  Un contre-exemple est donné par le mot $u = abbac(abbbbc)^\omega$ et le facteur $w = bb$. On a $E^-(bb) = \{a, b\}$ et $E^+(bb) = \{a, b, c\}$, donc $d^-(bb) = 2$ et $d^+(bb) = 3$, et $E(bb) = \{(a, b), (a, a), (b, b), (b, c)\}$. Ainsi, on a $m(w) = 4 - 2 - 3 + 1 = 0$ mais l'ensemble $E(bb)$ n'est pas de la forme adéquate pour que le facteur w soit qualifié d'ordinaire.

Proposition 2.20. *Pour tout $n \in \mathbb{N}$, on a*
$$\begin{aligned} p(n) &= 1 + \sum_{l=0}^{n-1} s(l) \\ &= 1 + (p(1) - 1)n + \sum_{m=0}^{n-1} (n - 1 - m)b(m) \quad . \end{aligned}$$

Cette proposition affirme que la fonction de complexité peut être entièrement calculée en connaissant soit la fonction s, soit à la fois $p(1) = \#(\mathrm{alph}(u))$ et la fonction b.

Démonstration. Montrons que $p(n) = 1 + \sum\limits_{l=0}^{n-1} s(l)$. On sait, par définition, que pour tout $l \in \mathbb{N}$, $s(l) = p(l + 1) - p(l)$ et $p(0) = 1$. Dès lors, on a
$$\begin{aligned} 1 + \sum_{l=0}^{n-1} s(l) =& 1 + s(0) + s(1) + \cdots + s(n - 1) \\ =& p(0) + p(1) - p(0) + p(2) - p(1) + \cdots + p(n) - p(n - 1) \\ =& p(n), \end{aligned}$$
la somme étant téléscopique.
De la même façon, vu que pour tout $m \in \mathbb{N}$, $b(m) = s(m + 1) - s(m)$ on obtient $s(l) = s(0) + \sum\limits_{m=0}^{l-1} b(m)$ pour tout $l \in \mathbb{N}$. Ainsi, en utilisant ce résultat intermédiaire dans l'égalité démontrée ci-avant, on en tire que

$$\begin{aligned}
p(n) &= 1 + \sum_{l=0}^{n-1} s(l) \\
&= 1 + \sum_{l=0}^{n-1} \left[s(0) + \sum_{m=0}^{l-1} b(m) \right] \\
&= 1 + n.s(0) + \sum_{l=0}^{n-1} \sum_{m=0}^{l-1} b(m) \\
&= 1 + n.(p(1) - 1) + \sum_{m=0}^{n-1} \sum_{l=m+1}^{n-1} b(m) \\
&= 1 + n.(p(1) - 1) + \sum_{m=0}^{n-1} (n - 1 - (m+1) + 1) b(m) \\
&= 1 + n.(p(1) - 1) + \sum_{m=0}^{n-1} (n - 1 - m) b(m) \quad .
\end{aligned}$$

□

La proposition précédente est particulièrement utile lorsque la fonction de complexité p croît lentement et donc que les fonctions s et b prennent des petites valeurs. Par exemple, par la première égalité de la Proposition 2.20, un mot infini u est un mot *Sturmien* ($p(n) = n + 1$) si la fonction s est la fonction constante 1. De la même façon, en utilisant cette fois la seconde égalité de la proposition, on en tire qu'un mot est *Sturmien* si l'on travaille sur un alphabet binaire ($p(1) = 2$) et si b est la fonction constante 0.

Nous pouvons désormais faire le lien entre les deux notions que nous venons de définir et étudier séparément. En effet, pour évaluer les fonction s et b, nous pouvons utiliser les facteurs spéciaux et bispéciaux.

Théorème 2.21. *Soit un mot infini $u \in \mathcal{A}^{\mathbb{N}}$. Pour tout $n \in \mathbb{N}$, on a*

(a) $s(n) = \sum_{w \in RS_n(u)} (d^+(w) - 1)$,

(b) $s(n) = \sum_{w \in LS'_n(u)} (d^-(w) - 1)$, *et*

(c) $b(n) = \sum_{w \in BS'_n(u)} m(w)$.

Démonstration. Soit un facteur w de longueur $n + 1$ de u. Le mot w est la donnée de son préfixe (resp. suffixe) de longueur n et de l'extension à droite (resp. gauche) de ce préfixe (resp. suffixe) dans w.

Ainsi, pour caractériser l'ensemble des facteurs de longueur $n+1$, on énumère ceux de longueur n et pour chacun, on regarde les façons de l'étendre à droite (resp. à gauche) dans u pour obtenir un facteur de longueur $n+1$. Remarquons ensuite que cette construction ne permet pas à deux facteurs différents de longueur n de produire le même facteur de longueur $n+1$ vu qu'ils en seront le préfixe (resp. suffixe).
On obtient

$$p(n+1) = \sum_{w\in L_n(u)} d^+(w) = \sum_{w\in L_n(u)} d^-(w).$$

Dès lors,

$$\begin{aligned} s(n) &= p(n+1) - p(n) \\ &= \sum_{w\in L_n(u)} d^+(w) - \sum_{w\in L_n(u)} 1 \\ &= \sum_{w\in L_n(u)} (d^+(w) - 1). \end{aligned}$$

Et de la même façon, on a

$$s(n) = \sum_{w\in L_n(u)} (d^-(w) - 1).$$

Cependant, le terme $d^+(w) - 1$ est non nul si et seulement si w est un facteur spécial à droite et $d^-(w)-1$ est non nul si et seulement si $w \in LS'_n(u)$. Ceci nous permet de conclure les points (a) et (b).
Pour le point (c), on remarque, par le même type de raisonnement que précédemment, que l'on a

$$p(n+2) = \sum_{w\in L_n(u)} \#\left(E(w)\right).$$

Ainsi, on obtient

$$\begin{aligned} b(n) &= s(n+1) - s(n) \\ &= p(n+2) - p(n+1) - p(n+1) + p(n) \\ &= \sum_{w\in L_n(u)} \#\left(E(w)\right) - \sum_{w\in L_n(u)} d^-(w) - \sum_{w\in L_n(u)} d^+(w) + \sum_{w\in L_n(u)} 1 \\ &= \sum_{w\in L_n(u)} m(w). \end{aligned}$$

La Proposition 2.13 nous permet de conclure, car la multiplicité bilatérale $m(w)$ de w est non nulle si et seulement si $w \in BS'_n(u)$.

□

Le Théorème 2.21 peut être traduit de la façon suivante :

(a) $s(n)$ est le nombre de facteurs spéciaux à droite w de u de longueur n, comptés avec une multiplicité égale à $d^+(w) - 1$.

(b) $s(n)$ est le nombre de facteurs spéciaux à gauche w de u de longueur n, ainsi que le préfixe unioccurent de cette longueur, s'il existe, comptés avec une multiplicité égale à $d^-(w) - 1$.

(c) $b(n)$ est le nombre de facteurs bispéciaux w de u de longueur n, ainsi que le préfixe exceptionnel de cette longueur, s'il existe, comptés avec une multiplicité (qui peut cette fois être négative) égale à $m(w)$.

Corollaire 2.22. *Soit u un mot binaire infini. Alors,*

(a) $s(n)$ est le nombre de ses facteurs spéciaux à droite de longueur n.

Et si u est récurrent,

(b) $s(n)$ est le nombre de ses facteurs spéciaux à gauche de longueur n, et

(c) $b(n) = Fb(n) - fb(n)$ où $Fb(n)$ (resp. $fb(n)$) est le nombre de facteurs bispéciaux forts (resp. faibles) de longueur n.

Démonstration. Le mot u étant binaire, pour tout facteur w de u, on a $d^+(w) \leq 2$ et $d^-(w) \leq 2$. De plus, par définition des facteurs spéciaux à droite (resp. à gauche), on a $d^+(w) \geq 2$ (resp. $d^-(w) \geq 2$). Ainsi, on a $d^+(w) - 1 = 1$ et $d^-(w) - 1 = 1$. La démonstration du point (a) est donc directe. Or, vu que le mot u est récurrent, il n'y a pas de préfixe unioccurent et on a $LS'_n(u) = LS_n(u)$. Ceci termine le point (b).
Il reste donc à montrer le point (c). Pour ce faire, notons $\mathcal{A} = \{a, b\}$ et démontrons que si w est un facteur bispécial du mot infini récurrent u, alors $m(w) \in \{-1, 0, 1\}$. Considérons le facteur bispécial w. Dès lors, on a $E^+(w) = \{a, b\} = E^-(w)$. Ainsi, les mots wa, aw, wb et bw sont des facteurs de u. Les facteurs aw et bw peuvent ne pas être spéciaux à droite mais vu que le mot u est infini, il existe au moins une extension à droite pour chacun d'entre eux. L'ensemble $E(w)$ contient donc au moins 2 éléments. Or, on a $E(w) \subseteq \{(a, a), (a, b), (b, a), (b, b)\}$ et donc on obtient $2 \leq \#(E(w)) \leq 4$. Ainsi, étant donné que pour tout facteur spécial à droite (resp. gauche) d'un mot infini binaire u on a $d^+(w) = 2$ (resp. $d^-(w) = 2$), par définition de la multiplicité bilatérale, on obtient $-1 \leq m(w) \leq 1$. Finalement, vu que le mot u est récurrent, on a $BS'_n(u) = BS_n(u)$. Ainsi, par le théorème précédent, on en tire que

$$\begin{aligned} b(n) &= \sum_{w \in BS_n(u)} m(w) \\ &= \sum_{\substack{w \in BS_n(u) \\ m(w) = -1}} m(w) + 0 + \sum_{\substack{w \in BS_n(u) \\ m(w) = 1}} m(w) \\ &= \sum_{\substack{w \in BS_n(u) \\ m(w) = -1}} (-1) + \sum_{\substack{w \in BS_n(u) \\ m(w) = 1}} 1 \\ &= -fb(n) + Fb(n). \end{aligned}$$

□

2.3 Seconde preuve du théorème de Morse et Hedlund

En utilisant l'outil des facteurs spéciaux et bispéciaux, on obtient une démonstration alternative du Théorème 2.2 de Morse et Hedlund.

Démonstration. Preuve du Théorème 2.2
Soient u un mot infini non ultimement périodique et n un entier naturel. Montrons que la fonction de complexité p de u est strictement croissante.
L'ensemble $L_n(u)$ est fini. Ainsi, il existe au moins un facteur $w \in L_n(u)$ qui apparaît au moins deux fois dans u. Notons i et j les positions de ces occurrences avec $i < j$. De plus, étant donné que le mot u n'est pas ultimement périodique, il existe $m \in \mathbb{N}$ tel que $u_{i+m} \neq u_{j+m}$ car sinon $(j-i)$ serait une période de u. Or, par définition de i et j, pour tout $k \in \{0, ..., n-1\}$, on a $u_{i+k} = u_{j+k}$. Dès lors, on a $m \geq n$.
Notons m_0 le plus petit entier vérifiant cette propriété et $w' = u_i u_{i+1} \cdots u_{i+m_0-1}$. Le suffixe de longueur n de w' est spécial à droite car il peut au moins s'étendre à droite par les lettres différentes u_{i+m_0} et u_{j+m_0}. Ainsi, le mot u a au moins un facteur spécial à droite de chaque longueur et par le Théorème 2.21 $p(n+1) - p(n) = s(n) \geq 1$. On en conclut que la fonction p est strictement croissante.

□

Chapitre 3

Graphes de Rauzy et mots lacunaires

3.1 Graphes de Rauzy

Les extensions et valences d'un facteur, ainsi que les facteurs spéciaux et bispéciaux, d'un mot $u \in \mathcal{A}^{\mathbb{N}}$ font partie des notions pouvant être visualisées sur un graphe représentant l'ensemble $L_n(u)$ des facteurs de longueur n de u. Cette méthode a été introduite en 1983 et tient le nom de son inventeur Gérard Rauzy (1928–2010). Dans la littérature, ces graphes sont parfois également appelés graphes des facteurs de u.

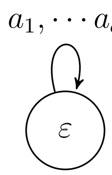
Définition 3.1. Le *graphe de* Rauzy d'ordre n du mot infini $u \in \mathcal{A}^{\mathbb{N}}$ est un graphe dirigé noté $G_n = (V, E)$ où :

- G_n a $p_u(n)$ nœuds dont les labels sont les facteurs de longueur n de u,
- il existe un arc du nœud v au nœud w si et seulement si il existe deux lettres $a, b \in \mathcal{A}$ telles que $vb = aw \in L_{n+1}(u)$. Dans ce cas, cet arc a un label correspondant à la lettre[1] a.

Formellement, on a

$$G_n = (L_n(u), \{(v, w) \in L_n(u) \times L_n(u) \mid \exists a, b \in \mathcal{A} : vb = aw \in L_{n+1}(u)\}) .$$

Remarque 3.2. Par définition, si l'on considère l'alphabet $\mathcal{A} = \{a_1, \cdots a_d\}$, pour tout mot $u \in \mathcal{A}^{\mathbb{N}}$ tel que $\mathcal{A} = \text{alph}(u)$, le graphe de Rauzy d'ordre 0, noté G_0, est le suivant :



En effet, le mot vide ε est l'unique facteur de u de longueur 0 et pour toute lettre $a_i \in \mathcal{A}$, $i \in \{1, \ldots, d\}$, on a $\varepsilon a_i = a_i \varepsilon \in L(u)$.

1. Selon l'application, il est parfois plus utile de travailler avec un graphe de Rauzy où les arcs ont un label correspondant à la lettre b ou au mot vb.

Par définition, à tout arc du graphe de RAUZY G_n correspond un mot de $L_{n+1}(u)$ et il y en a exactement $p(n+1)$. En effet, tout mot $z = z_0 \cdots z_n = z'z_n = z_0z'' \in L_{n+1}(u)$ correspond à un arc de label z_0 allant de son préfixe $z' \in L_n(u)$ vers son suffixe $z'' \in L_n(u)$. Le graphe de RAUZY G_n d'ordre n permet donc de définir univoquement les facteurs de longueur $n+1$ de u.

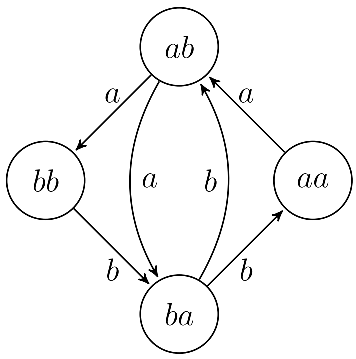
Exemple 3.3. Considérons le mot [2] de THUE-MORSE

$$t = abbabaabbaababba \cdots .$$

Étant donné que l'on a $L_2(t) = \{aa, ab, ba, bb\}$ et que le mot de THUE-MORSE ne contient aucun cube (Proposition I.1.27. [13]), on en tire que $p_t(3) = 6$ et :

$$\begin{aligned}
(aa)a &= a(aa) \notin L_3(t), \\
(aa)b &= a(ab) \in L_3(t), \\
(ab)a &= a(ba) \in L_3(t), \\
(ab)b &= a(bb) \in L_3(t), \\
(ba)a &= b(aa) \in L_3(t), \\
(ba)b &= b(ab) \in L_3(t), \\
(bb)b &= b(bb) \notin L_3(t), \\
(bb)a &= b(ba) \in L_3(t).
\end{aligned}$$

Le graphe de RAUZY d'ordre 2 noté G_2 est donné par :



Ce graphe a 6 arêtes, ce qui correspond effectivement à $p_t(3)$.

Remarque 3.4. Par définition du graphe de RAUZY, tout mot appartenant à $L_m(u)$ peut être écrit comme $z = ls$ où l correspond à un chemin de longueur $m-n$ dans le graphe de RAUZY G_n et s est le label du dernier nœud de ce chemin, $|s| = n$.
En effet, tout mot $z \in L_{n+1}(u)$ peut être écrit comme $z = ls$ où $l = z_0$ est le label d'un arc dans G_n et $s = \mathrm{Suff}_n(z)$ est la destination de cet arc. Et ce processus peut être appliqué de proche en proche en parcourant, de gauche à droite, le facteur de longueur m considéré avec des fenêtres de taille $n+1$.
Cependant, le graphe de RAUZY G_n d'ordre n d'un mot u ne permet pas de définir l'ensemble $L_m(u)$ pour $m > n+1$. Il existe des mots de longueur m pouvant être décomposés

2. La définition formelle de ce mot sera donnée dans le chapitre suivant.

de la même façon mais n'étant pas des facteurs de longueur m de u. Un contre-exemple est illustré par le graphe de Rauzy d'ordre 2 du mot de Thue-Morse présenté dans l'Exemple 3.3. En effet, le chemin défini par le parcours des arcs $(ab, ba), (ba, ab), (ab, ba)$ puis (ba, ab) correspond au label $l = abab$ et le dernier nœud s visité correspond à $s = ab$, dès lors le mot $z = ls = ababab$ pourrait être considéré comme un facteur de longueur $6 > 3$ du mot de Thue-Morse t. Ceci est absurde étant donné que le mot t ne contient pas de cube.

Il existe un arc entre les nœuds v et $w \in L_n(u)$ s'il existe deux lettres $a, b \in \mathcal{A}$ telles que $vb = aw \in L_{n+1}(u)$, ainsi le mot w est prolongeable à gauche en a dans le mot u. Ainsi, le degré entrant du nœud $w \in L_n(u)$ correspond à la définition de la valence à gauche $d^-(w)$ du facteur w dans u . De la même façon, le degré sortant du nœud $v \in L_n(u)$ correspond à la définition de la valence à droite $d^+(v)$ du facteur v dans u. Ceci explique les notations, étant habituelles en théorie des graphes, données à ces termes dans la section 2.2.1. De part cette correspondance, un facteur est spécial à droite (resp. gauche) si et seulement si le nœud correspondant dans le graphe de Rauzy a au moins deux arcs sortants (resp. entrants).

Dans le graphe de Rauzy G_n d'ordre n, si nous lisons les facteurs de longueur n de u dans leur ordre d'apparition dans u, c'est-à-dire en commençant par $u_0u_1 \cdots u_{n-1}$ puis $u_1u_2 \cdots u_n$ et ainsi de suite, nous obtenons un chemin infini dans G_n dont le label est u. Remarquons que les transitions entre ces facteurs sont toujours possibles par définition d'un arc entre deux nœuds de G_n. Si nous considérons les $(i+1)^{\text{ième}}$ et $(i+2)^{\text{ième}}$ facteurs énumérés dans cet ordre, l'arc correspondant dans G_n est le suivant.

$$u_iu_{i+1} \cdots u_{i+n-1} \xrightarrow{u_i} u_{i+1}u_{i+2} \cdots u_{i+n}$$

De la même façon, comme expliqué précédemment, si $z \in L_m(u)$ où $m \geq n$, alors il existe un chemin de longueur $|z| - n$ dans G_n, dont le label est le mot $l = \text{Pref}_{|z|-n}(z)$, commençant en le préfixe z' de longueur n de z et se terminant en le suffixe s de longueur n de z. Dans ces conditions, le facteur z de u peut être décomposé de la façon suivante : $z = ls$.

Définition 3.5. Un automate fini non déterministe est la donnée d'un quintuple

$$Aut = (Q, I, F, \mathcal{A}, \Delta)$$

où

- Q est un ensemble fini dont les éléments sont les états de Aut,
- $I \subseteq Q$ est l'ensemble des états initiaux,
- $F \subseteq Q$ désigne l'ensemble des états finaux,
- $\mathcal{A}$ est l'alphabet de l'automate,
- $\Delta \subseteq Q \times \mathcal{A}^* \times Q$ est une relation de transition (qu'on supposera finie).

Définition 3.6. Un mot $w = w_0 \cdots w_k$ est accepté par un automate non déterministe $Aut = (Q, I, F, \mathcal{A}, \Delta)$ s'il existe un état $q_0 \in I$, une suite d'états $q_1, \ldots, q_{n-1} \in Q$ et un état $q_n \in F$ où $n \in \mathbb{N}_0$ ainsi qu'un ensemble de mots $v_0, \ldots, v_{n-1} \in \mathcal{A}^*$ tels que

$$(q_0, v_0, q_1), (q_1, v_1, q_2), \ldots, (q_{n-1}, v_{n-1}, q_n) \in \Delta$$

et

$$w = v_0 \cdots v_{n-1}.$$

En d'autres termes, le mot w est accepté par l'automate Aut s'il existe un chemin initial et final dans le graphe associé à Aut dont le label est w.

Définition 3.7. Le langage accepté par un automate $Aut = (Q, I, F, \mathcal{A}, \Delta)$ est l'ensemble des mots acceptés par Aut. On le note $L(Aut)$.

Remarque 3.8. Un automate déterministe est défini de la même manière, à l'exception du fait qu'il n'existe qu'un seul état initial $I = \{q_0\}$ et que nous n'avons plus une relation de transition mais une fonction [3] de transition $\delta : Q \times \mathcal{A} \to Q$. Les labels sur les arcs sont désormais des lettres et non plus des mots. Les autres définitions s'adaptent donc aisément.

Au vu des définitions précédentes, le graphe de Rauzy G_n d'ordre n peut être vu comme un automate fini non déterministe

$$G_n = (Q, I, F, \mathcal{A}, \Delta)$$

où

- l'ensemble des états Q est $L_n(u)$,
- tous les états sont initiaux et finaux, c'est-à-dire $I = F = Q$,
- l'alphabet $\mathcal{A}$ de l'automate correspond à l'alphabet $\text{alph}(u)$ du mot infini u considéré,
- la relation de transition Δ est obtenue par la définition [4] 3.1.

Conformément à la définition, on note $L(G_n)$ le langage accepté par l'automate G_n.

Le non déterminisme de l'automate associé au graphe de Rauzy est une conséquence directe de la définition des transitions. En effet, pour tout $n \in \mathbb{N}_0$, d'un nœud $v \in L_n(u)$ de G_n, sortent uniquement des arcs dont le label est v_0. Cette caractéristique est illustrée dans l'Exemple 3.3 du graphe de Rauzy du mot de Thue-Morse. Par exemple, de l'état correspondant au mot ba (resp. ab) sortent deux arcs dont le label est la lettre b (resp. a).

Remarque 3.9. Par définition des arcs entre deux nœuds du graphe de Rauzy, pour tous entiers naturels m, n tels que $m \leq n$, tout chemin de longueur m dans G_n débutant en un nœud v a un label égal à $\text{Pref}_m(v)$. Ainsi, de tout nœud de G_n peut être construit un chemin de longueur n ayant le label du facteur correspondant à ce nœud. Dès lors, remarquons que la construction expliquée ci-avant pour la reconnaissance d'un mot z de longueur m dans G_n comme étant $z = ls$ où l est le label d'un chemin de longueur $m - n$ dans G_n et s le label du dernier nœud de ce chemin peut également être adaptée. Une variante serait de noter $z = l'$ où l' est le label d'un chemin de longueur m dans G_n, c'est-à-dire $z \in L(G_n)$. En effet, il suffit de continuer le procédé en parcourant le chemin du nœud s pour obtenir son label. La construction utilisée dépend de l'application considérée.

3. Ainsi, contrairement à précédemment, d'un état $q_i \in Q$, il ne peut sortir deux arcs de même label aboutissant à deux états différents.

4. En particulier, les labels des arcs sont des lettres et non des mots comme la définition l'autorise.

Proposition 3.10. *Soient un mot* $u \in \mathcal{A}^{\mathbb{N}}$ *et* G_n *le graphe de* RAUZY *d'ordre* $n \in \mathbb{N}$ *de* u. *On a*

(i) $L(G_{n+1}) \subseteq L(G_n)$,

(ii) $L(G_n) \cap \mathcal{A}^m = L_m(u)$, *pour tout* $m \leq n$ *et*

(iii) $L(u) = \bigcap\limits_{n\in\mathbb{N}} L(G_n)$.

Ainsi, la suite des langages $L(G_n)$ *approxime le langage* $L(u)$.

Démonstration. Considérons un mot $w \in L(u)$. Étant donné que u est un mot infini, pour tout $n \in \mathbb{N}$, le facteur w de u peut être étendu en un facteur $z = ws \in L_{|w|+n}(u)$. Ainsi, le facteur z de u définit un chemin dans G_n aboutissant dans s dont le label est w. Ainsi, on a $L(u) \subseteq L(G_n)$ pour tout $n \in \mathbb{N}$.

(i) Considérons un mot $w \in L(G_{n+1})$. Par définition du langage accepté par G_{n+1}, le mot w est le label d'un chemin

$$(v_0, w_0, v_1), (v_1, w_1, v_2), \ldots, (v_{|w|-1}, w_{|w|-1}, v_{|w|}),$$

où

$$v_i \in L_{n+1}(u), \forall i \in \{0, \ldots, |w|\}.$$

Par construction, du chemin dans G_{n+1}, pour tout $i \in \{0, \ldots, |w|-1\}$, il existe $a_i \in \mathcal{A}$ tel que $v_i a_i = w_i v_{i+1} \in L_{n+2}(u)$. Pour tout $i \in \{0, \ldots, |w|\}$, notons $v_i' \in L_n(u)$ le préfixe de longueur n de v_i et a_i' la dernière lettre de v_i. On a l'égalité

$$v_i' a_i' a_i = v_i a_i = w_i v_{i+1} = w_i v_{i+1}' a_{i+1}'.$$

En particulier, on a $v_i' a_i' = w_i v_{i+1}'$. Ainsi, il existe le chemin suivant dans G_n

$$(v_0', w_0, v_1'), (v_1', w_1, v_2'), \ldots, (v_{|w|-1}', w_{|w|-1}, v_{|w|}'),$$

ayant également le label w. On en tire que $w \in L(G_n)$ et $L(G_{n+1}) \subseteq L(G_n)$.

(ii) Considérons $m \leq n$ et un mot $w \in L(G_n) \cap \mathcal{A}^m$. Par définition, le mot w est le label d'un chemin de longueur m dans G_n commençant en un nœud $v \in L_n(u)$. Ainsi, le mot w est un préfixe de longueur m de v et, vu que v est un facteur de u, w en est également un. On en tire que $w \in L_m(u)$. Dès lors, pour tout $m \leq n$, on a

$$L(G_n) \cap \mathcal{A}^m \subseteq L_m(u).$$

Or, nous savons que $L(u) \subseteq L(G_n)$. Ainsi, on a

$$L_m(u) = L(u) \cap \mathcal{A}^m \subseteq L(G_n) \cap \mathcal{A}^m.$$

Par double inclusion, on obtient $L_m(u) = L(G_n) \cap \mathcal{A}^m$ pour tout $m \leq n$.

(iii) Notons $L = \bigcap_{n\in\mathbb{N}} L(G_n)$. On a

$$\begin{aligned}
L \cap \mathcal{A}^m &= \left(\bigcap_{n\in\mathbb{N}} L(G_n)\right) \cap \mathcal{A}^m \\
&= \bigcap_{n\in\mathbb{N}} \left(L(G_n) \cap \mathcal{A}^m\right) \\
&= \underbrace{\left(\bigcap_{\substack{n\in\mathbb{N}\\ m\leq n}} L(G_n) \cap \mathcal{A}^m\right)}_{\substack{=L_m(u)\\ \text{par le point } (ii)}} \quad \cap \quad \underbrace{\left(\bigcap_{\substack{n\in\mathbb{N}\\ m>n}} L(G_n) \cap \mathcal{A}^m\right)}_{\substack{\supseteq L_m(u)\\ \text{par la Remarque 3.4}}} = L_m(u).
\end{aligned}$$

Ainsi, on obtient $L = L(u)$.

□

3.1.1 Application aux mots Sturmiens

Rappelons qu'un mot binaire infini u est dit *Sturmien* si sa fonction de complexité en facteurs est telle que $p(n) = n + 1$ pour tout $n \in \mathbb{N}$.
Les graphes de Rauzy des mots *Sturmiens* peuvent être caractérisés. Une description de ceux-ci va être développée dans cette section.

Lemme 3.11. *Les mots* Sturmiens *sont récurrents.*

Démonstration. Procédons par contraposition. Considérons un mot infini u non récurrent. Par définition, il existe un facteur w de u apparaissant un nombre fini de fois dans u. Ainsi, il existe une dernière occurrence de w dans u. Décomposons le mot u en préfixe et suffixe apparaissant avant et après cette occurrence. On le note

$$u = u'wv$$

où $u' \in \mathcal{A}^*$, $v \in \mathcal{A}^{\mathbb{N}}$ et $w \notin L(v)$. En particulier, si l'on note $n = |w|$, étant donné que $L(v) \subseteq L(u)$ mais que $w \in L(u) \setminus L(v)$, on a $p_v(n) < p_u(n)$.
Si le mot v est ultimement périodique, on a $v = tz^\omega$ avec $t, z \in \mathcal{A}^+$. Ainsi, on obtient

$$u = u'wtz^\omega = t'z^\omega$$

où $t' = v'wt$ et u est également ultimement périodique. La fonction de complexité en facteurs p_u de u est donc bornée et u n'est pas *Sturmien*.
Dans le cas contraire, par le Théorème 2.2 de Morse et Hedlund, la fonction de complexité en facteurs p_v de v est strictement croissante, c'est-à-dire

$$n + 1 \leq p_v(n) < p_u(n)$$

et u n'est donc pas *Sturmien*.

□

Lemme 3.12. *Soient un mot* Sturmien u *et un nombre naturel* n. *Le mot* u *a exactement un facteur spécial à gauche de longueur* n *et un facteur spécial à droite de longueur* n. *De plus, le mot* u *contient au plus un seul facteur bispécial de longueur* n *et dans le cas de son existence, il est neutre.*

Démonstration. La fonction de complexité en facteurs de u est telle que $p(n) = n+1$ pour tout $n \in \mathbb{N}$. Ainsi, pour tout $n \in \mathbb{N}$, les première et secondes différences $s(n)$ et $b(n)$ sont respectivement égales à

$$s(n) = p(n+1) - p(n) = n+2-(n+1) = 1$$

et

$$b(n) = s(n+1) - s(n) = 1-1 = 0.$$

Or, par le Lemme 3.11, le mot u étant récurrent, on a $LS'_n(u) = LS_n(u)$ et $BS'_n(u) = BS_n(u)$. Ainsi, par le Théorème 2.21, on a

$$\begin{aligned} 1 = s(n) &= \sum_{w \in RS_n(u)} (d^+(w) - 1) \text{ , et} \\ 1 = s(n) &= \sum_{w \in LS_n(u)} (d^-(w) - 1). \end{aligned}$$

Dès lors, étant donné que le mot u est binaire et récurrent, dans les égalités précédentes, les termes de chaque somme sont égaux à 0 ou à 1. On en déduit qu'il n'y a qu'un seul facteur spécial à droite de longueur n et un seul facteur spécial à gauche de longueur n dans u. Ensuite, remarquons que le mot u n'a au plus qu'un seul facteur bispécial, cela a lieu lorsque les facteurs spéciaux à droite et à gauche coïncident. Dans ce cas, il est nécessairement neutre. En effet, par le Théorème 2.21, nous avons

$$0 = b(n) = \sum_{w \in BS_n(u)} m(w).$$

Ainsi, si nous notons w' l'unique facteur bispécial de w, on a $0 = m(w')$.

□

Définitions 3.13. Un *sous-graphe* d'un graphe $G = (V, E)$ est un graphe $G' = (V', E')$ défini tel que $V' \subseteq V$ et $E' = E \cap (V' \times V')$.
Une *composante fortement connexe* d'un graphe orienté $G = (V, E)$ est un sous-graphe $G' = (V', E')$ de G où V' est un sous-ensemble maximal de sommets de V tel que, pour tous $u, v \in V'$, il existe un chemin allant de u à v.
Un graphe est *fortement connexe* s'il est formé d'une seule composante fortement connexe.

Lemme 3.14. *Pour tout $n \in \mathbb{N}$, le graphe de* Rauzy *G_n d'ordre n d'un mot infini u récurrent est fortement connexe.*

Démonstration. Considérons deux nœuds $v, w \in G_n$. Par définition, les nœuds v et w sont des facteurs de longueur n de u. Dès lors, vu que u est récurrent, les facteurs v et w apparaissent une infinité de fois dans u. Le mot u peut donc être décomposé comme suit :

$$u = u_0 w u_1 v u_2 w u_3 v u_4 \cdots .$$

Considérons le mot $z = v u_2 w \in L(u)$. On a

$$|z| = |v u_2 w| = |v| + |u_2| + |w| = n + |u_2| + n = 2n + |u_2| \geq n.$$

Ainsi, il existe un chemin de longueur $|z| - n = n + |u_2|$ dans G_n débutant dans le nœud correspondant au préfixe de longueur n de z et aboutissant dans le nœud correspondant au suffixe de longueur n de z. Or, par définition, les préfixes et suffixes de longueur n de z sont respectivement v et w. De la même façon, le facteur $z' = w u_1 v$ permet de démontrer l'existence d'un chemin de w à v dans G_n.

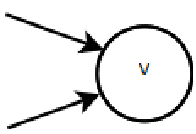
□

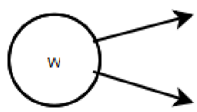
Remarquons que si le mot u n'est pas récurrent, il contient un préfixe exceptionnel u'. Dès lors, si l'on note $n = |u'|$, il n'existe aucun arc aboutissant dans le nœud du graphe de Rauzy G_n lui correspondant, car il n'existe aucune lettre $a \in \mathcal{A}$ telle que au' soit un facteur de longueur $n + 1$ de u.

Proposition 3.15. *Soient un mot* Sturmien *u, un entier naturel n ainsi que v et w les uniques facteurs de u de longueur n spéciaux à gauche et à droite respectivement. Le graphe de* Rauzy *G_n est de l'une des deux formes suivantes :*

- *Si $v = w$, alors G_n est composé de deux boucles disjointes autour du nœud correspondant à $v = w$.*
- *Si $v \neq w$, alors G_n a trois branches : une reliant v à w et deux reliant w à v.*

Démonstration. Étant donné que le facteur v (resp. w) est un facteur spécial à gauche (resp. à droite) du mot binaire u, il est représenté dans le graphe G_n par un nœud ayant un degré entrant (resp. sortant) égal à 2. On a

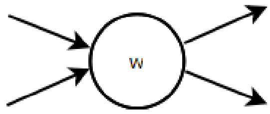


et


.

Par le Lemme 3.11, le mot u est récurrent et il ne peut donc pas y avoir de facteur ayant une valence à gauche égale à 0. Dans le graphe de Rauzy, cela se traduit par le fait qu'aucun nœud ne peut avoir un degré entrant nul. De plus, le Lemme 3.14 stipule que le graphe G_n est fortement connexe. Séparons la démonstration en deux cas possibles.

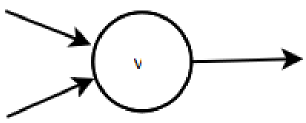
- Si $v = w$, tout nœud de G_n à l'exception de w a un degré entrant égal à 1 et un degré sortant égal à 1. De plus, le facteur w est bispécial car il a des degrés entrant et sortant égaux à deux.

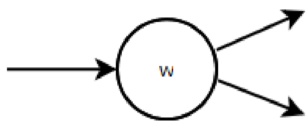
  

  .

  Ainsi, l'unique façon de connecter les nœuds de G_n est de créer deux boucles disjointes autour de w. En effet, en débutant de w, nous aboutissons dans un autre nœud w' en utilisant son unique arc entrant. Pour en sortir nous utilisons son unique arc sortant. Ainsi, nous dirons que ce nœud w' est *condamné*. En sortant de w' nous suivons le même procédé pour chaque nœud visité différent de w. Vu que le graphe G_n est fortement connexe, ce chemin aboutit de nouveau dans w. Ainsi, lorsque nous arrivons de nouveau dans w, nous avons condamné une partie des nœuds et créé une boucle autour de w. Or, les seconds arcs entrant et sortant de w doivent obligatoirement être utilisés. Dès lors, nous pouvons réitérer le procédé en utilisant le second arc sortant de w jusqu'à l'utilisation, par connexité, de tous les nœuds de G_n et le retour dans w. Qui plus est, cette boucle est disjointe de la première, étant donné que les sommets de la première boucle furent condamnés.

- Si $v \neq w$, alors le nœud v, ayant un degré entrant égal à 2 et un degré sortant égal à 1, est distingué du nœud w, ayant un degré entrant égal à 1 et un degré sortant égal à 2. On a

  et
  

  .

  Il existe dès lors deux façons de connecter v et w. En effet, rappelons que comme précédemment, les autres nœuds ont un degré entrant et un degré sortant égaux à 1 et considérons un arc sortant du nœud w. Par le même procédé, nous allons condamner des nœuds différents de w et v jusqu'au moment où nous revenons soit à v soit à w. Considérons le premier cas. Nous avons donc créé une branche de w à v et devons sortir de v par son unique arc sortant. Si nous aboutissons de nouveau dans v après le parcours d'autres nœuds différents de w, alors il n'y aura aucun arc de v à w et le graphe G_n n'est pas connexe, ceci est absurde par le Lemme 3.14. Ainsi, nous créons une branche de v à w. Ensuite, nous sortons du nœud w et devons utiliser les arcs restants, en particulier l'arc entrant dans le nœud v. Dès lors, le parcours se termine en v en ayant créé une nouvelle branche de w à v. En conclusion, cette première façon consiste à relier v à w et ensuite de relier les deux arcs sortants de w au deux arcs entrants de v.
  Considérons maintenant le second cas qui consiste à créer une première boucle en w. Nous devons continuer le parcours en utilisant le second arc sortant de w. Le sommet w est dès lors condamné car tous ses arcs ont été utilisés. Ainsi, nous allons aller de w à v et ensuite créer une boucle en v. Ce parcours n'admet donc aucun chemin de v à w, ceci est absurde par le Lemme 3.14.

□

Cette technique de caractérisation des graphes de RAUZY des mots *Sturmiens* pourrait être adaptée à d'autres mots dont la fonction de complexité en facteurs adopte certaines formes particulières. Lorsque la première différence s est connue, il existe un nombre fini de possibilité de graphes de RAUZY correspondants. Une étude similaire des graphes de RAUZY est faite dans l'article [1] et l'article de conférence [5] dans le cas où la fonction de complexité en facteurs est telle que $\lim\limits_{n\to+\infty} \frac{p(n)}{n} = 1$.

3.2 Mots de retour et mots lacunaires

Nous allons étudier une classe de mots qui est utile pour la construction d'autres mots ayant une faible complexité en facteurs. Nous allons, tout d'abord, introduire les mots de retour pour ensuite pouvoir définir les mots lacunaires.

Définition 3.16. Soit un mot $u \in \mathcal{A}^*$. Un mot $v \in \mathcal{A}^*$ est un *facteur interne* de u s'il existe $v' \in \mathcal{A}^+$ et $u' \in \mathcal{A}^+$ tels que $u = v'vu'$.

Définition 3.17. Soient un mot infini $u \in \mathcal{A}^{\mathbb{N}}$ et un facteur w de u. Un mot $v \in \mathcal{A}^+$ est un *mot de retour* de w dans u si les trois conditions suivantes sont vérifiées :

- vw est un facteur de u,
- w est un préfixe de vw, et
- w n'est pas un facteur interne de vw.

L'*ensemble des mots de retour* de w dans u est noté $\mathcal{R}_u(w)$.

En d'autres termes, le mot v est un mot de retour de w dans u si $vw \in L(u)$ et w est un préfixe et un suffixe de vw mais n'en est pas un facteur interne.

Une reformulation de cette définition est donnée par F. DURAND [7] :

Définition 3.18. Le mot $v \in \mathcal{A}^+$ est un mot de retour de w dans u si $v = u_i \cdots u_{j-1}$, où i et j sont deux positions d'occurrences successives de w dans u.

Exemple 3.19. Illustrons cette définition de mots de retour sur un préfixe du mot infini de THUE-MORSE

$$t = abbabaabbaababbabaababbaabbabaab \cdots$$

Étudions les mots de retour du facteur ab. Si l'on souligne les occurrences de ab et si l'on représente par un arc les mots de retour correspondant, on a

$$t = \overparen{\underline{ab}b}\,\overparen{\underline{ab}a}\,\overparen{\underline{ab}ba}\,\overparen{\underline{ab}}\,\overparen{\underline{ab}b}\,\overparen{\underline{ab}a}\,\overparen{\underline{ab}}\,\overparen{\underline{ab}ba}\,\overparen{\underline{ab}b}\,\overparen{\underline{ab}a}\,\underline{ab}\cdots$$

Définition 3.20. Un ensemble de mots $E = \{e_1, \ldots, e_k\}$ est un code si tout élément de E^* a une unique décomposition dans E. Formellement, l'ensemble E est un code si, pour tous mots $x, y \in E^*$ tels que $x = x_1 \cdots x_n$, $y = y_1 \cdots y_m$ et $x = y$, on a $m = n$ et $x_i = y_i$ pour tout $i \in \{1, \ldots, n\}$.

Par la Définition 3.18, nous pouvons aisément voir que l'ensemble $\mathcal{R}_u(w)$ est toujours un code. En effet, l'unique découpage d'un mot $x \in \mathcal{R}_u(w)^*$ doit être fait en coupant avant chaque occurrence de w.

Remarque 3.21. Les mots de retour sont toujours primitifs.
En effet, procédons par l'absurde. Supposons qu'il existe $z \in \mathcal{A}^+$ et $k \in \mathbb{N}$, $k \geq 2$, tels que $v = z^k$ soit un mot de retour d'un facteur w dans mot infini u. Par la Définition 3.18, on a $v = z^k = u_i \cdots u_{j-1}$ où i et j sont des occurrences successives de w. Or, le facteur w apparaît à la position $i + |z| < j$ étant donné que $k \geq 2$. Ceci est absurde par définition du mot de retour v.

Remarque 3.22. Dans le graphe de Rauzy G_n d'ordre $n = |w|$, les mots de retour de w dans u sont les labels de certains chemins de w à lui-même.
En effet, un mot de retour v de w est de la forme $u_i \cdots u_{j-1}$ où i et j sont des occurrences successives de w dans u et vw est un facteur de u. Ainsi, le chemin correspondant est la concaténation des arcs $(u_k \cdots u_{k+n-1},\ u_k\ , u_{k+1} \cdots u_{k+n})$ pour $k \in \{i, \ldots, j-1\}$. Par la Remarque 3.4, remarquons que la réciproque n'est pas nécessairement vraie.

Définition 3.23. Un mot infini $u \in \mathcal{A}^{\mathbb{N}}$ est dit *lacunaire* s'il existe une lettre $0 \in \mathcal{A}$ telle que pour tout $n \in \mathbb{N}$, $0^n \in L(u)$ et $\#(\mathcal{R}_u(0^n)) = 2$.

Notons que pour tout $a \in \mathcal{A} = \text{alph}(u)$, on a

- $a\varepsilon = a$ est un facteur de u,
- ε est un préfixe de $a\varepsilon = a = \varepsilon a$, et
- ε n'est pas interne à $a\varepsilon = a$.

Ainsi, on obtient $a \in \mathcal{R}_u(\varepsilon)$. De plus, pour tout mot v tel que $|v| \geq 2$, le mot vide est interne à v. En effet, par exemple pour un mot de longueur 2 noté $v = v_0 v_1$, on a en particulier $v = v_0 \varepsilon v_1$. Dès lors, on obtient $\mathcal{R}_u(\varepsilon) = \mathcal{A}$. Ainsi, en appliquant la définition avec $n = 0$, on a $\#(\mathcal{R}_u(0^n)) = \#(\mathcal{R}_u(\varepsilon)) = \#(\mathcal{A}) = 2$ et cela contraint donc un mot lacunaire à être binaire. Dès à présent, on considèrera donc l'alphabet $\mathcal{A} = \{0, 1\}$.

Exemple 3.24. Soit une suite strictement croissante $(e_k)_{k \in \mathbb{N}}$ de nombres naturels. Le mot $u = 10^{e_0}10^{e_1}10^{e_2}1 \cdots$ est lacunaire.
En effet, montrons que pour tout $n \in \mathbb{N}$ on a $\mathcal{R}_u(0^n) = \{0, 0^n 1\}$. Puisque l'alphabet est binaire, le cas $n = 0$ est vérifié par la remarque faite ci-avant. Ensuite, considérons le mot 0^n et remarquons tout d'abord que l'écriture $\mathcal{R}_u(0^n)$ a un sens étant donné que, pour tout $n \in \mathbb{N}$, le mot 0^n est un facteur de u car il existe $k' \in \mathbb{N}$ tel que $e_{k'} \geq n$ par définition de cette suite. Étudions dès à présent ses mots de retour. Le mot 0 est un mot de retour de 0^n dans u car

- $00^n = 0^{n+1}$ est un facteur de u puisque, de la même façon, il existe $k'' \in \mathbb{N}$ tel que $e_{k''} \geq n + 1$,
- 0^n est un préfixe de 0^{n+1} et
- 0^n n'est pas un facteur interne de 0^{n+1}.

Le mot $0^n 1$ est un mot de retour de 0^n dans u car

- $0^n 10^n$ est un facteur de u car si l'on considère $k' \in \mathbb{N}$ défini comme précédemment, le mot $0^n 10^n$ est un facteur de $0^{e_{k'}}10^{e_{k'+1}} \in L(u)$ car $e_{k'+1} > e_{k'} \geq n$,
- 0^n est un préfixe de $0^n 10^n$ et
- 0^n n'est pas un facteur interne de $0^n 10^n$.

De plus, il n'existe pas d'autre mot $v \in \{0,1\}^+ \setminus \{0, 0^n 1\}$ tel que v est un mot de retour de 0^n dans u. En effet, remarquons tout d'abord que le mot v ne peut pas débuter par le lettre 1 car dans ce cas nous n'aurions pas que 0^n est un préfixe de $v0^n$. Ainsi, le mot v est de la forme

(i) $v = 0^{m_1}$ où $m_1 \geq 2$,

(ii) $v = 0^{m_2}1$ où $m_2 > n$, ou bien

(iii) $v = 0^{m_3}1v'$ où $m_3 \geq n$, $v' \neq \varepsilon$ et $0^{m_3}1v'0^n \in L(u)$.

Ces cas sont tous à rejeter car, dans chacun d'eux, le mot 0^n est un facteur interne de $v0^n$. Ainsi, on a $\mathcal{R}_u(0^n) = \{0, 0^n 1\}$ et $\#(\mathcal{R}_u(0^n)) = 2$. On en tire que le mot u est lacunaire.

L'exemple suivant, connu sous le nom du *contre-exemple universel* illustre le fait que les mots lacunaires peuvent également être récurrents.

Exemple 3.25. Considérons une suite strictement croissante $(e_k)_{k\in\mathbb{N}}$ de nombres naturels. Notons $(u_k)_{k\in\mathbb{N}}$ la suite de mots définie récursivement comme suit :

$$\begin{cases} u_0 = 1 \\ u_{k+1} = u_k 0^{e_k} u_k, \ \forall k \in \mathbb{N}\,. \end{cases}$$

En guise d'illustration, les premier éléments de la suite sont

$$u_0 = 1,\ u_1 = 10^{e_0}1,\ u_2 = 10^{e_0}10^{e_1}10^{e_0}1 \text{ et } u_3 = 10^{e_0}10^{e_1}10^{e_0}10^{e_2}10^{e_0}10^{e_1}10^{e_0}1.$$

La suite $(u_k)_{k\in\mathbb{N}}$ converge vers un mot infini $u = \lim\limits_{k\to\infty} u_k$. Le mot u est récurrent car chaque facteur d'un élément de la suite $(u_k)_{k\in\mathbb{N}}$ se répète deux fois dans l'élément suivant. De plus, le mot binaire u est lacunaire.
En effet, montrons que pour tout $n \in \mathbb{N}$, on a $\mathcal{R}_u(0^n) = \{0, 0^n u_j\}$ où

$$j = \min\{k \in \mathbb{N} \mid e_k \geq n\}.$$

Tout d'abord, l'appartenance de 0 à $\mathcal{R}_u(0^n)$ se justifie par le même raisonnement que dans l'Exemple 3.24. Ensuite, le mot $0^n u_j$ est un mot de retour de 0^n dans u puisque

- $0^n u_j 0^n$ est un facteur de u puisqu'on a $u_{j+2} = u_{j+1}0^{e_{j+1}}u_{j+1} = u_j \underbrace{0^{e_j}u_j0^{e_{j+1}}}_{\supseteq\, 0^n u_j 0^n} u_j 0^{e_j} u_j$,
- 0^n est un préfixe de $0^n u_j 0^n$, et
- 0^n n'est pas un facteur interne de $0^n u_j 0^n$ car, par construction, les premières et dernières lettres de u_j sont des 1 et par minimalité de j.

De plus, il n'existe pas d'autre mot $v \in \{0,1\}^+ \setminus \{0, 0^n u_j\}$ tel que v est un mot de retour de 0^n dans u. En effet, de la même façon que dans l'Exemple 3.24, il ne peut y avoir un mot de retour v commençant par 1 étant donné que 0^n doit être un préfixe de $v0^n$. Ainsi, considérons les mots v de la forme

(i) $v = 0^{m_1}$ où $m_1 \geq 2$,

(ii) $v = 0^{m_2}u_j$ où $m_2 > n$,

(iii) $v = 0^{m_3}u_i$ où $m_3 \geq n$ et[5] $i > j$, ou bien

(iv) $v = 0^{m_4}1v'$ où $m_4 \geq n$, $v' \neq \varepsilon$ et $0^{m_4}1v' \in L(u)$.

Or, chaque cas peut être étudié pour en venir à la conclusion que le mot 0^n sera un facteur interne de $v0^n$. Ainsi, pour tout entier naturel n, l'ensemble des mots de retour du mot 0^n est donné par $\mathcal{R}_u(0^n) = \{0, 0^n u_j\}$ et $\#(\mathcal{R}_u(0^n)) = 2$. Cela signifie, dès lors, que le mot u construit ci-avant est lacunaire.

Les mots lacunaires ont une complexité en facteurs pouvant être étudiée facilement grâce à l'outil des mots spéciaux. La fin de cette section consiste à établir les lemmes et propositions nécessaires à l'évaluation de la fonction complexité en facteurs de cette classe de mots.

Lemme 3.26. *Soient un mot infini $u \in \mathcal{A}^{\mathbb{N}}$ et deux facteurs x et y de u. Si x est un préfixe de y, alors $\mathcal{R}_u(y) \subseteq \mathcal{R}_u(x)^+$.*

Démonstration. Le mot x étant un préfixe de y, notons $y = xy'$ où $y' \in \mathcal{A}^*$. Considérons $v \in \mathcal{A}^+$ un mot de retour de y dans u. Par définition, le mot $vy = vxy'$ est un facteur de u, $y = xy'$ est un préfixe de $vy = vxy'$ et n'en est pas un facteur interne. Ainsi, le mot vx est un facteur de u et le mot x est un préfixe de vx. Cependant, le mot x pourrait être interne à vx sans pour autant que y ne le soit pour le facteur[6] vy. Dès lors, étant donné les constructions de ces mots, un découpage de v avec le code $\mathcal{R}_u(x)$ est possible, unique et contient au moins un élément. Ainsi, on a $v \in \mathcal{R}_u(x)^+$. Cette découpe doit être faite avant chaque occurrence de x dans v.

□

Lemme 3.27. *Soit $u \in \{0,1\}^{\mathbb{N}}$ un mot lacunaire. Il existe une suite finie de mots $(z_n)_{n\in\mathbb{N}}$ telle que*

$$\begin{cases} z_0 = 1 \\ z_{n+1} \in z_n(0^n z_n)^*, \ \forall n \in \mathbb{N} \end{cases}$$

et $\mathcal{R}_u(0^n) = \{0, 0^n z_n\}$ pour tout $n \in \mathbb{N}$.

5. Le cas $i < j$ est impossible car $0^n u_i \notin L(u)$ dans ce cas.

6. En effet, un contre-exemple est donné par le mot de retour $v = abaaa$ de $y = ab$, en supposant que $abaaaab \in L(u)$ mais en considérant le préfixe $x = a$ de y.

Démonstration. Par définition d'un mot lacunaire, pour tout $n \in \mathbb{N}$, l'ensemble $\mathcal{R}_u(0^n)$ contient deux éléments et le mot 0^{n+1} est un facteur de u. Dès lors, un des deux mots de retour de 0^n doit obligatoirement être 0. Étudions le second mot de retour v de u. Remarquons tout d'abord que, par la Remarque 3.21, un mot de retour est primitif. Ainsi, il ne peut pas être de la forme 0^m avec $m \geq 2$. Or, vu que le mot 0^n doit être un préfixe de $v0^n$, le mot de retour v doit être de la forme $0^n z_n$ où $z_n \in \mathcal{A}^+$ et z_n débute par un 1. Nous remarquons que $z_0 = 1$ convient car $\mathcal{R}_u(\varepsilon) = \{0, 1\}$. Ensuite, étant donné que 0^n est un préfixe de 0^{n+1}, par le Lemme 3.26, on a $\mathcal{R}_u(0^{n+1}) \subseteq \mathcal{R}_u(0^n)^+$. En particulier, on a $0^{n+1}z_{n+1} \in \{0, 0^n z_n\}^+$. Or, le mot $0^{n+1}z_{n+1}0^{n+1}$ ne peut pas contenir de facteur interne 0^{n+1}, donc $0^{n+1}z_{n+1}$ ne peut pas contenir le mot $0(0^n z_n)$ en facteur interne et ne peut pas se terminer par 0. Ainsi, l'unique possibilité est que l'on ait $0^{n+1}z_{n+1} \in 0(0^n z_n)^+$, c'est-à-dire $z_{n+1} \in z_n(0^n z_n)^*$.

□

Proposition 3.28. *Soit $u \in \{0, 1\}^{\mathbb{N}}$ un mot lacunaire. Pour tout $n \in \mathbb{N}$, le mot u a au plus $n + 1$ facteurs spéciaux à droite de longueur n et donc*

$$p_n(n) \leq \frac{n^2 + n + 2}{2}.$$

Démonstration. En conservant les notations du Lemme 3.27, notons

$$S = \{\mathrm{Suff}_n(z_k 0^k)^m | m \geq 1 \text{ minimal tel que } |(z_k 0^k)^m| \geq n,\ k = 0, \ldots, n\}.$$

Démontrons que l'on a

$$RS_n(u) \subseteq S. \tag{3.1}$$

Soit w un facteur spécial à droite de u. Vu que le mot u est binaire, le facteur w a deux extensions à droite dans u. Dès lors, les mots $w0$ et $w1$ sont des facteurs de u. Considérons $k \in \mathbb{N}$ le plus grand entier tel que 0^k apparaît dans w et notons $w = w'0^k w''$ avec $|w''|$ minimal.

L'ensemble $\mathcal{R}_u(0^k)$ est un code, et au vu du Lemme 3.27 on a $\mathcal{R}_u(0^k) = \{0, 0^k z_k\}$. Ainsi, toute occurrence de 0^k est soit suivie par 0 soit par $z_k 0^k$. Cette construction permet de définir un autre code, de $\mathcal{R}_u(0^k)^*$, via les facteurs $\{0, z_k 0^k\}$ en utilisant un découpage après, et non avant, chaque occurrence de 0^k. Or, les mots $w'0^k w''0$ et $w'0^k w''1$ sont des facteurs de u. Ainsi, en notant $u^{(0)}, u^{(1)} \in \{0, 1\}^{\mathbb{N}}$ et $u'^{(0)}, u'^{(1)} \in \{0, 1\}^*$ tels que

$$u'^{(0)} w' 0^k w'' 0 u^{(0)} = u = u'^{(1)} w' 0^k w'' 1 u^{(1)},$$

on a $w'0^k w''0u^{(0)}$, $w'0^k w''1u^{(1)} \in \mathcal{R}_u(0^k)^{\mathbb{N}}$. Il résulte que les mots $w''0$ et $w''1$ sont des préfixes de mots dans $\{0, z_k 0^k\}^*$.

Supposons que le mot w'' soit non vide. Par maximalité de k, la première lettre de w'' ne peut pas être la lettre 0, sinon nous aurions $0^{k+1} \in L(w)$. De plus, par minimalité de $|w''|$, $z_k 0^k$ ne peut pas non plus être un préfixe de w''. Or, le facteur $w''1$ ne peut pas être un

préfixe de 0 et étant donné que $w'' \neq \varepsilon$, le facteur $w''0$ ne peut pas non plus être un préfixe de 0. Les deux facteurs $w''0$ et $w''1$ doivent donc être des préfixes de z_k0^k. Ceci est absurde vu que ces facteurs sont différents mais de même longueur. Dès lors, il résulte que le mot w'' est le mot vide et $w = w'0^k$.
Comme le facteur w est spécial à droite dans u, les mots $w0$ et $w1$ sont des facteurs de u et donc au moins deux occurrences de w sont présentes dans u. Ainsi, il existe une occurrence du mot 0^k commençant avant la deuxième occurrence de w dans u. Dans le pire cas, celle-ci correspond au facteur 0^k de la première occurrence de $w = w'0^k$. Le facteur de u débutant après cette occurrence de 0^k et se terminant par le deuxième facteur w est un facteur écrit sur le code $\{0, z_k0^k\}$, c'est-à-dire le facteur w est un suffixe d'un mot de $\{0, z_k0^k\}^*$. Or, par maximalité de l'entier naturel k, le mot 0^{k+1} n'est pas un facteur de w, ainsi le mot w est plus particulièrement un suffixe d'un mot $(z_k0^k)^m$, où $m \geq 1$ est le plus petit entier satisfaisant à cette propriété. Ceci termine la démonstration de l'inclusion d'ensemble (3.1).
Au vu du Corollaire 2.22, on en tire $s(n) \leq \#(S)$. Or, il y a autant d'éléments dans S que de différents k, ainsi $s(n) \leq n + 1$. Ainsi, par la Proposition 2.20, on a

$$\begin{aligned} p(n) &= 1 + \sum_{l=0}^{n-1} s(l) \leq 1 + \sum_{l=0}^{n-1}(l+1) \\ &= 1 + \frac{(n-1)n}{2} + n = \frac{n^2+n+2}{2}. \end{aligned}$$

□

Cette proposition nous permet de donner une borne supérieure pour la fonction de complexité en facteurs d'un mot lacunaire. Cependant, pour certains d'entre eux, leurs fonctions de complexité peuvent être étudiées plus précisément. L'exemple suivant illustre ce cas.

Exemple 3.29. Au vu de l'Exemple 3.24, si l'on considère $(e_k)_{k\in\mathbb{N}}$ une suite strictement croissante de nombre naturels, le mot $u = 10^{e_0}10^{e_1}10^{e_2}1\cdots$ est lacunaire. Démontrons que les facteurs spéciaux à droite dans le mot u sont les suffixes des facteurs de la forme $0^{e_{k-1}}10^{e_k}$, pour tout $k \in \mathbb{N}$, en posant $e_{-1} = 0$.
En effet, étant donné que la suite $(e_k)_{k\in\mathbb{N}}$ est strictement croissante, un mot de la forme 0^n est toujours spécial à droite car il existe $k' \in \mathbb{N}$ tel que $e_{k'} \geq n$ et ainsi 0^n1 est un facteur de u et 0^{n+1} est un facteur de u car $e_{k'+1} \geq n+1$. Les facteurs 0^n sont, en particulier, des suffixes de mots de la forme $0^{e_{k-1}}10^{e_k}$. Ensuite, notons qu'un mot contenant deux 1 n'a qu'une seule occurrence et ne peut pas être spécial à droite. Ainsi, un facteur spécial à droite doit contenir au plus un 1 et dans ce cas il s'écrit sous la forme 0^m10^n. Or, pour être spécial à droite, le facteur doit pouvoir être suivit d'un 1, ainsi $n = e_k$ pour un $k \in \mathbb{N}$ et $m \leq e_{k-1}$.
Le nombre de facteurs spéciaux à droite de longueur $n \in \mathbb{N}$ est donc égal au nombre de suffixes de longueur n de $0^{e_{k-1}}10^{e_k}$ pour tout $k \in \mathbb{N}$. Or, si $n \leq e_k$ alors $\mathrm{Suff}_n(0^{e_{k-1}}10^{e_k}) = 0^n$ et si $n \geq e_k + e_{k-1} + 2 = 1 + |0^{e_{k-1}}10^{e_k}|$, alors il n'existe aucun suffixe de longueur n de $0^{e_{k-1}}10^{e_k}$. Ainsi, pour tout $n \in \mathbb{N}$, les seuls facteurs spéciaux à droite de longueur n du mot u sont : le mot 0^n et les suffixes, tous différents, de longueur

n de $0^{e_{k-1}}10^{e_k}$, pour les entiers $k \in \mathbb{N}$ tels que $e_k + 1 \leq n \leq e_k + e_{k-1} + 1$. Formellement, on en tire

$$s(n) = 1 + \#(\{k \in \mathbb{N} \mid e_k \leq n - 1 \leq e_k + e_{k-1}\}).$$

Dans certains cas, le cardinal de cet ensemble peut être étudié plus précisément. Étudions le cas où $e_k \sim Ck^r$ pour un $C \in \mathbb{R}_{>0}$ et $r \in \mathbb{N}_0$. En considérant $\epsilon = \frac{1}{4}$ dans la Définition 1.16, nous avons

$$\exists k_0, \forall k \geq k_0, |e_k - Ck^r| < \frac{1}{4}|Ck^r|.$$

Ainsi, pour tout $k \geq k_0$, on a les inégalités

$$\left(\frac{3C}{4}\right) k^r < e_k < \left(\frac{5C}{4}\right) k^r. \tag{3.2}$$

Rappelons que nous voulons satisfaire aux inégalités

$$e_k \leq n - 1 \leq e_k + e_{k-1}. \tag{3.3}$$

Tout d'abord, notons que pour satisfaire simultanément aux inégalités (3.2) et (3.3), il faut que l'on ait $\left(\frac{3C}{4}\right) k^r < n - 1$, c'est-à-dire

$$k < \left(\frac{n-1}{\frac{3C}{4}}\right)^{\frac{1}{r}}.$$

Ainsi, on a [7] $s(n) = \mathcal{O}(\sqrt[r]{n})$. Ensuite, vérifions que tout entier naturel

$$k \in \left[\!\left[\left(\frac{n-1}{\frac{3C}{2}}\right)^{\frac{1}{r}}, \left(\frac{n-1}{\frac{5C}{4}}\right)^{\frac{1}{r}} \right]\!\right],$$

satisfait les inégalités (3.3). En effet, on a

$$\begin{aligned} &\frac{n-1}{\frac{3C}{2}} \leq k^r \leq \frac{n-1}{\frac{5C}{4}} \\ \Leftrightarrow\ &\frac{5C}{4} k^r \leq n - 1 \leq \frac{3C}{2} k^r. \end{aligned}$$

Or, on a $e_k < \frac{5C}{4} k^r$, $e_k \sim Ck^r$ et $e_{k-1} \sim C(k-1)^r$. En particulier, remarquons que l'on a $e_{k-1} \sim Ck^r$ et on obtient $\frac{3C}{2} k^r < 2Ck^r \sim e_k + e_{k-1}$. Les inégalités (3.3) sont donc vérifiées. Ainsi, vu que

$$\#\left(\left[\!\left[\left(\frac{n-1}{\frac{3C}{2}}\right)^{\frac{1}{r}}, \left(\frac{n-1}{\frac{5C}{4}}\right)^{\frac{1}{r}} \right]\!\right]\right) = \left(\left(\frac{4}{5}\right)^{\frac{1}{r}} - \left(\frac{2}{3}\right)^{\frac{1}{r}}\right) \left(\frac{n-1}{C}\right)^{\frac{1}{r}},$$

7. Voir les Définitions 5.1 pour les définitions formelles des notations asymptotiques de Landau $\mathcal{O}$, Ω et Θ.

on a $s(n) = \Omega(\sqrt[r]{n})$ et donc $s(n) = \Theta(\sqrt[r]{n})$. Or, par la Proposition 2.20, on a

$$p_v(n) = 1 + \sum_{l=0}^{n-1} s_v(l)$$

et donc le résultat $p_v(n) = \Theta(n\sqrt[r]{n})$ est vérifié. En effet, démontrons que $\sum\limits_{i=1}^{n-1} i^{\frac{1}{r}} \sim n\sqrt[r]{n}$ et le résultat asymptotique sera vérifié. Pour tout $i \in \{1, \ldots, n\}$, on a

$$\int_{i-1}^{i} t^{\frac{1}{r}} \; dt \leq i^{\frac{1}{r}} \leq \int_{i}^{i+1} t^{\frac{1}{r}} \; dt$$

donc

$$\int_{0}^{n-1} t^{\frac{1}{r}} \; dt \leq \sum_{i=1}^{n-1} i^{\frac{1}{r}} \leq \int_{1}^{n} t^{\frac{1}{r}} \; dt.$$

Étant donné que pour tous réels positifs a et b, on a $\int_a^b t^{\frac{1}{r}} = \left[\frac{t^{1+\frac{1}{r}}}{1+\frac{1}{r}}\right]_a^b$, le résultat voulu est démontré.

Chapitre 4

Morphismes

Ce chapitre définit la notion de morphisme, très importante en combinatoire des mots. Cette définition est nécessaire pour le théorème de Pansiot, qui sera étudié par la suite dans ce mémoire. Nous allons, tout d'abord, étudier quelques exemples très connus de mots construits par morphisme. Ensuite, nous allons démontrer quelques résultats, en lien avec la fonction de complexité en facteurs, obtenus lorsqu'on applique un morphisme à un mot infini $u \in \mathcal{A}^{\mathbb{N}}$.

4.1 Définitions générales

Définition 4.1. Soient $\mathcal{A}$ et $\mathcal{B}$ deux alphabets. Un *morphisme* est une application

$$\sigma : \mathcal{A}^* \to \mathcal{B}^*$$

telle que, pour tout $u, v \in \mathcal{A}^*$,

$$\sigma(uv) = \sigma(u)\sigma(v).$$

Un *endomorphisme* est un morphisme de $\mathcal{A}^*$ dans lui-même.

En particulier, pour tout morphisme σ, étant donné que l'on a

$$\sigma(\varepsilon) = \sigma(\varepsilon\varepsilon) = \sigma(\varepsilon)\sigma(\varepsilon),$$

on obtient $\sigma(\varepsilon) = \varepsilon$.

Remarque 4.2. Pour définir complètement un morphisme $\sigma : \mathcal{A}^* \to \mathcal{B}^*$, il suffit de définir l'image des lettres de $\mathcal{A}$ par σ.
En effet, par définition du morphisme, pour connaitre l'image d'un mot de $\mathcal{A}^*$, il suffit de concaténer les images des lettres qui le composent. Formellement, si l'on a un mot $u = u_0u_1 \cdots u_{|u|-1} \in \mathcal{A}^*$, alors l'image de ce mot est donnée par

$$\sigma(u) = \sigma(u_0)\sigma(u_1) \cdots \sigma(u_{|u|-1}).$$

Exemple 4.3. Considérons le morphisme $\sigma : \{a, b\} \to \{a, b, c\}$ défini tel que $\sigma(a) = bbc$, $\sigma(b) = a$ et $\sigma(c) = abc$. On a $\sigma(aabca) = bbcbbcaabcbbc$.

Définition 4.4. Soit un entier naturel $k \in \mathbb{N}$. Un morphisme $\sigma : \mathcal{A}^* \to \mathcal{B}^*$ est *k-uniforme* si pour toute lettre $a \in \mathcal{A}$, on a $|\sigma(a)| = k$. Un morphisme 1-uniforme est appelé un *code* ou *codage*. Dans ce cas, on le note généralement $\tau : \mathcal{A} \to \mathcal{B}$.

Définition 4.5. Soit un morphisme $\sigma : \mathcal{A}^* \to \mathcal{B}^*$. S'il existe une lettre $a \in \mathcal{A}$ telle que $\sigma(a) = \varepsilon$, le morphisme σ est dit *effaçant*. Dans le cas contraire, pour toute lettre $a \in \mathcal{A}$, on a $|\sigma(a)| > 0$ et le morphisme σ est dit *non effaçant* ou *positif*.

Si $\sigma : \mathcal{A}^* \to \mathcal{B}^*$ est un morphisme non effaçant, il peut être étendu en une application de l'ensemble $\mathcal{A}^{\mathbb{N}}$ dans l'ensemble $\mathcal{B}^{\mathbb{N}}$. En effet, si $x = x_0x_1 \cdots$ est un mot infini écrit sur l'alphabet $\mathcal{A}$ alors la suite de mots $(\sigma(x_0 \cdots x_{n-1}))_{n \geq 0}$ converge vers un mot de $\mathcal{B}^{\mathbb{N}}$ que l'on note $\sigma(x) = \sigma(x_0)\sigma(x_1) \cdots$. Cependant, par simplicité, nous définissons les morphismes sur $\mathcal{A}^*$, mais nous pouvons implicitement considérer leurs actions sur les mots de $\mathcal{A}^{\mathbb{N}}$.

Remarquons que si le morphisme σ est effaçant, une définition similaire peut être donnée mais l'image d'un mot infini peut être finie. Par exemple, considérons un morphisme effaçant $\sigma : \mathcal{A}^* \to \mathcal{B}^*$ défini tel que $\sigma(a) = \varepsilon$ pour une lettre $a \in \mathcal{A}$. Alors l'image par σ du mot infini a^ω est le mot vide ε.

Définition 4.6. Soit un endomorphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$. Un mot fini ou infini x est un *point fixe de* σ si $\sigma(x) = x$.

Définitions 4.7. Soit un endomorphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$. S'il existe $a \in \mathcal{A}$ et $u \in \mathcal{A}^+$ tels que $\sigma(a) = au$ et $\lim_{n \to +\infty} |\sigma^n(a)| = +\infty$, alors σ est dit *prolongeable (à droite) en a*.
Soit $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ un morphisme prolongeable en $a \in \mathcal{A}$. On a

$$\begin{aligned} \sigma(a) &= au \\ \sigma^2(a) &= au\sigma(u) \\ \sigma^3(a) &= au\sigma(u)\sigma^2(u) \\ &\vdots \end{aligned}$$

Étant donné que pour tout $n \in \mathbb{N}$, le mot $\sigma^n(a)$ est un préfixe de $\sigma^{n+1}(a)$ et que $|\sigma^n(a)|$ tend vers l'infini, la sequence $(\sigma^n(a))_{n \geq 0}$ converge vers le mot infini

$$\sigma^\omega(a) := \lim_{n \to +\infty} \sigma^n(a) = au\sigma(u)\sigma^2(u)\sigma^3(u) \cdots .$$

Ce mot infini est un point fixe de σ. Un mot infini obtenu d'une telle façon est dit *engendré par* σ ou, de manière plus générale, *purement morphique*. Soient $x \in \mathcal{A}^{\mathbb{N}}$ un mot purement morphique et un code $\tau : \mathcal{A} \to \mathcal{B}$, le mot $y = \tau(x) \in \mathcal{B}^{\mathbb{N}}$ est dit *morphique*.

Exemple 4.8. Soient un alphabet $\mathcal{A} = \{a, b, c\}$ et un endomorphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ défini par

$$\sigma(a) = acb, \sigma(b) = bb, \sigma(c) = b.$$

Le morphisme σ est prolongeable en a, et on a

$$\sigma^\omega(a) = acb\sigma(cb)\sigma^2(cb)\sigma^3(cb)\cdots = acb^\omega$$

Ainsi, le mot acb^ω est purement morphique. En utilisant le codage $\tau : \{a,b,c\} \to \{a,b\}$ défini par $\tau(a) = \tau(c) = a$ et $\tau(b) = b$, il résulte que le mot aab^ω est morphique.

Remarque 4.9. Soient un morphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ prolongeable en une lettre $a \in \mathcal{A}$ et le mot purement morphique $u = \sigma^\omega(a)$. Pour tout $n \in \mathbb{N}$, la puissance σ^n du mophisme σ est également prolongeable en a et engendre le même mot infini u.

Remarque 4.10. Comme nous l'avons déjà signalé, $\sigma^\omega(a)$ est nécessairement un point fixe de σ. Cependant, la réciproque n'est pas vérifiée.
En effet, un contre-exemple est donné par le mot $u = (ab)^\omega$. Il est un point fixe du morphisme $\sigma : \{a,b\}^* \to \{a,b\}^*$ défini par $\sigma(a) = \varepsilon$ et $\sigma(b) = ab$ qui n'est cependant pas prolongeable en a.

Remarque 4.11. La construction d'un mot purement morphique peut se voir comme étant l'avancée d'une tête de lecture et d'une tête d'écriture.
En effet, le mot $\sigma^\omega(a) = au\sigma(u)\sigma^2(u)\sigma^3(u)\cdots$ peut être construit de la manière suivante. Nous débutons l'écriture en ajoutant la lettre a comme première lettre du mot considéré et initialisons les têtes de lecture et d'écriture en dessous de cette occurrence de a. Dès lors, la construction de $\sigma^\omega(a)$ consiste à lire l'élément au-dessus de la tête de lecture, écrire, en avançant la tête d'écriture, son image par σ à partir la position de la tête d'écriture puis d'avancer les deux têtes considérées d'une position vers la droite. Cette construction provient du fait que, si l'on a $\sigma(a) = au$, alors pour tout $n \geq 1$,

$$\sigma^n(a) = \sigma^{n-1}(a)x_{n-1}$$

où

$$\begin{cases} x_0 = u \\ x_n = \sigma(x_{n-1}),\ \forall n \geq 1. \end{cases}$$

En effet, le cas de base $n = 1$ est trivialement vérifié car $\sigma^0 = id$ et par induction, on a

$$\sigma^{n+1}(a) = \sigma(\sigma^n(a)) = \sigma(\sigma^{n-1}(a)x_{n-1}) = \sigma^n(a)\sigma(x_{n-1}) = \sigma^n(a)x_n.$$

Définition 4.12. Pour tout morphisme $\sigma : \mathcal{A}^* \to \mathcal{B}^*$, on définit la *largeur de* σ par

$$\|\sigma\| := \max_{a\in\mathcal{A}} |\sigma(a)|.$$

En particulier, pour tout mot $w \in \mathcal{A}^*$, on a $|\,\sigma(w)| \leq \|\sigma\|\,|w|$.

Exemples 4.13. Nous allons maintenant donner quelques exemples classiques de mots morphiques et purement morphiques.

(i) **Mot de Thue-Morse**
Soit le morphisme 2-uniforme $\sigma : \{a,b\}^* \to \{a,b\}^*$ défini par $\sigma(a) = ab$ et $\sigma(b) = ba$. Les premières itérations de ce morphisme sont :

$$\begin{aligned} \sigma(a) &= ab \\ \sigma^2(a) &= abba \\ \sigma^3(a) &= abbabaab \\ &\vdots \end{aligned}$$

Or, cet endomorphisme est bien prolongeable en a, ainsi la notation $\sigma^\omega(a)$ a du sens, et on obtient

$$\begin{aligned} t := \sigma^\omega(a) &= ab\sigma(b)\sigma^2(b)\sigma^3(b)\cdots \\ &= abba\sigma(ba)\sigma^2(ba)\cdots \\ &= abbabaabbaababba\cdots. \end{aligned}$$

(ii) **Mot de Fibonacci**
Un autre exemple incontournable de mot purement morphique est celui du mot de Fibonacci. Définissons le morphisme $\sigma : \{a,b\}^* \to \{a,b\}^*$ par $\sigma(a) = ab$ et $\sigma(b) = a$. De la même façon, les premières itérations du morphisme sont :

$$\begin{aligned} \sigma(a) &= ab \\ \sigma^2(a) &= aba \\ \sigma^3(a) &= abaab \\ &\vdots \end{aligned}$$

et
$$f := \sigma^\omega(a) = abaababaabaababaababaabaab\cdots.$$

(iii) **La suite des carrés**
Soient l'alphabet $\mathcal{A} = \{a,b,c\}$ et le morphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ défini par $\sigma(a) = abcc$, $\sigma(b) = bcc$ et $\sigma(c) = c$. On obtient

$$\begin{aligned} \sigma(a) &= abcc \\ \sigma^2(a) &= abccbcccc \\ \sigma^3(a) &= abccbccccbcccccc \\ &\vdots \end{aligned}$$

et le mot purement morphique correspondant est

$$\sigma^\omega(a) = abccbccccbccccccbccccccccbccccccccccbc\cdots.$$

Dans le mot $\sigma^\omega(a)$, la distance entre la $n^{\text{ième}}$ et la $(n+1)^{\text{ième}}$ occurrence de la lettre b est égale à $2n+1$.

En effet, montrons qu'il y a $2n$ lettres c entre les $n^{\text{ième}}$ et $(n+1)^{\text{ième}}$ b. Par définition du morphisme, le $n^{\text{ième}}$ b produit deux lettres c entre les $(n+1)^{\text{ième}}$ et $(n+2)^{\text{ième}}$ occurrences de la lettre b. De plus, chaque lettre c présente entre les $n^{\text{ième}}$ et $(n+1)^{\text{ième}}$ b produit un c entre $(n+1)^{\text{ième}}$ et $(n+2)^{\text{ième}}$ occurrences de b. Dès lors, en connaissant ce pattern, et étant donné que le cas de base est trivialement vérifié, démontrons cette propriété par récurrence. Par induction, entre le $(n+1)^{\text{ième}}$ et le $(n+2)^{\text{ième}}$ b, il y a deux c produits par le $n^{\text{ième}}$ b et, par hypothèse de récurrence, les $2n+1$ lettres c présentes entre les $n^{\text{ième}}$ et $(n+1)^{\text{ième}}$ b vont chacune en produire une. Ainsi, il y a $2n+1+2 = 2(n+1)+1$ lettres c entre le $(n+1)^{\text{ième}}$ et le $(n+2)^{\text{ième}}$ b.
Dès lors, définissons le codage $\tau : \{a,b,c\} \to \{0,1\}$ tel que $\tau(a) = \tau(b) = 1$ et $\tau(c) = 0$. On a

$$\tau(\sigma^\omega(a)) = 1100100001000000100000000100000000010\cdots .$$

De plus, si l'on définit la suite caractéristique d'un ensemble $E \subseteq \mathbb{N}$ comme étant la suite $(x_n)_{n\geq 0}$ où

$$x_n = \begin{cases} 1 & \text{si } n \in E, \\ 0 & \text{sinon} \end{cases} .$$

Étant donné que l'on a

$$(n+1)^2 - n^2 = 2n+1,$$

nous venons de prouver que la suite caractéristique de l'ensemble des carrés est morphique.

(iv) **La suite des puissances de** 2
Soient le morphisme 2-uniforme défini sur l'alphabet $\{a,b,c\}$ par $\sigma : a \mapsto ab$, $b \mapsto bc, c \mapsto cc$ et le codage $\tau : a, c \mapsto 0, b \mapsto 1$. Les mots purement morphique et morphique correspondants sont respectivement :

$$\sigma^\omega(a) = abbcbcccbcccccccbccccccccccccccccbcc\cdots ,$$
$$\tau(\sigma^\omega(a)) = 011010001000000010000000000000000100\cdots .$$

Par un raisonnement analogue à celui établi dans l'exemple (iii), étant donné que le nombre de c se double à chaque itération du morphisme σ, la distance entre les occurrences du $n^{\text{ième}}$ et $(n+1)^{\text{ième}}$ b est égale à la différence $2^{n+1} - 2^n = 2^n$. On conclut que la suite caractéristique de l'ensemble des puissances de 2 est morphique.

Remarques 4.14.

- Tout mot infini périodique z^ω est purement morphique :
On a $z \in \mathcal{A}^+$ et notons $z = z_0 \cdots z_{n-1}$, $n \in \mathbb{N}_0$. Ainsi, z^ω est généré par l'endomorphisme σ de $\mathcal{A}^*$ défini tel que, pour toute lettre $a \in \mathcal{A}$, $\sigma(a) = z^2$. En effet, cet endomorphisme convient car $\sigma(z_0) = z^2 = z_0 z_1 \cdots z_{n-1} z$ et donc σ est prolongeable en z_0 ainsi

$$\sigma^\omega(z_0) = z_0 z_1 \cdots z_{n-1} z \underbrace{\sigma(z_1 \cdots z_{n-1} z)}_{=z^{2(n-1)} z^{2n}} \sigma^2(z_1 \cdots z_{n-1} z) \cdots = z^\omega .$$

Notons que le morphisme qui envoie toutes les lettres sur le mot z ne convient pas en toute généralité car, si $|z| = 1$, alors nous ne pourrions pas dire que σ est prolongeable en z_0.

- Tout mot infini ultimement périodique $u = tz^\omega$ est morphique :
En effet, u est l'image de 01^ω par le morphisme qui envoie 0 et 1 respectivement sur t et z. De plus, le mot 01^ω est purement morphique puisqu'il est engendré par l'endomorphisme $\sigma : \{0,1\} \to \{0,1\}$ tel que $\sigma(0) = 01$ et $\sigma(1) = 1$.

Exemple 4.15. Par la remarque précédente, le mot 001^ω est morphique. Cependant il n'est pas purement morphique. En effet, il serait impossible que ce mot soit purement morphique mais que la lettre 0 n'apparaisse pas une infinité de fois. Procédons par l'absurde. Considérons un morphisme σ prolongeable en 0 tel que $\sigma^\omega(0) = 001^\omega$. Le morphisme est prolongeable en 0 dès lors $\sigma(0) = 0u$ où $u \in \{0,1\}^+$. Pour que l'on ait $\sigma^\omega(0) = 001^\omega$, le mot u doit commencer par 0. Cependant, le morphisme σ défini tel que $\sigma(0) = 00$ ne convient pas étant donné que dans ce cas $\sigma^\omega(0) = 0^\omega$. De plus, un morphisme σ défini tel que $\sigma(0) = 0^n u'$ où $n \geq 3$ ne convient pas sinon le mot $\sigma^\omega(0)$ serait de la forme $0^n u' \sigma\, 0\sigma^2(0) \cdots$ avec autant de zéros de tête. Dès lors, il existe un mot $v \in \{0,1\}^*$ tel que $\sigma(0) = 001v$ et dans ce cas, le mot purement morphique construit est

$$\sigma^\omega(0) = 001v001v\sigma(1)\sigma(v)001v001v\sigma(1)\sigma(v)\sigma^2(1)\sigma^2(v)\cdots .$$

Ce mot n'est pas égal à 001^ω car il contient une infinité de 0.

Le théorème suivant dû à Cobham est un résultat fort. Il stipule que tout mot morphique peut être écrit comme étant un codage appliqué à un mot purement morphique où le morphisme en question est non effaçant. Ce théorème ne sera pas démontré dans le cadre de ce mémoire mais est consultable dans le livre [2] (théorème 7.5.1 et corollaire 7.7.5).

Théorème 4.16. (A. Cobham, 1968)

(i) *Un mot* $u \in \mathcal{A}^\mathbb{N}$ *est morphique si et seulement si il peut être écrit comme* $u = \tau(\sigma^\omega(x))$, *où :*

- $\mathcal{B}$ *est un alphabet,*
- $x \in \mathcal{B}$,
- $\sigma : \mathcal{B}^* \to \mathcal{B}^*$ *est un endomorphisme non effaçant et,*
- $\tau : \mathcal{B}^* \to \mathcal{A}^*$ *est un codage.*

(ii) *L'image d'un mot morphique par un morphisme est soit finie soit également morphique.*

4.2 Action d'un morphisme sur la complexité en facteurs

Soient $u \in \mathcal{A}^{\mathbb{N}}$ un mot infini et un morphisme [1] $\tau : \mathcal{A}^* \to \mathcal{B}^*$. Dans cette section, nous allons comparer les complexités en facteurs des mots u et $v = \tau(u)$.

Tout d'abord, formalisons le concept classique d'injectivité dans le cadre des morphismes.

Définition 4.17. Un morphisme $\tau : \mathcal{A}^* \to \mathcal{B}^*$ est injectif s'il n'existe pas deux mots x, y différents écrits sur l'alphabet $\mathcal{A}$ tels que $\tau(x) = \tau(y)$.

En particulier, un morphisme injectif est non effaçant. En effet, sinon s'il existe une lettre $a_0 \in \mathcal{A}$ telle que $\tau(a_0) = \varepsilon$, on a $\tau(a_0) = \varepsilon = \tau(\varepsilon)$.

Proposition 4.18. *Un morphisme $\tau : \mathcal{A}^* \to \mathcal{B}^*$ est injectif si et seulement si les lettres de $\mathcal{A}$ ont des images par τ différentes et si le langage $\tau(\mathcal{A})$ est un code.*

Démonstration. La condition est suffisante. En effet, supposons que le langage $\tau(\mathcal{A})$ soit un code et que les lettres de $\mathcal{A}$ sont envoyées sur des mots différents. Par la Définition 3.20 d'un ensemble définissant un code, s'il existe $x, y \in \mathcal{A}^*$ tels que $\tau(x) = \tau(y) \in (\tau(\mathcal{A}))^*$, alors la décomposition de $\tau(x) = \tau(y)$ dans $\tau(\mathcal{A})$ est unique. Or, les lettres de $\mathcal{A}$ sont envoyées sur des mots différents de $\mathcal{B}^*$. Ainsi, cette décomposition nous permet de construire le mot $x = y$ de $\mathcal{A}^*$ de manière unique.
Démontrons ensuite que la condition est nécessaire. Supposons que $\tau : \mathcal{A}^* \to \mathcal{B}^*$ est un morphisme injectif. Par définition, il n'existe pas deux mots différents x et y de $\mathcal{A}^*$ tels que $\tau(x) = \tau(y)$. Ainsi, en particulier, deux lettres de $\mathcal{A}$ ne peuvent pas avoir la même image par τ. De plus, si τ est injectif, pour tout mot de $t \in (\tau(\mathcal{A}))^*$ il n'existe qu'un unique mot $x \in \mathcal{A}^*$ tel que $\tau(x) = t$. Ceci signifie que $\tau(\mathcal{A})$ est un code.

□

Lemme 4.19. *Soient $\tau : \mathcal{A} \to \mathcal{B}$ un codage et un mot $u \in \mathcal{A}^{\mathbb{N}}$. Les fonctions de complexité des mots u et $\tau(u) \in \mathcal{B}^{\mathbb{N}}$ satisfont l'inégalité*

$$p_{\tau(u)}(n) \leq p_u(n),$$

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

1. Nous employons ici la notation τ, que nous avons jusqu'alors réservée aux codages car, dans la suite de ce mémoire, c'est en particulier à ce type de morphismes que nous appliquerons les résultats qui vont suivre.

Démonstration. Étant donné que τ est un codage, il code chaque lettre de $\mathcal{A}$ par une lettre de $\mathcal{B}$ et donc préserve les longueurs. Ainsi, on a $L_n(\tau(u)) = \tau(L_n(u))$ et

$$\#(L_n(\tau(u))) = \#(\tau(L_n(u))) \leq \#(L_n(u))$$

car le morphisme τ n'est pas nécessairement injectif donc des mots différents de $L_n(u)$ peuvent avoir la même image par τ. En d'autres termes, on obtient $p_{\tau(u)}(n) \leq p_u(n)$. □

Définition 4.20. Un *facteur propre* d'un mot $u \in \mathcal{A}^*$ est un facteur de u différent du mot vide et de u lui-même.

Une généralisation du Lemme 4.19 est la suivante :

Lemme 4.21. *Soient $\tau : \mathcal{A}^* \to \mathcal{B}^*$ un morphisme non effaçant et u un mot de $\mathcal{A}^{\mathbb{N}}$. Les fonctions de complexité des mots $u \in \mathcal{A}^{\mathbb{N}}$ et $\tau(u) \in \mathcal{B}^{\mathbb{N}}$ satisfont l'inégalité*

$$p_{\tau(u)}(n) \leq \|\tau\| \, p_u(n),$$

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

Démonstration. Notons $v = \tau(u)$. Considérons un entier naturel $n \in \mathbb{N}$ et un facteur de $w' \in L_n(v) \subseteq \mathcal{B}^n$. Notons x' le mot de $\mathcal{B}^* \cap \mathrm{Pref}(v)$ défini tel que $x'w' \in \mathrm{Pref}(v)$. Ensuite, considérons $x \in \mathcal{A}^*$ le plus long préfixe de u tel que $\tau(x)$ est un préfixe de x' et notons $w \in L_n(u)$ le mot de longueur n tel que $xw \in \mathrm{Pref}(u)$. Par construction vu que τ est non effaçant, on a $\tau(xw) \in \mathrm{Pref}(x'w')$. Finalement, notons $w = ay$ avec $a \in \mathcal{A}$ et $|y| = n - 1$. Soit $k = |x'| - |\tau(x)|$. Par construction, on a $k \geq 0$ et par maximalité de $|x|$, on a $\tau(xa) \notin \mathrm{Pref}(x')$. Dès lors, le mot x' est un préfixe propre de $\tau(xa)$ et on obtient

$$|x'| < |\tau(xa)| = |\tau(x)| + |\tau(a)|.$$

Ainsi, on a

$$k = |x'| - |\tau(x)| < |\tau(x)| + |\tau(a)| - |\tau(x)| = |\tau(a)| \leq \|\tau\|,$$

c'est-à-dire

$$0 \leq k < \|\tau\|.$$

De plus, on a

$$|\tau(xw)| = |\tau(xay)| = |\tau(xa)| + |\tau(y)|$$

avec $|\tau(xa)| \geq |x'| + 1$ et $|\tau(y)| \geq |y| = n - 1$ étant donné que τ est non effaçant. Ainsi

$$|\tau(xw)| \geq |x'| + n.$$

Or, on sait que $\tau(xw)$ est un préfixe de $x'w'$, donc on a

$$|\tau(xw)| \leq |x'w'| = |x'| + n.$$

Par double inégalité, on obtient $|\tau(xw)| = |x'| + n$, c'est-à-dire

$$|\tau(w)| = |x'| - |\tau(x)| + n = k + |w'|$$

et $\tau(xw) = x'w'$ car, par construction, $\tau(xw) \in \mathrm{Pref}(x'w')$ et nous venons de montrer que les deux mots ont la même longueur. Ainsi, le mot w' apparaît à la position k dans $\tau(w)$. Notons f la fonction qui à $w' \in L_n(v)$ associe $(w,k) \in L_n(u) \times [\![0, \|\tau\| - 1]\!]$. Par définition, la fonction f est injective car à un couple $(w,k) \in L_n(u) \times [\![0, \|\tau\| - 1]\!]$ donné, le mot $\tau(w)$ est univoquement déterminé et il existe un unique mot w' tel que w' apparaît à la position k dans $\tau(w)$. Ainsi, les cardinaux des ensembles vérifient

$$|L_n(\tau(u))| \leq |L_n(u)| \, \|\tau\| \, .$$

□

Dès à présent, nous voulons établir une borne inférieure pour la fonction $p_{\tau(u)}$. Pour ce faire, nous allons restreindre les catégories de morphismes considérées. En effet, par exemple, les morphismes où toute lettre est envoyée vers une puissance d'un même mot nous produit pour tout mot infini de départ, le même mot périodique. Nous voulons rejeter ce type de cas. En fait nous désirons plus précisément nous restreindre à des morphismes injectifs.

Lemme 4.22. *Si $\tau : \mathcal{A}^* \to \mathcal{B}^*$ est un morphisme injectif, les fonctions de complexité des mots $u \in \mathcal{A}^{\mathbb{N}}$ et $\tau(u) \in \mathcal{B}^{\mathbb{N}}$ satisfont l'inégalité*

$$\frac{1}{\|\tau\|} p_u \left(\left\lfloor \frac{n}{\|\tau\|} \right\rfloor \right) \leq p_{\tau(u)}(n),$$

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

Démonstration. Notons $v = \tau(u)$. Considérons un entier naturel $n \in \mathbb{N}$ et un facteur $w \in L_n(u)$. Soit E l'ensemble défini comme suit :

$$E := \{x \in L(u) \mid w \in \mathrm{Pref}(x) \text{ et } |\tau(x)| \leq \|\tau\| \, n\}.$$

Remarquons que l'ensemble E est non vide. En effet, le mot w appartient à E car $w \in L_n(u) \cap \mathrm{Pref}(w)$ et par définition de $\|\tau\|$ on a $|\tau(w)| \leq \|\tau\| \, |w| = \|\tau\| \, n$. De plus, vu que τ est injectif, il est en particulier non effaçant. Ainsi, si $x \in E$, on a $|x| \leq |\tau(x)| \leq \|\tau\| \, n$, et donc l'ensemble E est fini. Dès lors, choisissons un mot $y \in E$ de longueur maximale.
Le mot u étant infini, le mot $y \in L(u)$ admet une extension à droite dans u. C'est-à-dire qu'il existe une lettre $a \in \mathcal{A}$ telle que $ya \in L(u)$. Par maximalité de y, le mot ya n'appartient pas à l'ensemble E. Cependant, on a $ya \in L(u)$ et $w \in \mathrm{Pref}(y) \subseteq \mathrm{Pref}(ya)$. Cela implique que $|\tau(ya)| > \|\tau\| \, n$. Notons $w' \in \mathcal{B}^*$ le préfixe de $\tau(ya)$ dans v de longueur $\|\tau\| \, n$ et $k = \|\tau\| \, n - |\tau(y)|$. Étant donné que le mot y appartient à l'ensemble E, l'entier k est plus grand ou égal à 0. Par construction, on obtient également

$$k = \|\tau\| \, n - |\tau(y)| < |\tau(ya)| - |\tau(y)| = |\tau(a)| \leq \|\tau\| \, .$$

Ainsi, on en tire que $0 \leq k < |\tau(a)| \leq \|\tau\|$.
La fonction

$$f : L_n(u) \to L_{\|\tau\| n}(v) \times [\![0, \|\tau\| - 1]\!],\ w \mapsto (w', k)$$

est injective. En effet, en considérant le couple

$$(w', k) \in L_{\|\tau\| n}(v) \times [\![0, \|\tau\| - 1]\!],$$

le mot $\tau(y)$ peut être retrouvé en supprimant les k dernières lettres de w'. Ensuite, par injectivité de τ, nous trouvons y de manière unique et w en est son préfixe de longueur n. Par injectivité de f, on a

$$\begin{aligned} & |L_n(u)| \leq |L_{\|\tau\| n}(v)|\, \|\tau\| \\ \Leftrightarrow & p_u(n) \leq \|\tau\|\ \ p_v(\|\tau\|\, n) \\ \Leftrightarrow & \frac{1}{\|\tau\|} p_u(n) \leq p_{\tau(u)}(\|\tau\|\, n) \end{aligned}$$

Ainsi, on en tire que

$$\frac{1}{\|\tau\|} p_u\left(\left\lfloor \frac{n}{\|\tau\|} \right\rfloor\right) \leq \underbrace{p_{\tau(u)}\left(\|\tau\| \left\lfloor \frac{n}{\|\tau\|} \right\rfloor\right) \leq p_{\tau(u)}(n)}_{\text{par la Proposition 1.11}}.$$

□

Lemme 4.23. *Soient un mot $u \in \mathcal{A}^{\mathbb{N}}$ et un alphabet $\mathcal{B} \subset \mathcal{A}$ tels que $L(u) \cap \mathcal{B}^*$ est fini. Notons χ le morphisme effaçant les lettres de $\mathcal{B}$. Formellement,*

$$\chi : \mathcal{A}^* \to (\mathcal{A} \setminus \mathcal{B})^*,\ a \mapsto \begin{cases} a & \text{si } a \notin \mathcal{B} \\ \varepsilon & \text{si } a \in \mathcal{B}. \end{cases}$$

Le mot $\chi(u)$ est un mot infini de $(\mathcal{A} \setminus \mathcal{B})^{\mathbb{N}}$ et sa fonction de complexité satisfait l'inégalité

$$p_{\chi(u)}\left(\left\lfloor \frac{n}{M+1} \right\rfloor\right) \leq p_u(n),$$

pour tout $n \in \mathbb{N}$, où $M = \max\{|x| \mid x \in L(u) \cap \mathcal{B}^*\}$.

Démonstration. Par définition du morphisme χ et par le fait que $L(u) \cap \mathcal{B}^*$ soit fini, on conclut trivialement que le mot $\chi(u)$ est un mot infini écrit sur l'alphabet $\mathcal{A} \setminus \mathcal{B}$. Ensuite, pour la seconde partie de l'énoncé,pour tout $n \in \mathbb{N}$, définissons la fonction

$$f_n : \mathcal{A}^* \to (\mathcal{A} \setminus \mathcal{B})^{\leq n},\ x \mapsto \mathrm{Pref}_{\leq n}(\chi(x))$$

où $\mathrm{Pref}_{\leq n}(\chi(x))$ est le plus long préfixe de longueur inférieure ou égale à n dans $\chi(x)$.
Par définition de l'entier M, on a $f_n(L_{(M+1)n}(u)) = L_n(\chi(u))$. En effet, procédons par double inclusion.
Soit $x \in L_{(M+1)n}(u)$. Notons $x = x^{(1)} \cdots x^{(n)}$ où $x^{(i)}$ désigne le $i^{\text{ème}}$ facteur de x de longueur $M+1$. Par définition de M, pour tout $i \in \{1, ..., n\}$, les facteurs $x^{(i)}$ n'appartiennent pas à $\mathcal{B}^*$. Ainsi, on obtient $|\chi(x^{(i)})| \geq 1$ et

$$\begin{aligned} |\chi(x)| &= |\chi(x^{(1)} \cdots x^{(n)})| \\ &= |\chi(x^{(1)})| + \cdots + |\chi(x^{(n)})| \\ &\geq n. \end{aligned}$$

Dès lors, vu que $\chi(x) \in L(\chi(u))$, on a

$$f_n(x) = \mathrm{Pref}_{\leq n}(\chi(x)) = \mathrm{Pref}_n(\chi(x)) \in L_n(\chi(u)).$$

On en tire que $f_n(L_{(M+1)n}(u)) \subseteq L_n(\chi(u))$.
Soit $y \in L_n(\chi(u))$. Montrons qu'il existe $x \in \mathcal{A}^* \cap L_{(M+1)n}(u)$ tel que $f_n(x) = y$. Par définition de $\chi(u)$, si le mot y appartient à $L_n(\chi(u))$, il existe un facteur $x' \in L(u)$ tel que $\chi(x') = y$. Par définition de M, on a $|x'| \leq (M+1)n$. Si $|x'| = (M+1)n$, alors $x = x'$ convient. Sinon, étant donné que u est infini, notons x'' le facteur de u tel que $xx'' \in L(u)$ et $|x'x''| = (M+1)n$. On a $\chi(x'x'') = y\chi(x'')$ donc $f_n(x'x'') = f_n(x') = y$. Dès lors, le facteur $x = x'x''$ convient à démontrer que l'on a $f_n(L_{(M+1)n}(u)) \supseteq L_n(\chi(u))$.
Dès lors, vu que toute fonction est surjective sur son image, on a

$$\#(\, L_n(\chi(u))\,) = \#(f_n(L_{(M+1)n}(u))) \leq \#(L_{(M+1)n}(u)).$$

Ainsi, on en tire que

$$p_{\chi(u)}(n) \leq p_u((M+1)n).$$

Ensuite, en substituant n par $\left\lfloor \frac{n}{M+1} \right\rfloor$ dans cette inégalité, on obtient

$$p_{\chi(u)}\left(\left\lfloor \frac{n}{M+1} \right\rfloor\right) \leq \underbrace{p_u\left((M+1)\left\lfloor \frac{n}{M+1} \right\rfloor\right) \leq p_u(n)}_{\text{par la Proposition 1.11}}.$$

□

4.3 Complexité des mots automatiques

Définition 4.24. Soit un entier naturel $q \geq 2$. Un mot infini $u \in \mathcal{B}^{\mathbb{N}}$ est *q-automatique* s'il existe un morphisme q-uniforme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ prolongeable en une lettre $a \in \mathcal{A}$ et un codage $\tau : \mathcal{A} \to \mathcal{B}$ tels que $u = \tau(\sigma^\omega(a))$. L'alphabet $\mathcal{A}$ est appelé l'*alphabet interne* de u.

Nous allons étudier la fonction de complexité en facteurs des mots automatiques. Ces mots forment une sous-classe importante des mots morphiques. En effet, le fait d'avoir des images de longueur constante rend la description plus simple en recourant aux développement en base entière et aux automates. Les liens entre ces notions sont développés dans [2].

Une borne explicite sur la fonction de complexité d'un mot u qui est q-automatiques a été donnée par A. Cobham (1972), ([2], théorème 10.3.1) :

$$p_u(n) \leq qd^2n$$

pour tout $n \geq 1$, où $d = \#(\mathcal{A})$.
Le théorème suivant est une version plus forte de cette inégalité.

Théorème 4.25. *Soit u un mot q-automatique d'alphabet interne $\mathcal{A}$. Pour tout $n \geq 2$, on a*

$$p_u(n) \leq \left(d^2 + \frac{q-2}{2}d\right)(n-1),$$

où $d = \#(\mathcal{A})$. En particulier, on a $p_u(n) = \mathcal{O}(n)$.

Démonstration. Le mot u étant q-automatique, considérons $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ le morphisme q-uniforme, $\tau = \mathcal{A} \to \mathcal{B}$ le codage et a la lettre de $\mathcal{A}$ tels que $u = \tau(\sigma^\omega(a))$. Si l'on note $v = \sigma^\omega(a)$, on a $u = \tau(v)$. Par le Lemme 4.19, on a

$$p_u(n) = p_{\tau(v)}(n) \leq p_v(n).$$

Dès lors, démontrons l'inégalité voulue pour $p_v(n)$. Ainsi, pour simplifier les notations, dans le reste de la preuve, nous noterons $p = p_v$.
Tout d'abord, démontrons que, pour tous $m \in \mathbb{N}$ et $i \in [\![0, q]\!]$,

$$p(qm + i + 1) \leq (q - i)p(m + 1) + ip(m + 2). \tag{4.1}$$

Considérons un facteur w de longueur $qm + i + 1$ de v, et notons $k \in \mathbb{N}$ la position d'une de ses occurrences dans v. En effectuant la division euclidienne de k par q, l'entier k peut être écrit comme $k = qn + j$, où $0 \leq j < q$. En particulier, on a

$$i + j + 1 \leq 2q. \tag{4.2}$$

Notons z le mot de $\mathcal{A}^*$ défini par

$$z = \begin{cases} v_n v_{n+1} \cdots v_{n+m} & \text{si } i + j + 1 \leq q \\ v_n v_{n+1} \cdots v_{n+m+1} & \text{si } i + j + 1 > q \end{cases}.$$

Dans les deux cas, vu que $v = \sigma^\omega(a)$ et que z est un facteur de v, le mots $\sigma(z)$ en est un également. De plus, vu que la position de z dans v est n et que le morphisme σ est

q-uniforme, la position de $\sigma(z)$ dans v est qn. Ensuite, remarquons que la longueur du facteur $\sigma(z)$ est

$$|\,\sigma(z)| = q|z| = \begin{cases} q(m+1) & \text{si } i+j+1 \leq q \\ q(m+2) & \text{si } i+j+1 > q \end{cases}.$$

Ainsi, le facteur $\sigma(z)$ se termine à la position

$$\begin{aligned} & qn + \begin{cases} q(m+1) & \text{si } i+j+1 \leq q \\ q(m+2) & \text{si } i+j+1 > q \end{cases} \\ =& \begin{cases} q(n+m)+q & \text{si } i+j+1 \leq q \\ q(n+m)+2q & \text{si } i+j+1 > q \end{cases} \\ \geq& q(n+m)+i+j+1, \end{aligned}$$

où la dernière étape est triviale pour le premier cas, et pour le second, la justification est donnée par l'inégalité (4.2). Or, le facteur w apparaît à la position $qn+j$ dans v et est de longueur $qm+i+1$. Il se termine donc à la position $q(n+m)+i+j+1$ dans v. On en tire que le facteur $\sigma(z)$ couvre complètement w dans v. En d'autres termes, cela signifie que w est un facteur de $\sigma(z)$. De plus, étant donné que $\sigma(z)$ débute à la position qn, w est un facteur à la position j de $\sigma(z)$.

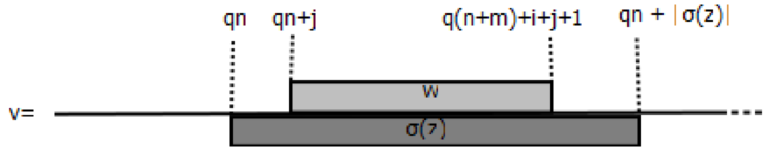


FIGURE 4.1 – Positions des facteurs w et $\sigma(z)$ dans le mot v.

Ainsi, le facteur w de v de longueur $qm+i+1$ est entièrement déterminé par le couple

$$(j, z) \in ([\![0, q-i-1]\!] \times L_{m+1}(v)) \cup ([\![q-i, q-1]\!] \times L_{m+2}(v))\,.$$

En effet, considérons un facteur $z \in L_{m+1}(v) \cup L_{m+2}(v)$ et notons n la position de ce facteur dans v. Par définition d'un morphisme q-uniforme, étant donné que $v = \sigma(v)$, le facteur $\sigma(z)$ est un facteur de v apparaissant à la position qn et l'entier $0 \leq j < q$ nous indique à quelle position de $\sigma(z)$ le facteur w de longueur $qm+i+1$ apparaît. Le mot w est entièrement déterminé étant donné qu'il est complètement couvert par $\sigma(z)$. De plus, tout facteur de longueur $qm+i+1$ de v peut être construit par cette procédure. Ainsi, la construction définit une surjection entre $([\![0, q-i-1]\!] \times L_{m+1}(v)) \cup ([\![q-i, q-1]\!] \times L_{m+2}(v))$ et $L_{qm+i+1}(v)$ et on a

$$\begin{aligned} & |L_{qm+i+1}(v)| \leq |[\![0, q-i-1]\!] \times L_{m+1}(v)| + |[\![q-i, q-1]\!] \times L_{m+2}(v)| \\ \Leftrightarrow & p(qm+i+1) \leq (q-i)p(m+1) + ip(m+2). \end{aligned}$$

Ceci termine la preuve de l'inégalité (4.1).
Notons $C = d^2 + \frac{q-2}{2}d$. Vu que $d = \#(\mathcal{A})$, par la Proposition 1.9, on a

$$p(2) \leq d^2 \leq C. \tag{4.3}$$

En considérant $m = 0$ et $2 \leq i \leq q$ dans l'inégalité (4.1), on a

$$\begin{aligned}
p(i+1) &\leq (q-i)p(1) + ip(2) \leq (q-i)d + id^2 \\
p(i+1) &\leq i\left(\frac{q-i}{i}d + d^2\right) \\
p(i+1) &\leq iC. \qquad (4.4)
\end{aligned}$$

Où la dernière étape se justifie comme suit :

$$\begin{aligned}
&\frac{q-i}{i} \leq \frac{q-2}{2} \\
\Leftrightarrow & 2(q-i) \leq i(q-2) \\
\Leftrightarrow & 2q \leq iq \\
\Leftrightarrow & 2 \leq i.
\end{aligned}$$

Ainsi, en utilisant l'inégalité (4.3) et en notant $i + 1 = n$ dans l'inégalité (4.4), on a $p(n) \leq (n-1)C$ pour tout $3 \leq n \leq q+1$.
Pour de plus grandes valeurs de n, procédons par induction en utilisant l'inégalité (4.1). Notons $n = qm + i + 1$, où $m \geq 1$ et $1 \leq i \leq q$ et supposons avoir l'hypothèse pour tout entier plus petit que n. On a $n = qm + i + 1 \geq 2m + 1 + 1 > m + 2$, ainsi l'hypothèse de récurrence s'applique à $p(m+1)$ et $p(m+2)$. On obtient

$$\begin{aligned}
p(n) = p(qm+i+1) &\leq (q-i)p(m+1) + ip(m+2) \\
&\leq (q-i)Cm + iC(m+1) \\
&= C(qm+i) \\
&= C(n-1).
\end{aligned}$$

□

Ce théorème nous permet de démontrer que l'inégalité du Lemme 4.21 ne tient plus si le morphisme τ est effaçant. En effet, considérons le morphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$, prolongeable en $a_0 \in \mathcal{A} = \{a_0, ..., a_k\}$, tel que $u = \sigma^\omega(a_0)$ est un mot infini pour lequel[2] $p_u(n) = \Theta(n^2)$. Construisons le nouvel alphabet $\mathcal{B} = \mathcal{A} \cup \{b\}$ avec $b \notin \mathcal{A}$ et considérons le morphisme $\sigma' : \mathcal{B}^* \to \mathcal{B}^*$ défini par

$$\begin{aligned}
&\sigma_{\mathcal{B}}(a_i) = \sigma(a_i) b^{\|\sigma\| - |\sigma(a_i)|}, \ \forall i \in \{0, ..., k\}, \text{ et} \\
&\sigma_{\mathcal{B}}(b) = b^{\|\sigma\|}.
\end{aligned}$$

Le morphisme σ est prolongeable en a_0, ainsi on a $\|\sigma\| \geq 2$. Par définition, le morphisme $\sigma_{\mathcal{B}}$ est $\|\sigma\|$-uniforme et prolongeable en a_0. Dès lors, en considérant le codage identité, le mot $v = \sigma_{\mathcal{B}}^\omega(a_0)$ est $\|\sigma\|$-automatique et par le Théorème 4.25, sa fonction de complexité en facteurs est en $\mathcal{O}(n)$. Ensuite, notons τ le morphisme effaçant la lettre b. Formellement,

2. L'existence d'un tel mot fera l'objet du chapitre suivant.

$$\tau(a_i) = a_i,\ \forall i \in \{0, ..., k\},\ \text{ et}$$
$$\tau(b) = \varepsilon.$$

Par construction, on a $\tau(v) = \tau(\sigma_{\mathcal{B}}^{\omega}(a_0)) = u$ mais l'inégalité suivante n'est pas vérifiée

$$p_{\tau(\sigma_{\mathcal{B}}^{\omega}(a_0))}(n) = p_u(n) \leq p_{\sigma_{\mathcal{B}}^{\omega}(a_0)}(n).$$

Exemple 4.26. Comme étudié dans l'Exemple 4.13, le mot de THUE-MORSE est un mot purement morphique engendré par le morphisme 2-uniforme $\sigma : \{a, b\}^* \to \{a, b\}^*$ défini par $\sigma(a) = ab$ et $\sigma(b) = ba$. Ainsi, par le Théorème 4.25, en utilisant le morphisme σ et le codage identité τ, le mot de THUE-MORSE

$$t = \sigma^{\omega}(a) = abaababaaabaababaababaabaab\cdots$$

est 2-automatique et a une complexité en facteurs en $\mathcal{O}(n)$. Plus particulièrement, en suivant les notations du Théorème 4.25, on a $q = 2$ et étant donné que l'alphabet interne de t est l'alphabet binaire $\{a, b\}$, l'entier d vaut également 2. Ainsi, on obtient

$$p_t(n) \leq \left(4 + \frac{2-2}{2}2\right)(n-1) = 4(n-1).$$

Chapitre 5

Théorème de Pansiot

Le but de ce chapitre est de donner une preuve complète du célèbre théorème de Pansiot (1984). Ce théorème stipule que la complexité en facteurs d'un mot purement morphique ne peut adopter que cinq comportements asymptotiques différents. La première preuve de ce théorème a été publiée par Pansiot [12]. Nous allons fournir la preuve de ce résultat établie dans l'article [6] de Cassaigne et Nicolas. Tout au long de ce chapitre, nous allons utiliser des classifications intermédiaires de morphismes afin de pouvoir au final obtenir le théorème de Pansiot en tant que conséquence presque immédiate des résultats préétablis. Pour ce faire, il nous faut rappeler les notations de Landau [1].

Définitions 5.1. Soient f et g deux fonctions numériques. On a :

(i) $f(n) = \mathcal{O}(g(n))$ si $\exists k > 0, \exists n_0 \geq 1,\ \forall n \geq n_0,\ |f(n)| \leq k.|g(n)|$, c'est-à-dire *si la fonction $|f|$ est majorée par la fonction $|g|$ asymptotiquement, à un facteur multiplicatif près.*

(ii) $f(n) = o(g(n))$ si $\forall \varepsilon > 0, \exists n_0 \geq 1,\ \forall n \geq n_0,\ |f(n)| \leq \varepsilon|g(n)|$, c'est-à-dire *si la fonction $|f|$ est négligeable par rapport à la fonction $|g|$ asymptotiquement.* Cela est équivalent à dire : $\lim\limits_{n\to\infty} \frac{f(n)}{g(n)} = 0$.

(iii) $f(n) = \Omega(g(n))$ si $\exists k > 0, \exists n_0 \geq 1,\ \forall n \geq n_0,\ k.|g(n)| \leq |f(n)|$, c'est-à-dire *si la fonction $|f|$ est minorée par la fonction $|g|$ asymptotiquement, à un facteur près.*

(iv) $f(n) = \Theta(g(n))$ si $\exists k_1, k_2 > 0, \exists n_0 \geq 1,\ \forall n \geq n_0,\ k_1.|g(n)| \leq |f(n)| \leq k_2.|g(n)|$, c'est-à-dire *si la fonction $|f|$ est majorée et minorée par la fonction $|g|$ asymptotiquement à des facteurs multiplicatifs près.*

(v) [2] $f \sim g$ si $\forall \epsilon > 0,\ \exists n_0 \geq 1,\ \forall n \geq n_0,\ |f(n) - g(n)| < \epsilon|g(n)|$ c'est-à-dire *si la fonction f asymptotiquement équivalente à la fonction g.* Cela est équivalent à dire : $\lim\limits_{n\to\infty} \frac{f(n)}{g(n)} = 1$.

Le théorème faisant l'objet de ce chapitre est le suivant :

---

1. Ces notations ont été largement utilisées par le mathématicien Edmund Landau(1877-1938), mais selon l'avis des historiens des sciences, l'origine exacte de ces notations serait due à Paul Bachmann(1837-1920).

2. Cette notation a déjà été introduite dans le chapitre 1 à la Définition 1.16.

Théorème 5.2. (J-J PANSIOT, 1984)
Soit u un mot purement morphique. L'une des affirmations suivantes est vérifiée :

(i) $p(n) = \Theta(1)$,

(ii) $p(n) = \Theta(n)$,

(iii) $p(n) = \Theta(n \log \log n)$,

(iv) $p(n) = \Theta(n \log n)$,

(v) $p(n) = \Theta(n^2)$.

Dans l'Exemple 4.26, il a été démontré que le mot de THUE-MORSE, qui est un mot purement morphique, a une complexité en facteurs en $\mathcal{O}(n)$. Or, ce mot n'est pas périodique car il ne contient, en particulier, pas de cube. Ainsi, par le Théorème 2.2 de MORSE et HEDLUND, sa fonction de complexité ne peut pas être en $\Theta(1)$. Par le Théorème 5.2 de PANSIOT, nous pouvons donc conclure que $p_t(n) = \Theta(n)$.

5.1 Quelques résultats d'analyse asymptotique

Pour démontrer le Théorème 5.2 de PANSIOT, quelques notions d'analyse asymptotique nous seront nécessaires. Cette section est constituée de résultats très utiles dans nos raisonnements futurs mais ne faisant pas appel à de la combinatoire des mots. C'est pourquoi, seuls les énoncés seront insérés dans cette section. Cependant, les détails des démonstrations sont disponibles dans l'annexe A de ce mémoire.

Définition 5.3. Soient deux couples (λ_1, μ_1) et $(\lambda_2, \mu_2) \in \mathbb{R} \times \mathbb{R}$. On dit que (λ_1, μ_1) est *lexicographiquement plus petit* que (λ_2, μ_2) si $\lambda_1 < \lambda_2$ ou bien si $\lambda_1 = \lambda_2$ et $\mu_1 < \mu_2$. On le note $(\lambda_1, \mu_1) <_l (\lambda_2, \mu_2)$

Proposition 5.4. *Soient deux couples de réels (β_1, α_1), $(\beta_2, \alpha_2) \in (\mathbb{R}_0 \times \mathbb{R}) \cup \{(0,0)\}$, on a*
$$k^{\alpha_1} \beta_1^k = o(k^{\alpha_2} \beta_2^k)$$
lorsque k tend vers l'infini si et seulement si $(\beta_1, \alpha_1) <_l (\beta_2, \alpha_2)$.

Considérons deux couples (β_1, α_1), $(\beta_2, \alpha_2) \in (\mathbb{R}_0 \times \mathbb{R}) \cup \{(0,0)\}$ tels que $(\beta_1, \alpha_1) \neq (\beta_2, \alpha_2)$. Étant donné que l'ordre lexicographique est total, au vu de la Proposition 5.4, on a $k^{\alpha_1} \beta_1^k = o(k^{\alpha_2} \beta_2^k)$ ou bien $k^{\alpha_2} \beta_2^k = o(k^{\alpha_1} \beta_1^k)$ lorsque k tend vers l'infini. Ainsi, les suites de la forme $(k^{\alpha} \beta^k)_{k \geq 1}$ avec $(\beta, \alpha) \in (\mathbb{R}_0 \times \mathbb{R}) \cup \{(0,0)\}$ peuvent être ordonnées totalement par leurs ordres de croissance asymptotique.

Définition 5.5. Soit V un espace vectoriel réel ou complexe. Une *norme* sur l'espace V est une application

$$\|\cdot\| : V \to \mathbb{R}$$

telle que les trois propriétés suivantes soient vérifiées pour tous $x, y \in V$ et $\lambda \in \mathbb{C}$:

(i) $\|x\| = 0 \Leftrightarrow x = 0$,

(ii) $\|\lambda x\| = |\lambda| \, \|x\|$, et

(iii) $\|x + y\| \leq \|x\| + \|y\|$.

Définition 5.6. La *norme de Manhattan* $\|\cdot\|_1$ sur l'espace $\mathbb{C}_d^d$ est définie telle que pour tout $\mathbf{X} \in \mathbb{C}_d^d$, $\|\mathbf{X}\|_1$ est égal à la somme des modules des entrées de $\mathbf{X}$.

Théorème 5.7. *Soient $d \in \mathbb{N}_0$ et $\|\cdot\|$ une norme définie sur $\mathbb{C}_d^d$. Pour toute matrice $\boldsymbol{M} \in \mathbb{C}_d^d$, il existe $(\beta, \alpha) \in (\mathbb{R}_0 \times \mathbb{R}) \cup \{(0,0)\}$ tel que $\left\|\boldsymbol{M}^k\right\| = \Theta(k^\alpha \beta^k)$ lorsque k tend vers l'infini.*

Nous allons maintenant définir un nouveau concept qui nous servira dans le Théorème 5.9, ainsi que dans son Corollaire 5.10, très utiles pour le Théorème 5.2 de Pansiot.

Définition 5.8. Soient f_1 et f_2 deux fonctions à valeurs réelles définies sur un ensemble X tel que $\mathbb{N} \subseteq X$. Pour tout $y \in \mathbb{N}$, notons

$$E_y(f_1, f_2) := \{k \in \mathbb{N} \mid f_1(k) \leq y \leq f_2(k)\}.$$

Théorème 5.9. *Soient $(\beta_1, \alpha_1), (\beta_2, \alpha_2) \in \mathbb{R}_{>1} \times \mathbb{R}$ et $f_1, f_2 : \mathbb{N} \to \mathbb{R}_{\geq 0}$ des fonctions telles que $f_1(k) = \Theta(k^{\alpha_1}\beta_1^k)$ et $f_2(k) = \Theta(k^{\alpha_2}\beta_2^k)$ lorsque k tend vers l'infini. Considérons la fonction $\Delta : \mathbb{R}_{>1} \to \mathbb{R}$ définie par*

$$\Delta(y) = -\left(\frac{1}{\log \beta_2} - \frac{1}{\log \beta_1}\right) \log y + \left(\frac{\alpha_2}{\log \beta_2} - \frac{\alpha_1}{\log \beta_1}\right) \log \log y$$

pour tout $y > 1$. Si (β_1, α_1) est lexicographiquement plus petit ou égal à (β_2, α_2), alors $\# \ (E_y(f_1, f_2)) = \Delta(y) + \mathcal{O}(1)$ lorsque y tend vers l'infini.

Corollaire 5.10. *En suivant les notations du Théorème 5.9, on obtient :*

(i) si $\beta_1 = \beta_2$ et $\alpha_1 = \alpha_2$, alors $\# \ (E_y(f_1, f_2)) = \mathcal{O}(1)$,

(ii) si $\beta_1 = \beta_2$ et $\alpha_1 < \alpha_2$, alors $\# \ (E_y(f_1, f_2)) = \Theta(\log \log y)$, et

(iii) si $\beta_1 < \beta_2$, alors $\# \ (E_y(f_1, f_2)) = \Theta(\log y)$.

Comme application de ce corollaire, évaluons le cardinal d'un ensemble de la forme donnée dans l'Exemple 3.29.

Exemple 5.11. Rappelons que dans l'Exemple 3.29, nous considérions $(e_k)_{k\in\mathbb{N}}$ une suite strictement croissante de nombres naturels et le mot lacunaire $u = 10^{e_0}10^{e_1}10^{e_1}1\cdots$. Nous avions établit l'égalité suivante

$$s(n) = 1 + \#\left\{k \in \mathbb{N} \mid e_k \leq n-1 \leq e_k + e_{k-1}\right\},$$

où l'on avait posé $e_{-1} = 0$. Considérons le cas où la suite $(e_k)_{k\in\mathbb{N}}$ est en plus de la forme $e_k = \Theta(\beta^k)$ pour un $\beta \in \mathbb{R}_{>0}$. On a

$$\#\left(\{k \in \mathbb{N} \mid e_k \leq n-1 \leq e_k + e_{k-1}\}\right) = \#\left(E_{n-1}(e_k, e_k + e_{k-1})\right).$$

Avec les notations du Théorème 5.9, on a $e_k = \Theta(\beta^k)$ et donc $(\beta_1, \alpha_1) = (\beta, 0)$. De la même façon, on a $e_k + e_{k-1} = \Theta(\beta^k) + \Theta(\beta^{k-1}) = \Theta(\beta^k)$ et $(\beta_2, \alpha_2) = (\beta, 0)$. Ainsi, par le Corollaire 5.10, on a $\#\left(E_{n-1}(e_k, e_k + e_{k-1})\right) = \mathcal{O}(1)$. On obtient $s(n) = \mathcal{O}(1)$ et finalement $p(n) = \mathcal{O}(n)$ par la Proposition 2.20.

Ensuite, énonçons les deux derniers résultats d'analyse asymptotique nécessaires à nos démonstrations futures.

Lemme 5.12. *Soit une suite réelle strictement positive $(f_k)_{k\in\mathbb{N}}$ définie telle que le quotient $\frac{f_{k+1}}{f_k}$ converge vers une limite réelle, notée β, lorsque k tend vers l'infini. Si $\beta > 1$, alors*

$$\sum_{j=0}^{k} f_j \sim \frac{\beta}{\beta - 1} f_k$$

lorsque k tend vers l'infini.

Lemme 5.13. *Soit $a > 1$, on a*

$$\sum_{i=2}^{n-1} \log_a \log_a(i) = \Omega(n \log_a \log_a n).$$

5.2 Théorème de Salomaa et Soittola

Cette section consiste à démontrer le théorème de Salomaa et Soittola énoncé ci-après. Il sera indispensable dans notre démarche pour aboutir à la preuve le Théorème 5.2 de Pansiot.

Théorème 5.14. (Salomaa et Soittola)
Pour tout endomorphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ et tout mot $x \in \mathcal{A}^*$, il existe un couple $(\beta, \alpha) \in (\mathbb{R}_{\geq 1} \times \mathbb{N}) \cup \{(0,0)\}$ tel que

$$|\sigma^k(x)| = \Theta(k^\alpha \beta^k)$$

lorsque k tend vers l'infini.

Remarque 5.15. Avec les notations du Théorème 5.14, si $(\beta, \alpha) = (0,0)$, alors on a $|\sigma^k(x)| = \Theta(0)$. Ainsi, cela implique que le mot x est effacé par une puissance du morphisme σ.

Pour obtenir la démonstration du Théorème 5.14, nous allons tout d'abord introduire de nouvelles définitions. En effet, la suite $(|\sigma^k(x)|)_{k \geq 0}$ peut être déterminée par les concepts de matrice d'incidence du morphisme σ et de vecteur de Parikh du mot x.

Définition 5.16. Soient un mot $u \in \mathcal{A}^*$ et un alphabet $\mathcal{B} \subseteq \mathcal{A}$. On note $|u|_\mathcal{B}$ le nombre de lettres de u appartenant à $\mathcal{B}$. En particulier, s'il existe une lettre $a \in \mathcal{A}$ telle que $\mathcal{B} = \{a\}$, on note $|u|_\mathcal{B} = |u|_a$.

Définition 5.17. Soit un morphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ où $\mathcal{A} = \{a_1, ..., a_d\}$. On définit la *matrice d'incidence* $\mathbf{M}_\sigma$ du morphisme σ comme suit :

$$(\mathbf{M}_\sigma)_{i,j} = |\sigma(a_j)|_{a_i}$$

pour tous $i, j \in \{1, \ldots, d\}$.

Exemples 5.18. En guise d'illustration, considérons les morphismes de Thue-Morse et de Fibonacci définis dans l'Exemple 4.13. Leurs matrices d'incidence sont respectivement

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad \text{et} \quad \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

Définition 5.19. Soient un alphabet $\mathcal{A} = \{a_1, ..., a_d\}$ et un mot $x \in \mathcal{A}^*$. Le *vecteur de* Parikh de x est le vecteur

$$\mathbf{P}(x) = \begin{pmatrix} |x|_{a_1} \\ \vdots \\ |x|_{a_d} \end{pmatrix}.$$

Nous allons maintenant démontrer des résultats liant ces nouvelles notions et la suite $(|\sigma^k(x)|)_{k \geq 0}$.

Proposition 5.20. *Soient un alphabet $\mathcal{A} = \{a_1, ..., a_d\}$, un mot $x \in \mathcal{A}^*$ et un morphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$. On a*

$$\boldsymbol{P}(\sigma(x)) = \boldsymbol{M}_\sigma \boldsymbol{P}(x).$$

Démonstration. Soit $i \in \{1, .., d\}$ et notons $x = x_0 \cdots x_{|x|-1}$. On a

$$\begin{aligned} |\sigma(x)|_{a_i} =& |\sigma(x_0) \cdots \sigma(x_{|x|-1})|_{a_i} \\ =& |\sigma(x_0)|_{a_i} + \cdots + |\sigma(x_{|x|-1})|_{a_i}. \end{aligned}$$

Ainsi, en regroupant les lettres identiques de x, on obtient

$$\begin{aligned} |\sigma(x)|_{a_i} =& \sum_{1 \le j \le d} |\sigma(a_j)|_{a_i} |x|_{a_j} \\ =& \sum_{1 \le j \le d} (\mathbf{M}_\sigma)_{ij} \mathbf{P}(x)_j \\ =& (\mathbf{M}_\sigma \mathbf{P}(x))_i. \end{aligned}$$

□

Corollaire 5.21. *Soient un alphabet $\mathcal{A} = \{a_1, ..., a_d\}$, un mot $x \in \mathcal{A}^*$ et un morphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$. Pour tout $k \in \mathbb{N}$, on a*

$$\boldsymbol{P}(\sigma^k(x)) = \boldsymbol{M}_\sigma^k \boldsymbol{P}(x).$$

Démonstration. Démontrons ce résultat par récurrence. Le cas de base $k = 0$ est vérifié. En effet, la matrice $\mathbf{M}_\sigma^0$ est la matrice identité de dimension d et σ^0 est le morphisme identité. Dès lors, supposons le résultat vrai pour $k \in \mathbb{N}$ et démontrons-le pour l'entier $k+1$. Par la Proposition 5.20, on a $\mathbf{P}(\sigma^{k+1}(x)) = \mathbf{M}_\sigma \mathbf{P}(\sigma^k(x))$ et par hypothèse de récurrence, on en tire que

$$\mathbf{P}(\sigma^{k+1}(x)) = \mathbf{M}_\sigma \mathbf{P}(\sigma^k(x)) = \underbrace{\mathbf{M}_\sigma \mathbf{M}_\sigma^k}_{= \mathbf{M}_\sigma^{k+1}} \mathbf{P}(x).$$

□

Par définition, la somme des composantes d'un vecteur de Parikh d'un mot est égale à la longueur de ce mot. Ainsi, un corollaire direct de la proposition précédente est le suivant.

Corollaire 5.22. *Soient un alphabet $\mathcal{A} = \{a_1, ..., a_d\}$, un mot $x \in \mathcal{A}^*$ et un morphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$. Pour tout $k \in \mathbb{N}$, on a*

$$|\sigma^k(x)| = \boldsymbol{U}\boldsymbol{M}_\sigma^k\boldsymbol{P}(x),$$

où $\boldsymbol{U}$ est le vecteur ligne de dimension d tel que $U_i = 1$, pour tout $i \in \{1, ..., d\}$.

Maintenant que nous avons le lien entre les concepts de matrice d'incidence, de vecteur de Parikh et de la suite $(|\sigma^k(x)|)_{k\geq 0}$, il nous suffit de démontrer le lemme suivant pour être en possession de tous les outils nécessaires à la démonstration du Théorème 5.14 de Salomaa et Soittola.

Lemme 5.23. *Soient un alphabet $\mathcal{A} = \{a_1, ..., a_d\}$, deux mots $x, y \in \mathcal{A}^*$ et un morphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$. S'il existe $k_0 \in \mathbb{N}$ tel que y est un facteur de $\sigma^{k_0}(x)$, alors*

$$|\sigma^k(y)| = \mathcal{O}(|\sigma^k(x)|)$$

lorsque k tend vers l'infini.

Démonstration. Si y est un facteur de $\sigma^{k_0}(x)$, alors on a $\sigma^{k_0}(x) = uyv$ où $u, v \in \mathcal{A}^*$. Ainsi, pour tout $k \in \mathbb{N}$, on a $\sigma^{k+k_0}(x) = \sigma^k(u)\sigma^k(y)\sigma^k(v)$. On en tire que

$$|\sigma^k(y)| \leq |\,\sigma^{k+k_0}(x)| = |\,\sigma^{k_0}(\sigma^k\,x)| \leq \left\|\sigma^{k_0}\right\| |\sigma^k(x)|,$$

par définition de la largeur d'un morphisme. □

Nous sommes dès lors prêts à démontrer le Théorème 5.14 de Salomaa et Soittola :

Démonstration. Preuve du Théorème 5.14 de Salomaa *et* Soittola
Notons $\mathcal{A} = \{a_1, ..., a_d\}$. Sans perte de généralité, on peut supposer que pour tout $a \in \mathcal{A}$, il existe $j \in \mathbb{N}$ tel que a apparaît dans $\sigma^j(x)$. En effet, si ce n'est pas le cas, notons $\mathcal{B}$ l'ensemble des lettres qui n'apparaissent dans aucun $\sigma^j(x)$, $j \in \mathbb{N}$. En particulier, les lettres de $\mathcal{B}$ ne sont pas présentes dans le mot x. Notons $\sigma_{\mathcal{A}\setminus\mathcal{B}}$ la restriction de l'endomorphisme σ à l'alphabet $\mathcal{A}\setminus\mathcal{B}$. Par définition, pour tout $k \in \mathbb{N}$, on a $\sigma^k(x) = \sigma^k_{\mathcal{A}\setminus\mathcal{B}}(x)$. Ainsi, le théorème s'appliquera à l'alphabet $\mathcal{A}\setminus\mathcal{B}$ et à l'endomorphisme $\sigma_{\mathcal{A}\setminus\mathcal{B}}$.
Considérons l'ensemble $\mathbb{C}^{\mathcal{A}}_{\mathcal{A}}$ étant l'ensemble des matrices à entrées complexes dont les lignes et colonnes sont indexées par l'alphabet $\mathcal{A}$. Soit $\|\cdot\|_1$ la *norme de Manhattan* sur $\mathbb{C}^{\mathcal{A}}_{\mathcal{A}}$ et, pour tout $k \in \mathbb{N}$, notons $S_k = \|\mathbf{M}_{\sigma^k}\|_1$. Par le Théorème 5.7, il existe $(\beta, \alpha) \in (\mathbb{R}_0 \times \mathbb{N}) \cup \{(0, 0)\}$ tel que $S_k = \Theta(k^\alpha\beta^k)$. Il nous reste donc à montrer que $|\sigma^k(x)| = \Theta(S_k)$. Tout d'abord, remarquons que

$$S_k = \|\mathbf{M}_{\sigma^k}\|_1 = \sum_{a,b\in\mathcal{A}} |\sigma^k(a)|_b.$$

Cependant, pour tout mot $w \in \mathcal{A}^*$, on a $\sum_{b\in\mathcal{A}} |w|_b = |w|$. Dès lors, on obtient

$$S_k = \sum_{a\in\mathcal{A}} |\sigma^k(a)|.$$

Remarquons que β appartient à $\{0\} \cup \mathbb{R}_{\geq 1}$ car $S_k \in \mathbb{N}$, pour tout $k \in \mathbb{N}$. En effet, si $S_{k_0} = 0$ pour un $k_0 \in \mathbb{N}$, alors pour tout $k \geq k_0$, on a $S_k = 0$ et $\beta = 0$ convient. Dans le cas contraire, $S_k \geq 1$ pour tout $k \in \mathbb{N}$. Par définition de la notation de LANDAU, il existe $\mu, \lambda > 0$ et $k_0 \geq 1$ tels que, pour tout $k \geq k_0$, $\mu k^\alpha \beta^k \leq S_k \leq \lambda k^\alpha \beta^k$. Or, si $\beta \in]0,1[$, il existe $k_1 \geq k_0$ tel que, pour tout $k \geq k_1$, $\mu k^\alpha \beta^k, \lambda k^\alpha \beta^k \in]0,1[$. Ceci est absurde.
Ensuite, vu que $|x|_a \leq |x|$ pour toute lettre $a \in \mathcal{A}$, on a

$$|\sigma^k(x)| = \sum_{a\in\mathcal{A}} |x|_a |\sigma^k(a)| \leq \sum_{a\in\mathcal{A}} |x||\sigma^k(a)| = |x| \sum_{a\in\mathcal{A}} |\sigma^k(a)| = |x| S_k.$$

On en tire que $|\sigma^k(x)| = \mathcal{O}(S_k)$. De plus, vu qu'on a supposé que pour toute lettre $a \in \mathcal{A}$, il existe $j \in \mathbb{N}$ tel que a apparaît dans $\sigma^j(x)$, le Lemme 5.23 stipule que pour toute lettre $a \in \mathcal{A}$, $|\sigma^k(a)| = \mathcal{O}(|\sigma^k(x)|)$. Dès lors, on obtient

$$S_k = \sum_{a\in\mathcal{A}} |\sigma^k(a)| = \mathcal{O}(|\sigma^k(x)|).$$

On conclut que $|\sigma^k(x)| = \Theta(S_k)$.

□

Ensuite, en combinant la Proposition 5.4 et le Théorème 5.14 de SALOMAA et SOITTOLA, on obtient le résultat suivant :

Lemme 5.24. *Pour tout $i \in \{1, 2\}$, soient $\mathcal{A}_i$ un alphabet, $\sigma_i : \mathcal{A}_i^* \to \mathcal{A}_i^*$ un endomorphisme et x_i un mot de $\mathcal{A}_i^*$. Au moins une des trois affirmations suivantes est vérifiée lorsque k tend vers l'infini :*

- $|\,\sigma_1^k(x_1)| = o(|\,\sigma_2^k(x_2)|)$,
- $|\,\sigma_1^k(x_1)| = \Theta(|\,\sigma_2^k(x_2)|)$,
- $|\,\sigma_2^k(x_2)| = o(|\,\sigma_1^k(x_1)|)$.

Démonstration. Par le Théorème 5.14, il existe $(\beta_1, \alpha_1), (\beta_2, \alpha_2) \in (\mathbb{R}_{\geq 1} \times \mathbb{N}) \cup \{(0, 0)\}$ tels que $|\,\sigma_1^k(x_1)| = \Theta(k^{\alpha_1}\beta_1^k)$ et $|\,\sigma_2^k(x_2)| = \Theta(k^{\alpha_2}\beta_2^k)$. Si $(\beta_1, \alpha_1) = (\beta_2, \alpha_2)$, alors on a $|\,\sigma_1^k(x_1)| = \Theta(|\,\sigma_2^k(x_2)|)$. Dans le cas contraire, on a $(\beta_i, \alpha_i) \leq_l (\beta_j, \alpha_j)$ où $i, j \in \{1, 2\}$ et $i \neq j$. Ainsi, par la Proposition 5.4, on obtient

$$k^{\alpha_i}\beta_i^k = o(k^{\alpha_j}\beta_j^k)$$

lorsque k tend vers l'infini, c'est-à-dire $|\,\sigma_i^k(x_i)| = o(|\,\sigma_j^k(x_j)|)$.

□

Remarque 5.25. En particulier, le lemme précédent implique que, pour tout endomorphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$, les lettres de $\mathcal{A}$ peuvent être ordonnées selon leurs ordres de croissance sous l'action du morphisme σ.
En effet, pour ordonner un alphabet $\mathcal{A}$, il suffit de considérer deux à deux les lettres de $\mathcal{A}$ et d'appliquer le lemme précédent aux alphabets $\mathcal{A}_1 = \mathcal{A}_2 = \mathcal{A}$, aux morphismes $\sigma_1 = \sigma_2 = \sigma$ et aux mots x_1, x_2 qui sont les deux lettres considérées.

Lemme 5.26. *Soient un endomorphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$, un mot $x \in A^*$ et un entier naturel $k_0 \in \mathbb{N}$. On a*
$$|\,\sigma^{k+k_0}(x)| = \Theta(|\sigma^k(x)|)$$
lorsque k tend vers l'infini.

Démonstration. Procédons par récurrence. Le cas $k_0 = 0$ est trivialement satisfait et vérifions le cas $k_0 = 1$. Par le Théorème 5.14 de Salomaa et Soittola, il existe un couple
$$(\beta, \alpha) \in (\mathbb{R}_{\geq 1} \times \mathbb{N}) \cup \{(0,0)\}$$
tel que $|\sigma^k(x)| = \Theta(k^\alpha \beta^k)$ lorsque k tend vers l'infini. Par la Remarque 5.15, si $(\beta, \alpha) = (0, 0)$, il existe $d \in \mathbb{N}$ tel que $\sigma^d(x) = \varepsilon$. Ainsi, pour tout $k \geq d$, on a
$$|\,\sigma^k(x)| = |\,\sigma^{k+1}(x)| = 0$$
et le résultat en découle. Ensuite, pour tout $(\beta, \alpha) \in (\mathbb{R}_{\geq 1} \times \mathbb{N})$, il est clair que $(k+1)^\alpha \beta^{k+1}$ est asymptotiquement équivalent à $\beta k^\alpha \beta^k$. Ainsi, par le Théorème 5.14 de Salomaa et Soittola, on obtient $|\,\sigma^{k+1}(x)| = \Theta(|\,\sigma^k(x)|)$.
Supposons, dès lors, le résultat vrai pour $k_0 \in \mathbb{N}_0$ et vérifions-le pour $k_0 + 1$. On a
$$|\,\sigma^{k+k_0+1}(x)| = \Theta(|\,\sigma^{k+k_0}(x)|) = \Theta(|\,\sigma^k(x)|).$$
□

Proposition 5.27. *Soient un endomorphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$, un mot $x \in \mathcal{A}^*$ et un entier naturel k_0 tel que $\sigma^{k_0}(x) \neq \varepsilon$. Les deux résultats suivants sont vérifiés lorsque k tend vers l'infini :*

(i) Pour toute lettre $a \in \mathrm{alph}(\sigma^{k_0}(x))$, on a $|\sigma^k(a)| = \mathcal{O}(|\sigma^k(x)|)$.

(ii) Il existe $b \in \mathrm{alph}(\sigma^{k_0}(x))$ tel que $|\sigma^k(b)| = \Theta(|\sigma^k(x)|)$.

Démonstration. L'affirmation (i) est une conséquence directe du Lemme 5.23 car la lettre a est un facteur particulier de $\sigma^{k_0}(x)$ pour un $k_0 \in \mathbb{N}$.
Démontrons l'affirmation (ii). Lorsque $k_0 = 0$, par la Remarque 5.25 nous avons un ordre sur les lettres de $\mathcal{A}$ selon leurs croissances sous l'action du morphisme σ. En particulier,

soit $b \in \mathcal{A}$ maximal selon cet ordre. Pour toute lettre $a \in \mathrm{alph}(x)$, il existe $l_a \geq 1$ tel que $|\sigma^k(a)| \leq l_a|\sigma^k(b)|$. Ainsi, en posant $l = \max\limits_{a \in \mathrm{alph}(x)} l_a$, on obtient $|\sigma^k(a)| \leq l|\sigma^k(b)|$ pour tout $a \in \mathrm{alph}(x)$. De plus, on a

$$|\sigma^k(x)| = \sum_{i=0}^{|x|-1} |\sigma^k(x_i)| \leq \sum_{i=0}^{|x|-1} l|\sigma^k(b)| = |x|l|\sigma^k(b)|.$$

On en déduit que $|\sigma^k(x)| = \mathcal{O}(|\sigma^k(b)|)$. Ainsi, en appliquant le résultat (i) à la lettre b, on obtient $|\sigma^k(b)| = \Theta(|\sigma^k(x)|)$.
Pour le cas général, appliquons le raisonnement précédent au mot $y = \sigma^{k_0}(x)$. Ainsi, il existe $b \in \mathrm{alph}(y) = \mathrm{alph}(\sigma^{k_0}(x))$ tel que $|\sigma^k(b)| = \Theta(|\sigma^k(y)|) = \Theta(|\sigma^k(\sigma^{k_0}(x))|) = \Theta(|\,\sigma^{k+k_0}(x)|)$. Or, par le Lemme 5.26 on a $|\,\sigma^{k+k_0}(x)| = \Theta(|\sigma^k(x)|)$. Ainsi, on obtient $|\sigma^k(b)| = \Theta(|\sigma^k(x)|)$.
□

Proposition 5.28. *Soient un endomorphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ et un réel $\beta \in \mathbb{R}_{\geq 1}$. S'il existe un mot $x \in A^*$ et un naturel α tels que $|\sigma^k(x)| = \Theta(k^\alpha \beta^k)$ lorsque k tend vers l'infini, alors il existe une lettre $b \in \mathcal{A}$ telle que $|\sigma^k(b)| = \mathcal{O}(\beta^k)$ lorsque k tend vers l'infini.*

Démonstration. Tout d'abord, remarquons que pour tout morphisme $\gamma : \mathcal{A}^* \to \mathcal{B}^*$ où $\mathcal{A}$ et $\mathcal{B}$ sont deux alphabets quelconques et tout mot $w \in A^+$, on a

$$|\gamma(w)| = \sum_{i=0}^{|w|-1} |\gamma(w_i)| \geq \sum_{i=0}^{|w|-1} \min_{a \in \mathcal{A}} |\gamma(a)| = |w| \min_{a \in \mathcal{A}} |\gamma(a)|.$$

Ainsi, vu que $|w| \neq 0$, on en tire que

$$\min_{a \in \mathcal{A}} |\gamma(a)| \leq \frac{|\gamma(w)|}{|w|}.$$

Appliquons ce résultat au cas particulier où le morphisme est σ^k et le mot est $\sigma^k(x)$, $k \in \mathbb{N}$. Ce mot est non vide étant donné que $|\sigma^k(x)| = \Theta(k^\alpha \beta^k)$ où $\beta \geq 1$. Ainsi, pour tout $k \in \mathbb{N}$, on obtient

$$\min_{a \in \mathcal{A}} |\sigma^k(a)| \leq \frac{|\sigma^k(w)|}{|w|} = \frac{|\sigma^k(\sigma^k(x))|}{|\sigma^k(x)|} = \frac{|\,\sigma^{2k}(x)|}{|\sigma^k(x)|}. \tag{5.1}$$

De plus, par hypothèse de convergence asymptotique, on a

$$\frac{|\,\sigma^{2k}(x)|}{|\sigma^k(x)|} = \Theta\left(\frac{(2k)^\alpha \beta^{2k}}{(k)^\alpha \beta^k}\right) = \Theta(2^\alpha \beta^k) = \Theta(\beta^k). \tag{5.2}$$

Ensuite, par le Lemme 5.24, il existe une lettre $b \in \mathcal{A}$ telle que pour toute lettre $a \in \mathcal{A}$, $|\sigma^k(b)| = \mathcal{O}(|\sigma^k(a)|)$. Ainsi, cette lettre b satisfait $|\sigma^k(b)| = \Theta(\min\limits_{a \in \mathcal{A}} |\sigma^k(a)|)$. Le résultat $|\sigma^k(b)| = \mathcal{O}(\beta^k)$ est obtenu en utilisant cette convergence asymptotique ainsi que les résultats (5.1) et (5.2).

□

5.2.1 Croissance exponentielle et borne polynomiale

Définition 5.29. Soient un endomorphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ et un mot $x \in \mathcal{A}^*$.

(i) **(Croissance exponentielle)**
Le mot x *croît de façon exponentielle* sous l'action du morphisme σ s'il existe un réel $\beta > 1$ tel que $|\sigma^k(x)| = \Omega(\beta^k)$ lorsque k tend vers l'infini.

(ii) **(Borne polynomiale)**
Le mot x est *borné polynomialement* sous l'action du morphisme σ s'il existe $\alpha \in \mathbb{N}$ tel que $|\sigma^k(x)| = \mathcal{O}(k^\alpha)$ lorsque k tend vers l'infini.

Remarque 5.30. Soient un mot $x \in \mathcal{A}^*$ et σ un endomorphisme de $\mathcal{A}^*$. Par le Théorème 5.14 de Salomaa et Soittola, soit x croît de façon exponentielle soit x est borné polynomialement sous l'action du morphisme σ.
En effet, il existe $(\beta, \alpha) \in (\mathbb{R}_{\geq 1} \times \mathbb{N}) \cup \{(0,0)\}$ tel que $|\sigma^k(x)| = \Theta(k^\alpha \beta^k)$. Dès lors, si $(\beta, \alpha) = (0,0)$, alors, par la Remarque 5.15, le mot x est effacé par une puissance de σ et donc x est borné polynomialement. Ensuite, si $(\beta, \alpha) \in \mathbb{R}_{\geq 1} \times \mathbb{N}$, il existe $\lambda, \mu > 0$ et $k_0 \in \mathbb{N}$ tels que, pour tout $k \geq k_0$,

$$\lambda k^\alpha \beta^k \leq |\sigma^k(x)| \leq \mu k^\alpha \beta^k.$$

Dès lors, si $\beta = 1$ alors $|\sigma^k(x)| = \Theta(k^\alpha)$ et x est borné polynomialement et si $\beta \neq 1$, alors $\lambda \beta^k \leq \lambda k^\alpha \beta^k \leq |\sigma^k(x)|$ et donc $|\sigma^k(x)| = \Omega(\beta^k)$.

Proposition 5.31. *Soient un endomorphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$, un mot $x \in \mathcal{A}^*$ ayant une croissance exponentielle sous l'action du morphisme σ et un couple $(\beta, \alpha) \in \mathbb{R}_{>1} \times \mathbb{N}$ tel que $|\sigma^k(x)| = \Theta(k^\alpha \beta^k)$ lorsque k tend vers l'infini. Soit $\mathcal{C} \subseteq \mathcal{A}$ l'alphabet défini tel que pour toute lettre $c \in \mathcal{C}$, $|\sigma^k(c)| = \Theta(k^\alpha \beta^k)$ lorsque k tend vers l'infini. On a*

$$|\sigma^k(x)|_{\mathcal{C}} = \Theta(\beta^k),$$

lorsque k tend vers l'infini.

Démonstration. Par la Proposition 5.27, remarquons que l'on a $|\sigma^k(x)|_{\mathcal{C}} > 0$ pour tout $k \in \mathbb{N}$. Ensuite, posons la fonction

$$f_{(\beta,\alpha)}(k) = \begin{cases} 1 & \text{si } k = 0 \\ k^\alpha \beta^k & \text{si } k \geq 1 \end{cases}.$$

Par définition, la fonction $f_{(\beta,\alpha)}$ est strictement positive et $|\sigma^k(c)| = \Theta(f_{(\beta,\alpha)}(k))$ pour tout $c \in \mathcal{C}$. Dès lors, pour tout $c \in \mathcal{C}$, il existe $\lambda_c, \mu_c > 0$ et $k_c > 0$ tels que, pour tout $k \geq k_c$,

$$\lambda_c f_{(\beta,\alpha)}(k) \leq |\sigma^k(c)| \leq \mu_c f_{(\beta,\alpha)}(k).$$

Posons $\lambda = \min_{c\in\mathcal{C}} \lambda_c$, $\mu = \max_{c\in\mathcal{C}} \mu_c$ et $k_{\mathcal{C}} = \max_{c\in\mathcal{C}} k_c$. Étant donné que l'alphabet est un ensemble fini et que $\mathcal{C} \subseteq \mathcal{A}$, pour toute lettre $c \in \mathcal{C}$, on a $\lambda > 0$, $\mu > 0$ et pour tout $k \geq k_{\mathcal{C}}$,

$$\lambda f_{(\beta,\alpha)}(k) \leq |\sigma^k(c)| \leq \mu f_{(\beta,\alpha)}(k). \tag{5.3}$$

De plus, le mot x est également tel que $|\sigma^k(x)| = \Theta(k^\alpha \beta^k) = \Theta(f_{(\beta,\alpha)}(k))$ lorsque k tend vers l'infini. Dès lors, il existe $\lambda', \mu' > 0$ et $k_x > 0$ tels que, pour tout $k \geq k_x$, on a

$$\lambda' f_{(\beta,\alpha)}(k) \leq |\sigma^k(x)| \leq \mu' f_{(\beta,\alpha)}(k). \tag{5.4}$$

Ainsi, en posant $k_0 = \max\{k_{\mathcal{C}}, k_x\}$, les inégalités (5.3) et (5.4) sont valables pour tout $k \geq k_0$.
Qui plus est, pour toute lettre $a \in \mathcal{A}$, en appliquant le Lemme 5.24 aux alphabets $\mathcal{A}_1 = \mathcal{A}_2 = \mathcal{A}$, aux morphismes $\sigma_1 = \sigma_2 = \sigma$, aux mots x_1 et x_2 étant respectivement la lettre a et le mot x, les croissances de $\sigma^k(a)$ et $\sigma^k(x)$ peuvent être comparées à l'infini. Pour toute lettre $a \in \mathcal{A}$, l'une des affirmations suivantes est vérifiée.

- $|\,\sigma^k(a)| = o(|\,\sigma^k(x)|)$,
- $|\,\sigma^k(a)| = \Theta(|\,\sigma^k(x)|)$,
- $|\,\sigma^k(x)| = o(|\,\sigma^k(a)|)$,

lorsque k tend vers l'infini. Or, par définition $|\,\sigma^k(x)| = \Theta(f_{(\beta,\alpha)}(k)) = \Theta(k^\alpha\beta^k)$. Dès lors, notons $\mathcal{B}$ le sous-ensemble de l'alphabet $\mathcal{A}$ défini tel que, pour tout $b \in \mathcal{B}$, $|\sigma^k(b)| = o(f_{(\beta,\alpha)}(k))$. Par définition, les lettres $c \in \mathcal{C}$ sont telles que $|\sigma^k(c)| = \Theta(f_{(\beta,\alpha)}(k))$ et les lettres $b \in \mathcal{B}$ sont telles que $|\sigma^k(b)| = o(f_{(\beta,\alpha)}(k))$. Ainsi, par le Lemme 5.24, l'ensemble $\mathcal{B} \cup \mathcal{C}$ correspond aux deux premières comparaisons asymptotiques sur les trois possibles de l'énoncé. Dès lors, $\mathcal{B} \cup \mathcal{C}$ est donc égal à l'ensemble des lettres de $a \in \mathcal{A}$ telles que $|\sigma^k(a)| = \mathcal{O}(|\sigma^k(x)|) = \mathcal{O}(f_{(\beta,\alpha)}(k))$.
Utilisons les éléments établis pour démontrer la thèse énoncée. Soient $j, k \in \mathbb{N}$, $j \geq k_0$, on a

$$\begin{aligned}
|\,\sigma^{j+k}(x)| &= \sum_{i=0}^{|x|-1} |\,\sigma^{j+k}(x_i)| \\
&= \sum_{a\in\mathcal{A}} |\,\sigma^j(a)||\sigma^k(x)|_a \\
&= \sum_{a\in\mathcal{B}\cup\mathcal{C}} |\,\sigma^j(a)||\sigma^k(x)|_a + \sum_{a\in\mathcal{A}\setminus(\mathcal{B}\cup\mathcal{C})} |\,\sigma^j(a)||\sigma^k(x)|_a.
\end{aligned}$$

Or, par la Proposition 5.27, on a $\sigma^k(x) \in (\mathcal{B} \cup \mathcal{C})^*$. On en tire que

$$\sum_{a\in\mathcal{A}\setminus(\mathcal{B}\cup\mathcal{C})} |\,\sigma^j(a)||\sigma^k(x)|_a = 0$$

et vu que $\mathcal{B} \cap \mathcal{C} = \emptyset$, on a

$$|\,\sigma^{j+k}(x)| = \sum_{b\in\mathcal{B}} |\,\sigma^j(b)||\sigma^k(x)|_b + \sum_{c\in\mathcal{C}} |\,\sigma^j(c)||\sigma^k(x)|_c. \tag{5.5}$$

Nous allons maintenant estimer chaque terme du membre de droite de l'égalité (5.5). Si l'on note $g_j = \max_{b\in\mathcal{B}} |\,\sigma^j(b)|$, on a

$$\begin{aligned} 0 \le \sum_{b\in\mathcal{B}} |\,\sigma^j(b)||\sigma^k(x)|_b \le & \sum_{b\in\mathcal{B}} g_j|\sigma^k(x)|_b \\ = & g_j \sum_{b\in\mathcal{B}} |\sigma^k(x)|_b \\ = & g_j|\sigma^k(x)|_{\mathcal{B}}. \end{aligned}$$

Ensuite, pour l'autre terme de la somme, par (5.3), on obtient

$$\begin{aligned} & \sum_{c\in\mathcal{C}} \lambda f_{(\beta,\alpha)}(j)|\sigma^k(x)|_c \le \sum_{c\in\mathcal{C}} |\,\sigma^j(c)||\sigma^k(x)|_c \le \sum_{c\in\mathcal{C}} \mu f_{(\beta,\alpha)}(j)|\sigma^k(x)|_c \\ \Leftrightarrow & \lambda f_{(\beta,\alpha)}(j) \underbrace{\sum_{c\in\mathcal{C}} |\sigma^k(x)|_c}_{=|\sigma^k(x)|_{\mathcal{C}}} \le \sum_{c\in\mathcal{C}} |\,\sigma^j(c)||\sigma^k(x)|_c \le \mu f_{(\beta,\alpha)}(j) \underbrace{\sum_{c\in\mathcal{C}} |\sigma^k(x)|_c}_{=|\sigma^k(x)|_{\mathcal{C}}}. \end{aligned}$$

De plus, en sachant que l'on a $\lambda' f_{(\beta,\alpha)}(j+k) \le |\,\sigma^{j+k}(x)| \le \mu' f_{(\beta,\alpha)}(j+k)$, et par les deux inégalités trouvées, on voit d'une part que

$$\begin{aligned} \lambda' f_{(\beta,\alpha)}(j+k) \le |\,\sigma^{j+k}(x)| = & \sum_{b\in\mathcal{B}} |\,\sigma^j(b)||\sigma^k(x)|_b + \sum_{c\in\mathcal{C}} |\,\sigma^j(c)||\sigma^k(x)|_c \\ \le & \; g_j|\sigma^k(x)|_{\mathcal{B}} + \mu f_{(\beta,\alpha)}(j)|\sigma^k(x)|_{\mathcal{C}} \end{aligned} \tag{5.6}$$

et d'une autre part que

$$0 + \lambda f_{(\beta,\alpha)}(j)|\sigma^k(x)|_{\mathcal{C}} \le |\,\sigma^{j+k}(x)| \le \mu' f_{(\beta,\alpha)}(j+k). \tag{5.7}$$

Ensuite, dans les inégalités (5.6), en soustrayant $g_j|\sigma^k(x)|_{\mathcal{B}}$ et en divisant ensuite chaque membre par $\mu f_{(\beta,\alpha)}(j)$, on obtient

$$\frac{\lambda' f_{(\beta,\alpha)}(j+k) - g_j|\sigma^k(x)|_{\mathcal{B}}}{\mu f_{(\beta,\alpha)}(j)} \le \frac{|\,\sigma^{j+k}(x)| - g_j|\sigma^k(x)|_{\mathcal{B}}}{\mu f_{(\beta,\alpha)}(j)} \le |\sigma^k(x)|_{\mathcal{C}}. \tag{5.8}$$

Or, en utilisant les extrêmes de l'inégalité (5.7), on a

$$|\sigma^k(x)|_{\mathcal{C}} \le \frac{\mu' f_{(\beta,\alpha)}(j+k)}{\lambda f_{(\beta,\alpha)}(j)}. \tag{5.9}$$

Ainsi, en combinant les extrêmes des inégalités (5.8) et (5.9), on obtient

$$\frac{\lambda' f_{(\beta,\alpha)}(j+k)}{\mu f_{(\beta,\alpha)}(j)} - \frac{g_j|\sigma^k(x)|_{\mathcal{B}}}{\mu f_{(\beta,\alpha)}(j)} \le |\sigma^k(x)|_{\mathcal{C}} \le \frac{\mu' f_{(\beta,\alpha)}(j+k)}{\lambda f_{(\beta,\alpha)}(j)}. \tag{5.10}$$

Or, à un $k \in \mathbb{N}$ fixé, on a

$$\lim_{j\to+\infty} \frac{f_{(\beta,\alpha)}(j+k)}{f_{(\beta,\alpha)}(j)} = \lim_{j\to+\infty} \frac{(k+j)^\alpha \beta^{k+j}}{j^\alpha \beta^j} = \beta^k$$

et par définition de $\mathcal{B}$ et de g_j, on a $g_j = o(f_{(\beta,\alpha)}(j))$, c'est-à-dire

$$\lim_{j\to+\infty} \frac{g_j}{f_{(\beta,\alpha)}(j)} = 0.$$

Ainsi, en prenant la limite pour j tendant vers l'infini dans chaque membre de l'inégalité (5.10), il résulte que

$$\frac{\lambda'}{\mu}\beta^k \leq |\sigma^k(x)|_{\mathcal{C}} \leq \frac{\mu'}{\lambda}\beta^k$$

avec $\frac{\lambda'}{\mu}$ et $\frac{\mu'}{\lambda} > 0$. On en conclut que $|\sigma^k(x)|_{\mathcal{C}} = \Theta(\beta^k)$.

□

La Proposition 5.31 affirme que la croissance exponentielle du mot x sous l'action du morphisme σ implique la croissance exponentielle également de $|\sigma^k(x)|_{\mathcal{C}}$ lorsque k tend vers l'infini. L'exemple suivant illustre la Proposition 5.31.

Exemple 5.32. Considérons l'alphabet binaire $\mathcal{A} = \{a,b\}$ et un morphisme $\sigma : \{a,b\}^* \to \{a,b\}^*$ défini par $\sigma(a) = aba$ et $\sigma(b) = bb$. Regardons les croissances des lettres a et b sous l'action du morphisme σ. Tout d'abord, nous pouvons remarquer que $\sigma^k(b) = b^{2^k}$. En effet, on a $\sigma^0(b) = b = b^{2^0}$ et

$$\sigma^{k+1}(b) = \sigma(\sigma^k(b)) = \sigma(b^{2^k}) = (\sigma(b))^{2^k} = (bb)^{2^k} = (b^2)^{2^k} = b^{2^{k+1}}.$$

Ainsi, on obtient $|\sigma^k(b)| = 2^k$. Ensuite, montrons que $|\sigma^k(a)| = \left(\frac{1}{2}k+1\right)2^k$. Par définition,

$$|\,\sigma^0(a)| = |a| = 1 = (\frac{1}{2}0+1)2^0$$

et, par récurrence,

$$\begin{aligned}
|\,\sigma^{k+1}(a)| &= |\sigma^k(aba)| \\
&= |\sigma^k(a)| + |\sigma^k(b)| + |\sigma^k(a)| \\
&= 2|\sigma^k(a)| + |\sigma^k(b)| \\
&= 2\left(\frac{1}{2}k+1\right)2^k + 2^k \\
&= (k+1)2^k + 2^{k+1} \\
&= \left(\frac{1}{2}(k+1)+1\right)2^{k+1}.
\end{aligned}$$

En conclusion, les lettres a et b sont telles que $|\sigma^k(a)| = \Theta(k2^k)$ et $|\sigma^k(b)| = \Theta(2^k)$. Elles croissent donc de façon exponentielle sous l'action du morphisme σ. Par la Proposition 5.31, on a $|\sigma^k(a)|_a = \Theta(2^k)$ et $|\sigma^k(b)|_b = \Theta(2^k)$. La seconde affirmation est trivialement vérifiée car on a $|\sigma^k(b)|_b = 2^k$ au vu de la définition du morphisme. L'affirmation $|\sigma^k(a)|_a = \Theta(2^k)$ est également vérifiée car, plus particulièrement, on a $|\sigma^k(a)|_a = 2^k$. En effet, en procédant par induction, on a $|\,\sigma^0(a)|_a = |a|_a = 2^0$ et

$$|\,\sigma^{k+1}(a)|_a = |\sigma^k(aba)|_a = |\sigma^k(a)|_a + |\sigma^k(b)|_a + |\sigma^k(a)|_a = 2^k + 0 + 2^k = 2^{k+1}.$$

Lemme 5.33. *Soient un endomorphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ et un mot $x \in \mathcal{A}^*$. Si x croît de façon exponentielle sous l'action du morphisme σ, alors*

$$\sum_{j=0}^{k-1} |\,\sigma^j(x)| = \Theta(|\sigma^k(x)|)$$

lorsque k tend vers l'infini.

Démonstration. Soit un couple $(\beta, \alpha) \in \mathbb{R}_{>1} \times \mathbb{R}$ et considérons la fonction $f_{(\beta,\alpha)}$ définie dans la preuve de la Proposition 5.31. Par définition, en notant $f_{(\beta,\alpha)}(k) = f_k$ pour tout $k \in \mathbb{N}$, la suite $(f_k)_{k\in\mathbb{N}}$ est strictement positive et, pour tout $k \geq 1$, on a

$$\lim_{k\to\infty} \frac{f_k}{f_{k-1}} = \lim_{k\to\infty} \frac{k^\alpha \beta^k}{(k-1)^\alpha \beta^{k-1}} = \beta.$$

Ainsi, on a $\lim\limits_{k\to\infty} \frac{f_k}{\beta f_{k-1}} = 1$ c'est-à-dire $\beta f_{k-1} \sim f_k$. Dès lors, en appliquant le Lemme 5.12, on obtient

$$\sum_{j=0}^{k-1} f_j \sim \frac{\beta}{\beta - 1} f_{k-1} \sim \frac{1}{\beta - 1} f_k = \Theta(f_k).$$

Jusqu'alors, le couple (β, α) était quelconque. Cependant, le théorème de Salomaa et Soittola 5.14 affirme qu'il peut être choisi de sorte que $|\,\sigma^j(x)| = \Theta(f_j)$ lorsque j tend vers l'infini. Dès lors, on en tire que

$$\sum_{j=0}^{k-1} |\,\sigma^j(x)| = \sum_{j=0}^{k-1} \Theta(f_j) = \Theta\left(\sum_{j=0}^{k-1} f_j\right) = \Theta(f_k) = \Theta(|\sigma^k(x)|).$$

□

5.3 Facteurs centraux

Une nouvelle notion va être introduite dans cette section afin d'établir un théorème caractérisant l'ensemble des facteurs d'un mot morphique.

Définition 5.34. Soit (x, y, z) un triplet de mots. Un mot w est un *facteur central* de (x, y, z) s'il existe un suffixe non vide x' de x et un préfixe non vide z' de z tels que $w = x'yz'$.

Au vu de la définition précédente, s'il existe un facteur central d'un triplet de mots (x, y, z), alors on a $x \neq \varepsilon$ et $z \neq \varepsilon$. Cependant, aucune condition sur le mot y n'est demandée. De plus, si w est un facteur central de (x, y, z), alors w est en particulier un facteur du mot xyz et si $y \neq \varepsilon$, alors y est un facteur propre de w.

Exemple 5.35. Considérons le triplet $(x, y, z) = (aba, aa, ba)$. Le mot xyz est $abaaaba$ et le mot $w = baaa$ est en particulier un facteur de xyz mais n'est pas un facteur central de (x, y, z). En effet, il contient le suffixe non vide ba de x concaténé avec le mot $y = aa$ mais le préfixe de z considéré est le mot vide. Étant donné que les suffixes non vides possibles de x sont $\{aba, ba, a\}$ et que les préfixes non vides possibles de z sont $\{ba, b\}$, l'ensemble des facteurs centraux de (x, y, z) est $\{abaaaba, abaaab, baaaba, baaab, aaaba, aaab\}$.

Remarque 5.36. Soient un entier $n \in \mathbb{N}_0$ et un triplet de mots (x, y, z). Il existe au plus $n - 1$ facteurs centraux de (x, y, z) de longueur n :
En effet, notons $m \in \mathbb{N}$ la longueur du mot y et considérons E l'ensemble des facteurs centraux de longueur n de (x, y, z). Par définition, pour tout $w \in E$, on a $w = x'yz'$ où x' (resp. z') est un suffixe (resp. préfixe) non vide de x (resp. z). Dès lors, l'inégalité $|w| = n \geq m + 2$ doit être vérifiée. Or, dans le mot w apparaît le facteur y. Ainsi, il faut partitionner la longueur du mot restant (égale à $n - m$) en un suffixe non vide x' de x et un préfixe non vide z' de z. Les possibilités sont les suivantes :

$\|x'\|$	$\|z'\|$
1	$n - m - 1$
2	$n - m - 2$
$\vdots$	$\vdots$
$n - m - 1$	1

Il y a $n - m - 1$ possibilités et on obtient $\#(E) = n - m - 1$. Ce cardinal est maximal lorsque $|y| = m = 0$, c'est-à-dire lorsque y est le mot vide.

Définition 5.37. Soient un endomorphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ et x, y, z des mots de $\mathcal{A}^*$. L'ensemble des mots $w \in \mathcal{A}^*$ tels qu'il existe un nombre naturel k pour lequel w est un facteur central de $(\sigma^k(x), \sigma^k(y), \sigma^k(z))$ est noté $L^\sigma(x, y, z)$. De plus, pour tout $n \in \mathbb{N}$, on note

$$L_n^\sigma(x, y, z) := L^\sigma(x, y, z) \cap \mathcal{A}^n$$

Remarque 5.38. Considérons des mots x, y_1, y_2 et z de $\mathcal{A}^*$ et un morphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$.

(i) Tout facteur central de (x, y_1y_2, z) est un facteur central de (x, y_1, y_2z) et de (xy_1, y_2, z). *En effet*, considérons un facteur central w de (x, y_1y_2, z). Par définition, il existe $x' \in \mathrm{Suff}(x) \setminus \{\varepsilon\}$ et $z' \in \mathrm{Pref}(z) \setminus \{\varepsilon\}$ tels que

$$w = x'(y_1y_2)z'.$$

En particulier, on a

$$w = (x'y_1)y_2z' \quad \text{et} \quad w = x'y_1(y_2z')$$

avec $x'y_1 \in \mathrm{Suff}(xy_1) \setminus \{\varepsilon\}$ et $y_2z' \in \mathrm{Pref}(y_2z) \setminus \{\varepsilon\}$. Ainsi, le mot w est un facteur central de (xy_1, y_2, z) et de (x, y_1, y_2z). Remarquons que la réciproque n'est pas vérifiée. En effet, un contre-exemple est établi en considérant les mots $x = ba$, $y_1 = baab$, $y_2 = aab$ et $z = aa$. Le mot $w = abaaba$ est un facteur central de $(x, y_1, y_2z) = (ba, baab, aabaa)$ et de $(xy_1, y_2, z) = (babaab, aab, aa)$ mais il n'est pas un facteur central de $(x, y_1y_2, z) = (ba, baabaab, aa)$.

(ii) L'ensemble $L_n^\sigma(x, y_1y_2, z)$ est un sous-ensemble de $L_n^\sigma(x, y_1, y_2z) \cap L_n^\sigma(xy_1, y_2, z)$.
En effet, par le même type de raisonnement, considérons un mot $w \in L_n^\sigma(x, y_1y_2, z)$. Par définition, il existe $k \in \mathbb{N}$ tel que w est un facteur central de

$$(\sigma^k(x), \sigma^k(y_1y_2), \sigma^k(z)) = (\sigma^k(x), \sigma^k(y_1)\sigma^k(y_2), \sigma^k(z)).$$

Or, par le point (i), tout facteur central de $(\sigma^k(x), \sigma^k(y_1)\sigma^k(y_2), \sigma^k(z))$ est un facteur central de

$$(\sigma^k(x), \sigma^k(y_1), \sigma^k(y_2)\sigma^k(z)) = (\sigma^k(x), \sigma^k(y_1), \sigma^k(y_2z))$$

et de

$$(\sigma^k(x)\sigma^k(y_1), \sigma^k(y_2), \sigma^k(z)) = (\sigma^k(xy_1), \sigma^k(y_2), \sigma^k(z)).$$

On en déduit que $w \in L_n^\sigma(xy_1, y_2, z) \cap L_n^\sigma(x, y_1, y_2z)$. Le même contre-exemple démontre que l'inclusion inverse n'est pas valide. En effet, il s'agit d'un cas particulier où le morphisme σ est l'identité.

Notation 5.39. Si x est un facteur de y, on note $x \sqsubseteq y$.

Munis de ces définitions, nous allons à présent prouver un théorème crucial pour la démonstration du théorème de Pansiot.

Théorème 5.40. *Soient un mot purement morphique $u \in \mathcal{A}^\mathbb{N}$ et un endomorphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ engendrant le mot u. Il existe un sous-ensemble fini G de $\mathcal{A} \times \mathcal{A}^+ \times \mathcal{A}$ tel que*

$$L_n(u) = \bigcup_{(a,w,a') \in G} L_n^\sigma(a, w, a')$$

pour tout $n \geq 3$.

Ce théorème peut être reformulé comme suit : pour tout mot $x \in \mathcal{A}^*$ de longueur au moins 3, x est un facteur de u si et seulement si il existe un triplet $(a, w, a') \in G$ tel que x appartient à $L_n^\sigma(a, w, a')$.

Démonstration. Posons G l'ensemble de tous les triplets $(a, w, a') \in \mathcal{A} \times \mathcal{A}^+ \times \mathcal{A}$ tels que $|w| \leq 2\,\|\sigma\| - 2$ et $awa' \in L(u)$.
Démontrons l'égalité d'ensembles désirée par double inclusion.
Tout d'abord, considérons un triplet $(a, w, a') \in G$ et fixons un naturel k. Le mot u est un point fixe du morphisme σ. Ainsi, si awa' est un facteur de u, alors $\sigma^k(awa')$ en est également un. Dès lors, tout facteur central du triplet $(\sigma^k(a), \sigma^k(w), \sigma^k(a'))$ est en particulier un facteur de $\sigma^k(awa')$ et donc un facteur de u.
Démontrons l'autre inclusion. Considérons un facteur x de u tel que $|x| \geq 3$. Notons u_0 la première lettre de u, on a $u = \sigma^\omega(u_0)$. Étant donné que x est un facteur de u, notons $r \in \mathbb{N}$ le plus petit entier tel que x soit un facteur de $\sigma^r(u_0)$. Définissons la suite $(x_i)_{0 \leq i \leq r}$ comme suit : $x_r = x$ et, pour tout $i \in [\![0, r-1]\!]$, x_i est le plus court facteur de $\sigma^i(u_0)$ tel que x_{i+1} est un facteur de $\sigma(x_i)$.

En particulier, on a $x = x_r \in L(\sigma^r(u_0))$. De plus, pour tout $i \in [\![0, r-1]\!]$, la construction des x_i est toujours possible de proche en proche. En effet, lorsque le mot x_{i+1} est défini, alors il est un facteur de $\sigma^{i+1}(u_0)$ et vu que $\sigma(\sigma^i(u_0)) = \sigma^{i+1}(u_0)$, le facteur x_i peut être construit de telle sorte que x_i est un facteur de longueur minimale de $\sigma^i(u_0)$, tel que x_{i+1} est un facteur de $\sigma(x_i) \sqsubseteq \sigma^{i+1}(u_0)$. En particulier, on a

$$x_r \sqsubseteq \sigma(x_{r-1}) \sqsubseteq \sigma(\sigma(x_{r-2})) = \sigma^2(x_{r-2}) \sqsubseteq \cdots \sqsubseteq \sigma^r(x_0).$$

Posons

$$q = 1 + \max\{i \in [\![0, r]\!] \mid |x_i| \leq 2\}.$$

L'entier q est bien défini et on a $1 \leq q \leq r$. En effet, remarquons tout d'abord que $q < 1+r$ car $\max\{i \in [\![0, r]\!] \mid |x_i| \leq 2\} \neq r$ étant donné que $x_r = x$ et $|x| \geq 3$. De plus, le maximum n'est jamais considéré sur un ensemble vide car, par construction, on sait que x_r est un facteur de $\sigma^r(u_0)$ et u_0 est un facteur de $\sigma^i(u_0)$ pour tout i vu que le morphisme σ est prolongeable en u_0. Ainsi, par minimalité à chaque étape, $x_0 = u_0$ et $|u_0| = 1$.
Pour tout $i \in [\![q, r]\!]$, par définition, on a $|x_i| \geq 3$. Dès lors, le mot x_i peut être écrit sous la forme $x_i = a_i w_i a'_i$ avec $(a_i, w_i, a'_i) \in \mathcal{A} \times \mathcal{A}^+ \times \mathcal{A}$. Par construction, pour tout $i \in [\![q, r-1]\!]$, le mot x_{i+1} est un facteur de

$$\sigma(x_i) = \sigma(a_i w_i a'_i) = \sigma(a_i)\, \sigma(w_i)\, \sigma(a'_i)$$

où x_i avait été choisi par minimalité, c'est-à-dire les mots $a_i w_i$ et $w_i a'_i$ n'auraient en particulier pas satisfaits à la propriété de définition du mot x_i. Ainsi, par définition, le mot x_{i+1} est un facteur central de $(\sigma(a_i), \sigma(w_i), \sigma(a'_i))$ pour tout $i \in [\![q, r-1]\!]$. De manière générale, le mot x_j est un facteur central de

$$(\sigma^{j-i}(a_i), \sigma^{j-i}(w_i), \sigma^{j-i}(a'_i)) \tag{5.11}$$

pour tous $i, j \in [\![q, r]\!]$ tels que $i \leq j$. En effet, le cas $i = j$ est vérifié car $x_i = a_i w_i a'_i$ est en particulier un facteur central de (a_i, w_i, a'_i) et le cas $j = i+1$ vient d'être considéré. Par récurrence, à un $i \in [\![q, r-1]\!]$ fixé, supposons donc avoir démontré la véracité de cette affirmation pour tout j tel que $i \leq j \leq l$, $l \in [\![q, r-1]\!]$ et démontrons-le pour $l+1 \leq r$. Par construction x_{l+1} est un facteur de $\sigma(x_l)$ et, par hypothèse de récurrence, x_l est un facteur central de $(\sigma^{l-i}(a_i), \sigma^{l-i}(w_i), \sigma^{l-i}(a'_i))$. En particulier, on a $x_l \sqsubseteq \sigma^{l-i}(a_i w_i a'_i)$. Ainsi,

$$x_{l+1} \sqsubseteq \sigma(x_l) \sqsubseteq \sigma(\sigma^{l-i}(a_i w_i a'_i)) = \sigma^{l+1-i}(a_i w_i a'_i) = \sigma^{l+1-i}(a_i)\, \sigma^{l+1-i}(w_i)\, \sigma^{l+1-i}(a'_i).$$

De plus, le fait que les suffixe et préfixe de $\sigma^{l+1-i}(a_i)$ et $\sigma^{l+1-i}(a'_i)$ soient non vides est garanti par la minimalité de la construction de la suite $(x_i)_{0 \leq i \leq r}$.
Dès lors, en appliquant (5.11) avec $i = q$ et $j = r$ et en posant $k = r - q$, on obtient que le mot $x_r = x$ est un facteur central de $(\sigma^k(a_q), \sigma^k(w_q), \sigma^k(a'_q))$ où (a_q, w_q, a'_q) est un triplet de G. En effet, par construction $(a_q, w_q, a'_q) \in \mathcal{A} \times \mathcal{A}^+ \times \mathcal{A}$ et le mot $a_q w_q a'_q$ est un facteur de $u = \sigma^\omega(u_0)$ car $x_q = a_q w_q a'_q$ et x_q est un facteur de $\sigma^q(u_0)$. Finalement, étant donné que x_q est un facteur de $\sigma(x_{q-1})$ et par définition de q, on a $|x_{q-1}| \leq 2$ et $|w_q| = |x_q| - 2 \leq |\sigma(x_{q-1})| - 2 \leq 2\,\|\sigma\| - 2$. Ainsi, le facteur x de u considéré est dans l'ensemble

$$\bigcup_{(a,w,a') \in G} L_n^\sigma(a, w, a').$$

□

Remarque 5.41. L'union du Théorème 5.40 n'est pas nécessairement disjointe.
En effet, un exemple est donné par le morphisme $\sigma : \{a,b\}^* \to \{a,b\}^*$ défini tel que $\sigma(a) = ab$ et $\sigma(b) = ab$. Le mot infini u engendré par σ en débutant par a est donné par $u = (ab)^\omega$. Les couples $(b,a,b), (a,b,a)$ appartiennent à l'ensemble G. Or, par exemple,

$$L_6^\sigma(b,a,b) \cap L_6^\sigma(a,b,a) \neq \emptyset$$

car le mot $ababab$ appartient aux deux ensembles car

$$(\sigma(b), \sigma(a), \sigma(b)) = (ab, ab, ab) = (\sigma(a), \sigma(b), \sigma(a)).$$

L'énoncé du Théorème 5.40 peut être légèrement modifié afin de prendre en compte également les mots de longueur 2.

Corollaire 5.42. *Avec les notations du Théorème 5.40, il existe un sous-ensemble fini H de $\mathcal{A} \times \mathcal{A}^* \times \mathcal{A}$ tel que*

$$L(u) \setminus (\mathcal{A} \cup \{\varepsilon\}) = \bigcup_{(a,w,a') \in H} L^\sigma(a,w,a').$$

Démonstration. Notons G_0 l'ensemble de triplets $(a, \varepsilon, a') \in \mathcal{A} \times \{\varepsilon\} \times \mathcal{A}$ tels que $aa' \in L_2(u)$. Démontrons, à nouveau par double inclusion, que l'ensemble $H = G_0 \cup G$ convient.
Considérons un mot $x \in L(u) \setminus (\mathcal{A} \cup \{\varepsilon\})$. Si $|x| \geq 3$, alors par le Théorème 5.40, on a

$$x \in \bigcup_{(a,w,a') \in G} L_{|x|}^\sigma(a,w,a').$$

Or, on a $L_{|x|}^\sigma(a,w,a') \subseteq L^\sigma(a,w,a')$ et $G \subseteq H$ donc

$$x \in \bigcup_{(a,w,a') \in H} L^\sigma(a,w,a').$$

Si $|x| = 2$, alors notons $x = x_0x_1 \in L_2(u)$. On a $(x_0, \varepsilon, x_1) \in G_0$ et x est en particulier un facteur central de (x_0, ε, x_1). Ainsi, en prenant le k de la Définition 5.37 égal à 0, le mot x appartient à $L^\sigma(x_0, \varepsilon, x_1)$.
De plus, l'autre inclusion est également vérifiée. En effet, considérons un mot

$$x \in \bigcup_{(a,w,a') \in H} L^\sigma(a,w,a') = \left(\bigcup_{(a,\varepsilon,a') \in G_0} L^\sigma(a,w,a') \right) \cup \left(\bigcup_{(a,w,a') \in G} L^\sigma(a,w,a') \right).$$

Par le Théorème 5.40, étant donné que $L^\sigma(a,w,a') = \bigcup_{n \in \mathbb{N}} L_n^\sigma(a,w,a')$, si $x \in \bigcup_{(a,w,a') \in G} L^\sigma(a,w,a')$, alors x est un facteur de longueur au moins 3 de u et en particulier

$x \in L(u) \setminus (\mathcal{A} \cup \{\varepsilon\})$. Ensuite, si $x \in \bigcup_{(a,\varepsilon,a') \in G_0} L^\sigma(a, w, a')$, remarquons alors que pour tous $(a, \varepsilon, a') \in G_0$ et $k \in \mathbb{N}$, on a $\sigma^k(a\varepsilon a') = \sigma^k(aa')$. Or, $aa' \in L_2(u)$ par définition de G_0, dès lors vu que u est engendré par σ, $\sigma^k(aa')$ et ses facteurs sont également des facteurs de u. $\square$

Exemple 5.43. Considérons le morphisme $\sigma : \{0,1\}^* \to \{0,1\}^*$ défini tel que $\sigma(0) = 01$ et $\sigma(1) = 10$ et notons t le mot de Thue-Morse engendré par ce morphisme [3]. Avec les notations du Théorème 5.40 et du Corollaire 5.42, construisons l'ensemble $H = G_0 \cup G$. L'ensemble G_0 est constitué des triplets $(0, \varepsilon, 1), (0, \varepsilon, 0), (1, \varepsilon, 0)$ et $(1, \varepsilon, 1)$. Ensuite, l'ensemble G est constitué des triplets dont la deuxième composante est un mot w de longueur inférieure ou égale à $2\|\sigma\| - 2 = 2$. Ainsi, les w possibles sont $0, 1, 00, 01, 10$ et 11 et étant donné que le mot de Thue-Morse ne contient pas de cube, les triplets appartenant à G sont donc :

(a, w, a')	$a = 0, a' = 0$	$a = 0, a' = 1$	$a = 1, a' = 0$	$a = 1, a' = 1$
$w = 0$	$\cancel{(0, 0, 0)}$	$(0, 0, 1)$	$(1, 0, 0)$	$(1, 0, 1)$
$w = 1$	$(0, 1, 0)$	$(0, 1, 1)$	$(1, 1, 0)$	$\cancel{(1, 1, 1)}$
$w = 00$	$\cancel{(0, 00, 0)}$	$\cancel{(0, 00, 1)}$	$\cancel{(1, 00, 0)}$	$(1, 00, 1)$
$w = 01$	$(0, 01, 0)$	$(0, 01, 1)$	$(1, 01, 0)$	$(1, 01, 1)$
$w = 10$	$(0, 10, 0)$	$(0, 10, 1)$	$(1, 10, 0)$	$(1, 10, 1)$
$w = 11$	$(0, 11, 0)$	$\cancel{(0, 11, 1)}$	$\cancel{(1, 11, 0)}$	$\cancel{(1, 11, 1)}$

L'ensemble H est entièrement déterminé et on a $L(t) \setminus (\mathcal{A} \cup \{\varepsilon\}) = \bigcup_{(a,w,a') \in H} L^\sigma(a, w, a')$.

5.4 Morphismes partout croissants

Une catégorie de morphismes, appelée *morphismes partout croissants*, va être étudiée dans cette section afin de pouvoir établir une première partie du théorème de Pansiot.

Définition 5.44. Soient $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ un endomorphisme et x un mot de $\mathcal{A}^*$.
Si $|\sigma^k(x)|$ tend vers l'infini lorsque k tend vers l'infini, on dit que le mot x est *croissant sous l'action du morphisme* σ ou que x *croît sous l'action du morphisme* σ. Sinon, la longueur $|\sigma^k(x)|$ est bornée et le mot x est dit *borné sous l'action du morphisme* σ.

Remarque 5.45. Par la Définition 4.7, si un endomorphisme σ est prolongeable en une lettre a, alors la lettre a est en particulier croissante sous l'action de σ.

Définissons maintenant la notion de périodicité d'une suite d'éléments quelconques.

Définition 5.46. Une suite $(a(n))_{n\in\mathbb{N}}$ est ultimement périodique s'il existe deux entiers $M, N \in \mathbb{N}_0$ tels que $M \geq N$ et pour tout $n > M$, $a(n) = a(n - N)$. Les entiers naturels N et $M - N$ sont appelés respectivement une période et une prépériode de la suite considérée.

3. Étant donné les notations (a, w, a') pour les triplets de l'ensemble H, le morphisme de Thue-Morse est défini sur l'alphabet binaire $\{0, 1\}$ et non pas $\{a, b\}$ comme habituellement dans ce mémoire.

Remarque 5.47. Dans le cas où le mot x est borné sous l'action du morphisme σ, la suite $(\sigma^k(x))_{k\geq 0}$ est ultimement périodique.
En effet, si la longueur $|\sigma^k(x)|$ est bornée, il existe $M \in \mathbb{N}$ tel que $|\sigma^k(x)| \leq M$ pour tout $k \in \mathbb{N}$. L'ensemble $\{\sigma^k(x) \,|\, k \in \mathbb{N}\}$ est un sous-ensemble de l'ensemble fini $\mathcal{A}^{\leq M}$. Dès lors, il existe $k_1, k_2 \in \mathbb{N}$ tels que $k_1 < k_2$ et $\sigma^{k_1}(x) = \sigma^{k_2}(x)$. Cela revient à dire que les entiers $k_2 - k_1$ et k_1 sont respectivement une période et une prépériode de la suite $(\sigma^k(x))_{k\geq 0}$.

Exemple 5.48. Considérons le morphisme $\sigma : \{a,b\}^* \to \{a,b\}^*$ défini par $\sigma(a) = ab$ et $\sigma(b) = b$. La lettre a est croissante sous l'action du morphisme σ car on a $\sigma^k(a) = ab^k$, ainsi $|\sigma^k(a)| = k+1$ tend vers l'infini lorsque k tend vers l'infini. Cependant, la lettre b est bornée sous l'action de σ car $|\sigma^k(b)| = |b| = 1$ pour tout $k \in \mathbb{N}$.

Définition 5.49. Un endomorphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ est dit *partout croissant* si toutes les lettres de l'alphabet $\mathcal{A}$ sont croissantes sous l'action de σ.

Exemple 5.50. Considérons le morphisme σ défini dans l'Exemple 5.48. Vu que la lettre b est bornée, l'endomorphisme σ n'est pas partout croissant. Cependant, par la Remarque 5.45, l'endomorphisme de THUE-MORSE $\sigma : \{a,b\}^* \to \{a,b\}^*$ défini par $\sigma(a) = ab$ et $\sigma(b) = ba$ est partout croissant.

Remarque 5.51. Soit $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ un endomorphisme partout croissant. Tout point fixe de σ est un mot infini purement morphique généré par σ.
En effet, considérons un mot x point fixe de σ. En particulier, on a $\sigma^k(x) = x$ pour tout $k \in \mathbb{N}$ et le mot x est infini car $|\sigma^k(x)|$ tend vers l'infini lorsque k tend vers l'infini. Ensuite, étant donné qu'on a $\sigma(x) = x$, l'image de x_0 par σ doit débuter par x_0 et on a $\sigma(x_0) = x_0 u$ où $u \in \mathcal{A}^+$ car la lettre x_0 est croissante sous l'action de σ. Or, pour tout k, le mot $\sigma^k(x_0)$, infini par croissance, est un préfixe de x, ainsi le mot infini purement morphique $x' = \lim\limits_{k\to+\infty} \sigma^k(x_0)$ est égal au mot x.
De plus, étant donné que la réciproque est donnée par la Définition 4.7, dans le cas d'un endomorphisme partout croissant, le critère est nécessaire et suffisant.
Dans le cas où une lettre de l'alphabet est bornée sous l'action de σ, le résultat est faux et un contre-exemple est donné par la lettre b dans l'Exemple 5.48.

Le but de cette section est de démontrer que la fonction de complexité d'un point fixe d'un morphisme partout croissant appartient à l'une des trois classes de complexités asymptotique suivantes :

$$\mathcal{O}(n), \Theta(n \log\log n) \text{ ou } \Theta(n \log n).$$

Rappelons que si la fonction de complexité d'un mot infini satisfait $p(n) = \mathcal{O}(n)$, alors par le Théorème 2.2 de MORSE et HELDLUND, on a $p(n) = \Theta(1)$ ou $p(n) = \Theta(n)$.

Définition 5.52. Un morphisme $\tau : \mathcal{A}^* \to \mathcal{B}^*$ est *expansif* si $|\tau(a)| \geq 2$ pour tout $a \in \mathcal{A}$.

Remarques 5.53.

(i) Tout endomorphisme expansif est partout croissant.
En effet, toute lettre $a \in \mathcal{A}$ est croissante sous l'action de σ car on a $|\sigma^k(a)| \geq 2^k$

pour tout $k \in \mathbb{N}$. Cependant, un endomorphisme peut être partout croissant sans pour autant être expansif. Un exemple est donné par l'endomorphisme de FIBONACCI $\sigma : \{a,b\}^* \to \{a,b\}^*$ défini tel que $\sigma(a) = ab$ et $\sigma(b) = a$. Cet endomorphisme n'est pas expansif car l'image de la lettre b est de longueur 1 mais vu que $\sigma(b) = a$, on a

$$\lim_{k \to +\infty} |\sigma^k(a)| = \lim_{k \to +\infty} |\sigma^k(b)| = +\infty.$$

(ii) Tout endomorphisme partout croissant admet une puissance expansive.
En effet, considérons un endomorphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ partout croissant. Par définition, pour toute lettre $a \in \mathcal{A}$, $|\sigma^k(a)|$ tend vers l'infini lorsque k tend vers l'infini. Ainsi, pour tout $a \in \mathcal{A}$, il existe $M_a \in \mathbb{N}$ tel que pour tout $m \geq M_a$, $|\sigma^m(a)| \geq 2$. En considérant $M = \max_{a \in \mathcal{A}} M_a$, on obtient

$$\forall a \in \mathcal{A}, \forall m \geq M, |\sigma^m(a)| \geq 2.$$

Ainsi, la puissance σ^M de σ est expansive.

De par ces deux remarques, il s'ensuit que les endomorphismes partout croissants génèrent les mêmes mots purement morphiques que les endomorphismes expansifs.

Proposition 5.54. *Soit $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ un endomorphisme partout croissant. Tout mot non vide écrit sur l'alphabet $\mathcal{A}$ croît de façon exponentielle sous l'action de σ.*

Démonstration. Procédons par contraposée. En appliquant la Proposition 5.28 avec $\beta = 1$, on obtient que si un endomorphisme admet un mot $x \in \mathcal{A}^*$ tel que $|\sigma^k(x)| = \Theta(k^\alpha)$ pour un $\alpha \in \mathbb{N}$ lorsque k tend vers l'infini, alors il admet également une lettre $b \in \mathcal{A}$ telle que $|\sigma^k(b)| = \mathcal{O}(1)$ lorsque k vers l'infini. En d'autres termes, cela signifie que si un endomorphisme admet un mot borné polynomialement, alors il admet une lettre bornée. $\square$

Définissons trois classes d'endomorphismes partout croissants.

Définitions 5.55.

(i) **(Quasi-uniforme)**
Soit $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ un endomorphisme. S'il existe $\beta \in \mathbb{R}_{\geq 1}$ tel que pour toute lettre $a \in \mathcal{A}$, on a $|\sigma^k(a)| = \Theta(\beta^k)$ lorsque k tend vers l'infini, alors le morphisme σ est dit *quasi-uniforme*.

(ii) **(Divergent polynomialement)**
Soit $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ un endomorphisme partout croissant. S'il existe $\beta \in \mathbb{R}_{>1}$ et une fonction non identiquement nulle $\alpha : \mathcal{A} \to \mathbb{N}$ telle que pour toute lettre $a \in \mathcal{A}$, on a $|\sigma^k(a)| = \Theta(k^{\alpha(a)}\beta^k)$ lorsque k tend vers l'infini, alors le morphisme σ est dit *divergent polynomialement*.

(iii) **(Divergent de façon exponentielle)**
Soit $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ un endomorphisme partout croissant. S'il existe deux lettres $a_1, a_2 \in \mathcal{A}$ et des couples $(\beta_1, \alpha_1), (\beta_2, \alpha_2) \in \mathbb{R}_{>1} \times \mathbb{N}$ tels que $\beta_1 \neq \beta_2$, $|\sigma^k(a_1)| = \Theta(k^{\alpha_1}\beta_1^k)$ et $|\sigma^k(a_2)| = \Theta(k^{\alpha_2}\beta_2^k)$ lorsque k tend vers l'infini, alors le morphisme σ est dit *divergent de façon exponentielle.*

Remarques 5.56. Les trois classes d'endomorphismes définies à la Définition 5.55 sont deux à deux disjointes. De plus, par le Théorème 5.14 de Salomaa et Soittola et par la Proposition 5.54, ces classes forment une partition de l'ensemble des endomorphismes partout croissants.
Si un endomorphisme est quasi-uniforme, alors pour tout mot $x \in \mathcal{A}^+$, on a $|\sigma^k(x)| = |\sigma^k(x_0)| + \cdots + |\sigma^k(x_{|x|-1})| = \Theta(\beta^k)$ lorsque k tend vers l'infini. De plus, pour tout $\beta \geq 1$, tout endomorphisme β-uniforme est en particulier quasi-uniforme. Remarquons ensuite qu'un endomorphisme quasi-uniforme n'est pas nécessairement partout croissant étant donné que β peut être égal à 1.
Si un endomorphisme partout croissant est divergent polynomialement, alors :

- pour tout mot $x \in \mathcal{A}^+$, il existe $\alpha(x) \in \mathbb{N}$ tel que $|\sigma^k(x)| = \Theta(k^{\alpha(x)}\beta^k)$. Il suffit de prendre $\alpha(x) = \max_{a \in \operatorname{alph} x} \alpha(a)$. Cette définition fournit une extension à $\mathcal{A}^+$ de la fonction α.
- il existe une lettre $a \in \mathcal{A}$ telle que $\alpha(a) = 0$. En effet, la fonction α n'est pas identiquement nulle, ainsi il existe une lettre $a_i \in \mathcal{A}$ telle que $|\sigma^k(a_i)| = \Theta(k^{\alpha(a_i)}\beta^k)$ et $\alpha(a_i) \neq 0$. Dès lors, par la Proposition 5.28, il existe une lettre $a \in \mathcal{A}$ telle que $|\sigma^k(a)| = \mathcal{O}(\beta^k)$. Ainsi, on obtient $|\sigma^k(a)| = \Theta(\beta^k)$ et $\alpha(a) = 0$.

Exemple 5.57. Pour tout $r \in \mathbb{N}$, considérons l'endomorphisme expansif $\sigma_r : \{a,b\}^* \to \{a,b\}^*$ défini par $\sigma_r(a) = aba^r$ et $\sigma_r(b) = bb$.
Si $r = 0$, alors le morphisme s'écrit $\sigma_0(a) = ab$ et $\sigma_0(b) = bb$. Dans ce cas, le morphisme est 2-uniforme et donc en particulier quasi-uniforme.
Si $r = 1$, alors le morphisme s'écrit $\sigma_1(a) = aba$ et $\sigma_1(b) = bb$. Ceci correspond à l'étude faite dans l'Exemple 5.32. Les résultats obtenus étaient : pour tout $k \in \mathbb{N}$, $|\sigma^k(a)| = \left(\frac{1}{2}k + 1\right) 2^k$ et $|\sigma^k(b)| = 2^k$. Ainsi, le morphisme σ_1 est divergent polynomialement avec $\alpha(a) = 1$ et $\alpha(b) = 0$.
Si $r \geq 2$, le résultat $|\,\sigma_r^k(b)| = 2^k$, obtenu précédemment, reste inchangé. Montrons ensuite que pour tout $k \in \mathbb{N}$, $|\,\sigma_r^k(a)| = \frac{r}{r-1}(r+1)^k - \frac{1}{r-1}2^k$. Par définition, on a

$$|\,\sigma_r^0(a)| = |a| = 1 = \frac{r}{r-1} - \frac{1}{r-1}$$

et par récurrence, on obtient

$$\begin{aligned}
|\,\sigma_r^{k+1}(a)| &= |\,\sigma_r^k(aba^r)| \\
&= (r+1)|\,\sigma_r^k(a)| + |\,\sigma_r^k(b)| \\
&= (r+1)\left(\frac{r}{r-1}(r+1)^k - \frac{1}{r-1}2^k\right) + 2^k \\
&= \frac{r}{r-1}(r+1)^{k+1} - \frac{1}{r-1}2^{k+1}.
\end{aligned}$$

Ainsi, l'endomorphisme σ_r est divergent de façon exponentielle car pour la lettre a on a $\beta_a = r+1 > 2$ et pour la lettre b, on a $\beta_b = 2$.

Nous pouvons maintenant énoncer le résultat principal de cette section.

Théorème 5.58. *Soient un mot $u \in \mathcal{A}^{\mathbb{N}}$ où $\mathcal{A} = \mathrm{alph}(u)$ et un endomorphisme partout croissant $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ tel que $\sigma(u) = u$.*

- *Si σ est quasi-uniforme, alors $p_u(n) = \mathcal{O}(n)$.*
- *Si σ est divergent polynomialement, alors $p_u(n) = \Theta(n \log\log n)$.*
- *Si σ est divergent de façon exponentielle, alors $p_u(n) = \Theta(n \log n)$.*

Remarque 5.59. L'hypothèse $\mathrm{alph}(u) = \mathcal{A}$ n'est pas à négliger. Par exemple, le mot f de Fibonacci, point fixe du morphisme $\sigma : a \mapsto ab,\ b \mapsto a$, est également obtenu en itérant le morphisme $\gamma : a \mapsto ab,\ b \mapsto a,\ c \mapsto \cdots$ où le choix de l'image de c dans $\{a,b,c\}^*$ est arbitraire. Dès lors, l'image de c par γ n'intervient en rien sur la complexité en facteurs de $f = \gamma^\omega(a) = \sigma^\omega(a)$ mais est indispensable pour la caractérisation du morphisme γ.
Plus particulièrement, tout endomorphisme partout croissant peut être obtenu comme la restriction d'un endomorphisme divergent de façon exponentielle.
En effet, considérons un endomorphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ partout croissant et une lettre $b \notin \mathcal{A}$. Notons $\mathcal{B} = \mathcal{A} \cup \{b\}$ et $\gamma : \mathcal{B}^* \to \mathcal{B}^*$ l'endomorphisme défini par

$$\gamma(a) = \sigma(a) \quad \forall a \in \mathcal{A},\ \text{et}$$
$$\gamma(b) = b^{\|\sigma\|+1}.$$

Pour tout $a \in \mathcal{A}$, on a $|\gamma^k(a)| = |\sigma^k(a)| \leq \|\sigma\|^k$. On en tire que $|\gamma^k(a)| = \mathcal{O}(\beta^k)$ où $\beta = \|\sigma\|$, et $|\gamma^k(b)| = (\|\sigma\|+1)^k = \Theta((\|\sigma\|+1)^k)$. Ainsi, l'endomorphisme γ est divergent de façon exponentielle. Or, par définition, l'endomorphisme σ est une restriction de γ. Dès lors, tout point fixe de σ est un point fixe de γ.

Désormais, le but de la suite de cette section est de démontrer le Théorème 5.58. Pour ce faire, par définition de la notation Θ de Landau, nous allons procéder en deux étapes en obtenant une borne supérieure et une borne inférieure nous permettant de conclure respectivement en les caractères asymptotiques $\mathcal{O}$ et Ω désirés.

5.4.1 Obtention de bornes supérieures

Définition 5.60. Soient un endomorphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ et des mots x, y, z écrits sur l'alphabet $\mathcal{A}$. Pour tout entier naturel $n \in \mathbb{N}$, notons

$$K_n^\sigma(x,y,z) := \{k \in \mathbb{N} \mid |\sigma^k(y)| + 2 \leq n \leq |\sigma^k(xyz)|\}.$$

Remarque 5.61. Si l'on note les fonctions f_1, f_2 les fonctions définies telles que, pour tout $k \in \mathbb{N}$, $f_1(k) = |\sigma^k(y)| + 2$ et $f_2(k) = |\sigma^k(xyz)|$, alors l'ensemble $K_n^\sigma(x,y,z)$ est égal à $E_n(f_1, f_2)$, défini dans la section 5.1. Ainsi, si le mot y croît de façon exponentielle, par le Théorème 5.14 de Salomaa et Soittola, il existe[4] $(\beta, \alpha) \in \mathbb{R}_{>1} \times \mathbb{N}$ tel que $|\sigma^k(y)| = \Theta(k^\alpha \beta^k)$ lorsque k tend vers l'infini. Or, si y croît de façon exponentielle, il en est de même pour le mot xyz et donc un résultat identique s'applique. En appliquant le Corollaire 5.10, nous pouvons conclure que sous ces hypothèses, on a

$$\begin{aligned}
\#(K_n^\sigma(x, y, z)) &= \mathcal{O}(1), \text{ou} \\
\#(K_n^\sigma(x, y, z)) &= \Theta(n \log\log n), \text{ou} \\
\#(K_n^\sigma(x, y, z)) &= \Theta(\log n).
\end{aligned}$$

Lemme 5.62. *Soient un mot $u \in \mathcal{A}^\mathbb{N}$ et un endomorphisme partout croissant $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ tel que $\sigma(u) = u$. Il existe un ensemble fini $G \subseteq \mathcal{A} \times \mathcal{A}^+ \times \mathcal{A}$ tel que, pour tout $n \geq 3$,*

$$p_u(n) \leq (n-1) \sum_{(a,w,a') \in G} \#(K_n^\sigma(a, w, a')).$$

Démonstration. Au vu du Théorème 5.40, il existe un sous-ensemble fini G de $\mathcal{A} \times \mathcal{A}^+ \times \mathcal{A}$ tel que, pour tout $n \geq 3$,

$$L_n(u) = \bigcup_{(a,w,a') \in G} L_n^\sigma(a, w, a').$$

Ainsi, par la Remarque 5.41, on a

$$p_u(n) = \# \left(\bigcup_{(a,w,a') \in G} L_n^\sigma(a, w, a') \right) \leq \sum_{(a,w,a') \in G} \#(L_n^\sigma(a, w, a')).$$

Dès lors, vérifions que, pour tous mots $x, y, z \in \mathcal{A}^*$, et tout $n \geq 1$, on a

$$\#(L_n^\sigma(x, y, z)) \leq (n-1)\#(K_n^\sigma(x, y, z)).$$

En particulier, cela sera vrai pour les triplets $(a, w, a') \in G$ et les entiers $n \geq 3$, et le résultat sera démontré.

Considérons un entier naturel $k \in \mathbb{N}$. Tout d'abord, par définition, pour qu'un mot w soit un facteur central de longueur n du triplet $(\sigma^k(x), \sigma^k(y), \sigma^k(z))$, il faut que w s'écrive $x'\sigma^k(y)z'$ où x' (resp. z') est un suffixe (resp. préfixe) non vide de $\sigma^k(x)$ (resp. $\sigma^k(z)$). Ainsi, on a

$$1 \leq |x'| \leq |\sigma^k(x)| \quad \text{et} \quad 1 \leq |z'| \leq |\sigma^k(z)|.$$

Or, on a $|x'\sigma^k(y)z'| = n$. Dès lors, on obtient

$$|\sigma^k(y)| + 2 \leq n \leq |\sigma^k(xyz)|.$$

4. Les cas correspondant aux couples $(\beta, \alpha) \in \{(0, 0)\} \cup (\{1\} \times \mathbb{N})$ donnés par le théorème sont à rejeter vu que y croît de façon exponentielle.

Ainsi, si $k \notin K_n^\sigma(x,y,z)$, alors $(\sigma^k(x), \sigma^k(y), \sigma^k(z))$ n'admet aucun facteur central de longueur n. Ensuite, par la Remarque 5.36, si $k \in K_n^\sigma(x,y,z)$, alors $(\sigma^k(x), \sigma^k(y), \sigma^k(z))$ admet au plus $(n-1)$ facteurs centraux de longueur n.

□

Théorème 5.63. *La fonction de complexité d'un mot infini u engendré par un endomorphisme partout croissant est en $\mathcal{O}(n \log n)$.*

Démonstration. Soient un endomorphisme partout croissant $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ et des mots $x, y, z \in \mathcal{A}^*$ tels que $y \neq \varepsilon$. Considérons $n \in \mathbb{N}$ et notons $K_n := \{k \in \mathbb{N} \mid |\sigma^k(y)| + 2 \leq n\}$. Par définition des ensembles, on a l'inclusion suivante :

$$K_n^\sigma(x,y,z) \subseteq K_n. \tag{5.12}$$

Par la Proposition 5.54, le mot non vide y croît de façon exponentielle sous l'action de σ. Ainsi, il existe $\beta > 1$ tel que $|\sigma^k(y)| = \Omega(\beta^k)$ lorsque k tend vers l'infini. Formellement,

$$\exists \mu > 0,\ \exists k_0 > 0,\ \forall k \geq k_0,\ \ \mu\beta^k \leq |\sigma^k(y)|.$$

Pour tout $k \in K_n$ tel que $k \geq k_0$, on obtient l'inégalité

$$\mu\beta^k \leq |\sigma^k(y)| < |\sigma^k(y)| + 2 \leq n,$$

c'est-à-dire

$$\mu\beta^k < n.$$

En passant au logarithme dans chaque membre, on a $k < \frac{\log(\frac{n}{\mu})}{\log(\beta)}$. Ainsi, le cardinal de l'ensemble K_n est en $\mathcal{O}(\log n)$. Par (5.12), il en est de même pour le cardinal de $K_n^\sigma(x,y,z)$. Or, par le Lemme 5.62, étant donné que la somme est finie, on a $p_u(n) = \mathcal{O}(n \log n)$.

□

Théorème 5.64. *Soient un mot infini $u \in \mathcal{A}^{\mathbb{N}}$ et un endomorphisme partout croissant $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ tel que $\sigma(u) = u$.*

- *Si σ est quasi-uniforme, alors $p_u(n) = \mathcal{O}(n)$.*
- *Si σ est divergent polynomialement, alors $p_u(n) = \mathcal{O}(n \log\log n)$.*

Démonstration.

- Supposons que σ soit quasi-uniforme. Il existe $\beta \in \mathbb{R}_{>1}$ tel que pour toute lettre $a \in \mathcal{A}$, $|\sigma^k(a)| = \Theta(\beta^k)$ lorsque k tend vers l'infini. Pour tout mot $w \in \mathcal{A}^+$, par la Remarque 5.56, on a également $|\sigma^k(w)| = \Theta(\beta^k)$ lorsque k tend vers l'infini. Ainsi, par la Remarque 5.61 et par le Corollaire 5.10, où les hypothèses $\beta_1 = \beta_2$ et $\alpha_1 = \alpha_2$ sont vérifiées, on a $\#(K_n^\sigma(x,y,z)) = \mathcal{O}(1)$, pour tout $(x,y,z) \in \mathcal{A} \times \mathcal{A}^+ \times \mathcal{A}$. Ainsi, par le Lemme 5.62, vu que la somme est finie, il s'ensuit que $p_n(n) = \mathcal{O}(n)$.

- Supposons que σ soit divergent polynomialement. Par la Remarque 5.56, il existe $\beta \in \mathbb{R}_{>1}$ et une fonction $\alpha : \mathcal{A}^+ \to \mathbb{N}$ tels que, pour tout mot non vide $w \in \mathcal{A}^+$, on a $|\sigma^k(w)| = \Theta(k^{\alpha(w)}\beta^k)$. Ainsi, étant donné que pour tous mots $x, y, z \in \mathcal{A}^*$ tels que $y \neq \varepsilon$, on a $|\sigma^k(y)| = \Theta(k^{\alpha(y)}\beta^k)$ et $|\sigma^k(xyz)| = \Theta(k^{\alpha(xyz)}\beta^k)$ avec le même β, par la Remarque 5.61 et le Corollaire 5.10, on a
$$\#(K_n^\sigma(x, y, z)) = \#(E_n(|\sigma^k(y)| + 2, |\sigma^k(xyz)|)) = \mathcal{O}(\log\log n).$$
Plus précisément, $\#(K_n^\sigma(x, y, z))$ sera en $\mathcal{O}(1)$ si $\alpha(y) = \alpha(xyz)$ et en $\Theta(\log\log n)$ dans le cas contraire. Comme précédemment, il s'ensuit que $p_u(n) = \mathcal{O}(n \log\log n)$.

$\square$

Ce théorème nous fournit une autre démonstration de la complexité en $\mathcal{O}(n)$ obtenue pour les mots q-automatiques dans le Théorème 4.25 de la section 4.3. En effet, considérons $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ un endomorphisme q-uniforme, $q \geq 2$, prolongeable en $a \in \mathcal{A}$ et $\tau : \mathcal{A} \to \mathcal{B}$ un codage. Notons $v = \sigma^\omega(a) \in \mathcal{A}^\mathbb{N}$ et $u = \tau(v) \in \mathcal{B}^\mathbb{N}$. L'endomorphisme σ est en particulier partout croissant et quasi-uniforme. Ainsi, par le Théorème 5.64, on en tire que la fonction de complexité en facteurs de v est en $\mathcal{O}(n)$. Finalement, par le Lemme 4.19, on a $p_u(n) \leq p_v(n)$. Cela démontre donc que la fonction de complexité du mot q-automatique u est en $\mathcal{O}(n)$.

5.4.2 Obtention de bornes inférieures

Avec les théorèmes démontrés ci-avant, nous avons pu établir des bornes supérieures pour les fonctions de complexité dans chacune des classes de morphismes définies. La suite de cette section consistera dès lors à obtenir des bornes inférieures. Le résultat principal à démontrer est le suivant.

Théorème 5.65. *Soient un mot infini $u \in \mathcal{A}^\mathbb{N}$, où $\mathcal{A} = \mathrm{alph}(u)$, et un endomorphisme partout croissant $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ tel que $\sigma(u) = u$.*

- *Si σ est divergent polynomialement, alors $p_u(n) = \Omega(n \log\log n)$.*
- *Si σ est divergent de façon exponentielle, alors $p_u(n) = \Omega(n \log n)$.*

Nous allons établir des résultats intermédiaires pour aboutir à la démonstration du Théorème 5.65.

Lemme 5.66. *Soient un mot infini binaire $v \in \{0, 1\}^\mathbb{N}$, une suite de mots finis binaires $(w_k)_{k\geq 0} \in (\{0, 1\}^*)^\mathbb{N}$ et une suite de naturels $(e_k)_{k\geq 0} \in \mathbb{N}^\mathbb{N}$ tels que pour tout $k \in \mathbb{N}$, le mot w_k est un suffixe de w_{k+1}, $e_k < e_{k+1}$ et le mot $w_k 1 0^{e_k} 1$ est un facteur de v. Pour tout $n \in \mathbb{N}$, on a*
$$s_v(n) \geq \#\left(\{k \in \mathbb{N} \mid 1 + e_k \leq n \leq 1 + e_k + |w_k|\}\right).$$

L'inégalité de ce lemme peut être reformulée comme suit : pour tout $n \in \mathbb{N}$, on a

$$s_v(n) \geq \# \left(E_n(f_1, f_2)\right),$$

où $f_1, f_2 : \mathbb{N} \to \mathbb{N}$ sont définies par $f_1(k) = 1 + e_k$ et $f_2(k) = 1 + e_k + |w_k|$, pour tout $k \in \mathbb{N}$.

Démonstration. Considérons un entier naturel $n \in \mathbb{N}$, et notons

$$K_n := \{k \in \mathbb{N} \mid 1 + e_k \leq n \leq 1 + e_k + |w_k|\}.$$

Pour tout $k \in K_n$, considérons x_k le suffixe de longueur n du mot $w_k 10^{e_k}$. Remarquons tout d'abord que, par définition de K_n, un suffixe d'une telle longueur existe. Ensuite, le mot x_k est un facteur spécial à droite de v. En effet, le mot $x_k 1$ est un facteur de v car $x_k 1$ est un suffixe de $w_k 10^{e_k} 1 \in L(v)$. De plus, le mot $x_k 0$ est un facteur de v car par définition w_k est un suffixe de w_{k+1}, ainsi $x_k 0$ est un suffixe de $w_{k+1} 10^{e_{k+1}} \in L(v)$. Ainsi, on obtient

$$s_v(n) \geq \#(\{x_k \mid k \in K_n\}).$$

Vérifions dès lors que tous les mots x_k construits de la sorte sont distincts, c'est-à-dire $\#(\{x_k \mid k \in K_n\}) = \#(K_n)$. Considérons $i, j \in K_n$ tels que $i < j$ et les mots x_i et x_j correspondants. Par définition, étant donné que $x_i = \mathrm{Suff}_n(w_i 10^{e_i})$, la lettre apparaissant à la position $n - e_i - 1$ dans x_i est un 1. Cependant, dans $x_j = \mathrm{Suff}_n(w_j 10^{e_j})$, vu que $e_i < e_j$, la lettre apparaissant à la position $n - e_i - 1$ est un 0. Ceci prouve que les mots x_k sont deux à deux distincts.

□

Lemme 5.67. *Soient un mot $u \in \mathcal{A}^{\mathbb{N}}$ et un endomorphisme partout croissant $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ tel que $\sigma(u) = u$. Pour toute lettre $a \in \mathrm{alph}(u)$, il existe une lettre $a' \in \mathrm{alph}(u)$ telle que a' apparaît infiniment souvent dans u et $|\sigma^k(a')| = \Theta(|\sigma^k(a)|)$.*

Démonstration. Fixons une lettre $a \in \mathrm{alph}(u)$ et notons $\mathcal{C} = \{c \in \mathcal{A} \mid |\sigma^k(c)| = \Theta(|\sigma^k(a)|)\}$. Par la Proposition 5.54, la lettre a croît de façon exponentielle sous l'action de σ. Ainsi, par définition et par le Théorème 5.14 de Salomaa et Soittola, il existe un réel $\beta > 1$ et un naturel $\alpha \in \mathbb{N}$ tels que $|\sigma^k(a)| = \Theta(k^\alpha \beta^k)$. Ensuite, par la proposition 5.31, lorsque k tend vers l'infini, on a $|\sigma^k(a)|_{\mathcal{C}} = \Theta(\beta^k)$, ainsi $\lim_{k\to\infty} |\sigma^k(a)|_{\mathcal{C}} = \infty$. Or, l'alphabet $\mathcal{A}$ étant fini, il en est de même pour $\mathcal{C}$ et il existe donc une lettre $a' \in \mathcal{C}$ telle que $\lim_{k\to\infty} |\sigma^k(a)|_{a'} = \infty$. Étant donné que $a \in \mathrm{alph}(u)$, et que $\sigma(u) = u$, il existe $v \in \mathcal{A}^*$ et $w \in \mathcal{A}^{\mathbb{N}}$ tels que

$$u = vaw = \sigma^k(v)\sigma^k(a)\sigma^k(w).$$

Le mot $\sigma^k(a)$ est donc un facteur de u pour tout $k \in \mathbb{N}$. On en tire que la lettre a' apparaît infiniment souvent dans le mot u.

□

Lemme 5.68. *Soient X un ensemble fini et $N = (\#(X))!$. Toute fonction $f : X \to X$ satisfait $f^N = f^{2N}$. En particulier, par induction, on a $f^N = f^{kN}$ pour tout $k \in \mathbb{N}_0$.*

Démonstration. Notons Y le sous-ensemble maximal de X tel que $f_{|Y}$ soit une permutation. En particulier, la permutation $f_{|Y}$ est un produit de cycles et si l'on note p le plus petit commun multiple des longueurs des cycles qui la composent, on a $(f_{|Y})^p = id$. Or, le nombre p divise N. Ainsi, on obtient $(f_{|Y})^N = id$. De plus, si $x \in X \setminus Y$, il existe $n < \#(X) \leq N$ tel que[5] $f^n(x) \in Y$. Soit $y \in Y$ tel que $f^n(x) = y$. On a

$$f^N(x) = f^{N-n}(f^n(x)) = f^{N-n}(y)$$

et

$$f^{2N}(x) = f^{2N-n}(f^n(x)) = f^{2N-n}(y) = f^{N-n}(f^N(y)).$$

Or, on a $f^N(y) = y$. Ainsi, on en tire que $f^N = f^{2N}$.

$\square$

Lemme 5.69. *Soient un endomorphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$, deux alphabets $\mathcal{B}$ et $\mathcal{C}$ formant une partition de $\mathcal{A}$ tels que $\sigma(\mathcal{B}) \subseteq \mathcal{B}^*$ et $\sigma(\mathcal{C}) \subseteq A^*\mathcal{C}\,\mathcal{A}^*$ et notons $N = (\#(\mathcal{C}))!$. Pour tout $c \in \mathcal{C}$, il existe deux lettres $c_1, c_2 \in \mathcal{C}$ telles que les trois assertions suivantes sont vérifiées :*

(i) $\sigma^N(c) \in \mathcal{B}^*\, c_1 A^* \cap \mathcal{A}^*\, c_2\, \mathcal{B}^*$,

(ii) $\sigma^N(c_1) \in \mathcal{B}^*\, c_1 A^*$,

(iii) $\sigma^N(c_2) \in A^* c_2\, \mathcal{B}^*$.

Par exemple, des alphabets $\mathcal{B}$ et $\mathcal{C}$ qui peuvent être utilisés dans ce lemme sont respectivement l'ensemble des lettres bornées et croissantes sous l'action de σ. En effet, ces ensembles forment une partition de l'alphabet de départ et, par définition, l'image d'une lettre bornée est formée de lettres bornées et l'image d'une lettre croissante sous l'action de σ contient au moins une lettre croissante sous l'action du morphisme σ.

Démonstration. Notons $f_1 : \mathcal{C} \to \mathcal{C}$ et $f_2 : \mathcal{C} \to \mathcal{C}$ les fonctions qui, à tout lettre c de $\mathcal{C}$, associent respectivement la première et la dernière lettre de $\sigma(c)$ appartenant à $\mathcal{C}$. Les fonctions f_1 et f_2 sont bien définies car, par hypothèse, on a $\sigma(\mathcal{C}) \subseteq \mathcal{A}^*\, \mathcal{C}\, \mathcal{A}^*$. Par définition, pour tout $c \in \mathcal{C}$, on a $\sigma(c) \in \mathcal{B}^*\, f_1(c)\, \mathcal{A}^* \cap \mathcal{A}^*\, f_2(c)\, \mathcal{B}^*$.
Pour tout $k \in \mathbb{N}_0$, on a

$$\sigma^k(c) \in \mathcal{B}^*\, f_1^k(c)\, \mathcal{A}^* \cap \mathcal{A}^*\, f_2^k(c)\, \mathcal{B}^* \,. \tag{5.13}$$

5. En effet, dans le cas contraire, si pour tout $n \in \mathbb{N}$, $f^n(x) \in X \setminus Y$, alors, vu que $X \setminus Y$ est fini, les éléments $\{z \in X \setminus Y \mid \exists n \in \mathbb{N}, f^n(x) = z\}$ formeraient un cycle. Ceci est absurde par définition de l'ensemble Y.

En effet, le cas $k = 1$ est satisfait par définition et vérifions cette affirmation par induction. Supposons avoir démontré l'exactitude de cette affirmation pour $k \in \mathbb{N}_0$ et vérifions-la pour $k+1$. Par hypothèse de récurrence, on a $\sigma^k(c) \in \mathcal{B}^* \, f_1^k(c) \, \mathcal{A}^*$ et $\sigma^k(c) \in \mathcal{A}^* \, f_2^k(c) \, \mathcal{B}^*$ où $f_1^k(c), f_2^k(c) \in \mathcal{C}$. Ainsi, vu que $\sigma(\mathcal{B}) \subseteq \mathcal{B}^*$, on a

$$\begin{aligned} \sigma^{k+1}(c) = \sigma(\sigma^k(c)) &\in \sigma(\mathcal{B}^*)\sigma(f_1^k(c))\sigma(A^*) \\ &\in \mathcal{B}^*(\mathcal{B}^* \, f_1(f_1^k(c)) \, \mathcal{A}^*) \, \mathcal{A}^* = \mathcal{B}^* \, f_1^{k+1}(c) \, \mathcal{A}^*, \end{aligned}$$

et, de la même façon,

$$\sigma^{k+1}(c) = \sigma(\sigma^k(c)) \in \mathcal{A}^* \, f_2^{k+1}(c) \, \mathcal{B}^* \, .$$

Or, par le Lemme 5.68, on a $f_1^N = f_1^{2N}$ et $f_2^N = f_2^{2N}$. Dès lors, en utilisant (5.13) avec $k = N \geq 1$ et en notant $c_1 = f_1^N(c)$ et $c_2 = f_2^N(c)$, la thèse est vérifiée. En effet, le point (i) est direct et on a

$$\sigma^N(c_1) = \sigma^N(f_1^N(c)) \in \mathcal{B}^* \, \underbrace{f_1^N(f_1^N(c))}_{=f_1^N(c)=c_1} \, \mathcal{A}^*,$$

et

$$\sigma^N(c_2) = \sigma^N(f_2^N(c)) \in \mathcal{A}^* \, \underbrace{f_2^N(f_2^N(c))}_{=f_2^N(c)=c_2} \, \mathcal{B}^* \, .$$

□

Remarque 5.70. Avec les notations du Lemme 5.69, pour tout triplet $(c, x, c') \in \mathcal{C} \times \mathcal{B}^* \times \mathcal{C}$ et tout $k \in \mathbb{N}$, il existe un unique triplet $(\overline{c}, \overline{x}, \overline{c}') \in \mathcal{C} \times \mathcal{B}^* \times \mathcal{C}$ tel que $\overline{c} \, \overline{x} \, \overline{c}'$ est un facteur central de $(\sigma^k(c), \sigma^k(x), \sigma^k(c'))$.
En effet, pour tout $k \in \mathbb{N}$, on a $\sigma^k(x) \in \mathcal{B}^*$ et en appliquant (5.13), de la démonstration précédent, à c et c' respectivement, on a $\sigma^k(c) \in \mathcal{A}^* \, f_2^k(c) \, \mathcal{B}^*$ et $\sigma^k(c') \in \mathcal{B}^* \, f_1^k(c') \, \mathcal{A}^*$. Dès lors, notons $f_2^k(c) = \overline{c}$ et $f_1^k(c') = \overline{c}'$. On a

$$\underbrace{\sigma^k(c)}_{\in \, \mathcal{A}^* \, \overline{c} \, \mathcal{B}^*} \quad \underbrace{\sigma^k(x)}_{\in \, \mathcal{B}^*} \quad \underbrace{\sigma^k(c')}_{\in \, B^* \overline{c}' \, \mathcal{A}^*} \quad = y \, \overline{c} \, \overline{x} \, \overline{c}' \, z,$$

où $y, z \in A^*$ et $\overline{x} \in \mathcal{B}^*$. Ainsi, par construction, seul le triplet $(\overline{c}, \overline{x}, \overline{c}') \in \mathcal{C} \times \mathcal{B}^* \times \mathcal{C}$ est tel que $\overline{c} \, \overline{x} \, \overline{c}'$ est un facteur central de $(\sigma^k(c), \sigma^k(x), \sigma^k(c'))$.

Les résultats préliminaires nécessaires à la démonstration du Théorème 5.65 ont désormais été mis en place. Ainsi, en utilisant les outils mis à notre disposition, démontrons le.

Démonstration. Preuve du Théorème 5.65

• Étudions tout d'abord le cas où le morphisme σ est divergent polynomialement. Par définition, il existe un réel $\beta \in \mathbb{R}_{>1}$ et une fonction $\alpha : \mathcal{A}^+ \to \mathbb{N}$ tels que pour tout mot $w \in \mathcal{A}^+$, $|\sigma^k(w)| = \Theta(k^{\alpha(w)}\beta^k)$ lorsque k tend vers l'infini. La fonction α étant non identiquement nulle, posons $\mathcal{B} = \{b \in \mathcal{A} \mid \alpha(b) = 0\}$ et $\mathcal{C} = \{c \in \mathcal{A} \mid \alpha(c) \geq 1\}$. Commençons par montrer que les trois affirmations suivantes sont vérifiées :

(i) $\mathcal{A} = \mathcal{B} \cup \mathcal{C}$ et $\mathcal{B} \cap \mathcal{C} = \emptyset$.

(ii) $L(u) \cap \mathcal{C}\,\mathcal{B}^+\,\mathcal{C} \neq \emptyset$.

(iii) $\sigma(\mathcal{B}) \subseteq \mathcal{B}^+$ et $\sigma(\mathcal{C}) \subseteq \mathcal{A}^*\,\mathcal{C}\,\mathcal{A}^*$.

En effet, par définition on a $\mathcal{A} = \mathcal{B} \cup \mathcal{C}$. De plus, les alphabets $\mathcal{B}$ et $\mathcal{C}$ sont disjoints. Ainsi, la propriété (i) est trivialement vérifiée . Remarquons qu'on a $\mathcal{C} \neq \emptyset$ car la fonction α n'est pas identiquement nulle et $\mathcal{B} \neq \emptyset$ par la Remarque 5.56. Ensuite, en appliquant le Lemme 5.67 avec le morphisme partout croissant σ ayant comme point fixe u et avec une lettre $c \in \mathcal{C}$, on obtient qu'il existe une lettre $c' \in \mathcal{A}$ telle que c' apparaît infiniment souvent dans u et $|\sigma^k(c')| = \Theta(|\sigma^k(c)|) = \Theta(k^{\alpha(c)}\beta^k)$ lorsque k tend vers l'infini. En particulier, la lettre c' appartient à l'alphabet $\mathcal{C}$ et apparaît infiniment souvent dans u. Or, vu que $\mathcal{A} = \mathrm{alph}(u) = \mathcal{B} \cup \mathcal{C}$ et $\mathcal{B} \neq \emptyset$, on a $L(u) \cap \mathcal{C}\,\mathcal{B}^+\,\mathcal{C} \neq \emptyset$, et la propriété (ii) est vérifiée. Finalement, démontrons la propriété (iii). Tout d'abord, appliquons le premier point de la Proposition 5.27 avec $k_0 = 1$ à une lettre $b \in \mathcal{B}$. Le morphisme σ est partout croissant, ainsi on a $\sigma(b) \neq \varepsilon$ et pour toute lettre $a \in \mathrm{alph}(\sigma(b)) \subseteq \mathcal{A}$, on a $|\sigma^k(a)| = \mathcal{O}(|\sigma^k(b)|) = \mathcal{O}(\beta^k)$. Or, en particulier $a \in \mathcal{A}$, et donc $|\sigma^k(a)| = \Theta(k^{\alpha(a)}\beta^k)$ avec $\alpha(a) = 0$ et a est une lettre de $\mathcal{B}$. On en tire que $\sigma(B) \subseteq \mathcal{B}^+$. Ensuite, appliquons le second point de cette même Proposition 5.27 avec $k_0 = 1$ à une lettre $c \in \mathcal{C}$. Il existe $a \in \mathrm{alph}(\sigma(c))$ telle que $|\sigma^k(a)| = \Theta(|\sigma^k(c)|) = \Theta(k^{\alpha(c)}\beta^k)$. En particulier, on a $\alpha(a) = \alpha(c)$, c'est-à-dire $a \in \mathcal{C}$ et vu que $a \in \mathrm{alph}(\sigma(c))$, on a $\sigma(c) \subseteq \mathcal{A}^*\,\mathcal{C}\,\mathcal{A}^*$. Ceci termine la preuve de la propriété (iii).

Notons $\chi : \mathcal{A}^* \to \{0,1\}^*$ le codage défini par

$$\chi(a) = \begin{cases} 0 & \text{si } a \in \mathcal{B} \\ 1 & \text{si } a \in \mathcal{C} \end{cases}.$$

La propriété (i) nous assure que l'application χ est bien définie. Si nous prouvons que $s_{\chi(u)}(n) = \Omega(\log\log n)$, le résultat $p_u(n) = \Omega(n \log\log n)$ sera vérifié. En effet, par la Proposition 2.20, on a $p_{\chi(u)}(n) = 1 + \sum_{l=0}^{n-1} s_{\chi(u)}(l)$ et donc, par le Lemme 5.13, le résultat $p_{\chi(u)}(n) = \Omega(n \log\log n)$ est vérifié. Ensuite, par le Lemme 4.19, on a $p_{\chi(u)}(n) \leq p_u(n)$, pour tout $n \in \mathbb{N}$. Ainsi, on en tire que p_u est en $\Omega(n \log\log n)$ également. Démontrons dès lors que $s_{\chi(u)}(n) = \Omega(\log\log n)$ en appliquant le Lemme 5.66.

Par la Remarque 5.53 (ii), il existe une puissance expansive de σ. Ainsi, soit $M \geq 1$ tel que σ^M est expansif. Notons $N = (\#(\mathcal{C}))!$ et $\tau = \sigma^{MN}$. On a $\tau(u) = u$ et vu que σ est partout croissant, aucune lettre n'est envoyée sur le mot vide par une puissance de σ. Dès lors, pour toute lettre $a \in \mathcal{A}$, on a $|\,\sigma^{MN}(a)| = |\,\sigma^N(\sigma^M(a))| \geq |\,\sigma^M(a)| \geq 2$. On en tire que le morphisme τ est également expansif. Pour tout mot $w \in \mathcal{A}^+$, on a

$$\begin{aligned} |\tau^k(w)| = |(\sigma^{MN})^k(w)| &= |\,\sigma^{kMN}(w)| \\ &= \Theta((kMN)^{\alpha(w)}\beta^{kMN}) \\ &= \Theta(k^{\alpha(w)}(\beta^{MN})^k). \end{aligned}$$

Ainsi, en considérant la même fonction α et le réel $\beta^{MN} > 1$, l'endomorphisme τ est divergent polynomialement.

Par la propriété (ii), il existe $c, c' \in \mathcal{C}$ et $x \in \mathcal{B}^+$ tels que cxc' soit un facteur de u. Par les propriétés (i) et (iii), les hypothèses du Lemme 5.69 sont vérifiées pour le morphisme σ^M. Ainsi étant donné que $\sigma(\mathcal{B}) \subseteq \mathcal{B}^+$, seule une lettre de $\mathcal{C}$ peut produire des occurrences de lettres de $\mathcal{C}$ dans u, et vu que $u = \sigma(u) = \sigma^M(u) = \sigma^{MN}(u)$, il existe $c'' \in \mathcal{C}$ tel que[6] $\sigma^{MN}(c'') = \tau(c'') \in \mathcal{B}^*\, c'\, \mathcal{A}^* \cup \mathcal{A}^*\, c\, \mathcal{B}^*$ et $(\sigma^M)^N(c) = \tau(c) \in \mathcal{A}^*\, c\, \mathcal{B}^*$ et $\tau(c') \in \mathcal{B}^*\, c'\, \mathcal{A}^*$. Soient $z, z' \in \mathcal{B}^*$ et $y, y' \in \mathcal{A}^*$ les mots tels que $\tau(c) = ycz$ et $\tau(c') = z'c'y'$.
Définissions les suites de mots $(w_k)_{k\in\mathbb{N}}$ et d'entiers $(e_k)_{k\in\mathbb{N}}$ par

$$w_k = \chi(\tau^{k-1}(y) \cdots \tau^2(y)\tau(y)y), \forall k \in \mathbb{N}$$

et

$$e_k = \left(\sum_{j=0}^{k-1} |\tau^j(z)|\right) + |\tau^k(x)| + \left(\sum_{j=0}^{k-1} |\tau^j(z')|\right), \forall k \in \mathbb{N}\,.$$

On a

$$\begin{aligned} w_{k+1} &= \chi(\tau^k(y)\tau^{k-1}(y) \cdots \tau^2(y)\tau(y)y) \\ &= \chi(\tau^k(y))\, \chi(\tau^{k-1}(y) \cdots \tau^2(y)\tau(y)y) \\ &= \chi(\tau^k(y))w_k. \end{aligned}$$

On en tire que w_k est un suffixe de w_{k+1}. Ensuite, on a

$$e_{k+1} - e_k = \underbrace{|\tau^k(z)| + |\tau^k(z')|}_{\geq 0} + \underbrace{|\tau^{k+1}(x)| - |\tau^k(x)|}_{>0 \text{ car } \tau \text{ est expansif et } x\in\mathcal{B}^+} > 0.$$

De plus, remarquons que l'on a

$$\begin{aligned} &\tau(c) = ycz, \\ &\tau^2(c) = \tau(y)\tau(c)\tau(z) = \tau(y)\, y\, c\, z\, \tau(z), \\ &\qquad\vdots \\ &\tau^k(c) = \tau^{k-1}(y)\tau^{k-2}(y) \cdots \tau(y)\, y\, c\, z\, \tau(z) \cdots \tau^{k-2}(z)\tau^{k-1}(z), \end{aligned}$$

et de la même façon,

$$\tau^k(c') = \tau^{k-1}(z')\tau^{k-2}(z') \cdots \tau(z')\, z'\, c'\, y'\, \tau(y') \cdots \tau^{k-2}(y')\tau^{k-1}(y').$$

Or, vu que $\sigma(\mathcal{B}) \subseteq \mathcal{B}^+$, $\tau = \sigma^{MN}$ et que $z, z' \in \mathcal{B}^*$, $x \in \mathcal{B}^+$, pour tout $k \in \mathbb{N}$, $\tau^k(x), \tau^k(z)$ et $\tau^k(z')$ sont des mots de $\mathcal{B}^*$. Dès lors, on a

6. Dans le cas où c est la première lettre de u, il suffit de remplacer cxc' par l'unique facteur central de $(\tau(c), \tau(x), \tau(c'))$ appartenant à $\mathcal{C} \times \mathcal{B}^+ \times \mathcal{C}$ en appliquant la Remarque 5.70 avec $k = N$ et le morphisme σ^M.

$$
\begin{aligned}
&\chi(\tau^k(cxc'))\\
=&\chi(\tau^k(c))\chi(\tau^k(x))\chi(\tau^k(c'))\\
=&\chi\left(\tau^{k-1}(y)\tau^{k-2}(y)\cdots\tau(y)\,y\,c\,z\,\tau(z)\cdots\tau^{k-2}(z)\tau^{k-1}(z)\right)\ \chi\left(\tau^k(x)\right)\\
&\chi\left(\tau^{k-1}(z')\tau^{k-2}(z')\cdots\tau(z')\,z'\,c'\,y'\,\tau(y')\cdots\tau^{k-2}(y')\tau^{k-1}(y')\right)\\
=&\underbrace{\chi\left(\tau^{k-1}(y)\tau^{k-2}(y)\cdots\tau(y)\,y\right)}_{=w_k}\underbrace{\chi(c)}_{=1}\\
&\underbrace{\chi\left(z\,\tau(z)\cdots\tau^{k-2}(z)\tau^{k-1}(z)\tau^k(x)\tau^{k-1}(z')\tau^{k-2}(z')\cdots\tau(z')\,z'\right)}_{=0^{e_k}}\underbrace{\chi(c')}_{=1}\\
&\chi\left(y'\,\tau(y')\cdots\tau^{k-2}(y')\tau^{k-1}(y')\right)\\
=&w_k10^{e_k}1\,\chi\left(y'\,\tau(y')\cdots\tau^{k-2}(y')\tau^{k-1}(y')\right).
\end{aligned}
$$

Or, le mot cxc' est un facteur de $u=\tau(u)$, donc on a $\tau^k(cxc')\in L(u)$, $\chi(\tau^k(cxc'))\in L(\chi(u))$ et $w_k10^{e_k}1\in L(\chi(u))$. Ainsi, par le Lemme 5.66, si l'on note

$$K_n=\{k\in\mathbb{N}\mid 1+e_k\leq n\leq 1+e_k+|w_k|\},$$

on a

$$s_v(n)\geq\#(K_n).$$

Démontrons dès lors que le cardinal de l'ensemble K_n est en $\Omega(\log\log n)$.
Le morphisme τ est partout croissant, ainsi par la Proposition 5.54, les mots z et z' ont une croissance exponentielle sous l'action du morphisme τ. Par le Lemme 5.33, on a $\sum_{j=0}^{k-1}|\tau^j(z)|=\Theta(|\tau^k(z)|)$ et $\sum_{j=0}^{k-1}|\tau^j(z')|=\Theta(|\tau^k(z')|)$ lorsque k tend vers l'infini. On obtient

$$
\begin{aligned}
e_k&=\Theta(|\tau^k(z)|)+|\tau^k(x)|+\Theta(|\tau^k(z')|)\\
&=\Theta(k^{\alpha(z)}(\beta^{MN})^k)+\Theta(k^{\alpha(x)}(\beta^{MN})^k)+\Theta(k^{\alpha(z')}(\beta^{MN})^k)\\
&=\Theta((\beta^{MN})^k),
\end{aligned}
$$

où la dernière étape se justifie par le fait que $\alpha(z)=\alpha(x)=\alpha(z')=0$ car ces mots appartiennent à $\mathcal{B}^*$. Ensuite, on a

$$
\begin{aligned}
|w_k|&=|\chi(\tau^k(c))|-|\chi(c)|-|\chi(z\,\tau(z)\cdots\tau^{k-2}(z)\tau^{k-1}(z))|\\
&=|\tau^k(c)|-1-\underbrace{\sum_{j=0}^{k-1}|\tau^j(z)|}_{=\Theta(|\tau^k(z)|)}\\
&=\Theta(k^{\alpha(c)}(\beta^{MN})^k)-1-\Theta((\beta^{MN})^k)\\
&=\Theta(k^{\alpha(c)}(\beta^{MN})^k).
\end{aligned}
$$

En notant f_1 et f_2 les fonctions définies par $f_1(k)=1+e_k$ et $f_2(k)=1+e_k+|w_k|$ pour tout $k\in\mathbb{N}$, on a $f_1(k)=\Theta((\beta^{MN})^k)$ et $f_2(k)=\Theta(k^{\alpha(c)}(\beta^{MN})^k)$ lorsque k tend vers l'infini. Ainsi, par le Corollaire 5.10, on obtient

$$\#(K_n)=\#(E_n(f_1,f_2))=\Theta(\log\log n).$$

Ceci conclut la preuve du cas où l'endomorphisme σ est divergent polynomialement.
• Étudions maintenant le cas où σ est divergent de façon exponentielle. Ce cas peut être étudié avec le même type d'arguments en considérant d'autres ensembles $\mathcal{B}$ et $\mathcal{C}$. Soit u_0 la première lettre de u. Par la propriété 5.54, vu que σ est partout croissant, la lettre u_0 croît de façon exponentielle sous l'action du morphisme σ. Notons $(\beta, \alpha) \in \mathbb{R}_{>1} \times \mathbb{N}$ le couple défini tel que $|\sigma^k(u_0)| = \Theta(k^\alpha \beta^k)$ lorsque k tend vers l'infini. Définissons les ensembles $\mathcal{B}$ et $\mathcal{C}$ comme suit.

$$\mathcal{B} = \{b \in \mathcal{A} \mid |\sigma^k(b)| = o(\beta^k)\}, \text{ et}$$
$$\mathcal{C} = \{c \in \mathcal{A} \mid |\sigma^k(c)| = \Omega(\beta^k)\}.$$

Démontrons tout d'abord que les propriétés (i), (ii) et (iii) du point précédent sont également vérifiées dans ce cas.
Vu que $\mathcal{A} = \text{alph}(u)$, pour tout $a \in \mathcal{A}$, il existe $k_0 \in \mathbb{N}$ tel que $a \in \text{alph}(\sigma^{k_0}(u_0))$. Ainsi, par la Proposition 5.27, pour tout $a \in \mathcal{A}$, on a $|\sigma^k(a)| = \mathcal{O}(|\sigma^k(u_0)|) = \mathcal{O}(k^\alpha \beta^k)$. C'est-à-dire il existe $\eta_a > 0$ et $k_0 > 0$ tels que pour tout $k \geq k_0$, $|\sigma^k(a)| \leq \eta_a k^\alpha \beta^k$. Or, par la propriété 5.54, toute lettre $a \in \mathcal{A}$ croît de façon exponentielle sous l'action du morphisme σ. Ainsi, il existe $(\beta_a, \alpha_a) \in \mathbb{R}_{>1} \times \mathbb{N}$ tel que $|\sigma^k(a)| = \Theta(k^{\alpha_a} \beta_a^k)$ lorsque k tend vers l'infini. Par conséquent, nous venons de démontrer que pour tout $a \in \mathcal{A}$, on a $\beta_a \leq \beta$. En particulier, si $\beta_a = \beta$ alors $a \in \mathcal{C}$ et si $\beta_a < \beta$ alors $a \in \mathcal{B}$ car dans ce cas $k^{\alpha_a} \beta_a^k = o(\beta^k)$. Cela signifie que $\mathcal{A} = \mathcal{B} \cup \mathcal{C}$ et étant donné les conditions disjointes sur β_a, les ensembles $\mathcal{B}$ et $\mathcal{C}$ sont disjoints. Ceci termine la preuve de la propriété (i). La propriété (ii) est une conséquence du Lemme 5.67 appliqué à la lettre u_0. En effet, la présence infiniment souvent d'une lettre $a' \in \mathcal{A}$ satisfaisant $|\sigma^k(a')| = \Theta(|\sigma^k(u_0)|)$ implique l'appartenance de cette lettre à l'ensemble $\mathcal{C}$. On conclut en utilisant le fait que l'ensemble $\mathcal{B}$ est non vide. Ensuite, par le second point de la propriété 5.27, il existe une lettre $a \in \mathcal{A}$ telle que $|\sigma^k(a)| = \Theta(|\sigma^k(u_0)|)$, c'est-à-dire $\mathcal{C} \neq \emptyset$. Qui plus est, l'ensemble $\mathcal{B}$ est non vide car par définition d'un morphisme divergent de façon exponentielle, il existe deux lettres de $\mathcal{A}$ n'ayant pas la même base d'exponentielle, ainsi toutes les lettres de $\mathcal{A}$ ne peuvent pas appartenir à l'ensemble $\mathcal{C}$. Par construction, on a $\sigma(\mathcal{B}) \subseteq \mathcal{B}^*$. En effet, sinon, s'il existe une lettre appartenant à $\mathcal{C}$ dans l'image d'une lettre $b \in \mathcal{B}$, alors b serait de l'ordre $\Omega(\beta^k)$ lorsque k tend vers l'infini. Ensuite, on a $\mathcal{C} \neq \emptyset$ donc il existe une lettre $c \in \mathcal{C}$ dans le mot u. Or, pour toute lettre dans $\mathcal{C}$, son image doit contenir une lettre de $\mathcal{C}$ pour que le critère à l'infini sur l'itération de σ en cette lettre soit vérifié. On en tire que $\sigma(\mathcal{C}) \subseteq \mathcal{A}^* \, \mathcal{C} \, \mathcal{A}^*$ et ceci termine la preuve de la propriété (iii).
Pour la suite de la preuve, considérons le même codage χ que précédemment et le même mot $\chi(u)$. Pour les mêmes raisons, démontrons que[7] $s_{\chi(u)}(n) = \Omega(\log n)$. Utilisons les mêmes constructions que précédemment pour obtenir M, N, τ, $cxc' \in (\mathcal{C} \, \mathcal{B}^+ \, \mathcal{C}) \cap L(u)$, $\tau(c) = ycz$, $\tau(c') = z'c'y'$ avec $z, z' \in \mathcal{B}^*$ et $y, y' \in \mathcal{A}^*$. En définissant également w_k et e_k de façon identique, les propriétés de ces suites restent valables et il suffit de vérifier que le cardinal de l'ensemble $K_n = \{k \in \mathbb{N} \mid 1 + e_k \leq n \leq 1 + e_k + |w_k|\}$ est en $\Omega(\log n)$. Pour tout mot $w \in \mathcal{B}^*$, on a $|\sigma^k(w)| = |\sigma^k(w_0)| + \cdots + |\sigma^k(w_{|w|-1})| = \Theta(k^{\alpha_w} (\beta_w^{MN})^k)$ si l'on note $(\beta_w, \alpha_w) = \max\{(\beta_{w_i}, \alpha_{w_i}) \mid i = 0 \ldots |w| - 1\}$ en considérant l'ordre lexicographique. Ainsi,

7. Le lemme correspondant n'a pas été démontré en annexe dans ce cas mais la démonstration est identique à celle du Lemme 5.13 en utilisant la fonction ln au lieu de ln ln.

par le Lemme 5.33, on a

$$\begin{aligned} e_k &= \Theta(|\tau^k(z)|) + |\tau^k(x)| + \Theta(|\tau^k(z')|) \\ &= \Theta(k^{\alpha_z}(\beta_z^{MN})^k) + \Theta(k^{\alpha_x}(\beta_x^{MN})^k) + \Theta(k^{\alpha_{z'}}(\beta_{z'}^{MN})^k) \\ &= \Theta(k^{\alpha_{e_k}}(\beta_{e_k}^{MN})^k) \end{aligned}$$

où $(\beta_{e_k}, \alpha_{e_k}) = \max\{(\beta_z, \alpha_z), (\beta_x, \alpha_x), (\beta_{z'}, \alpha_{z'})\}$.
De la même façon, étant donné que $c \in \mathcal{C}$, on a $\beta_c = \beta$ et on obtient

$$\begin{aligned} |w_k| &= |\tau^k(c)| - 1 - \Theta(|\tau^k(z)|) \\ &= \Theta(k^{\alpha_c}(\beta^{MN})^k) - 1 - \Theta(k^{\alpha_z}(\beta_z^{MN})^k) \\ &= \Theta(k^{\alpha_c}(\beta^{MN})^k), \end{aligned}$$

car, par définition, $\beta > \beta_z$. Ensuite, avec les mêmes notations pour les fonctions f_1 et f_2, on a $f_1(k) = \Theta(k^{\alpha_{e_k}}(\beta_{e_k}^{MN})^k)$ et $f_2(k) = \Theta(k^{\alpha_{e_k}}(\beta_{e_k}^{MN})^k) + \Theta(k^{\alpha_c}(\beta^{MN})^k) = \Theta(k^{\alpha_c}(\beta^{MN})^k)$ car $\beta_{e_k} < \beta$ lorsque k tend vers l'infini. Et donc par le Corollaire 5.10, on obtient finalement

$$\#(K_n) = \#(E_n(f_1, f_2)) = \Theta(\log n).$$

□

Nous venons donc d'établir des bornes supérieures et inférieures pour les fonctions de complexité de mots purement morphiques construits par des endomorphismes partout croissants. Cela nous permet d'obtenir le résultat du Théorème 5.58. En effet, le Théorème 5.58, qui est une première partie du théorème de Pansiot, est démontré en combinant les théorèmes 5.63, 5.64 et 5.65.

5.5 Morphismes contenant des lettres bornées

Pour terminer la preuve du Théorème 5.2 de Pansiot, il reste à étudier les cas où les endomorphismes contiennent des lettres bornées sous l'action du morphisme considéré. Tout au long de cette section, nous allons utiliser les notations suivantes :

Notations 5.71.

- σ est un endomorphisme de $\mathcal{A}^*$,
- u est un mot purement morphique engendré par σ,
- u_0 est la première lettre du mot u,
- $\mathcal{B}$ est l'ensemble des lettres bornées sous l'action de σ,
- $\mathcal{C}$ est l'ensemble des lettres croissantes sous l'action de σ.

Comme expliqué après l'énoncé du Lemme 5.69, on a $\mathcal{A} = \mathcal{B} \cup \mathcal{C}$, $\mathcal{B} \cap \mathcal{C} = \emptyset$, $\sigma(\mathcal{B}) \subseteq \mathcal{B}^*$, $\sigma(\mathcal{C}) \subseteq \mathcal{A}^* \mathcal{C} \mathcal{A}^*$ et plus généralement, les ensembles $\mathcal{B}^*$ et $\mathcal{A}^* \mathcal{C} \mathcal{A}^*$ sont respectivement les ensembles des mots bornés et croissants sous l'action de σ. En particulier, dans cette section, on a $\mathcal{B} \neq \emptyset$.

L'exemple suivant illustre le fait que les ensembles $\mathcal{B}$ et $\mathcal{C}$ ne sont pas complètement déterminés par le mot u.

Exemple 5.72. Considérons deux endomorphismes $\sigma_1, \sigma_2 : \{a,b,c,d\}^* \to \{a,b,c,d\}^*$ définis comme suit.

σ_1	σ_2
$a \mapsto acd$	$a \mapsto acd$
$b \mapsto b$	$b \mapsto b$
$c \mapsto cb$	$c \mapsto c$
$d \mapsto d$	$d \mapsto bd$

Pour tout $k \in \mathbb{N}$, on a $\sigma_1^k(cd) = cb^k d = \sigma_2^k(cd)$. En effet, le cas de base $k = 0$ est bien respecté : on a $\sigma_1^0(cd) = cd = cb^0 d = \sigma_2^0(cd)$. Par induction, supposons que la propriété a été prouvée pour $k \in \mathbb{N}$ et démontrons-la pour $k+1$. Pour tout $i \in \{1,2\}$, on a $\sigma_i^{k+1}(cd) = \sigma_i(\sigma_i^k(cd)) = \sigma_i(cb^k d) = \sigma_i(c)(\sigma_i(b))^k\, \sigma_i(d)$. Ainsi on a

- $\sigma_1^{k+1}(cd) = \sigma_1(c)(\sigma_1(b))^k\, \sigma_1(d) = (cb)b^k d = cb^{k+1}d$, et
- $\sigma_2^{k+1}(cb) = \sigma_2(c)(\sigma_2(b))^k\, \sigma_2(d) = cb^k(bd) = cb^{k+1}d$

On obtient

- $\sigma_1^w(a) = acd\,\sigma_1(cd)\,\sigma_1^2(cd)\,\sigma_1^3(cd)\cdots = \underbrace{acdcbdcb^2dcb^3d\cdots}_{:=u}$
- $\sigma_2^w(a) = acd\,\sigma_2(cd)\,\sigma_2^2(cd)\,\sigma_2^3(cd)\cdots = \overbrace{acdcbdcb^2dcb^3d\cdots}$.

Le mot purement morphique u est engendré par les morphismes σ_1 et σ_2. Cependant si l'on note $\mathcal{B}_1, \mathcal{C}_1$ et $\mathcal{B}_2, \mathcal{C}_2$ les ensembles des lettres bornées et croissantes sous σ_1 et σ_2 respectivement, on a

- $\mathcal{B}_1 = \{b,d\}, \mathcal{C}_1 = \{a,c\}$ et
- $\mathcal{B}_2 = \{b,c\}, \mathcal{C}_2 = \{a,d\}$.

En conclusion, nous remarquons que, lorsque nous considérons un mot purement morphique, nous devons toujours définir l'endomorphisme l'engendrant pour connaître les ensembles $\mathcal{B}$ et $\mathcal{C}$ correspondants.

5.5.1 Complexité sous-quadratique

Théorème 5.73. *Soient un mot purement morphique $u \in \mathcal{A}^{\mathbb{N}}$ et un endomorphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ tel que $\sigma(u) = u$. S'il y a un nombre fini de facteurs de u bornés sous l'action de σ, alors l'une des affirmations suivantes est vérifiée :*

(i) $p(n) = \Theta(1)$,

(ii) $p(n) = \Theta(n)$,

(iii) $p(n) = \Theta(n \log\log n)$,

(iv) $p(n) = \Theta(n \log n)$.

Démonstration. Sans perte de généralité, supposons $\mathcal{A} = \mathrm{alph}(u)$. L'idée de la preuve est d'utiliser le Théorème 5.58 pour obtenir une borne supérieure et une borne inférieure pour la fonction de complexité du mot u en utilisant les fonctions de complexité de deux mots engendrés par des endomorphismes partout croissants appartenant à la même classe.

• Établissons la borne supérieure :
Notons $X = (\mathcal{B}^* \mathcal{C} \mathcal{B}^*) \cap L(u)$. Par hypothèse, l'ensemble $\mathcal{B}^* \cap L(u)$, correspondant aux facteurs bornés de u, est fini. Ainsi, il y a au plus

$$(\#(\mathcal{B}^* \cap L(u)))^2 \, \#(\mathcal{C})$$

éléments dans X. Par conséquent, l'ensemble X est fini. Notons $\overline{\mathcal{A}}$ un alphabet de même cardinal que l'ensemble X et $\overline{\tau} : \overline{\mathcal{A}}^* \to \mathcal{A}^*$ un morphisme définissant une bijection de $\overline{\mathcal{A}}$ dans X. Ensuite, démontrons que l'on a $\sigma(X) \subseteq X^*$. Considérons un mot $x \in X$. On a $x = bcb' \in L(u)$ où $b, b' \in \mathcal{B}^*$ et $c \in \mathcal{C}$. Comme précisé avant la démonstration du Lemme 5.69, par définition des ensembles $\mathcal{B}$ et $\mathcal{C}$, et vu que $\sigma(u) = u$, on obtient

$$\sigma(x) = \underbrace{\sigma(b)}_{\in \mathcal{B}^*} \; \underbrace{\sigma(c)}_{\in \mathcal{A}^* \mathcal{C} \mathcal{A}^*} \; \underbrace{\sigma(b')}_{\in \mathcal{B}^*} \in L(u).$$

Ainsi, en notant $\sigma(c) = b_{i_1} c_{i_1} \cdots b_{i_n} c_{i_n} b_{i_{n+1}}$ tel que $b_{i_1}, \ldots, b_{i_{n+1}} \in \mathcal{B}^*$ et $c_{i_1}, \ldots, c_{i_n} \in \mathcal{C}$, un découpage possible dans X^+ est donné par la factorisation obtenue par les facteurs $\sigma(b) b_{i_1} c_{i_1}, b_{i_2} c_{i_2}, \ldots, b_{i_n} c_{i_n} b_{i_{n+1}} \sigma(b') \in X$. En utilisant cette factorisation [8] de $\sigma(x)$ dans X^+ pour tout $x \in X$, définissons le morphisme $\overline{\sigma} : \overline{\mathcal{A}}^* \to \overline{\mathcal{A}}^*$ qui associe à toute lettre $\overline{a} \in \overline{\mathcal{A}}$ le mot de $\overline{\mathcal{A}}^*$ correspondant à la factorisation de $\sigma(x)$ dans X^+ où $x = \tau(\overline{a})$. Par définition, on obtient

$$\overline{\tau} \circ \overline{\sigma} = \sigma \circ \overline{\tau}. \tag{5.14}$$

Remarquons que u_0 est une lettre de $\mathcal{C} \subseteq X$ et $\sigma(u_0) \in u_0 X^+$. En effet, le mot $u \in \mathcal{A}^{\mathbb{N}}$ est défini tel que $u = \sigma^\omega(u_0)$. En particulier, la lettre u_0 ne peut être bornée car σ est prolongeable en u_0. Ensuite, si $\sigma(u_0) \notin u_0 X^+$, alors on a $\sigma(u_0) \in u_0 \mathcal{B}^+$ et étant donné que $\sigma(\mathcal{B}) \subseteq \mathcal{B}^*$, on aurait $u \in u_0 \mathcal{B}^{\mathbb{N}}$ ce qui est absurde car u ne contient qu'un nombre fini de facteurs bornés. Dès lors, nous pouvons considérer que si l'on note $\overline{u_0}$ l'unique lettre de $\overline{\mathcal{A}}$ satisfaisant $\overline{\tau}(\overline{u_0}) = u_0 \in \mathcal{C} \subseteq X$, alors $\overline{u_0}$ est la première lettre de $\overline{\sigma}(\overline{u_0})$. En effet, on a

$$\sigma \circ \overline{\tau}(\overline{u_0}) = \sigma(u_0) \in u_0 X^+$$

et donc $\overline{\tau} \circ \overline{\sigma}(\overline{u_0}) \in u_0 X^+$. Ainsi, vu que $\overline{u_0}$ est l'unique lettre de $\overline{\mathcal{A}}$ satisfaisant $\overline{\tau}(\overline{u_0}) = u_0$, on en tire que l'on peut supposer que $\overline{u_0}$ est la première lettre de $\overline{\sigma}(\overline{u_0})$. Prouvons maintenant que le morphisme $\overline{\sigma}$ est partout croissant. Pour tout $k \in \mathbb{N}$, on a

$$\overline{\tau} \circ \overline{\sigma}^k = \sigma^k \circ \overline{\tau}. \tag{5.15}$$

Les cas $k = 0$ est trivialement vérifié et le cas $k = 1$ l'est également par l'équation (5.14). Supposons la formule vraie pour $k \in \mathbb{N}_0$ et démontrons-la par induction pour $k + 1$. On a

8. Un autre choix de découpage de $\sigma(x)$ dans X^+ mènerait à une définition différente de $\overline{\sigma}$. Cela ne changerait rien dans le reste du raisonnement vu l'abstraction dans la suite de la démonstration.

$$\begin{aligned}\overline{\tau} \circ \overline{\sigma}^{k+1} &= (\overline{\tau} \circ \overline{\sigma}^k) \circ \overline{\sigma} \\ &= (\sigma^k \circ \overline{\tau}) \circ \overline{\sigma} \\ &= \sigma^k \circ (\overline{\tau} \circ \overline{\sigma}) \\ &= \sigma^k \circ (\sigma \circ \overline{\tau}) = \sigma^{k+1} \circ \overline{\tau}.\end{aligned}$$

Ainsi, en considérant $\overline{c} \in \overline{\mathcal{A}}$, on a

$$|\,\sigma^k(\,\overline{\tau}(\overline{c})\,)| = |\overline{\tau}(\,\overline{\sigma}^k(\overline{c})\,)| \leq \|\overline{\tau}\|\,|\overline{\sigma}^k(\overline{c})|. \tag{5.16}$$

Or, l'élément $\overline{\tau}(\overline{c})$ appartient à $X = (\mathcal{B}^*\,\mathcal{C}\,\mathcal{B}^*) \cap L(u)$ et croît donc sous l'action de σ. Ainsi, vu l'inégalité (5.16), établie pour tout $k \in \mathbb{N}$, il s'ensuit que la lettre $\overline{c}$ croît sous l'action du morphisme $\overline{\sigma}$. Par conséquent, $\overline{\sigma}$ est partout croissant car la lettre $\overline{c}$ a été choisie arbitrairement. Ensuite, étant donné que $\overline{u_0}$ est la première lettre de $\overline{\sigma}\,(\overline{u_0})$ et que $\overline{u_0}$ croît sous l'action du morphisme $\overline{\sigma}$, le morphisme $\overline{\sigma}$ est prolongeable en $\overline{u_0}$. Notons $\overline{u} = \overline{\sigma}^{\omega}(\overline{u_0})$. Une conséquence de l'égalité (5.15) est que $\overline{\tau}(\overline{\sigma}^k(\overline{u_0})) = \sigma^k(\overline{\tau}(\overline{u_0})) = \sigma^k(u_0)$ pour tout $k \in \mathbb{N}$. Or, on a $\lim\limits_{k\to+\infty} \sigma^k(u_0) = u$ et $\lim\limits_{k\to+\infty} \overline{\sigma}^k(\overline{u_0}) = \overline{u}$. Ainsi, en tire $\overline{\tau}(\overline{u}) = u$. Finalement, par le Lemme 4.21, étant donné que $\varepsilon \notin X$, le morphisme $\overline{\tau}$ est non effaçant et on a

$$p_u(n) \leq \|\overline{\tau}\|\, p_{\overline{u}}(n), \tag{5.17}$$

pour tout $n \in \mathbb{N}$. On en tire que $p_u(n) = \mathcal{O}(p_{\overline{u}}(n))$. Ainsi, en appliquant le Théorème 5.63 au morphisme partout croissant $\overline{\sigma}$ engendrant le mot $\overline{u}$ on obtient que le mot $\overline{u}$ est en $\mathcal{O}(n \log n)$. On conclut cette partie de démonstration en utilisant l'inégalité (5.17) pour obtenir la complexité en $\mathcal{O}(n \log n)$ du mot u.

- Établissons la borne inférieure :

Notons $\chi : \mathcal{A}^* \to \mathcal{C}^*$ le morphisme effaçant les lettres de $\mathcal{B}$. Formellement,

$$\begin{aligned}\chi(b) &= \varepsilon,\ \forall b \in \mathcal{B}, \\ \chi(c) &= c,\ \forall c \in \mathcal{C}\,.\end{aligned}$$

Notons également $\underline{\sigma}$ la restriction du morphisme $\chi \circ \sigma$ à $\mathcal{C}^*$. Par définition de l'ensemble $\mathcal{C}$, l'image d'une lettre de $\mathcal{C}$ est dans $\mathcal{A}^*\,\mathcal{C}\,\mathcal{A}^*$. Ainsi, le morphisme $\underline{\sigma}$ est un endomorphisme non effaçant de $\mathcal{C}^*$. Comme démontré dans la première partie, on a $u_0 \in \mathcal{C}$ et $\sigma(u_0) \in u_0 X^+$. Ainsi, la première lettre de $\underline{\sigma}(u_0)$ est u_0. En effet, si l'on note $\sigma(u_0) = u_0 x_1 x_2 \cdots x_m$ pour un $m \in \mathbb{N}_0$ avec $x_i \in X$ pour tout $i \in \{1, \ldots, m\}$, alors on a

$$\underline{\sigma}(u_0) = \chi(\sigma(u_0)) = \chi(u_0 x_1 x_2 \cdots x_m) = \underbrace{\chi(u_0)}_{\in u_0\,\mathcal{C}^*}\underbrace{\chi(x_1)}_{\in \mathcal{C}^+}\underbrace{\chi(x_2)}_{\in \mathcal{C}^+}\cdots\underbrace{\chi(x_m)}_{\in \mathcal{C}^+}.$$

Pour tout $k \in \mathbb{N}$, on a

$$\chi \circ \sigma^k = \underline{\sigma}^k \circ \chi. \tag{5.18}$$

En effet, étant donné que $\mathcal{A} = \mathcal{B} \cup \mathcal{C}$ et qu'un morphisme est défini sur l'image des lettres, il nous suffit de vérifier cette égalité pour les lettres de chacun des ensembles $\mathcal{B}$ et $\mathcal{C}$.

En considérant $b \in \mathcal{B}$, on a $\underline{\sigma}^k(\chi(b)) = \underline{\sigma}^k(\varepsilon) = \varepsilon$ et, vu que $\sigma(\mathcal{B}) \subseteq \mathcal{B}^*$, $\chi(\underbrace{\sigma^k(b)}_{\in\mathcal{B}^*}) = \varepsilon$ également. Considérons dès lors une lettre $c \in \mathcal{C}$. Le cas de base $k = 0$ est trivialement vérifié et le cas de base $k = 1$ est également vrai car la définition de $\underline{\sigma}$ est équivalente à dire que l'on a

$$\chi \circ \sigma = \underline{\sigma} \circ \chi. \tag{5.19}$$

En effet, par définition, on a $\underline{\sigma}(\chi(c)) = \underline{\sigma}(c) = \chi(\sigma(c))$. Dès lors, en utilisant (5.19), le cas inductif de l'égalité (5.18) se démontre exactement de la même façon que dans la première partie de la preuve, lors de la démonstration de l'égalité (5.15).
En utilisant l'égalité (5.18), on obtient

$$\chi(\sigma^k(u_0)) = \underline{\sigma}^k(\chi(u_0)) = \underline{\sigma}^k(u_0).$$

On en tire que la lettre u_0 croît sous l'action du morphisme $\underline{\sigma}$, et donc que $\underline{\sigma}$ prolongeable en u_0. En effet, vu que u a un nombre fini de facteurs bornés sous l'action de σ, on a $|\underline{\sigma}^k(u_0)| = |\chi(\sigma^k(u_0))| \to +\infty$ lorsque k tend vers l'infini. Notons $\underline{u} = \underline{\sigma}^\omega(u_0)$. Le mot $\underline{u}$ est l'image du mot u par le morphisme χ. Ainsi, par le Lemme 4.23, on obtient

$$p_{\underline{u}}\left(\left\lfloor \frac{n}{M+1} \right\rfloor\right) \leq p_u(n), \tag{5.20}$$

pour tout $n \in \mathbb{N}$, où $M = \max\{|x| \mid x \in L(u) \cap \mathcal{B}^*\}$.

- Mise en commun des bornes établies :

Par définition du morphisme $\overline{\tau}$ induisant une bijection entre $\overline{\mathcal{A}}$ et $X = (\mathcal{B}^* \mathcal{C} \mathcal{B}^*) \cap L(u)$, considérons une paire $(c, \overline{a}) \in \mathcal{C} \times \overline{\mathcal{A}}$ telle que $\overline{\tau}(\overline{a}) \in \mathcal{B}^* c \mathcal{B}^*$. Pour tout $k \in \mathbb{N}$, on a

$$\overbrace{|\overline{\sigma}^k(\overline{a})| = |\sigma^k(c)|_{\mathcal{C}}}^{(i)} \; \hspace{-6em}\underbrace{\phantom{|\sigma^k(c)|_{\mathcal{C}}} = |\underline{\sigma}^k(c)|}_{(ii)} \tag{5.21}$$

En effet, pour démontrer l'égalité (i), notons $b_1, b_2 \in \mathcal{B}^*$ les mots tels que $\overline{\tau}(\overline{a}) = b_1 c b_2$. On a

$$\overline{\tau} \circ \overline{\sigma}^k(\overline{a}) = \sigma^k \circ \overline{\tau}(\overline{a}) = \sigma^k(b_1 c b_2) = \sigma^k(b_1)\, \sigma^k(c)\, \sigma^k(b_2),$$

Ainsi, on obtient

$$|\overline{\tau} \circ \overline{\sigma}^k(\overline{a})|_{\mathcal{C}} = |\,\sigma^k(b_1)|_{\mathcal{C}} + |\,\sigma^k(c)|_{\mathcal{C}} + |\,\sigma^k(b_2)|_{\mathcal{C}}.$$

Or, $\sigma(\mathcal{B}) \subseteq \mathcal{B}^*$ donc $|\,\sigma^k(b_1)|_{\mathcal{C}} = |\,\sigma^k(b_2)|_{\mathcal{C}} = 0$. De plus, on a $|\overline{\tau} \circ \overline{\sigma}^k(\overline{a})|_{\mathcal{C}} = |\overline{\sigma}^k(\overline{a})|$ car le morphisme $\overline{\tau}$ associe à chaque lettre de $\overline{\sigma}^k(\overline{a}) \in \overline{\mathcal{A}}^*$ un mot de X contenant exactement un élément de $\mathcal{C}$. On obtient donc $|\overline{\sigma}^k(\overline{a})| = |\,\sigma^k(c)|_{\mathcal{C}}$, ceci démontre l'égalité (i).
Ensuite, pour démontrer l'égalité (ii), étant donné que la lettre c appartient à l'alphabet $\mathcal{C}$, on a $\underline{\sigma}^k(c) = \underline{\sigma}^k \circ \chi(c) = \chi \circ \sigma^k(c)$. Dès lors, vu que χ est le morphisme effaçant les lettres de $\mathcal{B}$, on voit effectivement

$$|\underline{\sigma}^k(c)| = |\chi \circ \sigma^k(c)| = |\sigma^k(c)|_{\mathcal{C}}.$$

Vu que $\overline{\sigma}$ est partout croissant et qu'à tout lettre $c \in \mathcal{C}$ correspond une lettre $\overline{a} \in \overline{\mathcal{A}}$, les égalités (5.21) démontrent que le morphisme $\underline{\sigma}$ l'est également et ces deux endomorphismes appartiennent à la même classe. Nous allons devoir considérer les trois cas possibles suivants :

(a) Si $\overline{\sigma}$ et $\underline{\sigma}$ sont tous les deux quasi-uniformes, alors en particulier, le Théorème 5.58 affirme que la fonction de complexité en facteur du mot $\overline{u}$ est en $\mathcal{O}(n)$. Dès lors, de part l'inégalité (5.17), on obtient que la fonction de complexité du mot u est également en $\mathcal{O}(n)$. Ainsi, par le Théorème 2.2 de Morse et Hedlund, on conclut que [9] $p_u(n) = \Theta(1)$ ou $p_u(n) = \Theta(n)$.

(b) Si $\overline{\sigma}$ et $\underline{\sigma}$ sont tous les deux divergents polynomialement, alors le Théorème 5.58 affirme que les fonctions de complexité en facteurs de mots $\overline{u}$ et $\underline{u}$ sont toutes les deux en $\Theta(n \log\log n)$. Donc, en utilisant (5.17) et (5.20), on obtient $p_u(n) = \Theta(n \log\log n)$.

(c) Si $\overline{\sigma}$ et $\underline{\sigma}$ sont tous les deux divergents de façon exponentielle, alors on conclut de façon identique au cas (b) en sachant que le Théorème 5.58 affirme cette fois que les fonctions de complexité en facteurs de mots $\overline{u}$ et $\underline{u}$ sont toutes les deux en $\Theta(n \log n)$.

□

Remarque 5.74. En conservant les notations de la preuve du Théorème 5.73, remarquons que dans le cas (a), il se pourrait que $p_u(n) = \Theta(1)$ alors que $p_{\overline{u}}(n) = \Theta(n)$. En effet, lorsque le morphisme considéré est quasi-uniforme, le théorème 5.58 ne donne qu'une complexité asymptotique en $\mathcal{O}$ et non en Θ. Ainsi, l'inégalité (5.20) est inutile dans ce cas (contrairement aux cas (b) et (c)).
Un exemple illustrant cette situation est le suivant. Considérons l'alphabet $\mathcal{A} = \{0, 1, 2\}$, le morphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ défini par

$$\begin{aligned}\sigma(0) &= 01,\\ \sigma(1) &= 1212,\\ \sigma(2) &= \varepsilon,\end{aligned}$$

et le mot $u = \sigma^\omega(0)$. Étant donné que pour tout $k \in \mathbb{N}$, on a $\sigma^k(1) = (12)^{2k}$, on a $u = 01(12)^\omega$. Par définition du morphisme σ, on a $\mathcal{B} = \{2\}$ et $\mathcal{C} = \{0, 1\}$. L'ensemble $\mathcal{B}^* \cap L(u)$ est égal à $\{2\}$ car aucune puissance de 2 n'est présente dans le mot u. Remarquons que les facteurs 02 et 20 n'appartiennent pas à l'ensemble des facteurs de u. Ainsi, on obtient $X = \{0, 1, 12, 21, 212\}$. Pour plus de facilité dans les notations, définissons l'alphabet $\overline{\mathcal{A}} = \{\overline{a}_0, \overline{a}_1, \overline{a}_{12}, \overline{a}_{21}, \overline{a}_{212}\}$ en supposant, sans perte de généralité, que $\overline{\tau}(\overline{a}_x) = x$ pour tout $x \in X$. Ensuite, définissons le morphisme 2-uniforme $\overline{\sigma}$ comme suit [10] :

$$\begin{aligned}\overline{\sigma}(\overline{a}_0) &= \overline{a}_0\overline{a}_1,\\ \overline{\sigma}(\overline{a}_1) &= \overline{a}_1\overline{a}_{212}, \text{ et}\\ \overline{\sigma}(\overline{a}_{12}) &= \overline{\sigma}(\overline{a}_{211}) = \overline{\sigma}(\overline{a}_{21}) = \overline{a}_{212}\overline{a}_{12}.\end{aligned}$$

9. Voir Remarque 5.74.

10. Le choix fait pour la définition du morphisme $\overline{\sigma}$ ne correspond pas au découpage expliqué dans la preuve du Théorème 5.73. Cependant, comme précisé également dans cette preuve, le découpage explicite n'était pas unique.

Étudions maintenant le mot $\overline{u} = \overline{\sigma}^{\omega}(\overline{a}_0)$. On a $\overline{u} = \overline{a}_0\overline{a}_1\overline{\sigma}(\overline{a}_1)\overline{\sigma}^2(\overline{a}_1)\overline{\sigma}^3(\overline{a}_1)\cdots$. Or, pour tout $k \in \mathbb{N}_0$, en posant $e_k = 2^{k+1} - 2$, on a

$$\overline{\sigma}^k(\overline{a}_1) = \overline{a}_1\,\overline{a}_{212}\,(\overline{a}_{12})^{e_{k-1}}.$$

En effet, le cas de base où $k = 1$ est trivial car $e_0 = 2 - 2 = 0$, donc

$$\overline{\sigma}(\overline{a}_1) = \overline{a}_1\overline{a}_{212} = \overline{a}_1\overline{a}_{212}(\overline{a}_{12})^{e_0}.$$

Supposons dès lors que cette formule soit vérifiée pour $k \in \mathbb{N}_0$ et démontrons-le par induction pour $k + 1$. On a

$$\begin{aligned}
\overline{\sigma}^{k+1}(\overline{a}_1) &= \overline{\sigma}(\overline{\sigma}^k(\overline{a}_1)) \\
&= \overline{\sigma}(\overline{a}_1\,\overline{a}_{212}\,(\overline{a}_{12})^{e_{k-1}}) \\
&= \overline{\sigma}(\overline{a}_1)\,\overline{\sigma}(\overline{a}_{212})\,(\overline{\sigma}(\overline{a}_{12}))^{e_{k-1}} \\
&= \overline{a}_1\overline{a}_{212}(\overline{a}_{12}\overline{a}_{12})(\overline{a}_{12}\overline{a}_{12})^{e_k} \\
&= \overline{a}_1\overline{a}_{212}(\overline{a}_{12})^{2e_k+2}.
\end{aligned}$$

Or, on a $2e_k + 2 = 2(2^{k+1} - 2) + 2 = 2^{k+2} - 2 = e_{k+1}$ donc $\overline{\sigma}^{k+1}(\overline{a}_1) = \overline{a}_1\overline{a}_{212}(\overline{a}_{12})^{e_{k+1}}$. Ceci termine cette preuve par induction.
Dès lors, le mot

$$\begin{aligned}
\overline{u} &= \overline{a}_0\overline{a}_1\overline{\sigma}(\overline{a}_1)\overline{\sigma}^2(\overline{a}_1)\overline{\sigma}^3(\overline{a}_1)\cdots \\
&= \overline{a}_0\overline{a}_1\overline{a}_1\overline{a}_{212}(\overline{a}_{12})^{e_0}\overline{a}_1\overline{a}_{212}(\overline{a}_{12})^{e_1}\overline{a}_1\overline{a}_{212}(\overline{a}_{12})^{e_2}\cdots,
\end{aligned}$$

n'est pas périodique. Cependant, le mot $\overline{u}$ est engendré par le morphisme $\overline{\sigma}$ 2-uniforme, qui est donc en particulier quasi-uniforme. On obtient $p_{\overline{u}} = \Theta(n)$ contrairement au mot u qui lui est ultimement périodique et a donc une complexité en facteurs en $\Theta(1)$.

5.5.2 Complexité quadratique

Le premier but de cette section est de démontrer le Théorème 5.75, énoncé ci-après, qui stipule que la fonction de complexité en facteurs d'un mot morphique est en $\mathcal{O}(n^2)$. Ensuite, nous obtiendrons un résultat donnant la complexité asymptotique d'un mot purement morphique ayant, cette fois, une infinité de facteurs différents bornés sous l'action du morphisme considéré. Avec cette dernière classe de morphisme désormais étudiée, nous serons en mesure de conclure la démonstration du Théorème 5.2 de Pansiot.

Théorème 5.75. *(*A. Ehrenfeucht*,* K.P. Lee *et* G. Rozenberg*)*
La complexité en facteurs d'un mot purement morphique est en $\mathcal{O}(n^2)$.

Une démonstration alternative de ce théorème, détaillant la constante cachée derrière la notation asymptotique $\mathcal{O}$ de Landau, en fonction du cardinal de l'alphabet et de la largeur du morphisme l'engendrant, peut être consultée dans le livre [2] (théorème 10.4.7).

Par le Lemme 4.19, un corollaire du Théorème 5.75 peut directement être énoncé.

Corollaire 5.76. *La complexité en facteurs d'un mot morphique est en $\mathcal{O}(n^2)$.*

Nous allons maintenant établir les quelques résultats nécessaires à la démonstration du Théorème 5.75.

Lemme 5.77. *Pour tout endomorphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$, il existe un entier $M \geq 1$ tel que pour tout $x \in \mathcal{A}^*$, le mot x est borné sous l'action du morphisme σ si et seulement si $\sigma^{2M}(x) = \sigma^M(x)$.*

Démonstration. Soit $x \in \mathcal{A}^*$. S'il existe $M \geq 1$ tel que $\sigma^{2M}(x) = \sigma^M(x)$ alors, pour tout $q \in \mathbb{N}_0$, on a

$$\sigma^{qM}(x) = \sigma^M(x).$$

En effet, le cas $q = 1$ est trivialement vérifié et, par induction, on a

$$\sigma^{(q+1)M}(x) = \sigma^{qM}(\sigma^M(x)) = \sigma^M(\sigma^M(x)) = \sigma^{2M}(x) = \sigma^M(x).$$

Or, pour tout $k \geq M$, il existe $q \in \mathbb{N}_0$ et $0 \leq r < M$ tels que $k = qM + r$. Ainsi, on a

$$\sigma^k(x) = \sigma^{qM+r}(x) = \sigma^r(\sigma^{qM}(x)) = \sigma^r(\sigma^M(x)) = \sigma^{M+r}(x).$$

Dès lors, $|\sigma^k(x)|$ est borné lorsque k tend vers l'infini. Ainsi, aucun mot $x \notin \mathcal{B}^*$ ne peut satisfaire à la propriété, peu importe le choix de M.
Démontrons ensuite qu'un tel M existe pour l'ensemble des mots de $\mathcal{B}^*$. Notons $\mathcal{B}$ est l'ensemble des lettres bornées sous l'action de σ et $X = \{\sigma^k(b) \mid (b, k) \in \mathcal{B} \times \mathbb{N}\}$. Par la Remarque 5.47 pour tout $b \in \mathcal{B}$, la suite $(\sigma^k(b))_{k\in\mathbb{N}}$ est ultimement périodique. Dès lors, pour toute lettre $b \in \mathcal{B}$, on a un nombre fini d'éléments lui correspondant dans l'ensemble X. On en tire que l'ensemble X est fini. Ensuite, par définition, si $x \in X$, il existe $(b_x, k_x) \in \mathcal{B} \times \mathbb{N}$ tel que $x = \sigma^{k_x}(b_x)$. Ainsi, on a $\sigma(x) = \sigma(\sigma^{k_x}(b_x)) = \sigma^{k_x+1}(b_x) \in X$, c'est-à-dire $\sigma(X) \subseteq X$. Dès lors, par le Lemme 5.68 et vu que $\mathcal{B} \subseteq X$, la factorielle du cardinal de l'ensemble X est un naturel qui satisfait à la propriété voulue.

□

Remarque 5.78. Par le même raisonnement par induction que dans la preuve du Lemme 5.77, pour tout $x \in \mathcal{B}^*$, on a $\sigma^{kM}(x) = \sigma^M(x)$, pour tout $k \in \mathbb{N}_0$.

Lemme 5.79. *Soient deux ensembles W et X, deux fonctions $\sigma : W \to W$ et $f : W \to X$ et deux nombres $T \in \mathbb{N}_0$ et $n \in \mathbb{N}$. Les deux assertions suivantes sont équivalentes.*

(i) Pour tout $w \in W$, la suite $(f(\sigma^k(w)))_{k\geq 0}$ est ultimement périodique de période T et prépériode nT.

(ii) Pour tout $w \in W$, la suite $(f(\sigma^{kT}(w)))_{k\geq n}$ est constante.

Dans ce lemme, les fonctions et ensembles considérés ne sont pas restreints au domaine des morphismes et alphabets. Aucune propriété supplémentaire n'est nécessaire pour la démonstration de ce résultat. Cependant, dans la suite de ce mémoire, nous l'utiliserons dans le domaine de la combinatoire des mots en considérant des alphabets et des morphismes.

Démonstration. Remarquons tout d'abord que l'énoncé fait bien sens, au vu des domaines de définitions et des ensembles images des fonctions considérées.
Ensuite, démontrons plutôt que, pour tout élément w fixé de W, les deux assertions suivantes sont équivalentes.

(i') La suite $(f(\sigma^k(w)))_{k\geq 0}$ est ultimement périodique de période T et prépériode nT.

(ii') Pour tout $w' \in \{w, \sigma(w), \sigma^2(w), \ldots, \sigma^{T-1}(w)\}$, la suite $(f(\sigma^{kT}(w')))_{k\geq n}$ est constante.

L'équivalence de ces nouvelles assertions démontrera le lemme. Supposons avoir l'équivalence des assertions (i') et (ii') et démontrons celle des assertions (i) et (ii).
Si pour tout $w \in W$, la suite $(f(\sigma^k(w)))_{k\geq 0}$ est ultimement périodique de période T et prépériode nT, alors fixons un élément w_0 de W. Par les équivalences des assertions (i') et (ii'), en définissant l'ensemble $W_{w_0} = \{w_0, \sigma(w_0), \sigma^2(w_0), \ldots, \sigma^{T-1}(w_0)\}$, pour tout $w' \in W_{w_0}$ la suite $(f(\sigma^{kT}(w')))_{k\geq n}$ est constante. En particulier, $w_0 \in W_{w_0}$ et donc l'assertion (ii) est vérifiée pour w_0. Vu le choix arbitraire de cet élément, l'assertion (ii) est vérifiée pour tout $w \in W$.
Ensuite, si pour tout $w \in W$, la suite $(f(\sigma^{kT}(w)))_{k\geq n}$ est constante, alors considérons un $w_0 \in W$ particulier et conservons la notation de l'ensemble W_{w_0} correspondant. La propriété (ii) est vraie pour tout $w \in W$, or $\sigma(W) \subseteq W$, donc cette propriété est vraie pour tout $w' \in W_{w_0}$. Ainsi, en utilisant l'équivalence des assertions (ii') et (i'), on obtient que la suite $(f(\sigma^k(w_0)))_{k\geq 0}$ est ultimement périodique de période T et prépériode nT. Et ce raisonnement peut être fait pour tout élement $w_0 \in W$.
Démontrons dès lors l'équivalence des assertions (i') et (ii'). Pour ce faire, fixons un élément $w \in W$.
Si la suite $(f(\sigma^k(w)))_{k\geq 0}$ est ultimement périodique de période T et prépériode nT, alors, par la Définition 5.46, pour tout $k \geq (n+1)T$, on a $f(\sigma^k(w)) = f(\sigma^{k-T}(w))$. En particulier, en conservant la même notation pour l'ensemble W_w, considérons $w' \in W_w$. Il existe $0 \leq l < T$ tel que $w' = \sigma^l(w)$. On obtient

$$f(\sigma^{(n+1)T}(w')) = f(\sigma^{(n+1)T+l}(w)) = f(\sigma^{nT+l}(w)) = f(\sigma^{nT}(w')).$$

Et par induction, considérons $m \geq (n+1)$ et supposons avoir vérifié la propriété pour $m-1$, on a

$$f(\sigma^{mT}(w')) = f(\sigma^{mT+l}(w)) = f(\sigma^{(m-1)T+l}(w)) = f(\sigma^{(m-1)T}(w')) = f(\sigma^{nT}(w')).$$

L'assertion (ii') est dès lors vérifiée.
Finalement, si l'assertion (ii') est vérifiée, l'assertion (i') l'est également. En effet, une suite est ultimement périodique de période T et prépériode nT si les éléments aux positions relatives, après la prépériode, sont égaux. Or, en notant $w' = \sigma^l(w)$ où $0 \leq l < T$, pour tout $k \in \mathbb{N}$, on a $f(\sigma^{kT}(w')) = f(\sigma^{nT+l}(w))$.

$\square$

Lemme 5.80. *Pour tout endomorphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$, il existe un entier naturel $T \geq 1$ tel que, pour tout couple $(w, n) \in \mathcal{A}^* \times \mathbb{N}$, les suites $(\mathrm{Pref}_n(\sigma^k(w)))_{k\geq 0}$ et $(\mathrm{Suff}_n(\sigma^k(w)))_{k\geq 0}$ sont ultimement périodiques de période T et prépériode nT.*

Démonstration. Par le Lemme 5.77, il existe $M \geq 1$ tel que, pour tout mot $w \in \mathcal{B}^*$, $\sigma^{2M}(w) = \sigma^M(w)$. Notons $N = (\#(\mathcal{C}))!$, $\tau = \sigma^{MN}$ et $T = 3MN$. Vérifions que l'entier naturel T satisfait à la propriété.
En appliquant le Lemme 5.79 aux ensembles $\mathcal{A}^*$ et $\mathcal{A}^{\leq n}$ et aux fonctions σ et Pref_n (resp. Suff_n), il suffit de vérifier que la suite $(\mathrm{Pref}_n(\sigma^{kT}(w)))_{k\geq n}$ (resp.$(\mathrm{Suff}_n(\sigma^{kT}(w)))_{k\geq n}$) est constante. Or, on a

$$\sigma^{kT}(w) = \sigma^{k3MN}(w) = (\sigma^{MN})^{3k}(w) = \tau^{3k}(w).$$

Ainsi, il suffit de vérifier que pour tout $(w, n) \in \mathcal{A}^* \times \mathbb{N}$, les suites $(\mathrm{Pref}_n(\tau^{3k}(w)))_{k\geq n}$ et $(\mathrm{Suff}_n(\tau^{3k}(w)))_{k\geq n}$ sont constantes.
Le cas $n = 0$ est trivialement vérifié pour tout mot w car tout préfixe ou suffixe de longueur 0 est le mot vide. Dès lors, considérons $n \geq 1$. Démontrons que les suites $(\mathrm{Pref}_n(\tau^k(w)))_{k\geq n+2}$ et $(\mathrm{Suff}_n(\tau^k(w)))_{k\geq n+2}$ sont constantes et cela démontrera que les suites voulues, qui sont des sous-suites de celles-ci, car $3n \geq n + 2$ vu que $n \geq 1$, le sont.
Si le mot w appartient à l'ensemble $\mathcal{B}^*$, par la Remarque 5.78, on a

$$\tau^k(w) = \sigma^{MNk}(w) = \sigma^{(kN)M}(w) = \sigma^M(w)$$

pour tout $k \geq 1$. Ainsi, la suite $(\tau^k(w))_{k\geq 1}$ est constante. En particulier, les préfixes et suffixes de longueur n des éléments de cette suite sont également constants. Dès lors, étudions le cas où $w \in \mathcal{A}^*\,\mathcal{C}\,\mathcal{A}^*$.

- Étude de la suite $(\mathrm{Pref}_n(\tau^k(w)))_{k\geq n+2}$:

On a $w \in \mathcal{A}^*\,\mathcal{C}\,\mathcal{A}^*$ donc en considérant $c_0 \in \mathcal{C}$ la première lettre de $\mathcal{C}$ présente dans le mot w, il existe $u \in \mathcal{B}^*, v \in \mathcal{A}^*$ tels que $w = uc_0v$. Appliquons le Lemme 5.69 au morphisme σ^M, où $\sigma^M(\mathcal{B}) \subseteq \mathcal{B}^*$ et $\sigma^M(\mathcal{C}) \subseteq \mathcal{A}^*\,\mathcal{C}\,\mathcal{A}$, et à la lettre $c_0 \in \mathcal{C}$ où l'on a noté $N = (\#(\mathcal{C}))!$ et $(\sigma^M)^N = \tau$. Il existe des lettres $c_1, c_2 \in \mathcal{C}$ telles que

$$\begin{aligned}
&\tau(c_0) \in \mathcal{B}^*\,c_1\,\mathcal{A}^* \cap \mathcal{A}^*\,c_2\,\mathcal{B}^*,\\
&\tau(c_1) \in \mathcal{B}^*\,c_1\,\mathcal{A}^* \text{ et}\\
&\tau(c_2) \in \mathcal{A}^*\,c_2\,\mathcal{B}^*.
\end{aligned}$$

Notons $\tau(c_0) = u'c_1v'$ avec $u' \in \mathcal{B}^*$ et $v' \in \mathcal{A}^*$. En considérant $c = c_1$ et $x = \tau(u)u' \in \mathcal{B}^*$, le couple $(x, c) \in \mathcal{B}^* \times \mathcal{C}$ est défini tel que xc est un préfixe de $\tau(w)$. De plus, il existe $y \in \mathcal{B}^*$ et $z \in \mathcal{A}^*$ tels que $\tau(c) = ycz$. Ainsi défini, le mot yc est un préfixe de $\tau(c)$. Ensuite, remarquons que

$$\tau(\tau(x)) = \tau(x). \tag{5.22}$$

En effet, vu que $x \in \mathcal{B}^*$, en appliquant deux fois la Remarque 5.78, on a

$$\tau(\tau(x)) = \tau^2(x) = (\sigma^{MN})^2(x) = \sigma^{(2N)M}(x) = \sigma^M(x) = \sigma^{MN}(x) = \tau(x).$$

Ainsi, on en déduit que $\tau(x)\tau^{k-1}(c)$ est un préfixe de $\tau^k(w)$ pour tout $k \geq 2$. Vérifions le cas où $k = 2$, on a $\tau(x)\tau(c)$ est un préfixe de $\tau^2(w)$ car xc est un préfixe de $\tau(w)$. Et par induction, supposons la propriété vérifiée pour $k \geq 2$ et démontrons-le pour $k+1$. Par hypothèse de récurrence, le mot $\tau(x)\tau^{k-1}(c)$ est un préfixe de $\tau^k(w)$ donc en utilisant l'égalité (5.22), on a

$$\tau(\tau(x)\tau^{k-1}(c)) = \tau(\tau(x))\tau^k(c) = \tau(x)\tau^k(c)$$

est un préfixe de $\tau(\tau^k(w)) = \tau^{k+1}(w)$. Pour terminer la preuve de ce cas, séparons deux possibilités :

(a) Si $\tau(y) \neq \varepsilon$. Alors, pour tout $k \geq 1$, le mot $(\tau(y))^{k-1}yc$ est un préfixe du mot $\tau^k(c)$. En d'autres termes, en itérant le morphisme τ à la lettre c, on produit des puissances de $\tau(y)$ comme préfixe de $\tau^k(c)$. En effet, comme précédemment, utilisons le fait que vu que $y \in \mathcal{B}^*$, ainsi $\tau(\tau(y)) = \tau(y)$. Le cas $k = 1$ est vérifié par définition de y. De plus, par induction, on a $(\tau(y))^{k-1}yc$ est un préfixe du mot $\tau^k(c)$ et donc

$$(\tau(\tau(y))^{k-1}yc) = \tau((\tau(y)))^{k-1}\tau(\tau(y))\tau(c) = (\tau(y))^{k-1}\tau(y)ycz = (\tau(y))^k ycz$$

est un préfixe du mot $\tau^{k+1}(c)$. En particulier, le mot $(\tau(y))^k yc$ en est également un. En combinant les deux résultats que nous venons de prouver, on a

$$\begin{aligned} &\tau(x)\tau^{k-1}(c) \in \mathrm{Pref}(\tau^k(w)) \quad \forall k \geq 2 \ \text{ et} \\ &\underbrace{(\tau(y))^{l-1}yc \in \mathrm{Pref}(\tau^l(c)) \quad \forall l \geq 1}_{(\tau(y))^{k-2}yc\, \in \mathrm{Pref}(\tau^{k-1}(c)) \quad \forall k \geq 2} \\ \Rightarrow &\tau(x)(\tau(y))^{k-2}yc \in \mathrm{Pref}(\tau^k(w)) \quad \forall k \geq 2. \end{aligned}$$

Ainsi, pour tout $k \geq n+2$, $\tau(x)(\tau(y))^n$ est un préfixe de $\tau^k(w)$. Or, vu que $\tau(y) \neq \varepsilon$, on a $|\tau(x)(\tau(y))^n| \geq n$ et donc

$$\mathrm{Pref}_n(\tau^k(w)) = \mathrm{Pref}_n(\tau(x)(\tau(y))^n).$$

Le membre de droite étant indépendant de k, ce cas est démontré.

(b) Si $\tau(y) = \varepsilon$. Dans ce cas, pour tout $k \geq 1$, $\tau^k(c)$ est un préfixe de $\tau^{k+1}(c)$. En effet, le cas $k = 1$ est vérifié car, par définition, le mot yc est un préfixe de $\tau(c)$ et donc $\tau(y)\tau(c) = \tau(c)$ est un préfixe de $\tau^2(c)$. Et par induction, en supposant la propriété vraie pour $k \geq 1$, vérifions-la pour $k+1$. On a $\tau^k(c)$ est un préfixe de $\tau^{k+1}(c)$ donc $\tau(\tau^k(c)) = \tau^{k+1}(c)$ est un préfixe de $\tau^{k+2}(c)$. En particulier, on obtient $\tau(x)\tau^k(c)$ est un préfixe de $\tau(x)\tau^{k+1}(c)$ pour tout $k \geq 1$. Donc, vu que $n \geq 1$, en considérant $k \geq n+1$, on a

$$\tau(x)\tau^n(c) \in \mathrm{Pref}(\tau(x)\tau^{n+1}(c)) \in \ldots \in \mathrm{Pref}(\tau(x)\tau^{k-1}(c)) \in \mathrm{Pref}(\tau^k(w)).$$

Or, la lettre c croît sous l'action des morphismes σ et τ. Donc $\tau^n(c)$ est un préfixe propre de $\tau^{n+1}(c)$ pour tout $n \geq 1$ et on a $|\tau^n(c)| \geq n$. Ainsi, on obtient

$$\mathrm{Pref}_n(\tau^k(w)) = \mathrm{Pref}_n(\tau(x)\tau^n(c))$$

pour tout $k \geq n+1$. La conclusion est identique car le membre de droite est indépendant de k.

• Étude de la suite $(\mathrm{Suff}_n(\tau^k(w)))_{k\geq n+2}$:
Ce cas se démontre de façon symétrique. Dès lors, les justifications identiques ne seront pas explicitées. On a $w \in \mathcal{A}^*\,\mathcal{C}\,\mathcal{A}^*$ donc en considérant $c_0 \in \mathcal{C}$ la dernière lettre de $\mathcal{C}$ présente dans le mot w, il existe $u \in \mathcal{B}^*, v \in \mathcal{A}^*$ tels que $w = vc_0u$. Ensuite, de la même façon que précédemment, en appliquant le Lemme 5.69 au morphisme σ^M et à la lettre $c_0 \in \mathcal{C}$, notons $\tau(c_0) = v'c_2u'$ où $u' \in \mathcal{B}^*$ et $v' \in \mathcal{A}^*$, $c = c_2 \in \mathcal{C}$ et $x = u'\tau(u) \in \mathcal{B}^*$. Le couple $(x, c) \in \mathcal{B}^* \times \mathcal{C}$ est défini tel que cx est un suffixe de $\tau(w)$. De plus, il existe $y \in \mathcal{B}^*$ et $z \in \mathcal{A}^*$ tels que $\tau(c) = zcy$. Le mot cy est donc un suffixe de $\tau(c)$. Ensuite, l'égalité (5.22) est toujours vérifiée et on a $\tau^{k-1}(c)\tau(x)$ qui est un suffixe de $\tau^k(w)$ pour tout $k \geq 2$. Séparons les deux cas possibles :

(a) Si $\tau(y) \neq \varepsilon$. De façon symétrique au raisonnement établi précédemment, pour tout $k \geq 1$, le mot $cy(\tau(y))^{k-1}$ est un suffixe du mot $\tau^k(c)$. Or, $\tau^{k-1}(c)\tau(x)$ est un suffixe de $\tau^k(w)$ pour tout $k \geq 2$. Ainsi, on obtient $cy(\tau(y))^{k-2}\tau(x) \in \mathrm{Suff}(\tau^k(w))$ et donc pour tout $k \geq n+2$, on a $(\tau(y))^n\tau(x) \in \mathrm{Suff}(\tau^k(w))$. Or, la longueur de $(\tau(y))^n\tau(x)$ est au moins n vu que $\tau(y) \neq \varepsilon$ donc on obtient

$$\mathrm{Suff}_n(\tau^k(w)) = \mathrm{Suff}_n((\tau(y))^n\tau(x)).$$

(b) Si $\tau(y) = \varepsilon$, alors $\tau^k(c)$ est un suffixe propre de $\tau^{k+1}(c)$ pour tout $k \geq 1$ et donc $\tau^n(c)\tau(x)$ est un suffixe de $\tau^k(w)$ pour tout $k \geq n+1$. Ainsi, étant donné que la longueur de $\tau^n(c)$ est au moins n, on a

$$\mathrm{Suff}_n(\tau^k(w)) = \mathrm{Suff}_n(\tau^n(c)\tau(x))$$

pour tout $k \geq n+1$.

Ceci termine la démonstration du lemme.

□

Nous sommes désormais en mesure de démontrer le Théorème 5.75.

Démonstration. Preuve du Théorème 5.75
Soient $u \in \mathcal{A}^{\mathbb{N}}$ un mot purement morphique et $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ un endomorphisme qui engendre u. Démontrons que $p_u(n) = \mathcal{O}(n^2)$. Il nous suffit de prouver que, pour tous $x, y, z \in \mathcal{A}^*$, $\#(L_n^\sigma(x, y, z)) = \mathcal{O}(n^2)$. En effet, en appliquant le Corollaire 5.42, il existe un ensemble fini $H \subseteq \mathcal{A} \times \mathcal{A}^* \times \mathcal{A}$ tel que

$$L(u) \setminus (\mathcal{A} \cup \{\varepsilon\}) = \bigcup_{(a,w,a')\in H} L^\sigma(a, w, a').$$

Pour tout $n \geq 2$, on a

$$L_n(u) = \bigcup_{(a,w,a')\in H} L_n^\sigma(a, w, a').$$

Ainsi, on a

$$p_n(u) = \#\left(\bigcup_{(a,w,a')\in H} L_n^\sigma(a, w, a')\right) \leq \sum_{(a,w,a')\in H} \#(L_n^\sigma(a, w, a')).$$

Puisque l'ensemble H est fini, la somme est finie. Ainsi, en démontrant que, pour tout triplet de mots, on a $\#(L_n^\sigma(x,y,z)) = \mathcal{O}(n^2)$, le résultat sera en particulier vérifié pour tout triplet de H et on aura $p_u(n) = \mathcal{O}(n^2)$. Ensuite, par la Remarque 5.38 (ii), l'ensemble $L_n^\sigma(x,y,z)$ est un sous-ensemble de $L_n^\sigma(x,\varepsilon,yz)$. Dès lors, démontrons que, pour tous mots $x,y,z \in \mathcal{A}^*$, on a $\#(L_n^\sigma(x,\varepsilon,yz)) = \mathcal{O}(n^2)$.
Pour tous $n,k \in \mathbb{N}$, notons $x_{n,k} = \mathrm{Suff}_n(\sigma^k(x))$ et $y_{n,k} = \mathrm{Pref}_n(\sigma^k(yz))$. Et pour tout $n \in \mathbb{N}$, posons $X_n = \{(x_{n,k},\varepsilon,y_{n,k}) \mid k \in \mathbb{N}\}$. Par le Lemme 5.80, les suites $(x_{n,k})_{k\in\mathbb{N}}$ et $(y_{n,k})_{k\in\mathbb{N}}$ sont ultimement périodiques de période $T \geq 1$ et prépériode nT. Cela implique que l'ensemble X_n est fini et que $\#(X_n) \leq (n+1)T$, pour tout $n \in \mathbb{N}$. Or, les facteurs centraux de longueur $n \in \mathbb{N}$ de $(\sigma^k(x),\varepsilon,\sigma^k(yz))$ ne contiennent pas un préfixe de yz (resp. suffixe de x) de longueur plus grande que n. C'est-à-dire que ces derniers sont exactement les facteurs centraux de longueur n de $(x_{n,k},\varepsilon,y_{n,k})$. Ainsi, l'ensemble $L_n^\sigma(x,\varepsilon,yz)$ correspond à l'ensemble des facteurs centraux de longueur n de triplets de X_n. Or, par la Remarque 5.36, pour tout triplet de mots, il existe au plus $n-1$ facteurs centraux de longueur n correspondant à ce triplet. On obtient dès lors l'inégalité $\#(L_n^\sigma(x,\varepsilon,yz)) \leq n\,\#(X_n)$.
Ainsi, on a

$$\#(L_n^\sigma(x,y,z)) \leq \#(L_n^\sigma(x,\varepsilon,yz)) \leq n\,\#(X_n) \leq n(n+1)T.$$

On en tire bien que $\#(L_n^\sigma(x,y,z)) = \mathcal{O}(n^2)$.

□

Nous venons de démontrer que tout mot purement morphique a une complexité en facteurs en $\mathcal{O}(n^2)$. Or, nous savions que les classes de mots que nous avons déjà étudiés ont des complexités en $\Theta(1), \Theta(n), \Theta(n\log\log n)$ et $\Theta(n\log n)$. Dès lors, la suite de cette section correspond à la démonstration du comportement asymptotique en $\Theta(n^2)$ de la fonction de complexité de la dernière classe de mots purement morphiques non encore étudiée.

Proposition 5.81. *Soient un mot purement morphique $u \in \mathcal{A}^\mathbb{N}$, $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ un endomorphisme qui engendre u et notons $\mathcal{B}$ (*resp. *$\mathcal{C}$) l'ensemble des lettres bornées (*resp. *croissantes) sous l'action du morphisme σ. Il existe un sous-ensemble fini*

$$Q \subseteq \mathcal{C} \times \mathcal{B}^* \times \mathcal{B}^* \times \mathcal{B}^* \times \mathcal{B}^* \times \mathcal{B}^* \times \mathcal{C}$$

tel que

$$L(u) \cap \mathcal{C}\,\mathcal{B}^*\,\mathcal{C} = \{cy_1z_1^kxz_2^ky_2c' \mid (c,y_1,z_1,x,z_2,y_2,c') \in Q,\ k \in \mathbb{N}\}.$$

Démonstration. Par le Lemme 5.77, il existe un entier $M \geq 1$ tel que $\sigma^{2M}(w) = \sigma^M(w)$ pour tout $w \in \mathcal{B}^*$. Notons à nouveau $N = (\#(\mathcal{C}))!$ et $\tau = \sigma^{MN}$.
Le morphisme τ, qui est une puissance de σ, engendre également le mot u. Dès lors, en appliquant le Corollaire 5.42 au morphisme τ, vu que $L(u) \cap \mathcal{C}\,\mathcal{B}^*\,\mathcal{C}$ ne contient que des mots de longueur au moins 2, il existe un ensemble fini $H \subseteq \mathcal{A} \times \mathcal{A}^* \times \mathcal{A}$ tel que

$$\begin{aligned} L(u) \cap \mathcal{C}\,\mathcal{B}^*\,\mathcal{C} &= (L(u) \setminus (\mathcal{A} \cup \{\varepsilon\})) \ \cap\ \mathcal{C}\,\mathcal{B}^*\,\mathcal{C} \\ &= \left(\bigcup_{(a,w,a')\in H} L^\tau(a,w,a') \right) \ \cap\ \mathcal{C}\,\mathcal{B}^*\,\mathcal{C} \\ &= \bigcup_{(a,w,a')\in H} (L^\tau(a,w,a') \ \cap\ \mathcal{C}\,\mathcal{B}^*\,\mathcal{C})\,. \end{aligned} \tag{5.23}$$

Prouvons que pour tout triplet $(a, w, a') \in \mathcal{A} \times \mathcal{A}^* \times \mathcal{A}$, et donc en particulier pour tout triplet de H, il existe un ensemble fini $Q_{(a,w,a')} \subseteq \mathcal{C} \times \mathcal{B}^* \times \mathcal{B}^* \times \mathcal{B}^* \times \mathcal{B}^* \times \mathcal{B}^* \times \mathcal{C}$ tel que

$$L^\tau(a, w, a') \cap \mathcal{C}\,\mathcal{B}^*\,\mathcal{C} = \{cy_1 z_1^k x z_2^k y_2 c' \mid (c, y_1, z_1, x, z_2, y_2, c') \in Q_{(a,w,a')},\ k \in \mathbb{N}\}. \tag{5.24}$$

<u>Cas 1</u> : Si $(a, w, a') \notin \mathcal{C} \times \mathcal{B}^* \times \mathcal{C}$, alors on a

$$L^\tau(a, w, a') \cap \mathcal{C}\,\mathcal{B}^*\,\mathcal{C} = \emptyset \tag{5.25}$$

et l'ensemble $Q_{(a,w,a')} = \emptyset$ satisfait la propriété. En effet, démontrons l'égalité (5.25). Trois sous-cas possibles peuvent être étudiés pour que l'on ait $(a, w, a') \notin \mathcal{C} \times \mathcal{B}^* \times \mathcal{C}$.

- Si $a \in \mathcal{B}$, alors pour tout $k \in \mathbb{N}$, on a $\sigma^k(a) \in \mathcal{B}^*$. Or, un mot de $L^\tau(a, w, a')$ correspond à un facteur central d'un triplet $(\sigma^k(a), \sigma^k(w), \sigma^k(a'))$ pour un $k \in \mathbb{N}$ et donc commence par un suffixe non vide de $\sigma^k(a)$ appartenant à $\mathcal{B}^+$. Ainsi, on a $L^\tau(a, w, a') \subseteq \mathcal{B}\,\mathcal{A}^+$.
- Si $a' \in \mathcal{B}$, alors on conclut de façon symétrique en étudiant le préfixe de $\sigma^k(a')$ pour démontrer que $L^\tau(a, w, a') \subseteq \mathcal{A}^+\,\mathcal{B}$.
- Si $w \notin \mathcal{B}^*$, alors w ne peut pas être le mot vide et il existe une lettre de w appartenant à $\mathcal{C}$, c'est-à-dire $w \in \mathcal{A}^*\,\mathcal{C}\,\mathcal{A}^*$. Étant donné que $\sigma(\mathcal{C}) \subseteq \mathcal{A}^*\,\mathcal{C}\,\mathcal{A}^*$, pour tout $k \in \mathbb{N}$, on a $\sigma^k(w) \in \mathcal{A}^*\,\mathcal{C}\,\mathcal{A}^*$. Or, tout mot de $L^\tau(a, w, a')$ correspond à un facteur central d'un triplet $(\sigma^k(a), \sigma^k(w), \sigma^k(a'))$ pour un $k \in \mathbb{N}$. Ainsi, par définition, tout mot de $L^\tau(a, w, a')$ est de la forme $s\sigma^k(w)p$ où $k \in \mathbb{N}$ et s et p appartiennent à $\mathcal{A}^+$ puisque s (resp. p) est un suffixe (resp. préfixe) non vide de $\sigma^k(a)$ (resp. $\sigma^k(a')$). On en tire que $L^\tau(a, w, a') \in \mathcal{A}^+\,\mathcal{C}\,\mathcal{A}^+$.

<u>Cas 2</u> : Si $(a, w, a') \in \mathcal{C} \times \mathcal{B}^* \times \mathcal{C}$. Par la Remarque 5.70 appliquée à l'endomorphisme τ, pour tout $k \in \mathbb{N}$, étudions l'unique facteur central de $(\tau^k(a), \tau^k(w), \tau^k(a'))$ appartenant à $\mathcal{C}\,\mathcal{B}^*\,\mathcal{C}$. Lorsque $k = 1$, par définition, il existe $c, c' \in \mathcal{C}$ et $x_1, x_2 \in \mathcal{B}^*$ tels que cx_1 (resp. x_2c') est un suffixe (resp. préfixe) non vide de $\tau(a)$ (resp. $\tau(a')$) et l'unique facteur central de $(\tau(a), \tau(w), \tau(a'))$ s'écrit $cx_1\tau(w)x_2c'$. Ensuite, par le Lemme 5.69, il existe $y_1, y_2 \in \mathcal{B}^*$ tels que cy_1 (resp. y_2c') est un suffixe (resp. préfixe) de $\tau(c)$ (resp. $\tau(c')$). Par hypothèse, on a $\tau = \sigma^{MN}$. Ainsi, en appliquant deux fois la Remarque 5.78, pour tout $x \in \mathcal{B}^*$, on obtient

$$\tau^2(x) = \sigma^{2MN}(x) = \sigma^{(2N)M}(x) = \sigma^M(x) = \sigma^{MN}(x) = \tau(x).$$

Ainsi, pour tout $x \in \mathcal{B}^*$, on a

$$\tau(\tau(x)) = \tau(x). \tag{5.26}$$

Notons $z_1 = \tau(y_1) \in \mathcal{B}^*$ et $z_2 = \tau(y_2) \in \mathcal{B}^*$. Pour tout $k \geq 1$, démontrons les deux assertions suivantes :

(i) $cy_1z_1^{k-1}$ est un suffixe de $\tau^k(c)$,

(ii) $z_2^{k-1}y_2c'$ est un préfixe de $\tau^k(c')$.

Le cas $k = 1$ est vérifié par définition pour les deux affirmations. Démontrons-les par induction pour $k+1$ en supposant les avoir démontrées pour $k \in \mathbb{N}_0$.

(i) Par hypothèse de récurrence, le mot $cy_1z_1^{k-1}$ est un suffixe de $\tau^k(c)$. Dès lors, en appliquant le morphisme τ aux deux mots considérés, le mot $\tau(cy_1z_1^{k-1}) = \tau(c)\underbrace{\tau(y_1)}_{=z_1}\tau(z_1)^{k-1}$ est un suffixe de $\tau^{k+1}(c)$. Or, le mot cy_1 est un suffixe de $\tau(c)$ et par l'égalité (5.26), on a $\tau(z_1) = \tau(\tau(y_1)) = \tau(y_1) = z_1$. Ainsi, on obtient que le mot $cy_1z_1(z_1)^{k-1} = cy_1(z_1)^k$ est un suffixe de $\tau^{k+1}(c)$.

(ii) Par induction, le mot $z_2^{k-1}y_2c'$ est un préfixe de $\tau^k(c')$ donc le mot $\tau(z_2)^{k-1}\underbrace{\tau(y_2)}_{=z_2}\tau(c')$ est un suffixe de $\tau^{k+1}(c')$. Or, de la même façon, le mot y_2c' est un préfixe de $\tau(c')$ et $\tau(z_2) = z_2$ donc le mot $z_2^{k-1}z_2y_2c' = z_2^ky_2c'$ est un préfixe de $\tau^{k+1}(c')$.

De plus, pour tout $k \geq 2$, on a

(iii) $cy_1z_1^{k-2}\tau(x_1)$ est un suffixe de $\tau^k(a)$,

(iv) $\tau(x_2)z_2^{k-2}y_2c'$ est un préfixe de $\tau^k(a')$.

En effet, démontrons ces affirmations en utilisant les propriétés (i) et (ii) que nous venons d'établir.

(iii) Soit $k \geq 2$. Le mot cx_1 est un suffixe de $\tau(a)$ donc, en appliquant le morphisme τ^{k-1} à chacun de ces deux mots, on obtient que $\tau^{k-1}(cx_1) = \tau^{k-1}(c)\tau^{k-1}(x_1)$ est un suffixe de $\tau^k(a)$. Or, vu que $x_1 \in \mathcal{B}^*$, par l'égalité (5.26), on a $\tau^{k-1}(x_1) = \tau(x_1)$ pour tout $k \geq 2$. Ensuite, en appliquant la propriété (i) à $k-1 \geq 1$, le mot $cy_1z_1^{k-2}$ est un suffixe de $\tau^{k-1}(c)$. On obtient donc que le mot $cy_1z_1^{k-2}\tau(x_1)$ est un suffixe de $\tau^k(a)$.

(iv) De la même façon, le mot x_2c' est un préfixe de $\tau(a')$ donc, en considérant $k \geq 2$, appliquons le morphisme τ^{k-1} à chacun de ces deux mots. On obtient que $\tau^{k-1}(x_2)\tau^{k-1}(c') = \tau(x_2)\tau^{k-1}(c')$ est un préfixe de $\tau^k(a')$. Or, en appliquant la propriété (ii) à $k-1 \geq 1$, le mot $z_2^{k-2}y_2c'$ est un préfixe de $\tau^{k-1}(c')$ donc $\tau(x_2)z_2^{k-2}y_2c'$ est un préfixe de $\tau^k(a')$.

Notons, dès à présent, $x = \tau(x_1wx_2)$. Pour tout $k \in \mathbb{N}_0$, on a $x = \tau(x_1)\tau^k(w)\tau(x_2)$. En effet, pour tout $k \geq 1$, étant donné que $x_1wx_2 \in \mathcal{B}^*$, par l'égalité (5.26), on a

$$x = \tau^k(x_1wx_2) = \tau^k(x_1)\tau^k(w)\tau^k(x_2) = \tau(x_1)\tau^k(w)\tau(x_2).$$

Ainsi, pour tout $k \geq 2$, on obtient que le mot $cy_1z_1^{k-2}xz_2^{k-2}y_2c'$ est l'unique facteur central de $(\tau^k(a), \tau^k(w), \tau^k(a'))$ appartenant à $\mathcal{C}\,\mathcal{B}^*\,\mathcal{C}$. En effet, par les propriétés (iii) et (iv), le mot $cy_1z_1^{k-2}\tau(x_1)$ (resp. $\tau(x_2)z_2^{k-2}y_2c'$) est un suffixe (resp. préfixe) non vide de $\tau^k(a)$ (resp. $\tau^k(a')$) appartenant à $\mathcal{C}\,\mathcal{B}^*$ (resp. $\mathcal{B}^*\,\mathcal{C}$).

Ensuite, par définition, $L^\tau(a, w, a') \cap \mathcal{C}\,\mathcal{B}^*\,\mathcal{C}$ est l'ensemble des facteurs centraux des triplets $(\tau^k(a), \tau^k(w), \tau^k(a'))$, $k \in \mathbb{N}$, appartenant à $\mathcal{C}\,\mathcal{B}^*\,\mathcal{C}$. Or, si $k = 0$, le seul facteur central de (a, w, a') appartenant à $\mathcal{C}\,\mathcal{B}^*\,\mathcal{C}$ est le mot awa' lui-même. Et si $k = 1$, par définition, le seul facteur central de $(\tau(a), \tau(w), \tau(a'))$ appartenant à $\mathcal{C}\,\mathcal{B}^*\,\mathcal{C}$ est le mot $cx_1\tau(w)x_2c'$. Ensuite, si $k \geq 2$, nous venons de montrer que $cy_1z_1^{k-2}xz_2^{k-2}y_2c'$ est l'unique facteur central de $(\tau^k(a), \tau^k(w), \tau^k(a'))$ appartenant à $\mathcal{C}\,\mathcal{B}^*\,\mathcal{C}$. En résumé, on a

$$\begin{aligned} L^\tau(a,w,a') \cap \mathcal{C}\,\mathcal{B}^*\,\mathcal{C} &= \{awa', cx_1\tau(w)x_2c'\} \cup \{cy_1z_1^{k-2}xz_2^{k-2}y_2c' \mid k \geq 2\} \\ &= \{awa', cx_1\tau(w)x_2c'\} \cup \{cy_1z_1^jxz_2^jy_2c' \mid j \in \mathbb{N}\}. \end{aligned}$$

Dès lors, posons

$$Q_{(a,w,a')} := \{(a,\varepsilon,\varepsilon,w,\varepsilon,\varepsilon,a'), (c,x_1,\varepsilon,\tau(w),\varepsilon,x_2,c'), (c,y_1,z_1,x,z_2,y_2,c')\}.$$

Cet ensemble $Q_{(a,w,a')}$ satisfait la propriété (5.24) voulue. Par l'égalité d'ensemble (5.23), l'ensemble $Q = \bigcup\limits_{(a,w,a')\in H} Q_{(a,w,a')}$ satisfait la propriété de l'énoncé vu que H est fini.

$\square$

Proposition 5.82. *Soient $n \in \mathbb{N}$ et un mot binaire $w \in \{0,1\}^n$. Pour tout $i \in \{1,2\}$, notons $(q_i, e_i) \in \mathbb{N}^2$ tel que le facteur $10^{e_i}1$ apparaît à la position q_i dans w. Si les entiers e_1 et e_2 sont plus grands ou égaux à $\frac{1}{2}n - 1$, alors $(q_1, e_1) = (q_2, e_2)$.*

Démonstration. Tout d'abord, vu que $|10^{e_1}1| + |10^{e_2}1| = 4 + e_1 + e_2 \geq 2 + n$, le mot w ne peut pas être de la forme $x\,10^{e_i}1\,y\,10^{e_j}1\,z$ avec $i,j \in \{1,2\}$, $i \neq j$ et $x,y,z \in \{0,1\}^*$. De plus, par le même argument sur les longueurs, les facteurs $10^{e_1}1$ et $10^{e_2}1$ ne peuvent pas se chevaucher uniquement sur un 1 de leurs extrémités respectives. Ainsi, vu les blocs centraux de 0 et les extrémités de 1, le facteur $10^{e_i}1$ ne peut pas apparaître à une position $k \in \{1, \ldots, e_j\}$ dans $10^{e_j}1$, avec $i,j \in \{1,2\}$. On en tire que $10^{e_1}1 = 10^{e_2}1$ et que ce facteur apparaît à la position $q_1 = q_2$.

$\square$

La borne $\frac{1}{2}n - 1$ donnée par cette proposition ne peut pas être améliorée car si e_1 ou e_2 est plus petit strictement que cette borne, alors le mot w pourrait contenir une instance du mot $10^{e_1}10^{e_2}1$. Un contre-exemple est donné par le mot $w = 101001 \in (\{0,1\})^6$ et les triplets $(0,1)$ et $(2,2)$ car $e_1 = 1 < \frac{6}{2} - 1 = 2$.

Proposition 5.83. *Soient un mot binaire $v \in \{0,1\}^{\mathbb{N}}$, un réel $r \in \mathbb{R}_{\geq 1}$ et une suite $(e_k)_{k\in\mathbb{N}} \in \mathbb{N}^{\mathbb{N}}$ satisfaisant les propriétés suivantes :*

(i) Pour tout $k \in \mathbb{N}$, $10^{e_k}1$ est un facteur de v.

(ii) Pour tout $h,k \in \mathbb{N}$, si $h \neq k$, alors $e_h \neq e_k$.

(iii) Il existe $C > 0$ tel que $\lim\limits_{k\to\infty} \frac{e^k}{k^r} = C$.

La fonction de complexité en facteurs du mot v est en $\Omega(n^{1+\frac{1}{r}})$.

Démonstration. Considérons un entier naturel $n \in \mathbb{N}$ et définissons le mot $v_n = 0^n v$ et l'ensemble

$$K_n = \{k \in \mathbb{N} \mid \frac{1}{2}n - 1 \leq e_k \leq \frac{3}{4}n - 2\}.$$

Remarquons tout d'abord que, par la Remarque 1.8, on a

$$p_v(n) \leq p_{0^n v}(n) \leq p_v(n) + \underbrace{|0^n|}_{=n}$$

et donc, en particulier,

$$p_{v_n}(n) - n \leq p_v(n). \tag{5.27}$$

Remarquons ensuite que, par la propriété (iii), l'estimation du cardinal de l'ensemble K_n se ramène à une étude similaire à celle faite dans l'Exemple 3.29 en démontrant que tout entier naturel k appartenant à l'intervalle[11]

$$\left[\left[\left(\frac{n-2}{2C}\right)^{\frac{1}{r}}, \left(\frac{n-\frac{8}{3}}{\frac{4C}{3}}\right)^{\frac{1}{r}} \right]\right]$$

satisfait les inégalités définissant l'ensemble K_n. Cela nous permet donc d'affirmer que

$$\#(K_n) = \Omega(n^{\frac{1}{r}}). \tag{5.28}$$

Construisons maintenant une injection de l'ensemble $[[0, \lfloor \frac{n}{4} \rfloor]] \times K_n$ dans l'ensemble $L_n(v_n)$. Soit $(q, k) \in [[0, \lfloor \frac{n}{4} \rfloor]] \times K_n$, on a

$$q + |10^{e_k}1| = q + 2 + e_k \leq \frac{1}{4}n + 2 + \frac{3}{4}n - 2 = n. \tag{5.29}$$

Or, vu que pour tout $k \in \mathbb{N}$, le mot $10^{e_k}1$ est un facteur de v, il est également un facteur de v_n. L'occurrence de ce facteur dans v_n a dès lors au moins n lettres la précédant et donc en particulier au moins $q < n$ lettres la précédant également. Ainsi, par l'inégalité (5.29), cela a donc du sens de considérer un facteur de longueur n de v_n, que nous noterons $w_{q,k,n}$, défini tel que $10^{e_k}1$ apparaisse à la position q dans[12] $w_{q,k,n}$. Cette procédure définit une fonction de l'ensemble $[[0, \lfloor \frac{n}{4} \rfloor]] \times K_n$ dans l'ensemble $L_n(v_n)$. Or, étant donné que les entiers e_k sont tous différents, par la propriété (ii), et que $e_k \geq \frac{1}{2}n - 1$, la Proposition 5.82 affirme qu'un seul couple $(q, k) \in [[0, \lfloor \frac{n}{4} \rfloor]] \times K_n$ ne peut avoir pour image $w_{q,k,n}$ par cette fonction. On en tire que la fonction qui associe à tout couple (q, k) de $[[0, \lfloor \frac{n}{4} \rfloor]] \times K_n$, le facteur $w_{q,k,n} \in L_n(v_n)$, est injective. Ainsi, on a

$$p_{v_n}(n) = \#(L_n(v_n)) \geq \# \left(\left[\left[0, \left\lfloor \frac{1}{4}n \right\rfloor \right]\right] \times K_n \right) = (\left\lfloor \frac{n}{4} \right\rfloor + 1)\,\#(K_n) > \frac{n}{4}\,\#(K_n).$$

11. On a $\left(\frac{n-\frac{8}{3}}{\frac{4C}{3}}\right)^{\frac{1}{r}} \geq \left(\frac{n-2}{2C}\right)^{\frac{1}{r}} \Leftrightarrow n \geq 4$.

12. La construction du mot v_n est maintenant compréhensible car sans l'introduction de ce nouveau mot, le mot $w_{q,k,n}$ peut ne pas être un facteur de v. En effet, le mot $w_{q,k,n}$ est un facteur de v seulement si $10^{e_k}1$ apparaît à une position plus grande ou égale à q dans v.

Dès lors, par le comportement asymptotique (5.28), on obtient que $p_{v_n}(n) = \Omega(n^{1+\frac{1}{r}})$. Finalement, l'inégalité (5.27) permet de démontrer que la fonction de complexité en facteurs de v est en $\Omega(n^{1+\frac{1}{r}})$ également.

□

Lemme 5.84. *Soient un mot purement morphique $u \in \mathcal{A}^{\mathbb{N}}$ et un endomorphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ qui engendre u. Si le mot u n'est pas ultimement périodique, alors il existe une lettre $c \in \mathcal{A}$ telle que c croît sous l'action du morphisme σ et c apparaît infiniment souvent dans le mot u.*

Démonstration. Notons $u_0 \in \mathcal{A}$ la première lettre de u. Notons $x \in \mathcal{A}^+$ tel que $\sigma(u_0) = u_0x$. Deux cas sont possibles quant à la croissance du mot x sous l'action du morphisme σ.
Cas 1 : Si $x \in \mathcal{B}^+$. Notons $\tau = \sigma^M$ où $M \geq 1$ est l'entier naturel donné par le Lemme 5.77. Pour tout $w \in \mathcal{B}^*$, $\sigma^{2M}(w) = \sigma^M(w)$. Par hypothèse, le mot u est engendré par l'endomorphisme σ et donc l'est également par toute puissance de σ. En particulier, l'endomorphisme τ engendre le mot u et donc τ est prolongeable en u_0. Notons $y \in \mathcal{A}^+$ le mot défini par $\tau(u_0) = u_0y$. On a $\tau(u_0) = \sigma^M(u_0)$ donc on obtient $u_0y = u_0x\sigma(x)\,\sigma^2(x)\cdots\sigma^{M-1}(x)$, c'est-à-dire $y = x\sigma(x)\,\sigma^2(x)\cdots\sigma^{M-1}(x)$. Or, par définition de l'alphabet $\mathcal{B}$, on a $\sigma(\mathcal{B}) \subseteq \mathcal{B}^*$. Dès lors, on obtient que $y \in \mathcal{B}^+$ et, en particulier, $\tau(\tau(y)) = \sigma^{2M}(y) = \sigma^M(y) = \tau(y)$. Ensuite, démontrons que pour tout $k \geq 1$, on a

$$\tau^k(u_0) = \tau(u_0)(\tau(y))^{k-1}. \tag{5.30}$$

Le cas de base où $k = 1$ est trivialement vérifié et si $k = 2$, alors on obtient $\tau^2(u_0) = \tau(\tau(u_0)) = \tau(u_0y) = \tau(u_0)\tau(y)$. Supposons dès lors avoir démontré la véracité de l'égalité (5.30) pour $k \geq 2$ et démontrons-la pour $k+1$ par induction. On a

$$\begin{aligned}
\tau^{k+1}(u_0) &= \tau(\tau^k(u_0)) \\
&= \tau(\tau(u_0)(\tau(y))^{k-1}) \\
&= \tau^2(u_0)(\tau(\tau(y)))^{k-1} \\
&= \tau(u_0)\tau(y)(\tau(y))^{k-1} \\
&= \tau(u_0)(\tau(y))^k.
\end{aligned}$$

Dès lors, par le résultat (5.30), on obtient que le mot u, étant engendré par l'endomorphisme τ, peut être écrit de la forme $u = \tau^\omega(u_0) = \tau(u_0)(\tau(y))^\omega$. On en tire que le mot u est ultimement périodique.
Cas 2 : Si $x \in \mathcal{A}^*\,\mathcal{C}\,\mathcal{A}^*$. Par définition de l'alphabet $\mathcal{C}$, l'ensemble des lettres croissantes sous l'action du morphisme σ, on a $\sigma(\mathcal{C}) \subseteq \mathcal{A}^*\,\mathcal{C}\,\mathcal{A}^*$. Ainsi, pour tout $k \in \mathbb{N}$, il existe une lettre $c_k \in \mathcal{C}$ telle que c_k apparaît dans $\sigma^k(x)$. Or, l'alphabet $\mathcal{C}$ est un ensemble fini par définition. Ainsi, il existe une lettre $c \in \mathcal{C}$ telle que $c = c_k$ pour une infinité de $k \in \mathbb{N}$. Remarquons finalement que l'on a $u = u_0x\sigma(x)\,\sigma^2(x)\,\sigma^3(x)\cdots$ et donc cette lettre $c \in \mathcal{C}$ satisfait la propriété de l'énoncé.

□

Théorème 5.85. *Soient $u \in \mathcal{A}^{\mathbb{N}}$ un mot purement morphique et $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ un endomorphisme qui engendre u. Si le mot u n'est pas ultimement périodique et s'il existe une infinité de facteurs différents de u bornés sous l'action du morphisme σ, alors on a $p_u(n) = \Theta(n^2)$.*

Démonstration. Par le Théorème 5.75, pour démontrer que $p_u(n) = \Theta(n^2)$, il nous suffit de prouver que $p_u(n) = \Omega(n^2)$. Étant donné que le mot purement morphique u n'est pas ultimement périodique et que l'ensemble $L(u) \cap \mathcal{B}^*$ est infini, le Lemme 5.84 affirme que $L(u) \cap \mathcal{C}\,\mathcal{B}^*\,\mathcal{C}$ est infini. Or, par la Proposition 5.81, il existe un sous-ensemble fini $Q \subseteq \mathcal{C} \times \mathcal{B}^* \times \mathcal{B}^* \times \mathcal{B}^* \times \mathcal{B}^* \times \mathcal{B}^* \times \mathcal{C}$ tel que

$$L(u) \cap \mathcal{C}\,\mathcal{B}^*\,\mathcal{C} = \{cy_1z_1^kxz_2^ky_2c' \mid (c, y_1, z_1, x, z_2, y_2, c') \in Q,\ k \in \mathbb{N}\}.$$

Ainsi, étant donné que cet ensemble est infini, il existe un élément $(c, y_1, z_1, x, z_2, y_2, c') \in Q$ tel que au moins un des mots z_1 et z_2 est non vide. En effet, dans le cas contraire, si tous les éléments de Q sont tels que $z_1 = z_2 = \varepsilon$, alors on a $\#(L(u) \cap \mathcal{C}\,\mathcal{B}^*\,\mathcal{C}) = \#(Q)$. Ceci est absurde car l'ensemble Q est fini. Dès lors, il existe $(c, y_1, z_1, x, z_2, y_2, c') \in Q$ tel que $z_1z_2 \neq \varepsilon$ et $cy_1z_1^kxz_2^ky_2c'$ est un facteur de u pour tout $k \in \mathbb{N}$. Définissons le codage

$$\chi : \mathcal{A}^* \to \{0,1\}^*,\ a \mapsto \begin{cases} 0 & \text{si } a \in \mathcal{B} \\ 1 & \text{si } a \in \mathcal{C}\,. \end{cases}$$

De plus, pour tout $k \in \mathbb{N}$, si l'on note $e_k = |y_1z_1^kxz_2^ky_2| = |z_1z_2|k + |y_1xy_2|$, on a $\chi(cy_1z_1^kxz_2^ky_2c') = 10^{e_k}1$. Le mot $10^{e_k}1$ est un facteur de $\chi(u)$ pour tout $k \in \mathbb{N}$, la suite $(e_k)_{k\in\mathbb{N}}$ est telle que $e_k \neq e_h$ si $k \neq h$ car $|z_1z_2| \neq 0$. Qui plus est, on a $\lim_{k\to\infty} \frac{e_k}{k} = |z_1z_2| > 0$. Dès lors, par la Proposition 5.83, la fonction de complexité du mot $\chi(u) \in \{0,1\}^{\mathbb{N}}$ satisfait $p_{\chi(u)}(n) = \Omega(n^2)$. La conclusion découle du Lemme 4.19 qui affirme que, vu que χ est un codage, les fonctions de complexité de u et $\chi(u)$ satisfont l'inégalité

$$p_{\chi(u)}(n) \leq p_u(n),$$

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

$\square$

Exemple 5.86. Considérons l'endomorphisme $\sigma : \{0,1\}^* \to \{0,1\}^*$ défini tel que $\sigma(0) = 001$ et $\sigma(1) = 1$. Notons $u = \sigma^\omega(0)$. Pour tout $k \in \mathbb{N}$, la première lettre de $\sigma^k(01) = \sigma^k(0)\sigma^k(1)$ est 0 car σ est prolongeable en la lettre 0. Ensuite, démontrons par récurrence que 01^{k+1} est un suffixe de $\sigma^k(01)$. Le cas de base $k = 0$ est vérifié car $\sigma^0(01) = 01$. Procédons par récurrence. Supposons que cette affirmation est vérifiée pour $k \in \mathbb{N}$ et vérifions-la pour $k + 1$. Par hypothèse de récurrence, 01^{k+1} est un suffixe de $\sigma^k(01)$. En appliquant le morphisme σ à chacun de ces mots, on obtient que $\sigma(01^{k+1}) = \sigma(0)(\sigma(1))^{k+1} = 0011^{k+1} = 001^{k+2}$ est un suffixe de $\sigma^{k+1}(01)$. Ainsi, vu que $u = 0\,\sigma(01)\,\sigma^2(01)\,\sigma^3(01)\cdots$, on a que 01^k0 est un facteur de u pour tout $k \in \mathbb{N}$. En effet, le facteur 00 obtenu lors de la concaténation de 0 et de $\sigma(01)$, le facteur 010 apparaît dans l'occurrence de $\sigma^2(01) = 00100111$. Les facteurs 01^k0 pour $k \geq 2$ sont obtenus lors de la concaténation de $\sigma^{k-1}(01)$ et de $\sigma^k(0)$. Le mot u n'est dès lors pas ultimement périodique. De plus, la lettre 1 est bornée sous l'action du morphisme σ. Il existe donc une infinité de facteur différents de u bornés sous l'action du morphisme σ Par le Théorème 5.85, on conclut que $p_u(n) = \Theta(n^2)$.

5.6 Conclusion : utilisation des outils établis pour l'obtention du théorème de Pansiot

Rappelons que le but de ce chapitre était de démontrer le Théorème 5.2 de Pansiot affirmant que la complexité en facteurs d'un mot purement morphique ne peut adopter que l'un des comportements asymptotiques suivants : $\Theta(1), \Theta(n)$, $\Theta(n \log\log n), \Theta(n \log n)$ ou $\Theta(n^2)$. Nous avons dès lors tous les outils nécessaires à la démonstration de ce théorème.

Démonstration. Preuve du Théorème 5.2
Considérons un mot purement morphique $u \in \mathcal{A}^{\mathbb{N}}$ et un endomorphisme $\sigma : \mathcal{A}^* \to \mathcal{A}^*$ qui engendre u. Par définition, le morphisme σ est prolongeable en u_0, la première lettre de u, et $u = \sigma^\omega(u_0)$.
Si u est ultimement périodique, alors par la Proposition 1.23, on a $p_u(n) = \Theta(1)$. Supposons dès lors que le mot u n'est pas ultimement périodique. Dans ce cas, s'il y a un nombre fini de facteurs de u bornés sous l'action de σ, alors le Théorème 5.73 affirme que la fonction de complexité en facteurs du mot u est en $\Theta(n)$,$\Theta(n \log\log n)$ ou $\Theta(n \log n)$. Finalement, si au contraire, il existe une infinité de facteurs de u différents et bornés sous l'action du morphisme σ, alors le Théorème 5.85 démontre que la fonction de complexité du mot u est en $\Theta(n^2)$.

□

Chapitre 6

Exemples de calcul de complexité de mots morphiques

Ce chapitre est consacré à l'étude de l'expression analytique exacte de la fonction de complexité en facteurs des deux mots purement morphiques les plus fréquemment rencontrés dans le domaine de la combinatoire des mots : le mot de Thue-Morse et le mot de Fibonacci. Nous allons utiliser des applications du Théorème 5.2 de Pansiot pour obtenir au préalable des comportements asymptotiques et ensuite les outils mis en place dans la section 2.2.1. En effet, par le Théorème 2.21 et la Proposition 2.20, la fonction de complexité d'un mot infini peut entièrement être étudiée en connaissant l'ensemble de ses facteurs spéciaux ou bispéciaux de chaque longueur ainsi que leur multiplicité. Ces informations peuvent parfois être aisément connues. Ce sera notamment le cas dans les deux exemples étudiés.

6.1 Mot de Thue-Morse

Rappelons que le mot de Thue-Morse est le mot purement morphique

$$t = abbabaabbaababba \cdots$$

engendré par le morphisme 2-uniforme $\sigma : \{a, b\}^* \to \{a, b\}^*$ défini par $\sigma(a) = ab$ et $\sigma(b) = ba$ en débutant par la lettre a.

Étant donné que le morphisme σ est 2-uniforme, il est en particulier quasi-uniforme. Ainsi, par le Théorème 5.58, on a [1] $p_t(n) = \mathcal{O}(n)$. Ensuite, grâce au Théorème 5.2 de Pansiot, nous savons que les seules possibilités correspondant à ce cas sont $p_t(n) = \Theta(1)$ ou $p_t(n) = \Theta(n)$. Or, vu que le mot de Thue-Morse ne contient aucun cube, il ne peut donc pas être ultimement périodique. Ainsi, par le Théorème 2.2 de Morse et Hedlund, le comportement asymptotique $p_t(n) = \Theta(1)$ est à rejeter et on obtient $p_t(n) = \Theta(n)$.

La complexité en facteurs du mot de Thue-Morse nous est fournie de façon asymptotique par le Théorème 5.2 de Pansiot mais la valeur de la fonction p_t peut être étudiée plus précisément pour tout naturel $n \in \mathbb{N}$.

1. Cette complexité asymptotique a déjà été démontrée, par une autre méthode, dans l'Exemple 4.26

Lemme 6.1. *Le mot de* Thue-Morse *est récurrent.*

Démonstration. Démontrons tout d'abord qu'un mot $u \in \mathcal{A}^{\mathbb{N}}$ est récurrent si et seulement si tous ses préfixes apparaissent deux fois dans u.
La condition est trivialement nécessaire. Dès lors, supposons que tous les préfixes de u apparaissent deux fois dans u et démontrons que cela implique que le mot u est récurrent. Considérons w un facteur de u. Il suffit de démontrer que pour tout $i \in \mathbb{N}$ tel que le facteur w a une occurrence à la position i dans u, il existe un entier $j > i$ tel que w apparaît également à cette position. Pour ce faire, considérons le préfixe de u se terminant après l'occurrence de w ; $p = u_0 \cdots u_{i+|w|-1}$. Par hypothèse, le mot p apparaît une deuxième fois dans u à une position $k > 0$. On en conclut que le facteur w apparaît également une deuxième fois à la position $k + i > i$.

Il nous suffit dès lors de démontrer que tout préfixe du mot de Thue-Morse t apparaît deux fois. Soit p un préfixe de $t = \sigma^{\omega}(a)$. Il existe $n \in \mathbb{N}$ tel que p est un préfixe de $\sigma^n(a)$. Or, le mot $\sigma^{n+2}(a)$ est également un préfixe de t. Et on a

$$\sigma^{n+2}(a) = \sigma^n(\sigma^2(a)) = \sigma^n(abba) = \sigma^n(a)\,\sigma^n(b)\,\sigma^n(b)\,\sigma^n(a).$$

Donc p apparaît une deuxième fois dans t étant donné la seconde occurrence de $\sigma^n(a)$.

□

Nous venons de démontrer que le mot t de Thue-Morse est récurrent. Ainsi, vu que t est binaire, le Corollaire 2.22 pourra être utilisé en travaillant avec la seconde différence b et les facteurs bispéciaux du mot t.

Lemme 6.2. *Soit $w \in L(t)$, le mot w est de la forme $r_1\sigma(v)r_2$ où $v \in L(t)$ et $r_1, r_2 \in \{\varepsilon, a, b\}$. De plus, si $|w| \geq 5$, alors cette décomposition est unique.*

Démonstration. Considérons tout d'abord les facteurs du mot de Thue-Morse de longueur inférieure ou égale à 4. Le tableau suivant montre qu'ils admettent une décomposition de la forme demandée. La convention pour découper les facteurs dans cette illustration a été choisie comme suit : si le facteur contient aa ou bb alors la décomposition doit obligatoirement couper cette occurrence en deux parties car ces mots ne sont pas des images de lettres de t étant donné que $\{\sigma(a), \sigma(b)\} = \{ab, ba\}$, sinon on considère le découpage obtenu en parcourant le facteur de gauche à droite en considérant, dès lors, le plus long facteur possible v.

Facteurs	Décompositions	Facteurs	Décompositions
ε	$\varepsilon\,\sigma(\varepsilon)\,\varepsilon$	$aaba$	$a\,\sigma(a)\,a$
a	$a\,\sigma(\varepsilon)\,\varepsilon$	$aabb$	$a\,\sigma(a)\,b$
b	$b\,\sigma(\varepsilon)\,\varepsilon$	$abaa$	$a\,\sigma(b)\,a$
aa	$a\,\sigma(\varepsilon)\,a$	$abab$	$\varepsilon\,\sigma(aa)\,\varepsilon$
ab	$\varepsilon\,\sigma(a)\,\varepsilon$	$abba$	$\varepsilon\,\sigma(ab)\,\varepsilon$
ba	$\varepsilon\sigma(b)\varepsilon$	$baab$	$\varepsilon\sigma(ba)\varepsilon$
bb	$b\,\sigma(\varepsilon)\,b$	$baba$	$\varepsilon\,\sigma(bb)\,\varepsilon$
aab	$a\,\sigma(a)\,\varepsilon$	$babb$	$b\,\sigma(a)\,b$
aba	$\varepsilon\,\sigma(a)\,a$	$bbaa$	$b\,\sigma(b)\,a$
abb	$\varepsilon\,\sigma(a)\,b$	$bbab$	$b\,\sigma(b)\,b$
baa	$\varepsilon\,\sigma(b)\,a$		
bab	$\varepsilon\,\sigma(b)\,b$		
bba	$b\,\sigma(b)\,\varepsilon$		

Ensuite considérons un facteur de THUE-MORSE de longueur 5 et démontrons qu'il admet une telle décomposition et que celle-ci est unique. Tout d'abord, remarquons que $aa \in L(w)$ ou bien $bb \in L(w)$. En effet, dans le cas contraire, un a ne pourrait être suivi que par un b et vice-versa. Ainsi, le facteur w ne pourrait être que $ababa$ ou $babab$. Or, ces deux facteurs sont des chevauchements et on sait que le mot de THUE-MORSE ne peut contenir de tels mots. Ainsi, le facteur w a obligatoirement une occurrence de aa et/ou de bb. Or, étant donné que aa et bb ne sont pas des images de lettres par le morphisme σ, la décomposition du facteur w est obtenue en coupant au milieu du facteur aa ou bb et en découpant en facteurs de longueur 2 la suite du mot w, en partant de part et d'autre de cette coupe [2]. À chaque extrémité du mot w, il reste des facteurs de longueur au plus 1 ne pouvant plus être sélectionnés par le raisonnement établi. Étant donné que w est un facteur de t qui est un point fixe du morphisme σ, il est inclus dans l'image d'un facteur plus petit. Ainsi, le découpage explicité ci-avant est unique et de la forme demandée

$$r_1\sigma(v)r_2$$

où $v \in L(t)$ et $r_1, r_2 \in \{\varepsilon, a, b\}$.
Finalement, considérons les facteurs de longueur plus grande ou égale à 6. En utilisant le résultat prouvé ci-avant, leurs préfixes de longueur 5 admettent un découpage unique. Dès lors, celui-ci induit leurs découpages pour la suite du facteur.

□

En utilisant les concepts de facteurs bispéciaux forts, faibles et neutres introduits dans la Définition 2.16, on obtient le résultat suivant.

2. Dans le cas où les facteurs aa et bb sont tous les deux présents dans w, les seules possibilités sont $w = aabba$, $w = baabb$, $w = abbaa$ ou $w = bbaab$ vu que t est sans cube. Les découpages obtenu en commençant par "séparer" les deux occurrences successives de a ou bien les deux successives de b sont identiques.

Proposition 6.3. *Si l'on note respectivement F, f et N les ensembles de facteurs bispéciaux forts, faibles et neutres du mot t de* Thue-Morse*, on a*

(i) $F = \{\varepsilon\} \cup \{\sigma^k(ab), \sigma^k(ba) \mid k \in \mathbb{N}\}$,

(ii) $f = \{\sigma^k(aba), \sigma^k(bab) \mid k \in \mathbb{N}\}$,

(iii) $N = \{a, b\}$.

Démonstration. Considérons un facteur $w \in L(t)$ tel que $|w| \geq 5$. Par le Lemme 6.2, w admet une unique décomposition $w = r_1\sigma(v)r_2$ avec $v \in L(t)$, $r_1, r_2 \in \{\varepsilon, a, b\}$. Étudions les différents cas possibles, en fonctions des valeurs de r_1 et r_2.

- Si $r_2 = a$, alors $w = r_1\sigma(v)a$ et la dernière lettre de w est a. Si le mot wa est également un facteur de t, alors le Lemme 6.2 peut s'appliquer également à wa, c'est-à-dire il existe $v' \in L(t), r'_1, r'_2 \in \{\varepsilon, a, b\}$ tels que $wa = r'_1\sigma(v')r'_2$. Plus particulièrement, on a $r'_2 \in \{\varepsilon, a\}$ car le cas $r'_2 = b$ est à rejeter. Si $r'_2 = \varepsilon$, alors étant donné que la dernière lettre de w est a, le mot $\sigma(v')$ doit se terminer par aa. Ceci est absurde par définition de l'endomorphisme σ. Ainsi, on obtient obligatoirement que $r'_2 = a$. Cela implique que $wa = r'_1\sigma(v')a$ et donc $w = r'_1\sigma(v')\varepsilon$. Cette construction nous a fourni une deuxième décomposition du facteur w, c'est absurde vu l'unicité donnée par le Lemme 6.2. En conclusion, si $r_2 = a$, alors le facteur w est suivi par un b.
- Si $r_2 = b$, alors de la même façon, si le mot wb est également un facteur de t, alors considérons sa décomposition donnée par le Lemme 6.2 : $wb = r'_1\sigma(v')r'_2$. Le cas $r'_2 = a$ est également à rejeter. De plus, on ne peut pas avoir $r'_2 = \varepsilon$ car le mot w se termine par b et donc $wb = r'_1\sigma(v')\varepsilon$ impliquerait que $\sigma(v')$ se termine par bb, ce qui est impossible. Ainsi, on a $r'_2 = b$ et une deuxième décomposition $r'_1\sigma(v')\varepsilon$ du mot w est donnée. Ceci est également absurde. Ainsi, si $r_2 = b$, alors le facteur w est suivi par un a.

Les deux cas précédents impliquent que si le facteur w est spécial à droite, alors il faut forcément que r_2 soit le mot vide. Ainsi, les facteurs $wa, wb \in L(t)$ s'écrivent obligatoirement $wa = r_1\sigma(v)a$ et $wb = r_1\sigma(v)b$. Au vu des cas précédents, ces deux facteurs wa et wb de t peuvent uniquement être suivis par b et a respectivement. Or, si la factorisation de ces facteurs étaient de la forme $wab = r'_1\sigma(v')r'_2$ et $wba = r''_1\sigma(v'')r''_2$, où $r'_2, r''_2 \neq \varepsilon$, alors on a $r'_2 = b$, $r''_2 = a$ et $r'_1\sigma(v')$, $r''_1\sigma(v'')$ seraient des autres décompositions de wa et wb respectivement. Ceci est absurde par unicité de la factorisation. Ainsi, par définition du morphisme σ, on a $\sigma(a) = ab$ et $\sigma(b) = ba$, donc la décomposition de wab est donnée par $wab = r_1\sigma(va)\varepsilon$ et celle de wba par $wba = r_1\sigma(vb)\varepsilon$. De part ces constructions, les mots va et vb sont également des facteurs de t. On en conclut que v est également spécial à droite. En résumé, si l'on note $\overline{a} = b$ et $\overline{b} = a$, on obtient le résultat suivant :

(i) si $|w| \geq 5$ $y \neq \varepsilon$ et si $r_1\sigma(b)y$ est l'unique factorisation donnée par le Lemme 6.2 de w, alors wy n'est pas un facteur du mot de Thue-Morse et $r_1\sigma(wy)$ est l'unique factorisation valide (au sens du Lemme 6.2) de $w\overline{y}$.

Ainsi, nous venons de montrer que les facteurs spéciaux à droite de longueur au moins 5 de t sont tous de la forme $w = r_1\sigma(v)$ où v est lui-même un facteur spécial à droite de t. Par un raisonnement symétrique, on peut conclure que :

(ii) si $|w| \geq 5$ $y \neq \varepsilon$ et si $y\sigma(b)r_2$ est l'unique factorisation de w, alors yw n'est pas un facteur du mot de Thue-Morse et $\sigma(\overline{y}w)r_2$ est l'unique factorisation valide de $\overline{y}w$.

On en tire que les facteurs spéciaux à gauche de longueur au moins 5 de t sont tous de la forme $w = \sigma(v)r_2$ où v est lui-même un facteur spécial à gauche de t. Ainsi, les facteurs bispéciaux de longueur au moins 5 de t sont de la forme $w = \sigma(v)$ où v est également un facteur bispécial de t. De plus, les facteurs w et v ont la même multiplicité m calculée comme suit : $m(w) = \#\,(E(w)) - d^-(w) - d^+(w) + 1$. En effet, par définition des facteurs bispéciaux de mots binaires, on a $d^-(w) = d^+(w) = d^-(v) = d^+(v) = 2$ et il suffit dès lors de vérifier que $\#(E(w)) = \#(E(v))$. Posons $\overbrace{(x,y)} = (\overline{x}, y)$. Par définition, si $(x,y) \in (\{a,b\})^2$, alors $\overbrace{(x,y)}$ est le couple dont la première composante est la deuxième lettre de $\sigma(x)$ et la seconde composante est la première lettre de $\sigma(y)$.

— Si le couple $(x,y) \in (\{a,b\})^2$ appartient à $E(v)$, alors $\overbrace{(x,y)} \in E(w)$. En effet, si $(x,y) \in E(v)$, on a $xvy \in L(t)$. Vu que $\sigma(t) = t$, on obtient $\sigma(xvy) = \sigma(x)\sigma(v)\sigma(y) = \sigma(x)w\sigma(y) \in L(t)$. Ainsi, vu que $\sigma(x) = x\overline{x}$ et $\sigma(y) = y\overline{y}$, on a $(\overline{x}, y) = \overbrace{(x,y)} \in E(w)$.

— Si le couple $(x,y) \in (\{a,b\})^2$ appartient à $E(w)$, alors $\overbrace{(x,y)} \in E(v)$. En effet, si $(x,y) \in E(w)$, on a $xwy = x\sigma(v)y \in L(t)$. Or, par le Lemme 6.2, vu que xwy est un facteur de t de longueur au moins 5 et que $x\sigma(v)y$ en est une décomposition satisfaisant aux propriétés de l'énoncé, alors $x\sigma(v)y$ est l'unique décomposition de xwy. Ainsi, par les résultats (i) et (ii) établis ci-avant, ce facteur est obligatoirement précédé de $\overline{x}$ et suivi de $\overline{y}$. On en tire $\overline{x}x\sigma(v)y\overline{y} = \sigma(\overline{x})\sigma(v)\sigma(y)$ est un facteur de t et, par unicité, sa décomposition est donnée par $\sigma(\overline{x}vy)$. On obtient que $(\overline{x}, y) = \overbrace{(x,y)} \in E(v)$.

Ainsi, le facteur v est un facteur bispécial de t, de longueur inférieure à w vu que σ est 2-uniforme, tel que $m(v) = m(w)$.

Pour conclure, il suffit d'étudier les facteurs bispéciaux de longueur au plus 4 et les plus longs en seront les images par σ. Rappelons tout d'abord que le mot de Thue-Morse ne contient aucun cube, donc les facteurs en contenant, tels que $aaa, bbb, aaab, baaa, bbba, abbb$, $aaaa$ et $bbbb$, ne doivent pas être étudiés. En utilisant les concepts et les notations de la section 2.2.1, rappelons que les facteurs bispéciaux sont caractérisés comme forts, faibles ou neutres en fonction de leur multiplicité.

— ε est un facteur bispécial fort car $E^-(\varepsilon) = \{a,b\}$, $E^+(\varepsilon) = \{a,b\}$,
$E(\varepsilon) = \{(a,a),(a,b),(b,a),(b,b)\}$ donc $\#(E(\varepsilon)) = 4$, $d^-(\varepsilon) = 2 = d^+(\varepsilon)$ et $m(\varepsilon) = 1$.

— a est un facteur bispécial neutre car $E^-(a) = \{a,b\}$, $E^+(a) = \{a,b\}$,
$E(a) = \{(a,b),(b,a),(b,b)\}$ donc $\#(E(a)) = 3$, $d^-(a) = 2 = d^+(a)$ et $m(a) = 0$.

— b est un facteur bispécial neutre car $E^-(b) = \{a,b\}$, $E^+(b) = \{a,b\}$,
$E(b) = \{(a,a),(a,b),(b,a)\}$ donc $\#(E(b)) = 3$, $d^-(b) = 2 = d^+(b)$ et $m(b) = 0$.

— aa n'est pas un facteur bispécial car $aaa \notin L(t)$.

— ab est un facteur bispécial fort car $E^-(ab) = \{a, b\}$, $E^+(ab) = \{a, b\}$, $E(ab) = \{(a, a),(a, b),(b, a),(b, b)\}$ donc $\#(E(ab)) = 4$, $d^-(ab) = 2 = d^+(ab)$ et $m(ab) = 1$.

— ba est un facteur bispécial fort car $E^-(ba) = \{a, b\}$, $E^+(ba) = \{a, b\}$, $E(ba) = \{(a, a),(a, b),(b, a),(b, b)\}$ donc $\#(E(ba)) = 4$, $d^-(ba) = 2 = d^+(ba)$ et $m(ba) = 1$.

— bb et aab ne sont pas des facteurs bispéciaux car $bbb \notin L(t)$ et $aaab \notin L(t)$.

— aba est un facteur bispécial faible car $E^-(aba) = \{a, b\}$, $E^+(aba) = \{a, b\}$, $E(aba) = \{(a, b),(b, a)\}$. Remarquons que $(b, b) \notin E(aba)$ car $babab$ est un chevauchement et t ne peut pas contenir de tels mots. De plus, $(a, a) \notin E(aba)$ car $aabaa$ ne peut pas être présent dans t : en effet, le mot t ne contient aucun cube donc si ce facteur est présent, alors il est obligatoirement précédé par b et suivi par b. Or, le facteur $baabaab$ est un chevauchement. Donc $\#(E(aba)) = 2$, $d^-(aba) = 2 = d^+(aba)$ et $m(aba) = -1$.

— abb et baa ne sont pas des facteurs bispéciaux car $abbb \notin L(t)$ et $baaa \notin L(t)$.

— bab est un facteur bispécial faible car[3] $E^-(bab) = \{a, b\}$, $E^+(bab) = \{a, b\}$, $E(bab) = \{(a, b),(b, a)\}$ donc $\#(E(bab)) = 2$, $d^-(bab) = 2 = d^+(bab)$ et $m(bab) = -1$.

— bba, $aaba$, $aabb$, $abaa$ et $abab$ ne sont pas des facteurs bispéciaux car $bbba \notin L(t)$, $aaaba \notin L(t)$, $aaabb \notin L(t)$, $abaaa \notin L(t)$ et $ababa$ n'est pas un facteur de t car c'est un chevauchement.

— $abba$ est un facteur bispécial fort car $E^-(abba) = \{a, b\}$, $E^+(abba) = \{a, b\}$, $E(abba) = \{(a, a),(a, b),(b, a),(b, b)\}$ donc $\#(E(abba)) = 4$, $d^-(abba) = 2$, $d^+(abba) = 2$ et $m(abba) = 1$.

— $baab$ est un facteur bispécial fort car $E^-(baab) = \{a, b\}$, $E^+(baab) = \{a, b\}$, $E(baab) = \{(a, a),(a, b),(b, a),(b, b)\}$ donc $\#(E(baab)) = 4$, $d^-(baab) = 2$, $d^+(baab) = 2$ et $m(baab) = 1$.

— $baba$, $babb$, $bbaa$ et $bbab$ ne sont pas des facteurs bispéciaux car $babab$ est un chevauchement, $babbb \notin L(t)$, $bbaaa \notin L(t)$ et $bbbab \notin L(t)$.

Ainsi, on obtient que les facteurs bispéciaux forts, faibles et neutres de longueur plus petites ou égales à 4 sont respectivement $\{\varepsilon, ab, ba, abba, baab\}$, $\{aba, bab\}$ et $\{a, b\}$. Or, vu que $abba = \sigma(ab)$ et $baab = \sigma(ba)$, le résultat voulu est démontré.

□

Corollaire 6.4. *La seconde différence et la complexité en facteurs du mot de* Thue-Morse *sont données par :*

$$(i) \quad b(n) = \begin{cases} 1 & si\ n = 0 \\ 2 & si\ \exists k \in \mathbb{N}\ \ tel\ que\ n = 2.2^k \\ -2 & si\ \exists k \in \mathbb{N}\ \ tel\ que\ n = 3.2^k \\ 0 & sinon \end{cases}$$

3. On justifie le fait que $(a, a) \notin E(bab)$ et $(b, b) \notin E(bab)$ de manière symétrique à l'étude faite pour le facteur aba.

$$(ii) \quad p(n) = \begin{cases} 1 & si\ n = 0 \\ 2 & si\ n = 1 \\ 4 & si\ n = 2 \\ 4n - 2.2^k - 4 & si\ \exists k \in \mathbb{N}\ \ tel\ que\ 2.2^k < n \leq 3.2^k \\ 2n + 4.2^k - 2 & si\ \exists k \in \mathbb{N}\ \ tel\ que\ 3.2^k < n \leq 4.2^k. \end{cases}$$

Démonstration. (i) Étant donné que le morphisme σ est 2-uniforme, par la Proposition 6.3, les facteurs bispéciaux forts ont une longueur égale à 0 ou 2.2^k, pour un $k \in \mathbb{N}$, les facteurs bispéciaux faibles ont une longueur égale à 3.2^k, pour un $k \in \mathbb{N}$. Ainsi, en conservant les mêmes notations que dans la Proposition 6.3, on a

$$\#(F \cap \mathcal{A}^n) = \begin{cases} 1 & \text{si } n = 0 \\ 2 & \text{si } \exists k \in \mathbb{N} \text{ tel que } n = 2.2^k \text{ et} \\ 0 & \text{sinon} \end{cases}$$
$$\#(f \cap \mathcal{A}^n) = \begin{cases} 2 & \text{si } \exists k \in \mathbb{N} \text{ tel que } n = 3.2^k \\ 0 & \text{sinon} \end{cases}.$$

Dès lors, par le Corollaire 2.22, on obtient

$$b(n) = \#(F \cap \mathcal{A}^n) - \#(f \cap \mathcal{A}^n) = \begin{cases} 1 & \text{si } n = 0 \\ 2 & \text{si } \exists k \in \mathbb{N} \text{ tel que } n = 2.2^k \\ -2 & \text{si } \exists k \in \mathbb{N} \text{ tel que } n = 3.2^k \\ 0 & \text{sinon} \end{cases}.$$

(ii) Le fait que $p(0) = 1, p(1) = 2$ et $p(2) = 4$ est trivialement vérifié. Pour les derniers cas, utilisons la Proposition 2.20 tout en sachant que $p(1) = 2$. On a

$$p(n) = 1 + n + \sum_{m=0}^{n-1} (n - 1 - m) b(m). \tag{6.1}$$

• S'il existe $k \in \mathbb{N}$ tel que $2.2^k < n \leq 3.2^k$, alors on a $2.2^k \leq n - 1 < 3.2^k$ et, par la point (i), les seuls termes non nuls présents dans la somme de la formule (6.1) sont : $b(0) = 1$, $b(2.2^l) = 2$ pour tout $l \in \{0, \ldots, k\}$ et $b(3.2^l) = -2$ pour tout $l \in \{0, \ldots, k-1\}$. On obtient

$$\begin{aligned} p(n) &= 1 + n + (n - 1 - 0).1 + \sum_{l=0}^{k} (n - 1 - 2.2^l).2 - \sum_{l=0}^{k-1} (n - 1 - 3.2^l).2 \\ &= 2n + 2\sum_{l=0}^{k} (n - 1) - 2\sum_{l=0}^{k-1} (n - 1) - 4\sum_{l=0}^{k} 2^l + 6\sum_{l=0}^{k-1} 2^l \\ &= 2n + 2(n - 1) - 4\left(\frac{1 - 2^{k+1}}{1 - 2}\right) + 6\left(\frac{1 - 2^k}{1 - 2}\right) = 4n - 2.2^k - 4. \end{aligned}$$

• S'il existe $k \in \mathbb{N}$ tel que $3.2^k < n \leq 4.2^k$, alors on a $3.2^k \leq n-1 < 4.2^k$ et, de la même façon, les seuls termes non nuls présents dans la somme de la formule (6.1) sont : $b(0) = 1$, $b(2.2^l) = 2$ et $b(3.2^l) = -2$ pour tout $l \in \{0, \ldots, k\}$. Ainsi, on obtient

$$\begin{aligned} p(n) &= 1 + n + (n-1) + \sum_{l=0}^{k}(n-1-2.2^l)2 - \sum_{l=0}^{k}(n-1-3.2^l)2 \\ &= 2n - 4\sum_{l=0}^{k} 2^l + 6\sum_{l=0}^{k} 2^l \\ &= 2n - 4\left(\frac{1-2^{k+1}}{1-2}\right) + 6\left(\frac{1-2^{k+1}}{1-2}\right) = 2n + 4.2^k - 2. \end{aligned}$$

□

Les premières valeurs de la fonction de complexité du mot de Thue-Morse sont

$$1, 2, 4, 6, 10, 12, 16, 20, 22, 24, 28, 32, 36, 40, 42, 44, 46, 48, 52, 56, \ldots$$

Cette suite est référencée par A005942 dans l'OEIS [4]. Le graphique illustrant les premières valeurs de la fonction de complexité en facteurs du mot de Thue-Morse est le suivant.

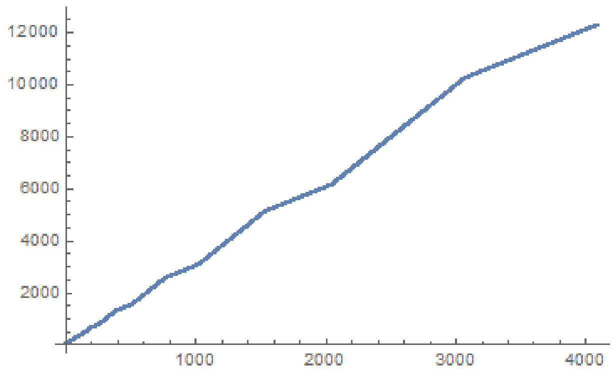


Figure 6.1 – Premières valeurs de $p_t(n)$.

6.2 Mot de Fibonacci

Le second exemple de complexité en facteurs de mot purement morphique que nous allons étudier en profondeur dans ce chapitre est celui du mot de Fibonacci. Comme introduit dans la section 4.1, le morphisme qui engendre le mot de Fibonacci est défini par $\sigma : \{a,b\}^* \to \{a,b\}^*$, $\sigma(a) = ab$, $\sigma(b) = a$ et on a

$$f := \sigma^\omega(a) = abaababaabaababaababaabaab \cdots .$$

4. The *On-line Encyclopedia of Integer Sequences*

Remarque 6.5. Le mot de FIBONACCI f ne contient pas le facteur bb.
En effet, le mot f est le point fixe du morphisme σ et est donc une concaténation d'éléments de $\{\sigma(a), \sigma(b)\} = \{ab, a\}$.

Vu que $\sigma(b) = a$, une première constatation que nous pouvons faire est qu'il n'existe pas de lettre bornée sous l'action du morphisme σ. Ainsi, par le Théorème 5.73, le cas $p(n) = \Theta(n^2)$ est à rejeter pour ce mot purement morphique. De plus, vu que $\sigma(b) = a$, les suites $|\sigma^k(a)|$ et $|\sigma^k(b)|$ ont le même comportement asymptotique. Ensuite, on peut remarquer que la longueur de la $n^{\text{ième}}$ itération du morphisme de FIBONACCI est égale au $(n+2)^{\text{ième}}$ nombre de FIBONACCI [5]. En effet, on a $|\,\sigma^0(a)| = |a| = 1$, $|\sigma(a)| = |ab| = 2$ et pour tout $n \in \mathbb{N}$, on a

$$\begin{aligned} |\,\sigma^{n+2}(a)| &= |\,\sigma^{n+1}(ab)| \\ &= |\,\sigma^{n+1}(a)| + |\,\underbrace{\sigma^{n+1}(b)}_{=\sigma^n(\sigma(b))}\,| \\ &= |\,\sigma^{n+1}(a)| + |\,\sigma^n(a)|. \end{aligned}$$

Ainsi, par la formule de BINET [6] on obtient que $|\sigma^k(a)|$ et $|\sigma^k(b)|$ sont en $\Theta\left(\phi^k\right)$, où $\phi = \frac{1+\sqrt{5}}{2}$ est le nombre d'or. On en conclut que le morphisme σ est quasi-uniforme et, par le Théorème 5.58, on a $p_f(n) = \mathcal{O}(n)$.

Nous allons maintenant étudier plus en profondeur la complexité en facteurs de ce mot et démontrer qu'il fait partie de la classe des mots *Sturmiens*. Remarquons que le caractère *Sturmien* du mot de FIBONACCI sera démontré grâce à des développements utilisant les propriétés de morphismes. Une autre possibilité serait d'utiliser la définition équivalente de mots *Sturmiens* en tant que mots de rotation d'angles irrationnels [10], dont f est un cas particulier pour lequel l'angle est égal à $2 - \phi$.

Définition 6.6. Le *miroir* d'un mot fini $u = u_0 \cdots u_{|u|-1} \in \mathcal{A}^*$ est le mot $u_{|u|-1} \cdots u_0$. On le note u^R.

Remarque 6.7. Soient $u, v \in \mathcal{A}^*$. Par définition, on a $(uv)^R = v^R u^R$.
En effet, on a $(uv)^R = (u_0 \cdots u_{|u|-1} v_0 \cdots v_{|v|-1})^R = v_{|v|-1} \cdots v_0 u_{|u|-1} \cdots u_0 = v^R u^R$.

Lemme 6.8. *Le langage $L(f)$ du mot de* FIBONACCI *est fermé pour l'opération miroir.*

Démonstration. Définissons le morphisme $\sigma_R : \{a,b\}^* \to \{a,b\}^*$ tel que $\sigma_R(a) = ba$ et $\sigma_R(b) = a$. Tout d'abord, remarquons que, pour tout mot $w \in \{a,b\}^*$, le miroir de son image par σ est égal à l'image de son miroir par σ_R. Formellement,

$$(\sigma(w))^R = \sigma_R(w^R). \tag{6.2}$$

5. Cette particularité explique le nom de cette suite.

6. La formule de BINET permet de calculer les termes de la suite de Fibonacci : si l'on note F_n le $n^{\text{ième}}$ nombre de FIBONACCI et ϕ (resp. ψ) le racine positive (resp. négative) de l'équation $x^2 = x + 1$, on a $F_n = \frac{1}{\sqrt{5}}(\phi^n - \psi^n)$.

En effet, procédons par induction sur la longueur du mot w. La propriété est vérifiée pour l'unique mot de longueur 0 étant le mot vide et pour les mots de longueur 1 par définition. Ensuite, par induction supposons avoir démontré la propriété pour les mots de longueur $n \in \mathbb{N}_0$ et démontrons-le pour $n+1$. En appliquant la Remarque 6.7 et en notant $w' = w_1 \cdots w_{|w|-1}$, par induction on a

$$(\sigma(w))^R = (\sigma(w_0)\sigma(w'))^R = (\sigma(w'))^R(\sigma(w_0))^R = \sigma_R(w'^R)\,\sigma_R(w_0^R) = \sigma_R(w'^R w_0^R) = \sigma_R(w^R).$$

Ensuite, démontrons que pour tout $w \in \{a,b\}^*$, on a

$$a\,\sigma_R(w) = \sigma(w)a. \tag{6.3}$$

Si w est le mot vide, a ou b, alors l'égalité est vérifiée. Dès lors, démontrons l'égalité pour des mots plus longs par induction. Supposons avoir démontré l'égalité pour des mots de longueur $n \in \mathbb{N}_0$ et démontrons-le pour les mots de longueur $n+1$. En conservant la même notation pour le mot w', on a

$$\begin{aligned} a\,\sigma_R(w) &= a\,\sigma_R(w_0 w') \\ &= (a\,\sigma_R(w_0))\,\sigma_R(w') \\ &= (\sigma(w_0)a)\,\sigma_R(w') \\ &= \sigma(w_0)(a\,\sigma_R(w')) \\ &= \sigma(w_0)\sigma(w')a = \sigma(w)a. \end{aligned}$$

Nous allons maintenant prouver, que pour tout $i \in \mathbb{N}$, le miroir de $f_i := \sigma^i(a)$ est un facteur de f, c'est-à-dire

$$(f_i)^R \in L(f).$$

Procédons par induction sur l'entier naturel i. Le cas $i = 0$ est trivialement vérifié donc supposons que $(f_i)^R \in L(f)$ et démontrons que $(f_{i+1})^R \in L(f)$. On a

$$f_{i+1} = \sigma^{i+1}(a) = \sigma(\sigma^i(a)) = \sigma(f_i),$$

et donc

$$(f_{i+1})^R = (\sigma(f_i))^R \underbrace{=}_{(6.2)} \sigma_R((f_i)^R). \tag{6.4}$$

On sait que le mot $(f_i)^R$ est un facteur de f par induction, donc ce facteur peut être étendu à droite en une lettre $l \in \{a,b\}$. Ainsi, $(f_i)^R l$ est également un facteur de f et son image par σ, $\sigma((f_i)^R l)$ également vu que f est un point fixe de σ. Or, on a $\sigma((f_i)^R l) = \sigma((f_i)^R)\sigma(l)$ et, pour tout $l \in \{a,b\}$, a est un préfixe de $\sigma(l)$. Ainsi, le mot

$$\sigma((f_i)^R)a \underbrace{=}_{(6.3)} a\,\sigma_R((f_i)^R) \underbrace{=}_{(6.4)} a(f_{i+1})^R$$

est un facteur de f, donc $(f_{i+1})^R$ est un facteur de f.

Or, tout facteur de f est un facteur d'un f_i pour un i assez grand. Ainsi, on conclut que $L(f)$ est fermé pour l'opération miroir.

$\square$

Lemme 6.9. *Soit $w \in \{a,b\}^*$. Si $\sigma(w)a$ est un facteur du mot de* FIBONACCI *f, alors w est également un facteur de f.*

Démonstration. Le cas où w est le mot vide est vérifié car $\varepsilon \in L(f)$. Dès lors, supposons $w \neq \varepsilon$. Vu que $f = \sigma^\omega(a)$, on a $f = \sigma(f)$. Pour que $\sigma(w)a$ soit un facteur de f avec $w \neq \varepsilon$, il faut qu'il existe $u \in L(f)$ tel que $\sigma(w)a$ soit un facteur de $\sigma(u)$. Considérons un tel u de longueur minimale. Le mot u est de longueur au moins 2. En effet, on a $w \neq \varepsilon$ donc $|\sigma(w)a| \geq 2$. Si $|\sigma(w)a| = 2$, on a obligatoirement $w = b$ et $\sigma(w)a = aa$. Les mots $u = a$ et $u = b$ ne conviennent dès lors pas. Si $|\sigma(w)a| \geq 3$, alors $|\sigma(u)| \geq 3$ et $|u| \geq 2$. De ce fait, notons $u = x_1u'x_2$ avec $x_1, x_2 \in \mathcal{A}$ tel que $\sigma(w)a = t_1\sigma(u')t_2$ où $t_1 \in \{a, b, ab\}$ est un suffixe non vide de $\sigma(x_1)$ et $t_2 \in \{a, ab\}$ est un préfixe non vide de $\sigma(x_2)$. Le cas où $t_1 = b$ est impossible car t_1 correspond à un préfixe de $\sigma(w)$ et toute image d'un mot par le morphisme σ commence par a. Ainsi, on a $t_1 = a$ ou bien $t_1 = ab$ et dans ces deux cas pour que t_1 soit un suffixe de $\sigma(x_1)$ il faut que $t_1 = \sigma(x_1)$. Le cas $t_2 = ab$ est également à rejeter car le mot $\sigma(w)a$ se termine par la lettre a. On obtient donc $\sigma(w)a = \sigma(x_1u')a$. Or, remarquons que l'ensemble $\{\sigma(a), \sigma(b)\} = \{ab, a\}$ est un code. En effet, l'unique découpage d'un mot $x \in \{ab, a\}^*$ est fourni en coupant avant chaque occurrence de la lettre a. Ainsi, par la Proposition 4.18, étant donné que $\sigma(a) \neq \sigma(b)$ et que l'ensemble $\{\sigma(a), \sigma(b)\}$ est un code, le morphisme σ est injectif. On en conclut que $w = x_1u' \in L(f)$.

□

Remarque 6.10. Dans l'énoncé du lemme précédent, la lettre a suivant l'occurrence de $\sigma(w)$ est indispensable.
En effet, un contre-exemple du lemme omettant la lettre a est donnée par le mot bb. Le facteur $\sigma(bb) = aa$ est effectivement un facteur du mot de FIBONACCI sans pour autant que bb en soit un.

Proposition 6.11. *Les facteurs spéciaux à gauche et à droite du mot de* FIBONACCI *sont respectivement ses préfixes et le miroir de ses préfixes.*

Démonstration. • Étudions les facteurs spéciaux à gauche de f. Une première constatation est que si w est spécial à gauche, alors $\sigma(w)$ l'est également. En effet, si aw et bw sont des facteurs de f, vu que $f = \sigma(f)$, on a $\sigma(aw), \sigma(bw) \in L(f)$. Or, on a $\sigma(aw) = \sigma(a)\sigma(w) = ab\sigma(w)$ et $\sigma(bw) = \sigma(b)\sigma(w) = a\sigma(w)$. Donc, on obtient que les mots $a\sigma(w)$ et $b\sigma(w)$ sont des facteurs de f, c'est-à-dire le facteur $\sigma(w)$ est spécial à gauche. En particulier, vu que le lettre a est spéciale à gauche, en itérant ce résultat, on obtient que, pour tout $i \in \mathbb{N}$, les facteurs $\sigma^i(a)$ sont spéciaux à gauche. Or, on a $f = \sigma^\omega(a)$ donc, tout préfixe de f est un préfixe d'un $\sigma^i(a)$ pour un $i \in \mathbb{N}$ assez grand. Ainsi, vu que tout préfixe d'un facteur spécial à gauche est lui-même spécial à gauche, on obtient que les préfixes de f sont spéciaux à gauche.
Démontrons maintenant que seuls les préfixes de f sont spéciaux à gauche. Supposons avoir

un facteur w spécial à gauche. En particulier, la lettre b est une extension à gauche de w donc, par la Remarque 6.5, le mot w commence par un a. Également par la Remarque 6.5, si la dernière lettre de w est un b, alors les occurrences de w dans f sont toujours suivies par un a et donc le facteur wa est également spécial à gauche. Ainsi, en notant $w' = w$ si w se termine par a, et $w' = wa$ sinon, le facteur spécial à gauche w' se termine par a et ne contient pas de bb donc il peut être factorisé sur l'ensemble $\{ab, a\}$. Par construction, le dernier élément de cette factorisation est un a. Ainsi, étant donné que cette factorisation correspond à une concaténation d'images de lettres par le morphisme σ, il existe $v \in \{a, b\}^*$ tel que $w' = \sigma(v)a$. Le Lemme 6.9 implique que le mot v est un facteur de f. De plus, le facteur w' est spécial à gauche, par définition, donc aw' et bw' sont des facteurs de f et ce dernier facteur[7] peut uniquement être précédé d'un a donc $abw' \in L(f)$. Ensuite, on a $aw' = a\sigma(v)a = \sigma(b)\sigma(v)a = \sigma(bv)a$ et $abw' = ab\sigma(v)a = \sigma(a)\sigma(v)a = \sigma(av)a$ donc en appliquant de nouveau le Lemme 6.9, on obtient que $av, bv \in L(f)$ et donc que le facteur v est spécial à gauche et ne peut débuter que par un a. Or, si $v \neq \varepsilon$, étant donné que la première lettre de v est un a, on a $|v| < |\sigma(v)|$ et, par définition, on a $|\sigma(v)| = |w'| - 1 \leq |w|$. En conclusion, les facteurs spéciaux à gauche sont nécessairement des préfixes de f. En effet, démontrons-le par induction sur la longueur des facteurs spéciaux à gauche. Avec les mêmes notations pour les mots v et w, on a $|v| < |w|$. Le plus court facteur spécial à gauche est le mot vide, et si $v = \varepsilon$, v est un préfixe de f et $w = a$ en est un également. Par induction, si v est un préfixe de f, alors il existe une lettre $l \in \{a, b\}$ telle que vl est un préfixe de f. Ainsi, étant donné que w est un préfixe de $\sigma(v)a$ qui est un préfixe de $\sigma(vl)$, et que $\sigma(vl)$ est un préfixe de f, car $f = \sigma(f)$, alors w est un préfixe de f.
• L'étude des facteurs spéciaux à droite découle d'une simple application du Lemme 6.8 au résultat que nous venons de trouver pour les facteurs spéciaux à gauche.

□

Corollaire 6.12. *Le mot de* FIBONACCI *est* Sturmien.

Démonstration. Étant donné qu'à une longueur fixée, il n'existe qu'un seul préfixe de f correspondant, la Proposition 6.11 stipule qu'il n'y a qu'un seul facteur spécial à droite de chaque longueur. Ainsi, par le Corollaire 2.22, on a $s(n) = 1$, pour tout $n \in \mathbb{N}$, et donc, par la Proposition 2.20, on a $p(n) = 1 + \sum_{l=0}^{n-1} 1 = n + 1$, pour tout $n \in \mathbb{N}$.

□

Nous venons donc de démontrer que le mot de FIBONACCI fait partie de la sous-classe des mots *Sturmiens* étant en particulier purement morphiques.

7. Le facteur bw' ne peut pas être un préfixe unioccurrent de f car il commence par b. Remarquons plus particulièrement que le mot de FIBONACCI est récurrent. La démonstration est identique à celle du Lemme 6.1 car $\sigma^{n+2}(a) = \sigma^n(aba) = \sigma^n(a)\,\sigma^n(b)\,\sigma^n(a)$.

Annexe A

Quelques compléments d'analyse asymptotique : démonstrations

Dans cette annexe, nous allons démontrer les résultats d'analyse asymptotique énoncés dans la section 5.1. Pour ce faire, des lemmes intermédiaires seront établis tout au long de cette section.

Proposition A.1. *Soient deux couples de réels* (β_1, α_1)*,* $(\beta_2, \alpha_2) \in (\mathbb{R}_0 \times \mathbb{R}) \cup \{(0,0)\}$*, on a*

$$k^{\alpha_1}\beta_1^k = o(k^{\alpha_2}\beta_2^k)$$

lorsque k *tend vers l'infini si et seulement si* $(\beta_1, \alpha_1) <_l (\beta_2, \alpha_2)$*.*

Démonstration. Si $(\beta_1, \alpha_1) <_l (\beta_2, \alpha_2)$, montrons que l'on a

$$\lim_{k\to+\infty} \frac{k^{\alpha_1}\beta_1^k}{k^{\alpha_2}\beta_2^k} = 0.$$

Si $\beta_1 < \beta_2$, on a

$$\lim_{k\to+\infty} k^{\alpha_1-\alpha_2}\left(\frac{\beta_1}{\beta_2}\right)^k = 0.$$

De plus, si $\beta_1 = \beta_2$ et $\alpha_1 < \alpha_2$, on a

$$\lim_{k\to+\infty} \frac{1}{k^{\alpha_2-\alpha_1}} = 0.$$

Ensuite, si $(\beta_1, \alpha_1) \geq_l (\beta_2, \alpha_2)$, le résultat n'est pas vérifié. En effet, si $(\beta_1, \alpha_1) = (\beta_2, \alpha_2)$, on a

$$\lim_{k\to+\infty} \frac{k^{\alpha_1}\beta_1^k}{k^{\alpha_2}\beta_2^k} = 1$$

et si $(\beta_1, \alpha_1) >_l (\beta_2, \alpha_2)$, alors par la première partie de la preuve, on a

$$\lim_{k\to+\infty} \frac{k^{\alpha_1}\beta_1^k}{k^{\alpha_2}\beta_2^k} = \infty.$$

□

Le théorème suivant et son corollaire, classiques dans le domaine de l'analyse fonctionnelle, ne seront pas démontrés dans le cadre de ce mémoire. Les preuves sont consultables dans le livre [9] (corollaire 3.14).

Dans cette annexe, nous considérerons un entier $d \in \mathbb{N}_0$.

Théorème A.2. (Équivalence des normes)
Soit V un espace vectoriel réel ou complexe. Si V est de dimension finie, alors toutes les normes sur V sont équivalentes : soient deux normes $\|\cdot\|_A$ et $\|\cdot\|_B$ définies sur V, il existe des nombres réels $0 < \lambda \leq \mu$ tels que, pour tout $x \in V$,

$$\lambda \|x\|_A \leq \|x\|_B \leq \mu \|x\|_A .$$

Corollaire A.3. *Pour toutes normes $\|\cdot\|_A$ et $\|\cdot\|_B$ définies sur $\mathbb{C}_d^d$ et toute matrice $\boldsymbol{M} \in \mathbb{C}_d^d$, on a $\|\boldsymbol{M}^k\|_A = \Theta(\|\boldsymbol{M}^k\|_B)$ lorsque k tend vers l'infini.*

Définition A.4. Soit $\mathbf{M}$ une matrice de $\mathbb{C}_d^d$, le *rayon spectral* de $\mathbf{M}$ est le réel

$$\rho(\mathbf{M}) = \max_{1\leq i\leq d} |\lambda_i|,$$

où $\lambda_1, \ldots, \lambda_d$ sont les valeurs propres de $\mathbf{M}$ comptées avec leur multiplicité.

Un théorème fondamental de Gelfand est le suivant :

Théorème A.5. *(Gelfand , 1941)*
Soient $\|\cdot\|$ une norme définie sur $\mathbb{C}_d^d$ et $\boldsymbol{M} \in \mathbb{C}_d^d$. Le réel $\sqrt[k]{\|\boldsymbol{M}^k\|}$ converge vers le rayon spectral $\rho(\boldsymbol{M})$ de $\boldsymbol{M}$ lorsque k tend vers l'infini. Formellement,

$$\rho(\boldsymbol{M}) = \lim_{k\to+\infty} \sqrt[k]{\|\boldsymbol{M}^k\|}$$

Rappelons que le premier outil primordial de la section 5.1 était le théorème suivant, qui est une variante au Théorème A.5.

Théorème A.6. *Soit $\|\cdot\|$ une norme définie sur $\mathbb{C}_d^d$. Pour toute matrice $\boldsymbol{M} \in \mathbb{C}_d^d$, il existe $(\beta, \alpha) \in (\mathbb{R}_0 \times \mathbb{R}) \cup \{(0,0)\}$ tel que $\left\|\boldsymbol{M}^k\right\| = \Theta(k^\alpha \beta^k)$ lorsque k tend vers l'infini.*

Démonstration. Soit $\mathbf{Q} \in \mathbb{C}_d^d$ une matrice inversible telle que $\mathbf{QMQ}^{-1}$ est sous la forme normale de Jordan. Il existe une matrice diagonale $\mathbf{D} \in \mathbb{C}_d^d$ et une matrice nilpotente $\mathbf{N} \in \mathbb{C}_d^d$ telles que $\mathbf{QMQ}^{-1}$=$\mathbf{D}$+$\mathbf{N}$ et $\mathbf{DN}$=$\mathbf{ND}$.
Considérons une application sur $\mathbb{C}_d^d$ définie par

$$\|\cdot\|_{\mathbf{Q}} : \mathbb{C}_d^d \to \mathbb{R}\ ,\ \ \mathbf{X} \mapsto \max_{i,j \in \{1,\ldots,d\}} |(\mathbf{QXQ}^{-1})_{ij}|.$$

Vérifions que $\|\cdot\|_{\mathbf{Q}}$ définit une norme sur $\mathbb{C}_d^d$. Pour tout $\mathbf{X}$, $\mathbf{Y} \in \mathbb{C}_d^d$ et $\lambda \in \mathbb{C}$, on a

(i) $\|\mathbf{X}\|_{\mathbf{Q}} = 0 \Leftrightarrow \max_{i,j \in \{1,\ldots,d\}} |(\mathbf{QXQ}^{-1})_{ij}| = 0 \Leftrightarrow |(\mathbf{QXQ}^{-1})_{ij}| = 0\ \forall i,j \in \{1,...,d\}$
$\Leftrightarrow \mathbf{QXQ}^{-1} = 0 \Leftrightarrow \mathbf{X} = 0$ car la matrice $\mathbf{Q}$ est inversible.

(ii) $\|\lambda\mathbf{X}\|_{\mathbf{Q}} = \max_{i,j \in \{1,\ldots,d\}} |(\mathbf{Q}(\lambda\mathbf{X})\mathbf{Q}^{-1})_{ij}| = \max_{i,j \in \{1,\ldots,d\}} |\lambda(\mathbf{QXQ}^{-1})_{ij}|$
$= \max_{i,j \in \{1,\ldots,d\}} |\lambda||(\mathbf{QXQ}^{-1})_{ij}| = |\lambda|\,\|\mathbf{X}\|_{\mathbf{Q}}$.

(iii) $\|\mathbf{X}+\mathbf{Y}\|_{\mathbf{Q}} = \max_{i,j \in \{1,\ldots,d\}} |(\mathbf{Q}(\mathbf{X}+\mathbf{Y})\mathbf{Q}^{-1})_{ij}| = \max_{i,j \in \{1,\ldots,d\}} |(\mathbf{QXQ}^{-1})_{ij} + (\mathbf{QYQ}^{-1})_{ij}|$
$\leq \max_{i,j \in \{1,\ldots,d\}} \left(|(\mathbf{QXQ}^{-1})_{ij}| + |(\mathbf{QYQ}^{-1})_{ij}|\right)$
$\leq \max_{i,j \in \{1,\ldots,d\}} |(\mathbf{QXQ}^{-1})_{ij}| + \max_{i,j \in \{1,\ldots,d\}} |(\mathbf{QYQ}^{-1})_{ij}| = \|\mathbf{X}\|_{\mathbf{Q}} + \|\mathbf{Y}\|_{\mathbf{Q}}$

Pour tous $i, j \in \{1,...,d\}$, définissons la fonction

$$m_{i,j} : \mathbb{N} \to \mathbb{C}\ ,\ \ k \mapsto (\mathbf{QM}^k\mathbf{Q}^{-1})_{ij}.$$

Dès lors, pour tout $k \in \mathbb{N}$, on a $\left\|\mathbf{M}^k\right\|_{\mathbf{Q}} = \max_{i,j \in \{1,\ldots,d\}} |m_{i,j}(k)|$. Par le Corollaire A.3, il nous suffit de montrer le résultat énoncé pour cette norme et il sera vérifié pour toute norme définie sur $\mathbb{C}_d^d$. Pour tout $k \in \mathbb{N}$, on a, vu que les matrices $\mathbf{N}$ et $\mathbf{D}$ commutent,

$$\begin{aligned}\mathbf{QM}^k\mathbf{Q}^{-1} &= \mathbf{QMQ}^{-1}\mathbf{QMQ}^{-1} \cdots \mathbf{QMQ}^{-1}\\ &= (\mathbf{D}+\mathbf{N})^k\\ &= \sum_{h=0}^{k} \binom{k}{h} \mathbf{D}^{k-h}\mathbf{N}^h.\end{aligned}$$

Cependant, par définition, la matrice $\mathbf{N}$ est nilpotente d'indice de nilpotence $n_{\mathbf{N}} \leq d$. Ainsi, la matrice $\mathbf{N}^h$ est nulle pour tout $h \geq d$. Dès lors, pour tout $k \geq d-1$, on obtient

$$\mathbf{QM}^k\mathbf{Q}^{-1} = \sum_{h=0}^{d-1} \binom{k}{h} \mathbf{D}^{k-h}\mathbf{N}^h.$$

Pour tout $k \geq d$, on en tire

$$m_{i,j}(k) = (\mathbf{Q}\mathbf{M}^k\mathbf{Q}^{-1})_{ij} = \left(\sum_{h=0}^{d-1} \binom{k}{h} \mathbf{D}^{k-h}\mathbf{N}^h\right)_{ij} = \sum_{h=0}^{d-1} \binom{k}{h} \left(\mathbf{D}^{k-h}\mathbf{N}^h\right)_{ij}.$$

Or, étant donné que la matrice $\mathbf{D}$ est diagonale, il en résulte que

$$\left(\mathbf{D}^{k-h}\mathbf{N}^h\right)_{ij} = \sum_{l=0}^{d} \mathbf{D}_{il}^{k-h}\mathbf{N}_{lj}^h = \mathbf{D}_{ii}^{k-h}\mathbf{N}_{ij}^h.$$

Ainsi, si l'on note $\lambda_i = \mathbf{D}_{ii}$ une des valeurs propres de $\mathbf{D}$, on obtient

$$m_{i,j}(k) = \sum_{h=0}^{d-1} \binom{k}{h} \lambda_i^{k-h}\mathbf{N}_{ij}^h = \lambda_i^k \sum_{h=0}^{d-1} \binom{k}{h} \lambda_i^{-h}\mathbf{N}_{ij}^h = \lambda_i^k f_{i,j}(k),$$

où $f_{i,j}$ est un polynôme à coefficients complexes.
Il s'ensuit que pour tous $i, j \in \{1, ..., d\}$, il existe $(\beta_i, \alpha_{i,j}) \in (\mathbb{R}_0 \times \mathbb{R}) \cup \{(0,0)\}$ tel que $|m_{i,j}(k)| = \Theta(k^{\alpha_{i,j}}\beta_i^k)$:

— si $f_{i,j}$ n'est pas le polynôme constant 0, alors $\beta_i = |\lambda_i|$ et $\alpha_{i,j}$ est le degré du polynôme $f_{i,j}$,

— si $f_{i,j}$ est identiquement nul, alors $(\beta_i, \alpha_{i,j}) = (0,0)$.

Dès lors, posons $(\beta, \alpha) = \max\limits_{i,j\in\{1,...,d\}} (\beta_i, \alpha_{i,j})$ en suivant l'ordre lexicographique. Par la Proposition A.1, on obtient que $\left\|\mathbf{M}^k\right\|_{\mathbf{Q}} = \Theta(k^\alpha \beta^k)$.

□

Remarque A.7. Comme $\lim\limits_{k\to+\infty} \sqrt[k]{k} = 1$, par le Théorème A.5 de GELFAND et le Théorème 5.7, on a $\beta = \rho(\mathbf{M})$.

Ensuite, en suivant les notations de la Définition 5.8, nous allons établir deux résultats intermédiaires nécessaires à la démonstration du Théorème 5.9.

Lemme A.8. *Soient deux bijections croissantes $f_1, f_2 : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$. On a*

$$\#\ (E_y(f_1, f_2)) = \max\{f_1^{-1}(y) - f_2^{-1}(y), 0\} + \mathcal{O}(1)$$

lorsque y tend vers l'infini.

Démonstration. Puisque les fonctions f_1 et f_2 sont des bijections croissantes, pour tout réel $y \geq 0$, on a[1]

1. Dans la suite de la démonstration, si $a > b$, la notation $[\![a, b]\!]$ désignera l'intervalle entier vide.

$$\begin{aligned} E_y(f_1,f_2) =&\{k\in\mathbb{N} \mid f_1(k)\leq y\leq f_2(k)\} \\ =&\{k\in\mathbb{N} \mid k\leq f_1^{-1}(y) \text{ et } f_2^{-1}(y)\leq k\} \\ =&[\![f_2^{-1}(y), f_1^{-1}(y)]\!]. \end{aligned}$$

De plus, pour tous réels x_1, x_2, on a

$$\max\{x_2-x_1,0\}\leq \#\ ([\![x_1,x_2]\!])\leq \max\{x_2-x_1,0\}+1.$$

Ainsi, on obtient les inégalités suivantes :

$$\max\{f_1^{-1}(y)-f_2^{-1}(y),0\}\leq \#\ (E_y(f_1,f_2))\leq \max\{f_1^{-1}(y)-f_2^{-1}(y),0\}+1.$$

Ceci nous permet de conclure.

□

Dans les hypothèses du lemme précédent, si les fonctions f_1 et f_2 sont telles que $f_1(x)\leq f_2(x)$ pour les réels x positifs assez grands, alors on obtient

$$\#\ (E_y(f_1,f_2)) = f_1^{-1}(y)-f_2^{-1}(y)+\mathcal{O}(1)$$

lorsque y tend vers l'infini. Ainsi, asymptotiquement, estimer le cardinal de l'ensemble $E_y(f_1,f_2)$ se réduit simplement à l'étude des fonctions f_1^{-1} et f_2^{-1}.

Lemme A.9. *Soient $\beta,\alpha\in\mathbb{R}$ avec $\beta>1$ et une bijection croissante $f:\mathbb{R}_{\geq 0}\to\mathbb{R}_{\geq 0}$. Si $f(x)=\Theta(x^\alpha\beta^x)$ lorsque x tend vers l'infini, alors*

$$f^{-1}(y)=\frac{1}{\log\beta}\log y\ -\ \frac{\alpha}{\log\beta}\log\log y\ +\mathcal{O}(1)$$

lorsque y tend vers l'infini.

Démonstration. Tout d'abord, remarquons que pour toutes fonctions f_1, f_2, strictement positives à l'infini, on a

$$f_1(x)=\Theta(f_2(x)) \text{ si et seulement si } \log(f_1(x))=\log(f_2(x))+\mathcal{O}(1), \tag{A.1}$$

lorsque x tend vers l'infini.
En effet, la condition est nécessaire :

$$\begin{aligned} &f_1(x)=\Theta(f_2(x)) \\ \Leftrightarrow &\exists k_1,k_2>0, \exists x_0>0,\ \forall x\geq x_0,\ k_1.|f_2(x)|\leq|f_1(x)|\leq k_2.|f_2(x)|. \end{aligned}$$

Ainsi, vu que les fonctions f_1 et f_2 sont positives à l'infini, pour tout x suffisamment grand, on a

$$k_1.|f_2(x)|\leq|f_1(x)|\leq k_2.|f_2(x)|$$

$$
\begin{aligned}
&\Leftrightarrow k_1.f_2(x) \leq f_1(x) \leq k_2.f_2(x)\\
&\Leftrightarrow \log(k_1.f_2(x)) \leq \log(f_1(x)) \leq \log(k_2.f_2(x))\\
&\Leftrightarrow \log(k_1) + \log(f_2(x)) \leq \log(f_1(x)) \leq \log(k_2) + \log(f_2(x)).
\end{aligned}
$$

Or, le fait que $f_1(x), f_2(x), k_1$ et k_2 soient strictement positifs n'implique pas que $\log(f_1(x))$, $\log(f_2(x)), \log(k_1)$ et $\log(k_2)$ le soient également. Cependant,

$$
\begin{aligned}
&\log(k_1) + \log(f_2(x)) \leq \log(f_1(x)) \leq \log(k_2) + \log(f_2(x))\\
\Leftrightarrow &\log(k_1) \leq \log(f_1(x)) - \log(f_2(x)) \leq \log(k_2)\\
\Rightarrow &|\log(f_1(x)) - \log(f_2(x))| \leq \max\{|\log(k_1)|, |\log(k_2)|\}.
\end{aligned}
$$

On en tire $\log(f_1(x)) - \log(f_2(x)) = \mathcal{O}(1)$. De plus, la condition est suffisante. En effet, on a

$$
\begin{aligned}
&\log(f_1(x)) = \log(f_2(x)) + \mathcal{O}(1)\\
\Leftrightarrow &\log(f_1(x)) - \log(f_2(x)) = \mathcal{O}(1)\\
\Leftrightarrow &\exists k > 0, \exists x_0 > 0, \forall x \geq x_0, |\log(f_1(x)) - \log(f_2(x)|) \leq k\\
\Leftrightarrow &\exists k > 0, \exists x_0 > 0, \forall x \geq x_0, -k \leq \log(f_1(x)) - \log(f_2(x)) \leq k.
\end{aligned}
$$

Dès lors, en définissant $k_1 > 0, k_2 > 0$ par $\log(k_1) = -k$ et $\log(k_2) = k$, on en tire que

$$\exists k_1, k_2 > 0, \exists x_0 > 0, \forall x \geq x_0, \log(k_1) \leq \log(f_1(x)) - \log(f_2(x)) \leq \log k_2.$$

Ainsi, en utilisant les premières équivalences obtenues dans la preuve de la condition nécessaire, on obtient $f_1(x) = \Theta(f_2(x))$.
On sait que $f(x) = \Theta(x^\alpha \beta^x)$. Dès lors, en utilisant le critère (A.1) que nous venons d'établir, il en résulte que

$$\log f(x) = \log(x^\alpha\beta^x) + \mathcal{O}(1) = \log(x^\alpha) + \log(\beta^x) + \mathcal{O}(1) = \alpha\log(x) + x\log\beta + \mathcal{O}(1) = \Theta(x).$$

En appliquant de nouveau ce critère, on a

$$\log\log f(x) = \log(x) + \mathcal{O}(1) \text{ et donc } \alpha\log\log f(x) = \alpha\log(x) + \mathcal{O}(1).$$

En combinant les deux résultats que nous venons de trouver, on obtient

$$
\begin{array}{rrl}
 & \alpha\log\log f(x) - \alpha\log(x) + \mathcal{O}(1) & = 0\\
- & \log f(x) - x\log\beta - \alpha\log(x) + \mathcal{O}(1) & = 0\\
\hline
 & \alpha\log\log f(x) - \log f(x) + x\log\beta + \mathcal{O}(1) & = 0.
\end{array}
$$

Maintenant, remarquons que par hypothèse sur f, on a $\lim_{y\to\infty} f^{-1}(y) = \infty$. Ainsi, en substituant x par $f^{-1}(y)$ dans le résultat obtenu ci-avant, on a

$$
\begin{aligned}
&\alpha\log\log y - \log y + f^{-1}(y)\log\beta = \mathcal{O}(1)\\
\Leftrightarrow &f^{-1}(y) = \frac{1}{\log\beta}\log y \; - \; \frac{\alpha}{\log\beta}\log\log y \; + \mathcal{O}(1)
\end{aligned}
$$

lorsque y tend vers l'infini.

□

Avec ces résultats établis, nous sommes, dès lors, en mesure de démontrer le Théorème 5.9 ainsi que son Corollaire 5.10 rappelés ci-après.

Théorème A.10. *Soient* $(\beta_1, \alpha_1), (\beta_2, \alpha_2) \in \mathbb{R}_{>1} \times \mathbb{R}$ *et* $f_1, f_2 : \mathbb{N} \to \mathbb{R}_{\geq 0}$ *des fonctions telles que* $f_1(k) = \Theta(k^{\alpha_1}\beta_1^k)$ *et* $f_2(k) = \Theta(k^{\alpha_2}\beta_2^k)$ *lorsque* k *tend vers l'infini. Considérons la fonction* $\Delta : \mathbb{R}_{>1} \to \mathbb{R}$ *définie par*

$$\Delta(y) = -\left(\frac{1}{\log\beta_2} - \frac{1}{\log\beta_1}\right)\log y + \left(\frac{\alpha_2}{\log\beta_2} - \frac{\alpha_1}{\log\beta_1}\right)\log\log y$$

pour tout $y > 1$*. Si* (β_1, α_1) *est lexicographiquement plus petit ou égal à* (β_2, α_2)*, alors* $\#\,(E_y(f_1, f_2)) = \Delta(y) + \mathcal{O}(1)$ *lorsque* y *tend vers l'infini.*

Démonstration. Pour tout $i \in \{1, 2\}$, posons $g_i : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ une bijection croissante telle que $g_i(x) = x^{\alpha_i}\beta_i^x$ pour tout x suffisamment grand. Étant donné que $f_i(k) = \Theta(k^{\alpha_i}\beta_i^k)$, il existe $\lambda_i, \mu_i > 0$ et $k_i \in \mathbb{N}$ tels que, pour tout $k \geq k_i$,

$$\lambda_i g_i(k) \leq f_i(k) \leq \mu_i g_i(k).$$

Posons $k_0 = \max\{k_1, k_2\}$. Ainsi, pour tout $k \geq k_0$, on a

$$\lambda_1 g_1(k) \leq f_1(k) \leq \mu_1 g_1(k) \text{ et } \lambda_2 g_2(k) \leq f_2(k) \leq \mu_2 g_2(k). \tag{A.2}$$

Soient un réel y et un entier k tels que $k \in E_y(\mu_1 g_1, \lambda_2 g_2) \setminus [\![0, k_0 - 1]\!]$. Par définition, on a

$$\mu_1 g_1(k) \leq y \leq \lambda_2 g_2(k) \text{ et } k \geq k_0.$$

Ainsi, par les inégalités (A.2), on obtient

$$\lambda_1 g_1(k) \leq f_1(k) \leq \mu_1 g_1(k) \leq y \leq \lambda_2 g_2(k) \leq f_2(k) \leq \mu_2 g_2(k).$$

En particulier, on en déduit que $f_1(k) \leq y \leq f_2(k)$, c'est-à-dire $k \in E_y(f_1, f_2)$.
Ensuite, considérons un entier $k \in E_y(f_1, f_2)$, on a

$$f_1(k) \leq y \leq f_2(k).$$

En particulier, si $k \notin [\![0, k_0 - 1]\!]$, par les inégalités (A.2), on a

$$\lambda_1 g_1(k) \leq y \leq \mu_2 g_2(k).$$

Dès lors, on obtient

$$\begin{aligned} E_y(f_1, f_2) &= (E_y(f_1, f_2) \cap [\![0, k_0 - 1]\!]) \cup (E_y(f_1, f_2) \cap [\![k_0, +\infty[\![) \\ &\subseteq [\![0, k_0 - 1]\!] \cup E_y(\lambda_1 g_1, \mu_2 g_2). \end{aligned}$$

Ainsi, on a

$$E_y(\mu_1 g_1, \lambda_2 g_2) \setminus [\![0, k_0 - 1]\!] \subseteq E_y(f_1, f_2) \subseteq E_y(\lambda_1 g_1, \mu_2 g_2) \cup [\![0, k_0 - 1]\!].$$

Or, on a

$$\#(E_y(\mu_1 g_1, \lambda_2 g_2) \setminus [\![0, k_0 - 1]\!]) \geq \#\left(E_y(\mu_1 g_1, \lambda_2 g_2)\right) - k_0$$

et

$$\#(E_y(\lambda_1 g_1, \mu_2 g_2) \cup [\![0, k_0 - 1]\!]) \leq \#\left(E_y(\lambda_1 g_1, \mu_2 g_2)\right) + k_0.$$

On en tire

$$\#\left(E_y(\mu_1 g_1, \lambda_2 g_2)\right) - k_0 \leq \#\left(E_y(f_1, f_2)\right) \leq \#\left(E_y(\lambda_1 g_1, \mu_2 g_2)\right) + k_0.$$

Or, par les lemmes A.8 et A.9, puisque les fonctions $\mu_1 g_1, \lambda_1 g_1$ sont en $\Theta(x^{\alpha_1}\beta_1^x)$ et les fonctions $\lambda_2 g_2, \mu_2 g_2$ en $\Theta(x^{\alpha_2}\beta_2^x)$, on a

$$\begin{aligned}
&\#\ \left(E_y(\mu_1 g_1, \lambda_2 g_2)\right) = \#\ \left(E_y(\lambda_1 g_1, \mu_2 g_2)\right) \\
=&\max\{\frac{1}{\log \beta_1}\log y \ - \ \frac{\alpha_1}{\log \beta_1}\log\log y \ - \ \frac{1}{\log \beta_2}\log y \ + \ \frac{\alpha_2}{\log \beta_2}\log\log y \ + \mathcal{O}(1), 0\} + \mathcal{O}(1) \\
=&\Delta(y) + \mathcal{O}(1).
\end{aligned}$$

En effet, la dernière étape se justifie par le fait que lorsque (β_1, α_1) est lexicographiquement plus petit ou égal à (β_2, α_2), on a $\Delta(y) \geq 0$. De fait, si $(\beta_1, \alpha_1) = (\beta_2, \alpha_2)$, alors on a $\Delta(y) = 0$ et si $(\beta_1, \alpha_1) <_l (\beta_2, \alpha_2)$, alors $\lim\limits_{y\to\infty} \Delta(y) = \infty$. Ainsi, on obtient $\#\ \left(E_y(f_1, f_2)\right) = \Delta(y) + \mathcal{O}(1)$.

□

Corollaire A.11. *En suivant les notations du Théorème A.10, on obtient :*

(i) si $\beta_1 = \beta_2$ et $\alpha_1 = \alpha_2$, on a $\#\ \left(E_y(f_1, f_2)\right) = \mathcal{O}(1)$,

(ii) si $\beta_1 = \beta_2$ et $\alpha_1 < \alpha_2$, on a $\#\ \left(E_y(f_1, f_2)\right) = \Theta(\log\log y)$, et

(iii) si $\beta_1 < \beta_2$, on a $\#\ \left(E_y(f_1, f_2)\right) = \Theta(\log y)$.

Démonstration. La démonstration de (i) est immédiate car, avec ces hypothèses, on a

$$\Delta(y) = 0.$$

Pour la partie (ii), en posant $\beta = \beta_1 = \beta_2$,

$$\Delta(y) = \left(\frac{\alpha_2}{\log \beta} - \frac{\alpha_1}{\log \beta}\right)\log\log y$$

et donc

$$\begin{aligned}\#\ (E_y(f_1, f_2)) &= \Delta(y) + \mathcal{O}(1)\\ &= \left(\frac{\alpha_2}{\log \beta} - \frac{\alpha_1}{\log \beta}\right) \log\log y + \mathcal{O}(1)\\ &= \Theta(\log\log y).\end{aligned}$$

De la même façon, pour la partie (iii), on obtient

$$\#\ (E_y(f_1, f_2)) = -\left(\frac{1}{\log \beta_2} - \frac{1}{\log \beta_1}\right) \log y + \left(\frac{\alpha_2}{\log \beta_2} - \frac{\alpha_1}{\log \beta_1}\right) \log\log y + \mathcal{O}(1)$$

c'est-à-dire

$$\#\,(E_y(f_1, f_2)) = \Theta(\log y).$$

□

Finalement, démontrons les deux derniers lemmes énoncés dans la section 5.1.

Lemme A.12. *Soit une suite réelle strictement positive $(f_k)_{k\in\mathbb{N}}$ définie telle que le quotient $\frac{f_{k+1}}{f_k}$ converge vers une limite réelle, notée β, lorsque k tend vers l'infini. Si $\beta > 1$, alors*

$$\sum_{j=0}^{k} f_j \sim \frac{\beta}{\beta - 1} f_k$$

lorsque k tend vers l'infini.

Démonstration. Notons $(F_k)_{k\geq 0}$ la suite des sommes partielles de la suite $(f_k)_{k\in\mathbb{N}}$. Formellement, pour tout $k \in \mathbb{N}$,

$$F_k = \sum_{j=0}^{k} f_j.$$

Puisque $\frac{1}{\beta} \in]0, 1[$, soient $r, s \in \mathbb{R}$ tels que $0 \leq r < \frac{1}{\beta} < s < 1$. On sait que

$$\lim_{k\to+\infty} \frac{f_k}{f_{k+1}} = \frac{1}{\beta}.$$

Ainsi, il existe $k_0 \in \mathbb{N}$ tel que, pour tout $k \geq k_0$,

$$r \leq \frac{f_k}{f_{k+1}} \leq s. \tag{A.3}$$

Considérons deux entiers k_1, k_2 tels que $k_0 \leq k_1 \leq k_2$. Pour tout $k \in [\![k_1, k_2 - 1]\!]$, les inégalités (A.3) sont valides. Dès lors, en multipliant membre à membre ces $k_2 - k_1$ inéquations, on obtient

$$r^{k_2-k_1} \leq \underbrace{\frac{f_{k_1}}{f_{k_1+1}} \frac{f_{k_1+1}}{f_{k_1+2}} \cdots \frac{f_{k_2-1}}{f_{k_2}}}_{=\frac{f_{k_1}}{f_{k_2}}} \leq s^{k_2-k_1}. \tag{A.4}$$

L'inégalité (A.4) est valide pour tous k_1, k_2 tels que $k_0 \leq k_1 \leq k_2$. Dès lors, fixons $k_2 \geq k_0$ et sommons membre à membre ces inéquations pour tout $k_1 \in [\![k_0 + 1, k_2]\!]$, on obtient

$$\sum_{k_1=k_0+1}^{k_2} r^{k_2-k_1} \leq \sum_{k_1=k_0+1}^{k_2} \frac{f_{k_1}}{f_{k_2}} \leq \sum_{k_1=k_0+1}^{k_2} s^{k_2-k_1}. \tag{A.5}$$

Or, on a

$$\sum_{k_1=k_0+1}^{k_2} r^{k_2-k_1} = \sum_{j=0}^{k_2-k_0-1} r^j = \frac{1 - r^{k_2-k_0}}{1-r},$$

et, de la même façon,

$$\sum_{k_1=k_0+1}^{k_2} s^{k_2-k_1} = \frac{1 - s^{k_2-k_0}}{1-s}.$$

De plus, on sait que

$$\sum_{k_1=k_0+1}^{k_2} \frac{f_{k_1}}{f_{k_2}} = \frac{1}{f_{k_2}} \sum_{k_1=k_0+1}^{k_2} f_{k_1} = \frac{1}{f_{k_2}} \left(\sum_{k_1=0}^{k_2} f_{k_1} - \sum_{k_1=0}^{k_0} f_{k_1} \right) = \frac{1}{f_{k_2}} \left(F_{k_2} - F_{k_0} \right).$$

Ainsi, pour tout $k_2 \geq k_0$, les inégalités (A.5) deviennent

$$\frac{1 - r^{k_2-k_0}}{1-r} \leq \frac{1}{f_{k_2}} F_{k_2} - \frac{1}{f_{k_2}} F_{k_0} \leq \frac{1 - s^{k_2-k_0}}{1-s}. \tag{A.6}$$

Ensuite, remarquons que lorsque k_2 tend vers l'infini, $r^{k_2-k_0}$ et $s^{k_2-k_0}$ tendent vers 0 puisque r et s sont inférieurs à 1. Il en est de même pour le quotient $\frac{F_{k_0}}{f_{k_2}}$. En effet, en utilisant la seconde inégalité de (A.4) avec $k_1 = k_0 + 1$, on obtient

$$\begin{aligned} &\frac{f_{k_0+1}}{f_{k_2}} \leq s^{k_2-k_0-1} \\ \Leftrightarrow &\frac{1}{f_{k_2}} \leq s^{k_2} \frac{s^{-k_0-1}}{f_{k_0+1}}. \end{aligned}$$

Dès lors, on a $\lim\limits_{k_2 \to \infty} \frac{1}{f_{k_2}} = 0$ et vu que F_{k_0} est une constante par rapport à k_2, on a également $\lim\limits_{k_2 \to \infty} \frac{F_{k_0}}{f_{k_2}} = 0$. Ainsi, lorsque k_2 tend vers l'infini, les inégalités (A.6) deviennent :

$$\frac{1}{1-r} \leq \liminf_{k_2 \to \infty} \frac{F_{k_2}}{f_{k_2}} \leq \limsup_{k_2 \to \infty} \frac{F_{k_2}}{f_{k_2}} \leq \frac{1}{1-s}.$$

Pour conclure, remarquons que les réels r et s ont été choisis arbitrairement et peuvent donc être aussi proches que l'on souhaite de $\frac{1}{\beta}$, il s'ensuit que

$$\begin{aligned}
&\lim_{k_2\to\infty} \frac{F_{k_2}}{f_{k_2}} = \frac{1}{1-\frac{1}{\beta}} = \frac{\beta}{\beta-1} \\
\Leftrightarrow &\lim_{k_2\to\infty} \frac{F_{k_2}}{\frac{\beta}{\beta-1} f_{k_2}} = 1.
\end{aligned}$$

Ainsi, la fonction $F_{k_2} = \sum_{j=0}^{k_2} f_j$ est asymptotiquement équivalente à $\frac{\beta}{\beta-1} f_{k_2}$ par la Définition 1.16 des notations de LANDAU.

□

Lemme A.13. *Soit $a > 1$, on a*

$$\sum_{i=2}^{n-1} \log_a \log_a(i) = \Omega(n \log_a \log_a n).$$

Démonstration. Démontrons ce résultat pour le logarithme népérien ln et, par un simple changement de base de logarithme, le résultat sera démontré. Dans ce cas, le résultat à démonter est

$$\sum_{i=2}^{n-1} \ln \ln i = \Omega(n \ln \ln n).$$

Tout d'abord, comme $f : t \mapsto \ln \ln t$ est une fonction croissante, en utilisant une approximation de l'intégrale par la somme de RIEMANN, on obtient une approximation par excès. Formellement,

$$\sum_{i=2}^{n-1} \ln \ln i \geq \int_2^{n-1} \ln \ln t \; dt.$$

Démontrons ensuite que l'on a $\int_2^{n-1} \ln \ln t \; dt \sim n \ln \ln n$. En appliquant le théorème de l'HOSPITAL sur l'intervalle ouvert $]2, +\infty[$, grâce au théorème fondamental du calcul intégral, on a

$$\begin{aligned}
&\lim_{n\to+\infty} \frac{\int_2^{n-1} \ln \ln t \; dt}{n \ln \ln n} \\
= &\lim_{n\to+\infty} \frac{\ln \ln (n-1)}{\ln \ln n + \frac{1}{\ln n}} = 1.
\end{aligned}$$

Ainsi, il existe $N \in \mathbb{N}$ tel que, pour tout $n \geq N$,

$$\frac{\int_2^{n-1} \ln \ln t \; dt}{n \ln \ln n} > \frac{1}{2}.$$

En conclusion, pour tout $n \geq N$, on obtient

$$\sum_{i=2}^{n-1} \ln\ln i \geq \int_2^{n-1} \ln\ln t \; dt > \frac{1}{2} n \ln\ln n.$$

On en tire, alors, que la somme $\sum_{i=2}^{n-1} \ln\ln i$ est en $\Omega(n \ln\ln n)$.

$\square$

Ceci termine l'ensemble des preuves des résultats énoncés dans la section 5.1.

Bibliographie

[1] Ali Aberkane, *Words whose complexity satisfies* $\lim \frac{p(n)}{n} = 1$, Theoret. Comput. Sci. **307** (2003), n° 1, 31–46.

[2] Jean-Paul Allouche et Jeffrey Shallit, *Automatic Sequences : Theory, Applications, Generalizations*, Cambridge University Press, Cambridge, 2003.

[3] Valérie Berthé et Michel Rigo, *Preliminaries*, Combinatorics, Automata and Number Theory, Encyclopedia Math. Appl., vol. 135, Cambridge Univ. Press, Cambridge, 2010, p. 1–33.

[4] Julien Cassaigne, *Complexité et facteurs spéciaux*, Bull. Belg. Math. Soc. Simon Stevin **4** (1997), n° 1, 67–88, journées montoises (mons, 1994).

[5] Julien Cassaigne, *Words with complexity* $p(n) = n + o(n)$, Workshop on Words and Complexity, Lyon, 2018, disponible via l'URL `<https://wordscomplexity.sciencesconf.org/data/pages/Cassaigne.pdf>`.

[6] Julien Cassaigne et François Nicolas, *Factor complexity*, Combinatorics, Automata and Number Theory, Encyclopedia Math. Appl., vol. 135, Cambridge Univ. Press, Cambridge, 2010, p. 163–247.

[7] Fabien Durand, *A characterization of substitutive sequences using return words*, Discrete Math. **179** (1998), n° 1-3, 89–101.

[8] N. Pytheas Fogg, *Substitutions in dynamics, arithmetics and combinatorics*, Lecture Notes in Mathematics, vol. 1794, Springer-Verlag, Berlin, 2002, edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel.

[9] Serge Lang, *Real analysis*, 2e éd., Addison-Wesley Publishing Company, Advanced Book Program, Reading, MA, 1983.

[10] M. Lothaire, *Algebraic combinatorics on words*, Encyclopedia of Mathematics and its Applications, vol. 90, Cambridge University Press, Cambridge, 2002.

[11] Marston Morse et Gustav A. Hedlund, *Symbolic Dynamics*, Amer. J. Math. **60** (1938), n° 4, 815–866.

[12] Jean-Jacques Pansiot, *Complexité des facteurs des mots infinis engendrés par morphismes itérés*, Automata, languages and programming (Antwerp, 1984), Lecture Notes in Comput. Sci., vol. 172, Springer, Berlin, 1984, p. 380–389.

[13] Michel Rigo, *Théorie des automates et langages formels*, Université de Liège, 2009–2010, note de cours de l'Université de Liège.