Motifs inevitables

Adrien Boiret

9 janvier 2008

1 Motifs et notions d'évitabilité

1.1 1.1 Définitions de base

On commencera par introduire les notions nécessaire à l'étude d'évitabilité des motifs, à savoir les **mots infinis**, les **motifs** et les diverses notions d'**évitabilité**.

Définition 1. Un mot fini w sur un alphabet Σ est une suite finie de lettres (éléments de Σ). Un mot infini **u** sur Σ est une suite infinie $(u_n)_{n\in\mathbf{N}}$ de lettres.

Exemple 2. Par exemple, la suite $(u_n)_{n\in\mathbf{N}}$; $u_{2i} = a$ et $u_{2i+1} = b$ définit le mot **u**=ababababab.... sur l'alphabet $\Sigma = \{a, b\}$.

Définition 3. Soit Δ un alphabet disjoint de Σ . Ses éléments sont appellés les variables. On appelle motif un mot sur cet alphabet.

Par souci de notation, les lettres de Σ seront symbolisées par des minuscules, et celles de Δ par des majuscules.

Il faut comprendre les variables comme pouvant prendre des valeurs dans Σ^* par un morphisme non-effaçant h : $\Delta^* \to \Sigma^*$

Définition 4. Soit w un mot sur Σ et p un motif sur Δ. On dit que w contient p s'il existe un morphisme $h : \Delta^* \to \Sigma^*$ tel que h(p) soit un facteur de w. Si w ne contient pas p, on dit qu'il l'évite.

Exemple 5. aabbaabb contient AA (avec h(A)=aabb), mais pas AAA.

On définit ensuite la notion d'évitabilité, qui consiste à être évité par un langage infini.

Proposition 6. *Soit Σ un alphabet fini, p un motif. Les trois énoncés suivants sont équivalents :*
- *Pour tout $n \in \boldsymbol{N}$, il existe un mot w de longueur n ou plus qui évite p.*
- *Il existe un langage infini $L \in P(\Sigma^*)$ qui évite p.*
- *Il existe un mot infini **u** qui évite p.*

Un tel motif est dit évitable. S'il n'a pas ces propriétés, il est naturellement qualifié d'inévitable

Démonstration :

1 ⇔ 2 Sur un alphabet fini, un langage est infini si et seulement si il contient des mots de longueur arbitraire.

3 ⇒ 2 Si **u** mot infini évite p, le langage infini des facteurs de **u** l'évite également.

2 ⇒ 3 Soit L langage infini évitant p. L'alphabet étant fini, on peut construire un mot infini **u** tel que tout préfixe de **u** soit préfixe d'une infinité d'éléments de L. Si **u** contient p, un de ses préfixes u contient p, donc tous les éléments de L admettant u comme préfixe (il y en a une infinité) contiennent p, or L évite p.

La notion d'indépendandance est bien évidemment dépendante de l'alphabet choisi, ou plus exactement de son cardinal. Ainsi on dira qu'un motif est *k-évitable* s'il est évitable sur les alphabets de cardinal k, et *k-inévitable* dans le cas contraire.

Un motif est finalement dit *évitable* s'il l'est pour un alphabet, et *inévitable* s'il l'est pour tout alphabet fini.

Il vient immédiatemment la hiérarchisation suivante :

2-évitable ⇒ 3-évitable ⇒ ... ⇒ k-évitable ⇒ ... ⇒ évitable
inévitable ⇒ ...⇒ k-inévitable ⇒... ⇒ 3-inévitable ⇒ 2-inévitable

On définit alors *l'indice d'évitabilité* $\alpha(p)$ d'un motif p comme le plus petit k tel que p est k-évitable.

Exemple 7.
- ABA est inévitable : sur un alphabet à k lettres, on rencotre forcément 3 fois la même lettre dans un mot de longueur 2k+1. On prend h(A)=a, avec a cette lettre, et h(B) le mot compris entre la première et la troisième occurence de a (non vide : il y a au moins la seconde occurence de a).
- AA est 2-inévitable, mais 3-évitable (voir les mots de Thue-Morse)

1.2 1.2 Propriétés simples

Les motifs ont des propriétés évidentes pour l'évitabilité :

Proposition 8. *Si un motif p su l'alphabet Δ est évitable, toutes ses images par permutations de Δ seront également évitables.*

Définition 9. On dit qu'un motif p divise un autre motif q s'il existe un morphisme non-effaçant $h : \Delta^* \to \Delta^*$ tel que h(p) est un facteur de q

Proposition 10. *Si un motif p évitable divise un motif q, q est évitable*

Exemple 11. Par exemple, vu que AA est 3-évitable, le motif DACBACBD est 3-évitable, car il contient ACBACB, image de AA par h : A$\mapsto ACB$.

Définition 12. On dit qu'un langage L est répétitif si pour tout entier n, il existe un mot de L de la forme wu^nv.

Proposition 13. *Un langage répétitif n'évite aucun motif*

Démonstration : Si $\|p\| = n$, on choisit un mot x de L tel que $x = wu^n v$, et le morphisme

$h : A \in \Delta \mapsto u$. x n'évite pas p, d'où L n'évite pas p

Notons que du coup, les langages rationnels, qui sont répétitifs dès qu'ils sont infinis, n'apportent rien dans la recherche d'un langage évitant un motif.

1.3 1.3 Motifs avec constantes, ensembles de motifs, formules

On présentera ici les généralisations les plus naturelles de la notion de motif :

Définition 14. Un motif avec constante est un mot sur l'alphabet $(\Sigma \cup \Delta)^*$. On dira qu'un mot w sur l'alphabet Σ contient un motif avec constante p s'il existe un morphisme non-effaçant $h : (\Sigma \cup \Delta)^* \to \Sigma^*$ qui fixe Σ tel que h(p) soit facteur de q .

Définition 15. On appelle formule simple un ensemble fini de motifs $p_i, note f = p_1.p_2....p_m$. On dira que w contient la formule simple $f = p_1.p_2....p_m$ s'il existe un morphisme non-effaçant $h : \Delta^* \to \Sigma^*$ tel que $h(p_1), h(p_2)...h(p_m)$ soient facteurs de w.

On peut lire le point entre les motifs comme un "Et" logique. On considèrera donc que $f.f = f$. Il est évident que$f$ n'est inévitable que si tous les p_i le sont.

Définition 16. On dit qu'un ensemble fini de formules est contenu dans un mot w si l'un de ses éléments au moins est contenu dans w. On note$f_1 + f_2 + ... + f_m$ l'ensemble $\{f_1, f_2, ..., f_m\}$

On peut considérer le + comme un "Ou" logique. On considèrera donc que $f + f = f$.Il est évident qu'un ensemble de formules n'est évitable que si tous les f_i le sont.

Proposition 17. *Soient f et f' deux formules, et W une variable qui n'apparait ni dans f ni dans f'. on a alors que fWf' est inévitable si et seulement si f.f' est inévitable.*

Il est évident que l'inévitabilité de fWf' entraine celle de p.p'. À l'inverse, si p.p' est inévitable, cela entraine que tout mot infini **u** contient p et p'.

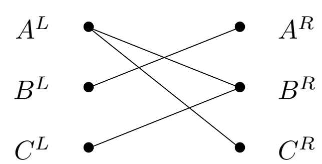
Soit **u'** le mot obtenu en ne considérent que les lettres de **u** après la première occurrence du motif p, plus une autre lettre p' apparait lui aussi dans **u'**. On a bien montré la présence du motif pWp' dans **u**.

Ainsi l'évitabilité de ABBWCABXBBAYBCBZAAB est équivalente à celle de la formule f=AAB.CAB.BBA.BCB(.AAB, mais AAB est déjà dans f).

2 Évitabilité sur un alphabet non fixé, algorithme de Zimin

On s'intéressera dans cette partie au problème d'évitabilité sur un alphabet arbitrairement grand. On y présentera quelques résultats, ainsi qu'un algorithme dû à Zimin qui décide de l'évitabilité d'un motif en passant par les graphes d'adjacence.

Définition 18. Soit p un motif sur Δ. Son graphe d'adjacence est un graphe bipartie avec comme sommets deux copies de Δ, une à gauche et une à droite (i.e. pour chaque variable $X \in \Delta$ on aura deux sommets : X^L à gauche et X^R à droite), et comme arrêtes les $(X, Y) \in \Delta^2$ tels que XY est facteur de p.



Exemple ici du graphe d'adjacence de ABACBA

On aura besoin de quelques notations relatives aux graphes. Pour p un motif et P un ensemble de sommets du graphe d'adjacence, on note $C(P, p)$ la "composante connexe engendrée" par P dans le graphe d'adjacence de p. $C_L(P, p)$ (respectivement $C_R(P, p)$) est l'ensemble des $X \in \Delta$ tels que X^L (respectivement X^R) soit dans $C(P, p)$.

On dira qu'une partie P de l'ensemble S des sommets d'un tel graphe est *libre* s'il n'existe aucun sommet de type (X^L, Y^R) , $X \in P$, $y \in P$

Exemple 19. Les parties $\{A\}$ et $\{B\}$ sont libres pour le motif ABACBA

Définition 20. On dit qu'un motif p se réduit en une étape en q s'il existe $P \subset \Delta$ telle que :

- P est libre pour p
- En effaçant dans p les occurences des variables de P, on trouve exactement q

On dit que p se réduit en q, et on note $p \to q$ quand on trouve $p = p_0, p_1...p_m = q$ tels que $\forall 0 \leq i \leq m-1$, p_i se réduit en une étape en p_{i+1}. On dit que p est réductible s'il se réduit en ε

L'algorithme de Zimin repose sur une unique propriété :

Théorème 21. *Un motif est évitable si et seulement si il est irréductible*

Exemple 22. Le motif ABACBA se réduit en AACA et en CBC. **On remarque** que le motif AACA n'est pas réductible alors que CBC l'est : pour savoir qu'un motif est irréductible, il faut essayer toutes ses réductions !

Partant de là, Zimin propose un algorithme très simple pour savoir si un motif p est évitable :

– On cherche les parties libres du graphe d'adjacence du motif p
– On en choisit une de manière non-déterministe !
– On réduit p par rapport à cette partie libre, puis on recommence avec le motif réduit obtenu

2.1 Démonstration

:

2.1.1 Réductible $\Rightarrow$ Inévitable

Pour démontrer ce point, on se servira du lemme suivant :

Lemme 23. *Si $p \to p$, et q inévitable, alors p est inévitable*

ε étant clairement inévitable, on aurait bien la propriété cherchée.

Démonstration : Soit p réduit par supression de P en q. On va raisonner par récurence sur la taille de l'alphabet Σ. Pour Σ de cardinal 1, c'est évidemment vrai, car tout motif est inévitable.

Considérons maintenant que l'on a le résultat pour Σ, et que l'on cherche à le prouver pour $\Sigma' = \Sigma \cup \{a\}\Sigma$.

Soit q inévitable sur Σ', d'où p inévitable sur Σ. Notons L_p l'ensemble (fini) des mots évitant p sur Σ, et M l'ensemble des mots sur Σ' évitant p et commençant par un a. Ces mots sont soit des puissances de a (il n'y en a qu'un nombre fini), soit un mot du type N^n avec $N = \{a^i w a^j\ ,\ w \in L_p\ ,\ i,j \leq |p|\}$ (également fini).

On peut donc considérer un mot de M comme un mot sur "l'alphabet N" (ou une puissance de a). Soit $X \in \Delta$ une variable n'apparaissant pas dans p. On a immédiatement qX inévitable. Il existe donc n_0 tel que tout mot de longueur $\geq n_0$ n'évite pas qX, sur l'alphabet N.

On a donc , pour de tels w, un morphisme h pour lequel h(qX) est facteur de w (vu comme mot sur Σ). Notons, pour tout $Y \in \Delta\ ,\ h(Y) \in aA^+$

On définit alors le morphisme g suivant :

– $g(X) = a$ si $X \in P$
– $g(X) = a^{-1}h(X)a$ si $X \in C_L(P^L, p) \cap C_R(P^L, p)$
– $g(X) = h(X)a$ si $X \in C_L(P^L, p) \setminus (C_R(P^L, p) \cup P)$
– $g(X) = a^{-1}h(X)$ si $X \in C_R(P^L, p) \setminus (C_L(P^L, p) \cup P)$
– $g(X) = h(X)$ comme mot sur Σ si ni X^L ni X^R ne sont dans C(P,p)

Ce g sert à construire des mots à partir de motifs tels que g(p) soit facteur de h(qX)(ce qui donnerait que w comme mot de Σ n'évite pas p).

On montrera pour commencer que pour p_k préfixe de p de longueur k, et que $p_k \to q_k$ pour P, avec q_k préfixe de q, alors

$(a\ ou\ \varepsilon)g(p_k) = h(q_k)(a\ ou\ \varepsilon)$, selon que p_k commence (resp. termine) par une lettre de $C_R(P^L, p)$ (resp. $C_L(P^L, p)$).

- k=1 : Cela découle de la construction de g par rapport à h.
- Soit $(a\ ou\ \varepsilon)g(p_k) = h(q_k)(a\ ou\ \varepsilon)$, et $p_{k+1} = p_k Y$.
  Si on a $(a\ ou\ \varepsilon)g(p_k) = h(q_k)a$, c'est que p_k se termine par une lettre de $C_L(P^L, p)$, et donc que $Y \in C_R(P^L, p)$, ce qui efface le a qui gênait. De même, on n'a $(a\ ou\ \varepsilon)g(p_{k+1}) = h(q_{k+1})a$ que si $Y \in C_L(P^L, p)$.

Ainsi on a que $(a\ ou\ \varepsilon)g(p) = h(q)(a\ ou\ \varepsilon)$, d'où évidemment g(p) facteur de h(qX).

Cela donne que les mots assez longs (sur l' "alphabet" N) n'évitent pas p. M est par conséquent fini, ce qui veut dire que p est inévitable.

2.1.2 Irréductible ⇒ Évitable

Cette partie de la démonstration requiert la démonstration de quelques lemmes.

Lemme 24. *Soient p, q, tels que q divise p (i.e. on a f non effacant tel que h(q) facteur de p. Soit P partie libre pour p, et P' l'ensemble des X tels que $h(X) \in P^+$. On a que P' est libre. De plus, si $p \to p'$ et $q \to q'$ par réduction par P et P', h'(q') est facteur de p', avec h' le morphisme qui à $X \in (\Delta \setminus F')$ associe son image h(X) débarassé de ses variables dans P.*

Démonstration : On définit pour $X \in \Delta$ $h_1(X)$ *et* $h_2(X)$ la première et la dernière lettre de h(X). On a évidemment que si XY facteur de q, $h_1(X)h_2(Y)$ est facteur de p. Le graphe d'adjacence de q est donc "envoyé par h" sur un sous-graphe du graphe d'adjacence de p, en collant X^L su $h_1(X)^L$ et X^R sur $h_2(X^R)$. Aussi, si X et y sont dans P', ils ne seront pas liés dans le graphe de q, puisque leurs images ne le sont pas dans le graphe de p. P' est donc libre. La fin du lemme s'obtient en comparant les processus de réduction dans p et de h' sur q.

On notera qu'en réalité, $X \in P'$ a une image de longueur 1, vu que deux lettres de F ne peuvent se suivre dans p.

Lemme 25. *Soient p, p', q tels que $p \to p'$ et h(q) facteur de p, avec f non-effaçant. On aura q', h' tels que $q \to q'$ et h'(q') facteur de p', avec pour tout X apparaissant dans q mais pas dans q', h(X) est constitué de variable apparaissant dans p mais pas dans p'.*

La démonstration de ce lemme est immédiate, il suffit d'itérer le lemme précédent autant de fois qu'il faut de réductions pour passer de p à p'.

Lemme 26. *Soit $q = \delta_V(p)$ le motif obtenu en effaçant dans p toute apparition des variables $X \in V$, avec V pas nécessairement libre pour p.*

S'il existe r motif et h morphisme non-effaçant tels que $r \to q$ et h(p) facteur de r, avec en plus $X \in V \Leftrightarrow h(X)$ est fait de variables n'apparaissant pas dans q, alors $r \to q$.

Démonstration :

On sait que $r \to q$ et que h(p) est facteur de r. On applique le lemme 25 pour obtenir p' motif et h' non-effaçant tels que $p \to p'$ et h'(p') facteur de q, tel que si X n'apparait pas dans p', h(X) est constitué de variable apparaissant dans r mais pas dans q. Par hypothèse, cela entraine que de tels X (qui constituent exactement la partie libre effacée lors du passage de p à p') appartiennent à V.

On a donc $\delta_V(p') = \delta_V(p) = q$. Or, on a h' non effaçant tel que h'(p') facteur de q. Par conséquent |p'|=|q|, d'où $\delta_V(p') = p' = q$, ce qui donne $p \to q$.

On s'intéresse maintenant à des mots infinis bien particuliers définis sur les alphabets de longueur 4k, avec de très forte propriétés d'évitabilité.

Définition 27. Pour k entier, on note $A_k = \{a_0, a_1, ..., a_{2k-1}, b_0, b_1, ..., b_{2k-1}\}$ l'alphabet à 4k lettres, et on définit sur A_k^* le morphisme φ_k avec pour i allant de 0 à 2k-1 :

- $\varphi_k(a_i) = a_0 b_i a_1 b_{i+1} ... a_{k-1} b_{i+k-1}$
- $\varphi_k(b_i) = a_k b_i a_{k+1} b_{i+1} ... a_{2k-1} b_{i+k-1}$

avec la convention d'indices pris modulo 2k-1.

On définit ensuite $w^{(k)}$ le mot infini produit par itération de φ_k une infinité de fois sur a_0, i.e. son point fixe.

On montrera que tout motif irréductible est évité par un certain $w^{(k)}$.

Lemme 28. *Soit v un facteur de taille 2 ou plus de $w^{(k)}$. On aura $0 \leq i \leq 4k-1$ et $x \in A_k$ tel que, si v apparait à la position i, on a :*

- *$n \equiv i\,[4k]$*
- *la lettre de $w^{(k)}$ à la position $n' = \lfloor \frac{n}{2k} \rfloor$ sera x*

Démonstration :

On se contentera de montrer ce lemme pour le cas $|v| = 2$, qui entraine de manière évidente le cas général. En remarquant qu'à chaque application de φ_k, on conserve l'alternance des a et des b dans le mot, on obtient que chaque occurence d'un $\varphi_k(a_i)$ est suivie d'un a_k, et chaque occurence d'un $\varphi_k(b_i)$ est suivie d'un a_0. Ainsi, on a toujours $w^{(k)}_{2mk+2i} = a_i$. On peut donc connaitre sa position dans $w^{(k)}$ modulo 4k à l'aide des a_i. Or, un motif de deux lettres contiendra toujours un a_i, ce qui fixe bien sa position dans $w^{(k)}$; on a donc montré le premier point du lemme.

De même, chaque image étant de longueur 2k, si v apparait à la position p, c'est qu'il débute dans l'image de $x = w^{(k)}_{n'}$. On déduit cette lettre de manière simple, par la construction de φ_k :

- $v = a_i b_j$: Si $0 \leq i \leq k-1,\ x = a_{j-i}$; sinon, $x = b_{j-i+k}$
- $v = b_j a_i$: Si $1 \leq i \leq k,\ x = a_{j-i+1}$; sinon, $x = b_{j-i+k+1}$

v détermine son "antécedent" par φ_k, et donc la lettre en position n'.

Lemme 29. *Soit p motif, et k entier tel qu'il y ait moins de 2k variables différentes de* Δ *dans p. Soit v facteur de* $w^{(k)}$ *tel que* $\varphi_k(v)$ *contient p. On a alors un motif q tel que v contient q et* $p \to q$.

Démonstration :

$\varphi_k(v)$ contient p, on a donc un morphisme non-effaçant h pour lequel h(p) est facteur de $\varphi_k(v)$. Le lemme 28 nous enseigne que pour toute lettre x, il y a 2k mots de deux lettres qui caractérisent l'appartenance à $\varphi_k(x)$, tous terminant par une lettre différente.

Or, il y a moins de 2k variables différentes dans p. Un de ces 2k mots de deux lettres finira donc par une lettre qui n'est la première lettre d'aucun h(X), pour X apparait dans p. Appelons ce mot d_x, le mot décisif pour x. On remarquera que si d_x apparait dans h(p), il est contenu entièrement dans un des h(X), avec X variable apparaissant dans p.

On notera V l'ensemble des variables X telles que h(X) ne contient aucun mot décisif, et $q = \delta_V(p)$. On construit le morphisme $h' : (\Delta \setminus V)^* \to \Sigma^*$, qui envoie X sur $x_1 ... x_m$, avec x_i les lettres telles que d_{x_i} facteurs de h(X), dans leur ordre d'apparition. Ce morphisme est donc non-effaçant. Vu que chaque mot décisif correspond à une lettre de v, on a que h'(v) est facteur de q, et donc v contient q.

Reste à prouver que $p \to q$, c'est-à-dire, que V est libre. On définit le morphisme $f : \Delta^* \to (\Sigma \cup \Delta^*)$:

- Si $X \in V$, f(X)=h(X)
- Si X ne conitient ni a_0, ni a_k, mais un (et un seul) mot décisif d_x, on a que $\varphi_k(x) = a_{0 \ ou \ k} v_1 h(X) v_2$. On prend alors $f(X) = h(X) v_1 X v_2 h(X)$
- Si $h(X) = v_1 \varphi_k(w) a_{0 \ ou \ k} v_2$ avec w le plus long possible, on notera a_i la première lettre de $\varphi_k(w) a_{0 \ ou \ k}$, x_1 *et* x_2 les première et dernière lettres de h(X). On prendra $f(X) = v_1 v_1' X v_2' v_2$, où v_1 vaut ε si d_{x_1} est dans $v_1 a_1$, $\varphi_k(x_1)$ sinon, et de même pour $v'2$, ε si d_{x_2} est dans $a_j v_2$, $\varphi_k(x_1)$, avec j=0 ou k, ou $(a_{k-j})^{-1} \varphi_k(x_2) a_j$ sinon.

On remarquera qu'en effaçant de f(p) les lettres de Σ, on retombe exactement sur q. De plus, deux variables consécutives X_1 et X_2 dans q seront séparées par $v_1 a_i v_2$, avec $a_{k-i} v_1$ l'image par φ_k de la dernière lettre de $h'(X_1)$, et $a_i v_2$ l'image par φ_k de la première lettre de $h'(X_2)$.

On réduit alors f(p) (en tant que motif sur $\Sigma \cup \Delta$) comme suit :

- On supprime tous les b_i, qui forment évidemment une partie libre pour f(p). On obtient p_0.
- On supprime les unes après les autres les parties $\{a_i, a_{k+i}\}$, pour i de 1 à k-1. On obtien p_{k-1}, alternance de lettres de Δ et de $\{a_0, a_k\}$.
- On supprime la partie libre $\{a_0, a_k\}$, pour obtenir q.

On a donc $f(p) \to q$. On est donc dans les conditions du lemme 26, qui nous donne $p \to q$.

Lemme 30. *Si $w^{(k)}$ contient p contenant plus de 2k variables distinctes, p est réductible.*

Démonstration :

Pour m assez grand, $\varphi_k^m(a_o)$ contient p. Le lemme 29 nous donne un motif p_1 tel que $p \rightarrow p_1$ et que $\varphi_k^{m-1}(a_o)$ contient p_1.

On répète l'opération jusqu'à trouver p_m, tel que a_0 contient p_m, ce qui veut dire que p_m est vide ou réduit à une variable, donc réductible. On a par ailleurs $p \rightarrow p_m$, d'où p réductible.

Démonstration de Inévitable $\Rightarrow$ Réductible :

Si p est inévitable, on aura que tous les $w^{(k)}$ contiennent p, dont un k suffisamment petit pour pouvoir appliquer le lemme 30, ce qui nous donne p réductible.

On a donc démontré la correction de l'algorithme de Zimin, sa terminaison étant évidente. Il établit la décidabilité du problème d'évitabilité.