Les lemmes d'itération

Pour les rationnels

L'idée du lemme d'itération dans le cas des langages rationnels est la suivante : soit M un automate qui reconnaît un langage L. Soit N le nombre d'états de M et w un mot de L de longueur plus grande que N. Puisque w est dans L, il est reconnu par M. Un chemin d'étiquette w est de longueur supérieure ou égale à N et contient donc une boucle. Cette boucle peut être répétée autant de fois qu'on veut, le mot obtenu sera toujours reconnu et donc dans L.

Formellement, on écrira :

Soit $L \in \text{Rat}(\Sigma^*)$, il existe un entier N tel que tout mot w de L avec $|w| \geq N$ a une décomposition en $w = xuy$ avec $u \neq \varepsilon$, $|xu| \leq N$ et $xu^iy \in L$ pour tout $i \geq 0$.

On appelle u le facteur itérant.

Le lemme d'itération est utilisé pour montrer qu'un langage n'est pas rationnel. On suppose qu'un langage L est rationnel. Il vérifie donc le lemme d'itération. Soit N l'entier du lemme et w un mot (généralement un mot particulier) de longueur plus grande que N. On montre alors qu'on ne peut pas placer dans w le facteur itérant ce qui contredit le fait que L est rationnel.

Exemple :

Soit $L = \{a^nb^n, n \geq 0\}$, N l'entier du lemme et $w = a^Nb^N$. D'après le lemme d'itération, il existe une décomposition $w = xuy$ avec $u \neq \varepsilon$, $|xu| \leq N$ et $xu^iy \in L$ pour tout $i \geq 0$. Comme $|xu| \leq N$, $u = a^j$. $xu^2y = a^ka^{2j}a^{N-(k+j)}b^N \in L$, donc $k + 2j + N - (k + j) = N + j = N \Rightarrow j = 0$ ce qui contredit $u \neq \varepsilon$. Donc L ne peut pas être rationnel.

Pour montrer qu'un langage est ou n'est pas rationnel, on utilise aussi les propriétés de fermeture de $\text{Rat}(\Sigma^*)$.

Propriétés de fermeture de Rat(Σ^*)

Nous avons déjà vu que $\text{Rec}(\Sigma^*)$ (et donc $\text{Rat}(\Sigma^*)$) est fermé par union finie, intersection finie, complémentaire, produit de concaténation et étoile.

On peut ajouter à ces propriétés, que $\text{Rat}(\Sigma^*)$ est fermé par morphisme et morphisme inverse. Rappelons qu'un morphisme est une application

$\mu : \Sigma^* \longrightarrow \Delta^*$

qui vérifie les propriétés suivantes :

$$\begin{array}{lcl} \mu(u \cdot v) & = & \mu(u) \cdot \mu(v) \\ \mu(\varepsilon) & = & \varepsilon \end{array}$$

Ces propriétés permettent de définir les morphismes uniquement sur les lettres d'où le nom de morphismes alphabétiques.

Exemple : Soit $\mu : \Sigma = \{a, b\} \longrightarrow \Delta^* = \{0, 1\}^*$ défini par $\mu(a) = 01$ et $\mu(b) = \varepsilon$
$\mu(a^*b^*) = (01)^*$ où a^*b^* et $(01)^*$ sont rationnels.
$\mu(\{a^nb^n, n \geq 0\}) = (01)^n, n \geq 0 = (01)^*$ où $\{a^nb^n, n \geq 0\}$ n'est pas rationnel et $(01)^*$ est rationnel par contre $\mu^{-1}((01)^*) = a^*$ qui est rationnel.

Pour les algébriques

L'idée du lemme d'itération dans le cas des langages algèbriques est la suivante : soit G une grammaire sous forme normale de Chomsky réduite qui engendre un langage L. Soit N le nombre de non-terminaux de G et w un mot de L de longueur plus grande que N. Puisque w est dans L, il est donc engendré par G et une dérivation de S en w contient plus de N non-terminaux. Il existe donc au moins un non-terminal qui apparaît 2 fois dans la dérivation. On peut appliquer la même série de règles à partir de cette nouvelle occurrence et rester dans le langage L. Comme la grammaire est sous forme de Chomsky, on peut alors faire apparaître 2 facteurs itérants dans le mot.

Formellement, on écrira :

Soit $L \in \text{Alg}(\Sigma^*)$, il existe un entier N tel que tout mot w de L avec $|w| \geq N$ a une décomposition en $w = xuyvz$ avec $u \neq \varepsilon$ ou$v \neq \varepsilon$, $|uyv| \neq N$ et $xu^iyv^iz \in L$ pour tout $i \geq 0$.

u et v sont deux facteurs itérants. Si $u = \varepsilon$ ou $v = \varepsilon$, on retrouve la version du lemme pour les rationnels.

Comme pour les langages rationnels, le lemme d'itération est utilisé pour montrer qu'un langage n'est pas algèbrique. On suppose qu'un langage L est algèbrique donc il vérifie le lemme d'itération. Soit N l'entier du lemme et w un mot (généralement un mot particulier) de longueur plus grande que N. On montre alors qu'on ne peut pas placer dans w les deux facteurs itérants ce qui contredit le fait que L est algèbrique.

Exemple : Soit $L = \{a^nb^nc^n, n \geq 0\}$, N l'entier du lemme et $w = a^Nb^Nc^N$. D'après le lemme d'itération, il existe une décomposition $w = xuyvz$ avec $u \neq \varepsilon$ ou$v \neq \varepsilon$, $|uyv| \neq N$ et $xu^iyv^iz \in L$ pour tout $i \geq 0$. Si u ou v contient des lettres différentes, par exemple $u = a^ib^j$, alors en itérant u, on va faire apparaître un facteur ba et le mot obtenu ne sera pas dans L. u et v sont donc des puissances de lettre. Mais on itère au plus deux facteurs alors que le mot contient trois types de lettres, on ne peut donc pas conserver le même nombre de a que de b que de c.

Pour montrer qu'un langage est ou n'est pas algèbrique, on utilise aussi les propriétés de fermeture de $\text{Alg}(\Sigma^*)$.

Propriétés de fermeture de Alg(Σ^*)

Nous avons déjà vu que $\text{Alg}(\Sigma^*)$ est fermé par union finie, produit de concaténation et étoile.

On peut ajouter à ces propriétés, que $\mathrm{Alg}(\Sigma^*)$ est fermé par morphisme et morphisme inverse et image miroir par contre il n'est fermé ni par intersection ni par complémentation.

Il existe cependant une propriété qui est fort utile comme nous le verrons dans les exercices : l'intersection d'un algèbrique et d'un rationnel est algèbrique.

On peut encore donner quelques propriétés de fermeture sur les langages déterministes $\mathrm{Det}(\Sigma^*)$.

$\mathrm{Det}(\Sigma^*)$ est fermé par complémentation, morphisme inverse mais pas direct. Il n'est fermé ni par union ni par intersection. Par contre l'intersection d'un déterministe et d'un rationnel est déterministe.