La complexité de l'algorithme de Brzozowski est super-polynomiale génériquement.

Sven De Félice, Cyril Nicaud

Laboratoire d'informatique de Gaspard Monge, Université Paris-Est Marne-la-Vallée, CNRS

Aléa 2013

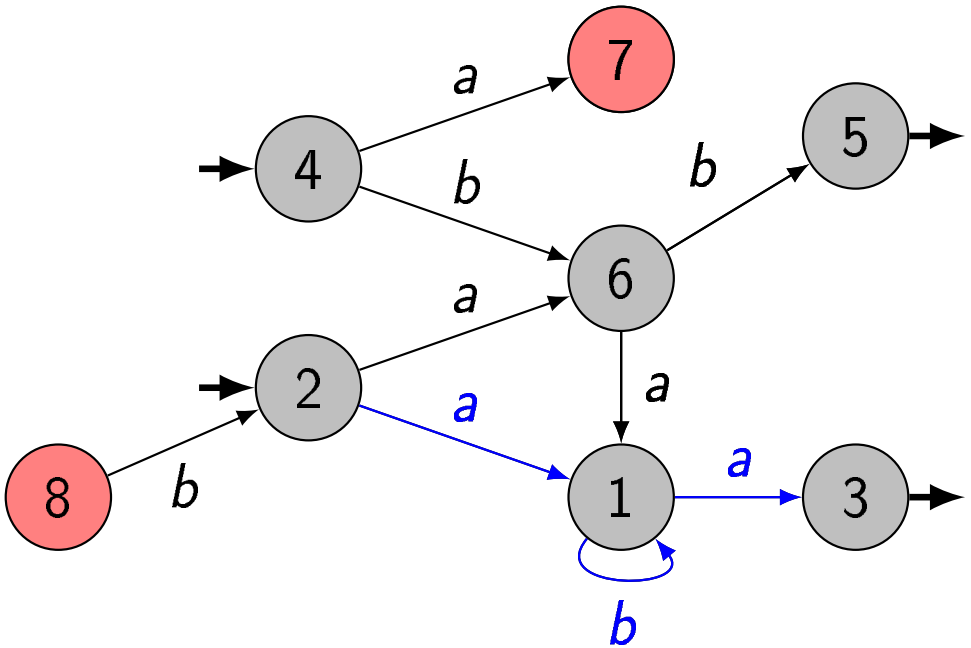
Algorithme de Brzozowski

- Algorithme de minimisation de Brzozowski.
- Adapté pour les automates non-déterministes.
- Pas efficace pour les automates déterministes avec distribution uniforme (sujet de l'exposé).

Brzozowski's algorithm [edit]

As Brzozowski (1963) observed, reversing the edges of a DFA produces a non-deterministic finite automaton (NFA) for the reversal of the original language, and converting this NFA to a DFA using the standard powerset construction (constructing only the reachable states of the converted DFA) leads to a minimal DFA for the same reversed language. Repeating this reversal operation a second time produces a minimal DFA for the original language. The worst-case complexity of Brzozowski's algorithm is exponential, as there are regular languages for which the minimal DFA of the reversal is exponentially larger than the minimal DFA of the language,[5] but it frequently performs better than this worst case would suggest.[4]

Automate

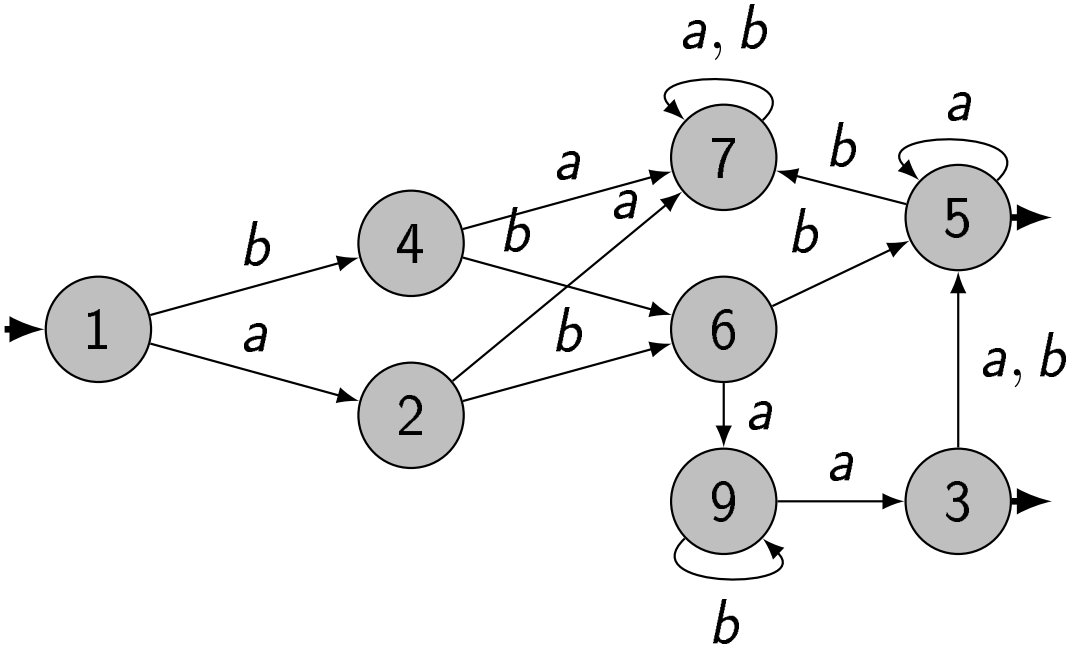


- $\Sigma = \{a, b\}$, $|\Sigma| = k$, nombre d'états : n.
- Langage reconnu : $\{aa, ab, bb, aaa, aba, baa, ...\}$

Automate émondé

Un automate est émondé si pour chaque état i il existe un chemin allant d'un état initial à un état terminal passant par i.

Automate déterministe



Déterministe (complet)

Un automate est déterministe si :

- Il possède un unique état initial.
- Pour chaque état e et chaque lettre ℓ il existe une et une seule transition sortante de e étiquetée par ℓ.

k fonctions de $[n]$ dans $[n]$ $\longrightarrow$ une pour chaque lettre.

Automate minimal

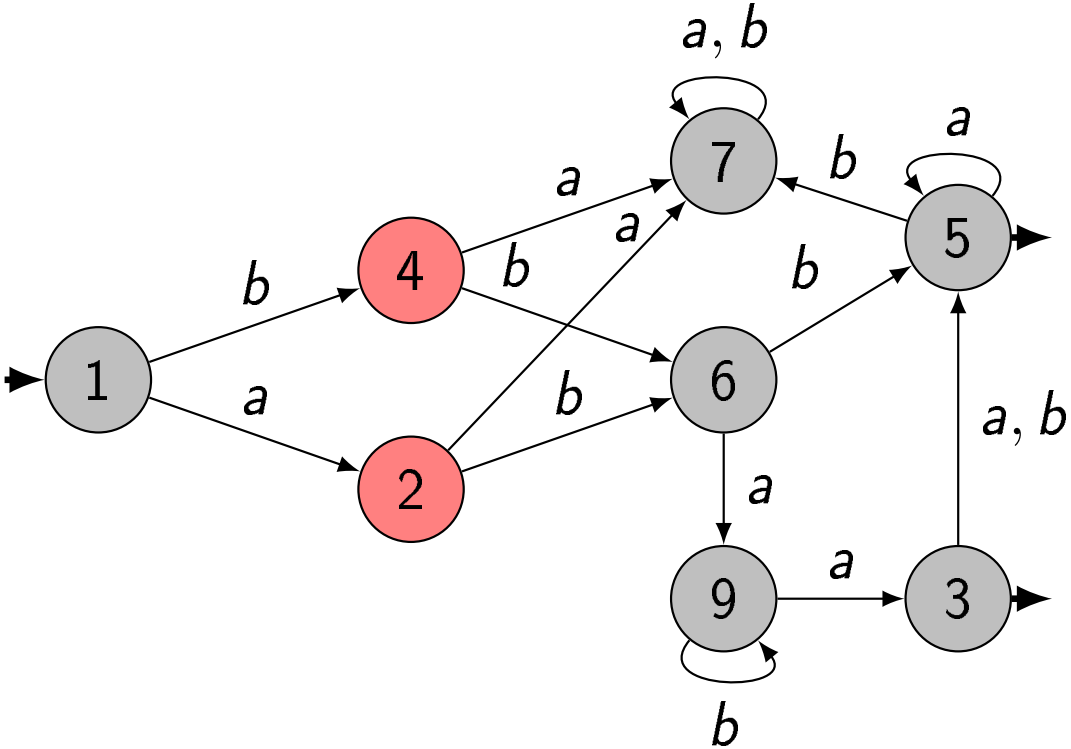
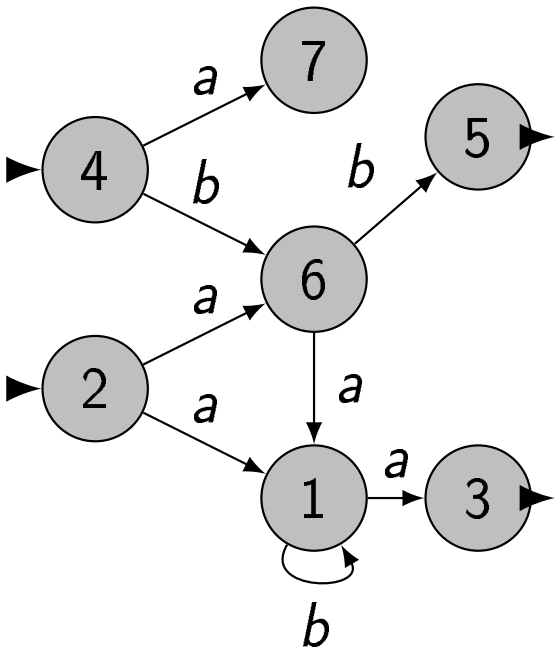


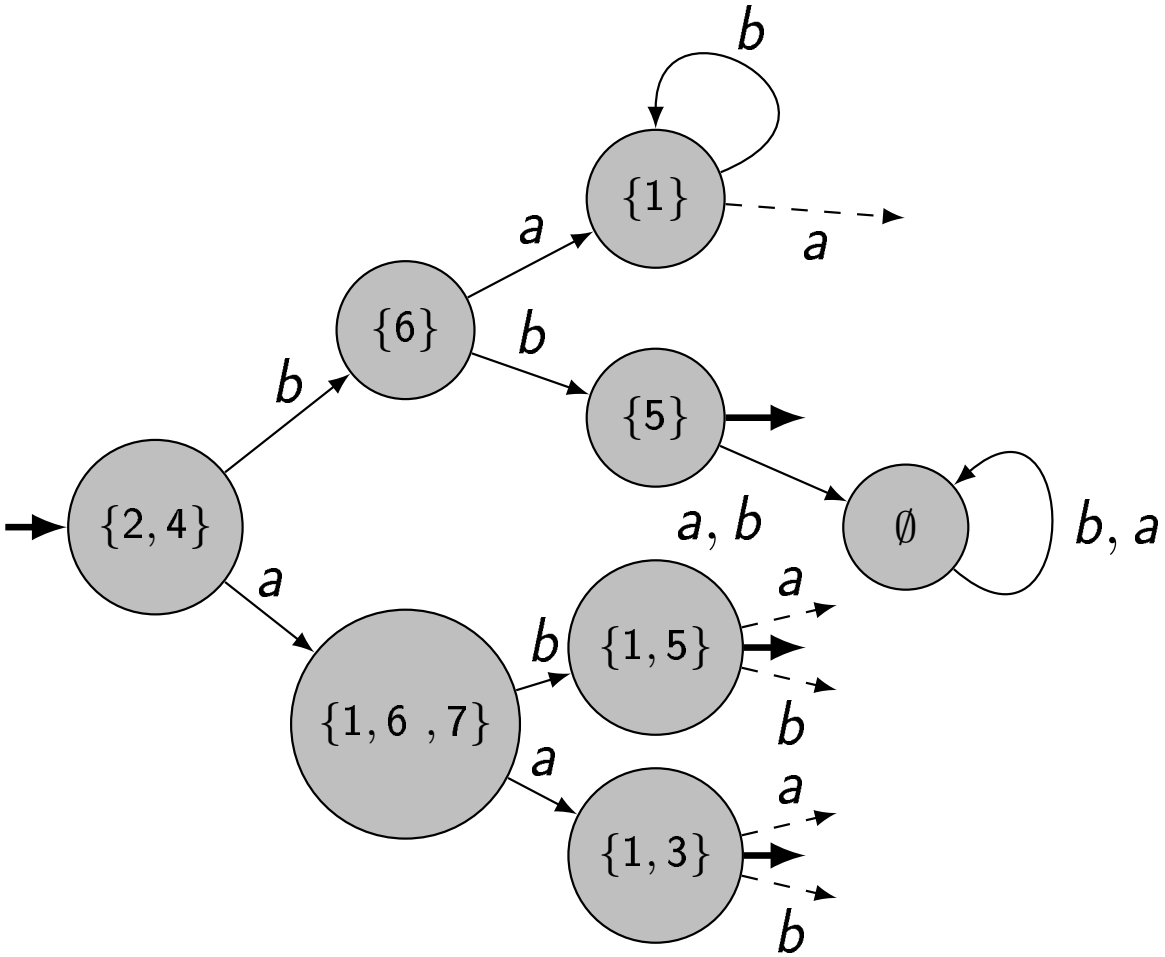
Figure: automate déterministe non minimal

Minimal

Un automate déterministe est minimal si il est le plus petit automate déterministe reconnaissant un langage.

Subset Construction





Dans le pire des cas le déterminisé est de taille 2^n.

Définitions

- Un état est accessible si il existe un chemin allant d'un état initial à cet état.
- Le miroir d'un automate est obtenu en permutant les états initiaux et les état terminaux et en inversant l'orientation des transitions.
- Un automate est co-déterministe si son miroir est déterministe.

Proposition

Le miroir d'un automate reconnaît exactement les mots renversés de l'automate initial.

Le renversé du mot *aabba* est *abbaa*.

Algorithme de Brzozowski

Théorème [Brzozowski,1962]

Le déterminisé (par subset construction) d'un automate co-déterministe émondé est minimal.

Algorithme de Brzozowski :

1. On émonde l'automate.
2. On déterminise son miroir (explosion des états).
3. On le retourne.
4. On le déterminise à nouveau.

Algorithmes de minimisation d'automates déterministes

1. Moore
   - Pire des cas $O(k \cdot n^2)$
   - En moyenne $\Theta(nlog(logn))$ pour la distribution uniforme [David,2012].
2. Hopcroft
   - Pire des cas $O(kn\log(n))$.
   - On peut faire aussi bien que Moore pour la distribution uniforme [David,2012].
3. Brzozowski
   - Pire des cas $O(2^n)$.
   - Génériquement $\Omega(e^{\alpha \log^2(n)})$.

Générique

Une propriété est générique si sa probabilité tend vers 1 lorsque n grandit.

Modèle

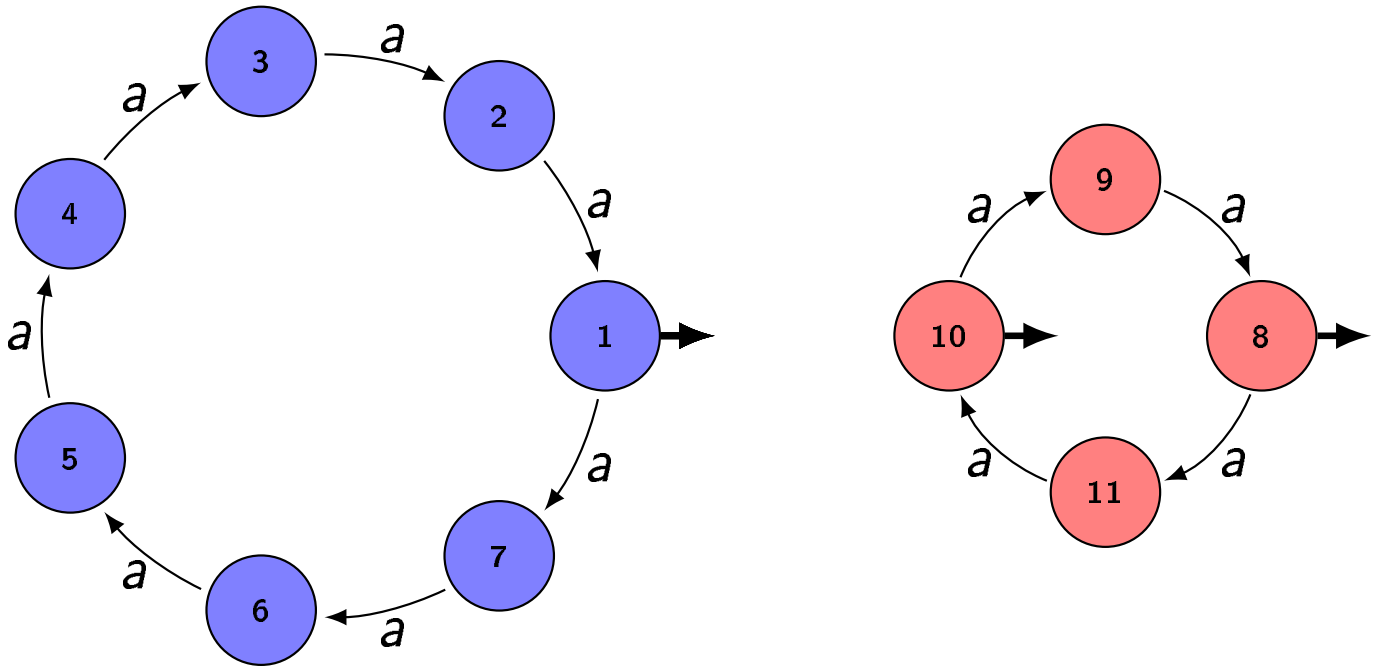
Distribution :

- Les k fonctions des transitions sont tirées uniformément parmi les fonctions de $[n]$ dans $[n]$.
- L'état initial est 1.
- Un état est terminal avec probabilité $p \in]0, 1[$.

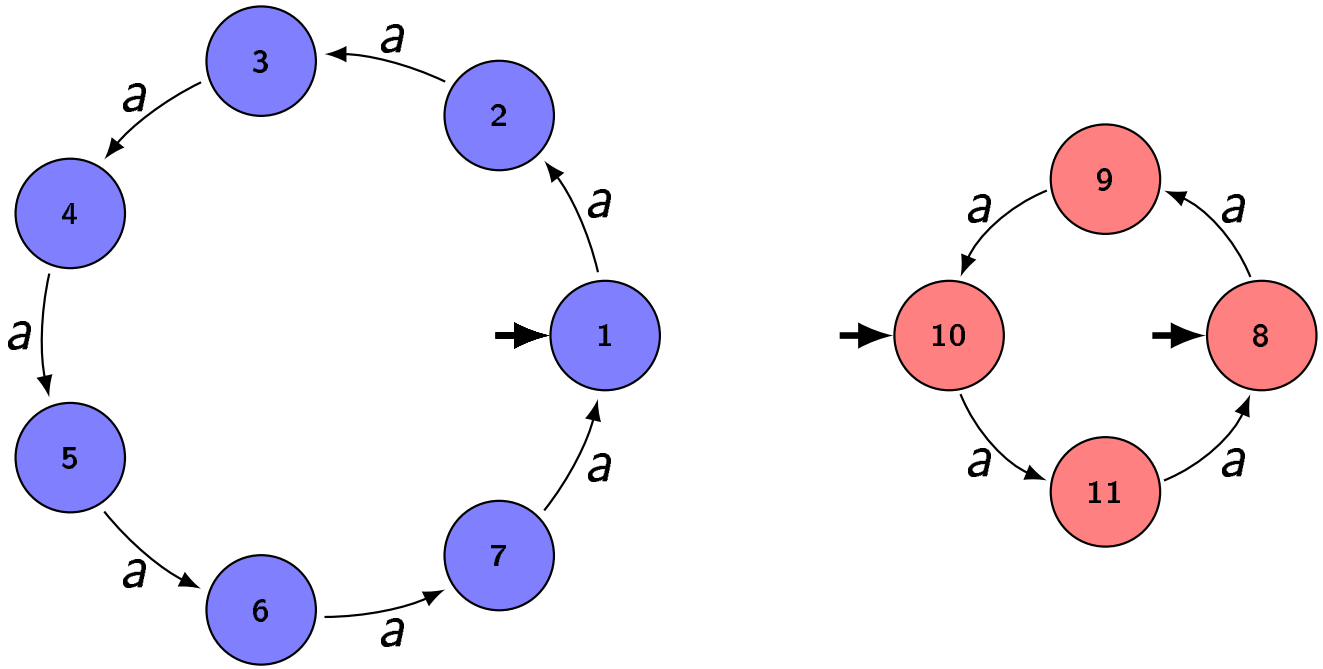
Théorème

La complexité en état du déterminisé du miroir est $\Omega(e^{\alpha \log^2 n})$ génériquement.

Exemple sur deux cycles



Exemple sur deux cycles



$$\{1,8,10\} \xrightarrow{a} \{2,9,11\} \xrightarrow{a} \{3,8,10\} \xrightarrow{a} \{4,9,11\} \cdots\cdots \cdots \{1,8,10\}$$

La longueur du cycle engendré est le ppcm des périodes des cycles : $\text{ppcm}(7,2) = 14$

Idée de la preuve

Primitif

Un cycle est primitif si sa période est égale à sa longueur.

On regarde les "grands" cycles de la lettre a, ceux qui ont une longueur supérieure à $\log(n)$. On va montrer que génériquement :

- Ces grands cycles sont accessibles.
- Le ppcm de leurs longueurs dépasse $e^{\alpha \log^2(n)}$.
- Ils sont primitifs.

La trace sur ces cycles nous permettra de conclure.

Accessibilité des grands cycles

Proposition [Sportiello]

Génériquement, tous les cycles de taille supérieure à $\log(n)$ sont accessibles.

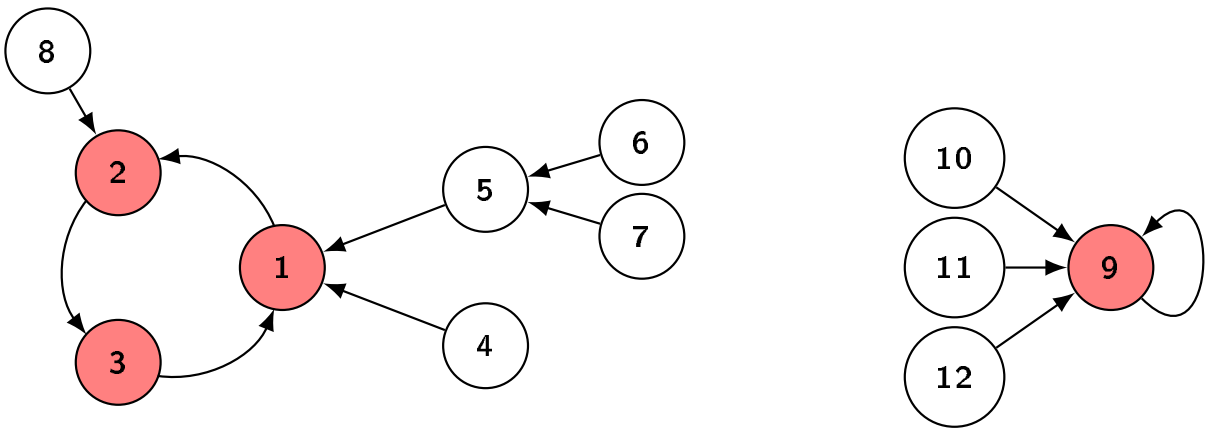
Lemme [Carayol,Nicaud,2012]

La proportion des états accessibles d'un automate du modèle est génériquement comprise dans un intervalle $[\alpha, \beta]$ $0 < \alpha < \beta < 1$.

Fonction aléatoire

Une fonction de $[n]$ dans $[n]$ est décomposable en un ensemble de cycles d'arbres.

On peut utiliser des méthodes de combinatoire analytique [Flajolet,Odlyzko,1990].



- Génériquement, le nombre de points cycliques dépasse $n^{\frac{1}{3}}$.
- A taille fixée, la partie cyclique d'une fonction aléatoire se comporte comme une permutation uniforme.

Ppcm des grands cycles

Théorème [Erdös,Turán,1965]

Génériquement, le ppcm des longueurs des cycles d'une permutation de taille m est plus grande que $e^{\frac{1}{3}\log^2(m)}$.

Proposition [Landau,1903]

Le ppcm des longueurs des cycles d'une permutation de taille m est inférieur à $2e^{\sqrt{m\log(m)}}$.

Proposition

Le nombre de cycle d'une permutation de taille m ne dépasse pas $2\log(m)$ génériquement.

Théorème

Génériquement, le ppcm des longueurs des grands des cycles d'une fonction aléatoire est plus grand que $e^{\alpha\log^2(n)}$

Primitivité des grands cycles

Proposition

Génériquement, tous les cycles de longueur supérieure à $\log(n)$ sont primitifs.

Dans le cas uniforme ($p = \frac{1}{2}$) :

- Regarder si un cycle de longueur ℓ est primitif revient à regarder la primitivité des mots de longueur ℓ sur deux lettres.
- Un mot w n'est pas primitif s'il existe un mot u tel que $w = u^d$, $d > 1$.

La probabilité qu'un mot de longueur ℓ soit le carré d'un autre est $2^{\frac{-\ell}{2}}$.

La somme sur tous les d majore la probabilité pour un cycle de n'être pas primitif.

Pour le cas non uniforme, voir [Chassaing,Zohoorian Azad,2010].

Récapitulatif

Génériquement :

- Le ppcm des a-cycles plus grands que $\log(n)$ dépasse $e^{\alpha \log^2(n)}$.
- Les a-cycles plus grands que $\log(n)$ sont primitifs.
- Les a-cycles plus grands que $\log(n)$ sont accessibles.

Théorème

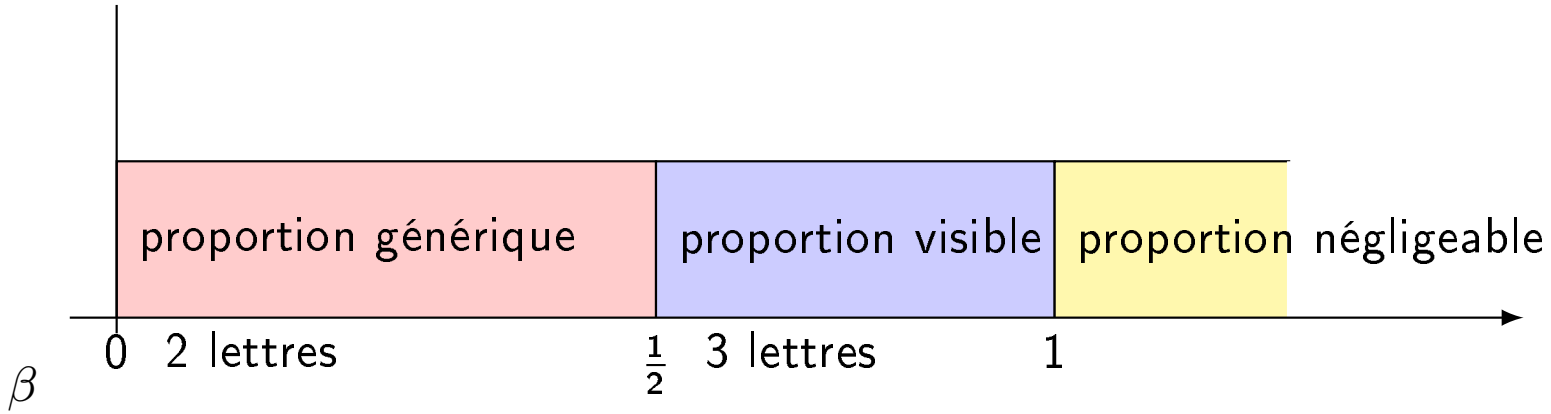
La trace du déterminisé du miroir sur les cycles plus grands que $\log(n)$ est supérieure à $e^{\alpha \log^2(n)}$ génériquement.

Corrollaire

L'automate minimal du langage miroir est génériquement de longueur supérieure à $e^{\alpha \log^2(n)}$.

Horizon

Si la probabilité d'être terminal dépend de n : $p_n = \Theta(\frac{1}{n^\beta})$.



Visible

Une propriété A_n est visible si il existe $\alpha > 0$ tel que $\mathbb{P}(A_n) > \alpha$ pour tout n à partir d'un certain rang.

Super-polynomial

Une fonction $f(n)$ est super-polynomiale si pour tout $q \in \mathbb{N}$ $n^q = O(f(n))$.