Trucs sur Kolakoski

Michaël Rao

CNRS - ENS Lyon
LIP - équipe MC2
michael.rao@ens-lyon.fr
http://perso.ens-lyon.fr/michael.rao/

La question de Kolakoski

Question de W. Kolakoski, (problème 5304, American Mathematical Monthly 1965)

Donnez une règle simple pour construire la séquence:

$$1221121221221121122121121221\ldots$$

Quel est le i-ème terme ? Est ce que la séquence est périodique ?

$$\underbrace{1}\,\underbrace{22}\,\underbrace{11}\,\underbrace{2}\,\underbrace{1}\,\underbrace{22}\,\underbrace{1}\,\underbrace{22}\,\underbrace{11}\,\underbrace{2}\,\underbrace{11}\,\underbrace{22}\,\underbrace{1}\,\underbrace{2}\,\underbrace{11}\,\underbrace{2}\,\underbrace{1}\,\underbrace{22}\ldots$$

$$\underbrace{1}_{1}\,\underbrace{22}_{2}\,\underbrace{11}_{2}\,\underbrace{2}_{1}\,\underbrace{1}_{1}\,\underbrace{22}_{2}\,\underbrace{1}_{1}\,\underbrace{22}_{2}\,\underbrace{11}_{2}\,\underbrace{2}_{1}\,\underbrace{11}_{2}\,\underbrace{22}_{2}\,\underbrace{1}_{1}\,\underbrace{2}_{1}\,\underbrace{11}_{2}\,\underbrace{2}_{1}\,\underbrace{1}_{1}\,\underbrace{22}_{2}\ldots$$

Définition

Soit w un mot (infini ou non).

- Un *bloc* est une occurrence d'un facteur $x \dots x$ qui n'est pas suivi, ni précédé par un x. Exemples: $abacccbda$, $aabd$.
- L'ensemble des blocs de w forme une décomposition de w:

$$w = B_1 B_2 \dots = x_1^{k_1} x_2^{k_2} \dots$$

  avec $k_i > 0$ et $x_i \neq x_{i+1}$.
- $\Delta(w) = k_1 k_2 \dots$.
- $\Delta(w)$ est un mot sur l'alphabet $\mathbb{N}^*$.

Exemples :
$\Delta(\epsilon) = \epsilon, \quad \Delta(1) = 1, \quad \Delta(abbccc) = 123, \quad \Delta((12)^\omega) = 1^\omega.$

Définition

Pour le mot de Kolakoski:

- On se place sur l'alphabet $\{1, 2\}$.
- Un mot est *dérivable* s'il ne contient pas 111 ni 222.
- Si un mot est dérivable, $\Delta(w) \in \{1, 2\}^*$.
- Δ a deux points fixes parmi les mots infinis:
  $\kappa = 1221121\ldots$ et $\kappa' = 1^{-1}\kappa = 221121\ldots$.

Oldenburger [1939] définit “le mot égal à sa propre trajectoire”.
Il s'agit du mot bi-infini $\widetilde{\kappa}1\kappa$.

Premières questions

Premières questions:

- κ est il périodique ? $\rightarrow$ NON
- κ est il image par un point fixe d'un morphisme ? $\rightarrow$ NON
- κ est il récurent ? $\rightarrow$ Conjecturé

Premières questions

Autres conjectures:

- La densité de 1 dans κ existe, et est $\frac{1}{2}$ [Keane 91].
- L'ensemble des facteurs de κ est fermé par complémentation et par miroir [Dekking 81].
- L'ensemble des facteurs de κ est exactement l'ensemble des mots lisses finis.

$\delta(w)$ est $\Delta(w)$ auquel on a enlevé la première et/ou la dernière lettre si elle est 1.

Ex: $\delta(1221) = 2 \quad \delta(2212212) = 2121 \quad \delta(21) = \epsilon$

Un mot fini est *lisse* (C^∞) si pour tout $k \geq 0$, $\delta^k(w)$ est dérivable.

Mesure de Dekking

Conjecture généralisée:

Conjecture (Dekking 95)

Pour tout mot fini $w \in C^\infty$, la fréquence d'apparition de w dans κ existe, et est:

$$p(w) = \frac{1}{s(w) \cdot 3^{h(w)}}$$

où:

- $s(\epsilon) = 1$
- $s(1) = s(2) = 2,$
- $s(w) = s(\delta(w)).$

- $h(1) = h(2) = h(\epsilon) = 0$
- $h(w) = 1 + h(\delta(w))$

Théorème (Chvátal)

La densité supérieure de 1 (et la densité supérieure de 2) est au plus 0.50084

Transducteur

Un *transducteur* est un n-uplet $(Q, \Sigma, \Sigma', \eta)$, où:

- Q est l'ensemble (fini) des *états*,
- Σ est un ensemble fini (l'*alphabet d'entrée*),
- Σ' est un ensemble fini (l'*alphabet de sortie*),
- $\eta : (Q \times \Sigma) \to (Q \times \Sigma'^*)$ est la *fonction de transition*.

On étend η à η^* sur les mots, où $\eta(q, uv) = (q'', wt)$ avec $\eta(q, u) = (q', w)$ et $\eta(q', v) = (q'', t)$.

Un *point fixe* d'un transducteur T avec l'état initial i est un mot infini w tel que $\lim_{k\to\infty} w_k$ converge vers w, où $\eta^*(i, w[1..k]) = (q_k, w_k)$ pour tout $k \geq 1$.

Kolakoski par un transducteur

κ et κ' sont les deux points fixes du transducteur suivant :

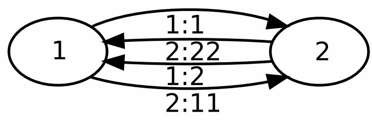


Figure: Le transducteur $\mathbb{T}$.

Produit de transducteurs

Soit $A = (Q, \Sigma, \Sigma', \eta)$ et $B = (Q', \Sigma', \Sigma'', \eta')$.

$B \cdot A = (Q \times Q', \Sigma, \Sigma'', \eta'')$, avec $\eta''((p,p'),x) = ((q,q'),w'')$ où $\eta'^*(p',w') = (q',w'')$ et $\eta(p,x) = (q,w')$.

Ce produit correspond à la composition de A et B: la sortie de $B \cdot A$ est la sortie de B quand il prend en entrée la sortie de A.

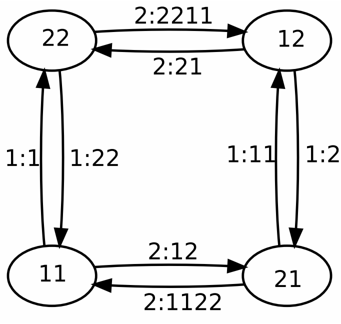


Figure: $\mathbb{T}^2$.

Notez que si w est un point fixe de T avec état initial i, et un point fixe de T' avec état initial i', alors w est un point fixe de $T' \cdot T$ avec état initial $i \times i'$.

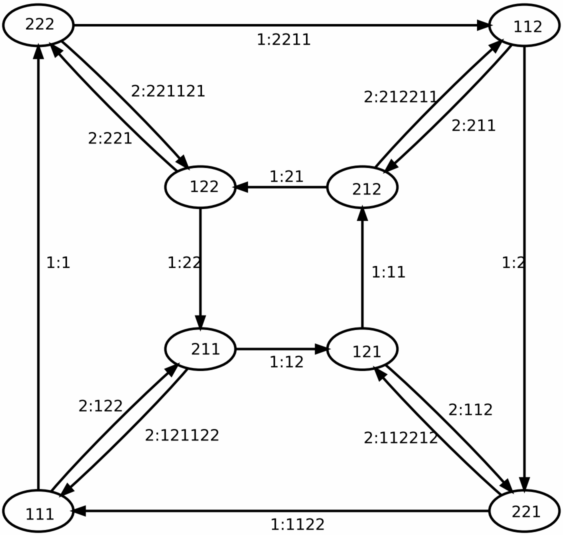
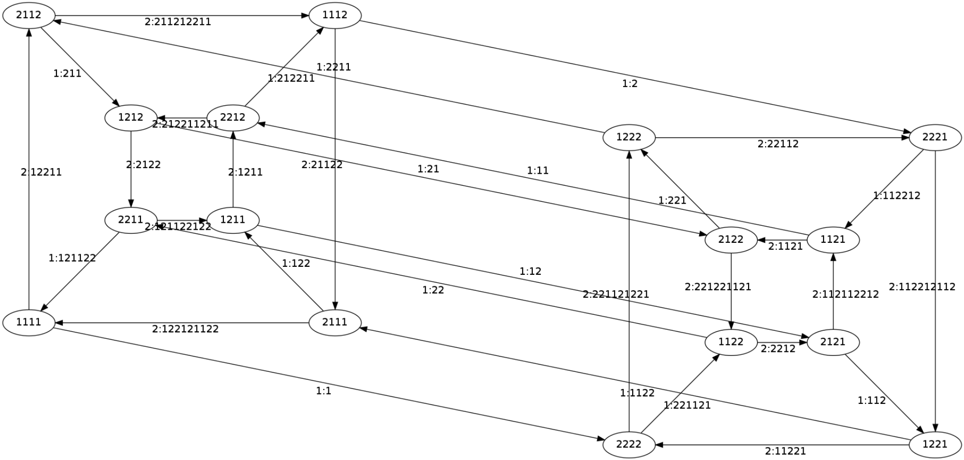


Figure: $\mathbb{T}^3$.


2112
1112
2:211212211
1:211
1:2211
1:212211
1:2
1212
2212
2:212211211
1222
2221
2:22112
2:2122
2:1211
2:21122
1:21
1:11
1:112212
2:12211
1:221
2211
1211
2:121122122
2122
1121
2:1121
1:121122
1:122
1:12
1:22
2:221121221
2:221221121
2:112112212
2:112212112
1111
2111
2:122121122
1122
2121
2:2212
1:1
1:1122
1:221121
1:112
2222
1221
2:11221

Autres points fixes des $\mathbb{T}^k$

- κ est également un point fixe de $\mathbb{T}^k$.
- Mais κ^k avec $k \geq 2$ a d'autres points fixes !
- $\mathbb{T}^2$ a 4 points fixes: κ, κ', $\kappa_2 = 112112212212112212121\ldots$ et $\kappa'_2 = 212212112212211211212211\ldots$ (qui sont les deux points limites de la séquence A111090).
- Presque toutes les questions ouvertes sur κ restent pertinentes sur les autres point fixes de $\mathbb{T}^n$.
- Ces autres points fixes sont des cas spéciaux de *mots lisses infinis*.

Conjectures sur $\mathbb{T}^k$

- $\mathbb{T}^k$ est “réversible”.
- $\mathbb{T}^k$ est fortement connexe.
- Il existe un chemin valide entre toute paire de sommets.
- Si un chemin valide infini passe par un état v, alors il repassera par v (“récurent”).
- Tout chemin infini valide passe une infinité de fois par tous les états de $\mathbb{T}^k$.
- Tout chemin infini valide passe une infinité de fois par toutes les transitions de $\mathbb{T}^k$.
- Il existe un entier k_n tel que tout chemin valide de taille k_n passe par tous les états de $\mathbb{T}^k$.

Cycle moyen (mean cycle)

Soit G un graphe orienté, avec deux pondérations sur les arcs: le *poids* w et la *taille* l.
Le cycle moyen maximum ("mean cycle") de G est:

$$m(G) = \max_C \frac{w(C)}{l(C)}$$

où le max est sur tous les cycles C de G.

- Ce paramètre est bien connu en optimisation.
- Il existe des algorithmes polynomiaux et efficaces pour le calculer [Karp, Howard...]

Notamment $m(G)$ est une borne supérieure asymptotique de $\frac{w(P)}{l(P)}$ pour tout chemin P dans G.

$m(\mathbb{T}^k)$

On prend $\mathbb{T}^k$ comme un graphe orienté.

- $l(a)$ est la taille de l'image de l'arc a.
- $w(a)$ est le nombre de 1 dans l'image de l'arc a.

$m(\mathbb{T}^k)$ est une borne supérieure sur la densité supérieure de 1 dans tout mot infini dérivable k fois.

Notamment, $m(\mathbb{T}^k)$ est une borne supérieure sur la densité supérieure de 1 dans κ.

n	mean cycle	decimal form
1	2/3	0.666666666
2	2/3	0.666666666
3	5/9	0.555555555
4	8/15	0.533333333
5	12/23	0.521739130
6	12/23	0.521739130
7	64/123	0.520325203
8	64/123	0.520325203
9	26/51	0.509803921
10	26/51	0.509803921
11	312/617	0.505672609

Table: Mean cycle of $\mathbb{T}^n$.

n	mean cycle	decimal form
12	386/765	0.504575163
13	249/494	0.504048582
14	10333/20535	0.503189676
15	590/1173	0.502983802
16	3304/6579	0.502203982
17	28157/56097	0.501934149
18	1475/2941	0.501530091
19	10118/20185	0.501263314
20	94877/189367	0.501021825
21	10456/20877	0.500838243
22	14573/29103	0.500738755

Table: Mean cycle of $\mathbb{T}^n$.

n	mean cycle	decimal form
23	923194/1844197	0.500594025
24	1029052/2056123	0.500481731
25	2126206/4249033	0.500397619
26	252577/504818	0.500332793
27	9857581/19704826	0.500262270
28	3232826/6462849	0.500216854
29	14006410/28002869	0.500177678
30	147516178/294946897	0.500144871
31	737731/1475104	0.500121347
32	3845003/7688497	0.500098133
33	13815783/27627163	0.500079686

Table: Mean cycle of $\mathbb{T}^n$.

Nouvelle sur la densité

Théorème

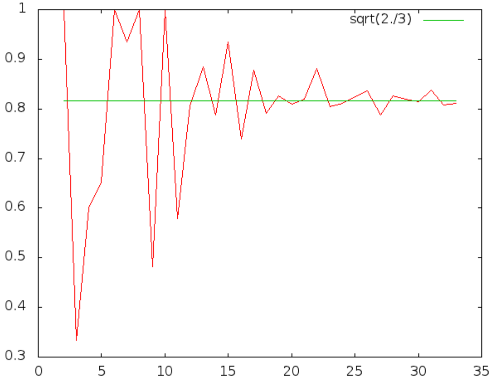
La densité supérieure de 1 et de 2 dans κ est au plus 0.50008.

Chvátal utilisait également le cycle moyen maximum dans une famille de graphes pour prouver la borne de 0.50084.

Il ne s'agit pas des mêmes graphes, mais il existe un morphisme $V(C_{n+1}) \to V(\mathbb{T}^n)$.

On a $m(C_{n+1}) = m(V(\mathbb{T}^n))$.

La différence entre le cycle moyen maximum et $\frac{1}{2}$ semble être en $O((2/3)^{n/2}) = O(0.816497^n)$.



Conjecture

$\lim_{k\to\infty} (m(\mathbb{T}^{k+1}) - \frac{1}{2})/(m(\mathbb{T}^k) - \frac{1}{2}) = \sqrt{2/3}$.

Algorithmes

La vue par transducteurs nous donne directement un algorithme en espace $O(log(n))$ pour calculer les n premières lettres de κ.

L'utilisation de $\mathbb{T}^k$ peut également servir à faire des grands sauts dans le calcul du nombre de 1 dans les x premières lettres de κ.

La taille moyenne des longueurs des arcs est $(3/2)^k$.

$\rightarrow$ On conjecture un gain de vitesse de 3/2 en utilisant le double de mémoire.

10^0	1
10^1	5
10^2	49
10^3	502
10^4	4996
10^5	49972
10^6	499986
10^7	5000046
10^8	50000675
10^9	500001223
10^{10}	4999997671
10^{11}	50000001587
10^{12}	500000050701
10^{13}	5000000008159
10^{14}	50000000316237
10^{15}	500000000977421
10^{16}	4999999994637728
10^{17}	49999999977479348
10^{18}	499999999944465105

Table: Nombre de 1 dans les 10^x premières lettres de κ.

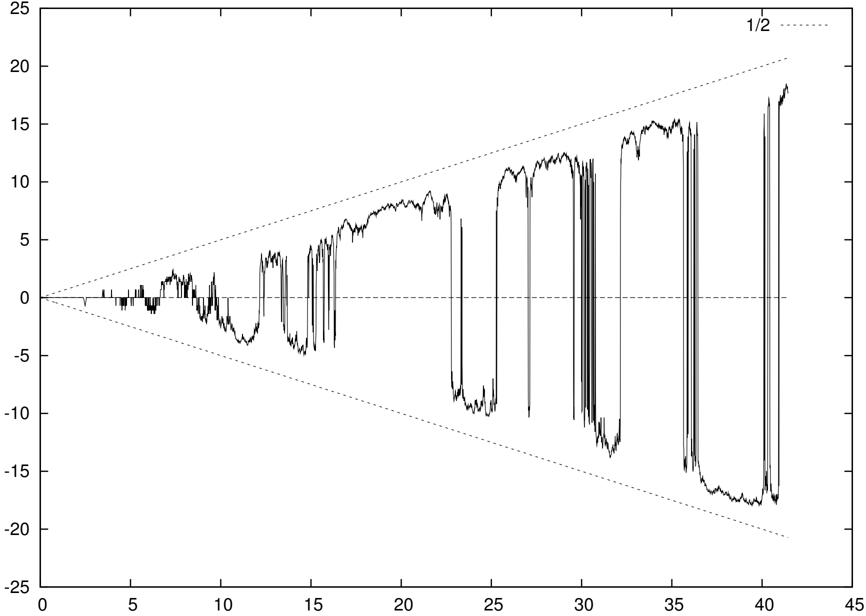


Figure: Différence en échelle logarithmique.

Différence en $O(\sqrt{n})$

Le graphique précédent appuie la conjecture que la différence entre le nombre de 1 et de 2 dans un préfixe de taille n de κ est en $O(\sqrt{n})$.

Soit

$$f(n) = \max_{w \in C^\infty, |w|=n} \big| |w|_1 - |w|_2 \big|$$

(*càd* la différence maximum entre le nombre de 1 et de 2 dans un mot lisse de taille n).

Conjecture (Carpi)

$f(n) = O(\sqrt{n})$.

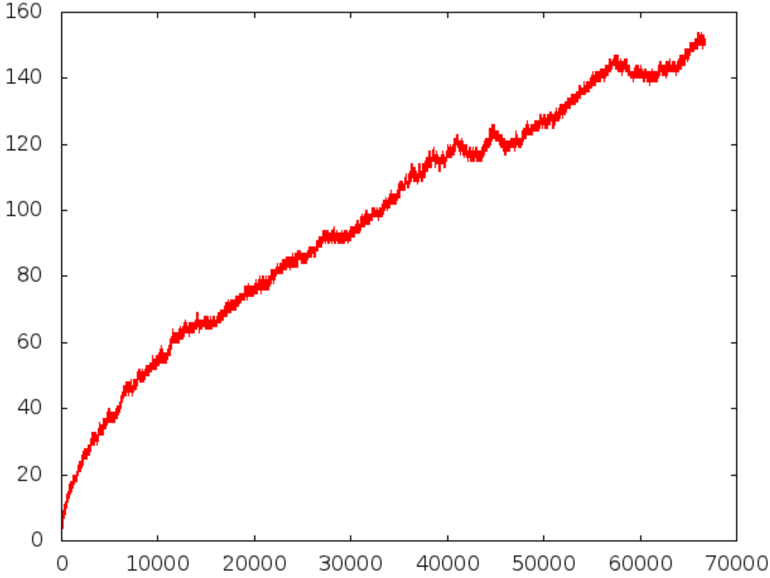

160
140
120
100
80
60
40
20
0
0
10000
20000
30000
40000
50000
60000
70000

Figure: $f(n)$