Walnut ?

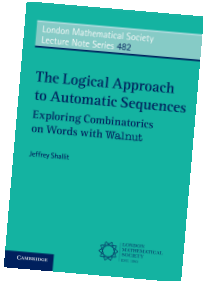
motivation, introduction

Nicolas Ollinger (LIFO, Université d'Orléans)

Squirrel Club, 28 février 2023

Squirrel club : le programme

Groupe de travail autour de l'outil **Walnut**, en s'appuyant sur le **Walnut book**.



1. pré-requis : chapitres 1 à 6 (mots, automates, logique, . . .) ;
2. l'outil Walnut : session TP sur chapitres 7 à 9 ;
3. sous le capot : les entrailles de Walnut ;
4. Walnutiser PresTAF ? motivation, discussions.

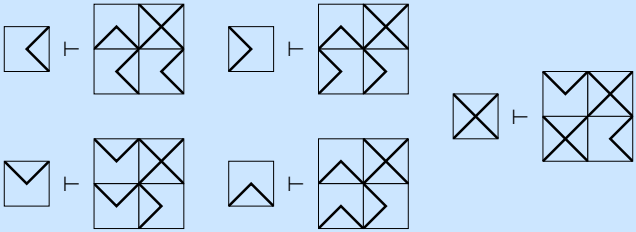
Aujourd'hui

13h30 – 13h50. Introduction

Pourquoi s'intéresser à Walnut ? Que peut-on faire avec ? Quels sont les théorèmes sous-jacents ?

14h – 17h. Pré-requis

Principales définitions et théorèmes des 6 premiers chapitres du livre de Walnut : mots, automates finis, logique du premier ordre, suites automatiques, arithmétiques de Presburger et de Büchi.



1. Motivation

Scenario typique

Un objet k-automatique, c'est-à-dire un coloriage C de $\mathbb{N}^d$ défini par un **automate fini**, qui associe une couleur $c \in \Sigma$ à tout point $z \in \mathbb{N}^d$ donné par ses coordonnées exprimées en **base k**.

Une propriété de cet objet, exprimable au **premier ordre**. Les variables parcourent des positions sur $\mathbb{N}$, des prédicats permettent d'obtenir la couleur en tout point, l'arithmétique est restreinte à l'addition et à l'ordre $\langle \mathbb{N}, +, <, C \rangle$.

Vérifier la validité de la propriété pour cet objet *(si elle est close)*.

Identifier l'ensemble des valeurs qui satisfont la propriété *(si elle possède des variables libres)*.

Thue-Morse : sans recouvrement

Mot de Thue-Morse

$$\begin{aligned} T &: a \mapsto ab, \quad b \mapsto ba \\ \mathcal{T} &= abbabaabbaababbabaababbaabbabaab\ldots \end{aligned}$$

Propriété Ce mot est sans recouvrement, c'est-à-dire qu'il n'admet aucun facteur de la forme $xuxux$ avec x non vide.

$$\neg\exists i, m, n \quad m > 0 \wedge \mathcal{T}[i, i+2m+n] = \mathcal{T}[i+m+n, i+3m+2n]$$

Thue-Morse : synchronisation

Mot de Thue-Morse

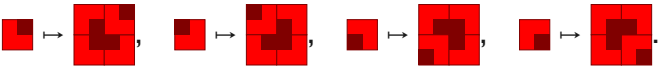
$$\begin{aligned} T &: a \mapsto ab, \qquad b \mapsto ba \\ \mathcal{T} &= abbabaabbaababbabaababbaabbabaab\ldots \end{aligned}$$

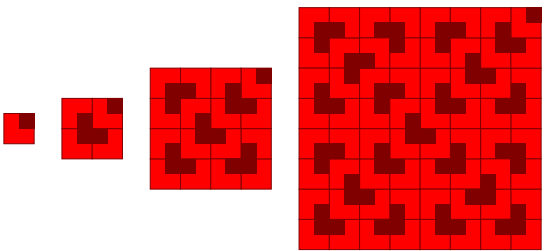
Propriété Il existe un taille de facteur k telle que

$$\forall i, j \quad \mathcal{T}[i, i+k] = \mathcal{T}[j, j+k] \longrightarrow i = j \pmod 2$$

Apériodicité

$\Sigma =$ 

$s :$ 



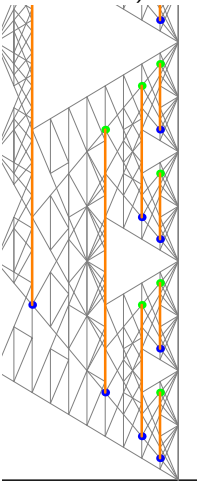
Propriété Le coloriage engendré est apériodique

$$\forall i, j \quad i + j > 0 \longrightarrow \exists x, y \quad C[x][y] \neq C[x + i][y + j]$$

Computation windows

For each state of a signal, the set of positions that contains it in this space-time diagram is 2-recognizable (computable by a finite automaton that takes as an input the couple of coordinates).

- Using signals at speed 2 and 3, some 2×1 pattern characterizes the positions $(3.2^h - 1[14.2^h], 2^h), h \in \mathbb{N}$;
- Using signals at speed 5 and 6, some 2×1 pattern characterizes the positions $(12.2^h - 1[14.2^h], 2^h), h \in \mathbb{N}$.



5.4 2-recognizability

Our construction relies on the ability to identify some specific sets of spacetime positions. We achieve this goal by considering products of independent P-signals. To avoid boring calculations, we rely on the following lemma to assert some regularity of P-signals and rely on the theory of p-recognizable sets of tuples of integers [5] to characterize these positions.

Lemma 9. *The spacetime sequence $\Delta : \mathbb{N}^2 \to \mathbb{Z}_{n+1}$ of the (n,k)-embedding $\mathcal{F}$ of the XOR, where $\Delta(x,y) = \mathcal{F}^y(^\omega 0.10^\omega)_{-x}$, is 2-recognizable.*

Proof. The spacetime sequence Δ of the (n,k)-embedding $\mathcal{F}$ is generated by the 2-substitution $s : \mathbb{Z}_{n+1} \to \mathbb{Z}_{n+1}^2$ uniquely defined, as per Fig. 2, by

$$s(0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \qquad s\begin{pmatrix} n \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} n & \cdots & (k+1) & k & \cdots & 1 & n & \cdots & (k+1) & \cdots & 1 \\ 0 & \cdots & 0 & n & \cdots & \cdots & \cdots & 1 & 0 & \cdots & 0 \end{pmatrix}^{\intercal}$$

Indeed, the substitution rule is compatible with the local rule of the OCA. ∎

5.5 Computation windows

Let $\mathcal{F}$ denote the OCA we are constructing. We use 4 P-signals of speeds 2, 3, 5 and 6 as foundations of $\mathcal{F}$ — they allow us to build computation windows. Let c be the initial configuration which is null everywhere except for $c(0) = (1, 1, 1, 1)$. We use Lemma 9 twice for each of the following lemmas, that is, we have a characterization of the set of positions where some specific state pairs appear. Both lemmas are illustrated in Fig. 3, 4 and 5.

Lemma 10.

$$\forall m \leq 0, p \geq 0, \begin{cases} \mathcal{F}^p(c)_m = (1, 0, \_, \_) \\ \mathcal{F}^p(c)_{m+1} = (0, 3, \_, \_) \end{cases} \Leftrightarrow \begin{cases} \exists h \in \mathbb{N}^*, m = -8^h \\ p \equiv -3m - 1[-14m] \end{cases}$$

Lemma 11.

$$\forall m \leq 0, p \geq 0, \begin{cases} \mathcal{F}^p(c)_m = (\_, \_, 7, 6) \\ \mathcal{F}^p(c)_{m+1} = (\_, \_, 5, 0) \end{cases} \Leftrightarrow \begin{cases} \exists h \in \mathbb{N}^*, m = -8^h \\ p \equiv -12m - 1[-14m] \end{cases}$$

Hence we add a fifth layer using alphabet $\{0, 1\}$ with the rule:

$$\begin{aligned} \delta_5\left((1, 0, \_, \_, x), (0, 3, \_, \_, \_)\right) &= 1 - x \\ \delta_5\left((\_, \_, 7, 6, x), (\_, \_, 5, 0, \_)\right) &= 1 - x \\ \delta_5\left((\_, \_, \_, \_, x), (\_, \_, \_, \_, \_)\right) &= x \quad \text{otherwise.} \end{aligned}$$

Actually, the same arguments work for columns $-2 \cdot 8^h, h \in \mathbb{N}^*$ and $-4 \cdot 8^h, h \in \mathbb{N}^*$, hence we use them all. We therefore call computation windows the vertical segments in the spacetime diagram where the fifth layer contains 1:

$$(m, p) \text{ in a computation window} \Leftrightarrow \begin{array}{l} \exists h \in \mathbb{N}^*, m = -2^h,\ p' \equiv p[14 \cdot 2^h] \\ \text{and } 3 \cdot 2^h \leq p' < 12 \cdot 2^h. \end{array}$$

Walnut : la calculatrice qu'il vous faut !

$$\neg\exists i,m,n \quad m > 0 \wedge \mathcal{T}[i, i+2m+n] = \mathcal{T}[i+m+n, i+3m+2n]$$

```
[Walnut]$ eval nooverlap
          "~ Ei,m,n (m>0 & Ak (k < 2*m+n => T[i+k] = T[i+m+n+k]))":
computed ~:1 states - 9ms
m>0:2 states - 19ms
 k<((2*m)+n):4 states - 2ms
  T[(i+k)]=T[(((i+m)+n)+k)]:20 states - 11ms
   (k<((2*m)+n)=>T[(i+k)]=T[(((i+m)+n)+k)]):73 states - 8ms
    (A k (k<((2*m)+n)=>T[(i+k)]=T[(((i+m)+n)+k)])):5 states - 162ms
     (m>0&(A k (k<((2*m)+n)=>T[(i+k)]=T[(((i+m)+n)+k)]))):1 states - 1m
      (E i , m , n (m>0&(A k (k<((2*m)+n)=>T[(i+k)]=T[(((i+m)+n)+k)]))
       ~(E i , m , n (m>0&(A k (k<((2*m)+n)=>T[(i+k)]=T[(((i+m)+n)+k)])
Total computation time: 215ms.

____
TRUE
```

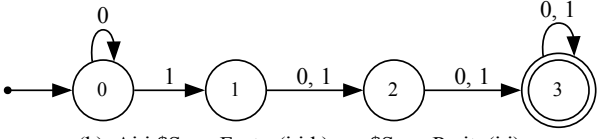
Walnut : la calculatrice qu'il vous faut !

$$\forall i, j \quad \mathcal{T}[i, i+k] = \mathcal{T}[j, j+k] \longrightarrow i = j \pmod 2$$

```
[Walnut]$ def SameFactor "Ak k<n => T[i+k] = T[j+k]";

[Walnut]$ def SameParity "Ea,b,c i = 2*a+c & j = 2*b+c";

[Walnut]$ eval syncT "Ai,j $SameFactor(i,j,k) => $SameParity(i,j)":
(SameFactor(i,j,k))=>SameParity(i,j))):19 states - 1ms
 (A i , j (SameFactor(i,j,k))=>SameParity(i,j)))):4 states - 0ms
Total computation time: 2ms.
```



(k): Ai,j $SameFactor(i,j,k) => $SameParity(i,j)

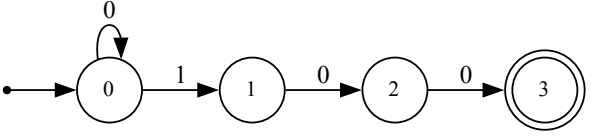
Walnut : la calculatrice qu'il vous faut !

```
[Walnut]$ def SameFactor "Ak k<n => T[i+k] = T[j+k]";

[Walnut]$ def SameParity "Ea,b,c i = 2*a+c & j = 2*b+c";

[Walnut]$ def syncT "Ai,j $SameFactor(i,j,k) => $SameParity(i,j)";

[Walnut]$ eval smallest "$syncT(n) & ~ $syncT(n-1)":
computed ~:1 states - 15ms
computed ~:2 states - 2ms
~syncT((n-1))):4 states - 0ms
 (syncT(n))&~syncT((n-1)))):4 states - 1ms
Total computation time: 22ms.
```



(n): $syncT(n) & ~ $syncT(n-1)

LOGIC AND p-RECOGNIZABLE SETS OF INTEGERS

Véronique Bruyère* Georges Hansel Christian Michaux*†
Roger Villemaire‡

Abstract

We survey the properties of sets of integers recognizable by automata when they are written in p-ary expansions. We focus on Cobham's theorem which characterizes the sets recognizable in different bases p and on its generalization to $\mathbb{N}^m$ due to Semenov. We detail the remarkable proof recently given by Muchnik for the theorem of Cobham-Semenov, the original proof being published in Russian.

2. Théorèmes sous-jacents

WEAK SECOND-ORDER ARITHMETIC AND FINITE AUTOMATA[1])

By J. Richard Büchi in Ann Arbor, Michigan

Büchi 1960

Introduction

The formalism of regular expressions was introduced by S. C. Kleene [6] to obtain the following basic theorems.

Synthesis. *To every regular expression $\mathfrak{E}$ one can effectively obtain a finite automata $\mathfrak{A}$ with binary output $\mathfrak{U}$ such that $\mathfrak{E}$ denotes the behavior of $\langle \mathfrak{A}, \mathfrak{U} \rangle$.*

Analysis. *To every finite automaton $\mathfrak{A}$ with binary output $\mathfrak{U}$ one can effectively construct a regular expression $\mathfrak{E}$ such that the behavior of $\langle \mathfrak{A}, \mathfrak{U} \rangle$ is denoted by $\mathfrak{E}$.*

For simplified expositions of Kleene's theory see Copi-Elgot-Wright [4], Rabin-Scott [13], and Myhill [8].

It will be shown here that a more conventional formalism, a weak second-order arithmetic, can be used in place of the formalism of regular expressions. Our theorems 1, 2 section 4 are equivalent to Kleene's synthesis and analysis theorems. This result is of interest for automata theory because formulas of weak second-order arithmetic seem to be more convenient than regular expressions for formalizing conditions on the behavior of automata. In addition, our synthesis and analysis theorems yield rather complete information on the strength of weak second-order arithmetic (see Section 5), thus providing an example of applying automata theory to logic.

Automates et logique

Les travaux de **Büchi** *et al* dans les années 1960 ont initié l'exploration des liens entre **logique**, **automates finis**, **arithmétique**...

L'état de l'art **logic and p-recognizable sets of integers** de **Bruyère** *et al* synthétise de manière très lisible l'ensemble des résultats classiques sur la caractérisation des ensembles de nombres reconnaissables par automates en base p.

Quatre modes de reconnaissance

Soit $p \geqslant 2$ un entier et $s : \mathbb{N} \to A$ un coloriage de $\mathbb{N}$ à l'aide d'un alphabet fini A.

p-substition Une **p-substitution** est une application $f : B \to B^p$. Elle peut être étendue en un morphisme f sur B^*. Si $f(b)$ commence par b alors la suite $(f^n(b))$ converge vers un point fixe $f^\omega(b)$. Soit $g : B \to A$, le point fixe engendre le coloriage $g(f^\omega(b))$ qui est dit **généré par p-substitution**.

p-automate Un coloriage s est **p-reconnaissable** s'il existe un automate fini à sortie, c'est-à-dire un automate fini muni d'une fonction de sortie $h : Q \to A$ et qui, pour toute écriture $u \in \{0, \ldots, p-1\}^*$ d'un entier i en base p atteint un état $q \in Q$ tel que $h(q) = s(i)$.

p-définissable Soit V_p la fonction qui à $x \neq 0$ associe la plus grande puissance de p qui divise x. Un coloriage s est **p-définissable** si pour toute lettre $a \in A$, il existe une formule du premier ordre φ_a de $\langle \mathbb{N}, +, V_p \rangle$ telle que

$$s^{-1}(a) = \left\{ n \in \mathbb{N} \,\middle|\, \langle \mathbb{N}, +, V_p \rangle \vDash \varphi_a(n) \right\}$$

p-algébrique Dans le cas où p est premier, on identifie tout coloriage s à une série formelle sur un corps K de caractéristique p :

$$S(x) = \sum_{n \geqslant 0} s_n x^n \quad \in K[[x]]$$

Un coloriage s est **p-algébrique** si $S(x)$ est algébrique sur $K[[x]]$.

Théorème d'équivalence

Théorème 4.1 Soit $p \geqslant 2$ un entier et $s : \mathbb{N} \to A$ un coloriage de $\mathbb{N}$. Les affirmations suivantes sont équivalentes :

(1) s est engendré par une **p-substitution** ;

(2) s est **p-reconnaissable** ;

(3) s est **p-définissable** ;

(4) s est **p-algébrique**. (uniquement si p premier)

Le théorème se **généralise** aux coloriages en **dimensions supérieures** $s : \mathbb{N}^d \to A$. C'est le **théorème 5.1**.

Pour Walnut, c'est essentiellement $(2) \Leftrightarrow (3)$ qui est exploité *et* $(1) \Rightarrow (2)$

Bonus : théorème de Cobham-Semenov

Théorème Soient $p, q \geqslant 2$ deux entiers multiplicativement indépendants et $s : \mathbb{N}^d \to A$ un coloriage. Si s est **p-reconnaissable** et **q-reconnaissable** alors s est **définissable** dans $\langle \mathbb{N}, + \rangle$.

Autrement dit, les ensembles d'entiers reconnaissables en toute base sont exactement les ensembles définissables dans l'**arithmétique de Presburger**.

Les ensembles d'entiers reconnaissables en base p sont les ensembles définissables dans l'**arithmétique de Büchi en base p**.

On peut généraliser le résultat à d'autres systèmes de numération plus exotiques...