

Equations profinies

Jean-Éric Pin¹

¹LIAFA, CNRS et Université Paris Diderot

Mai 2011, Réunion FREC

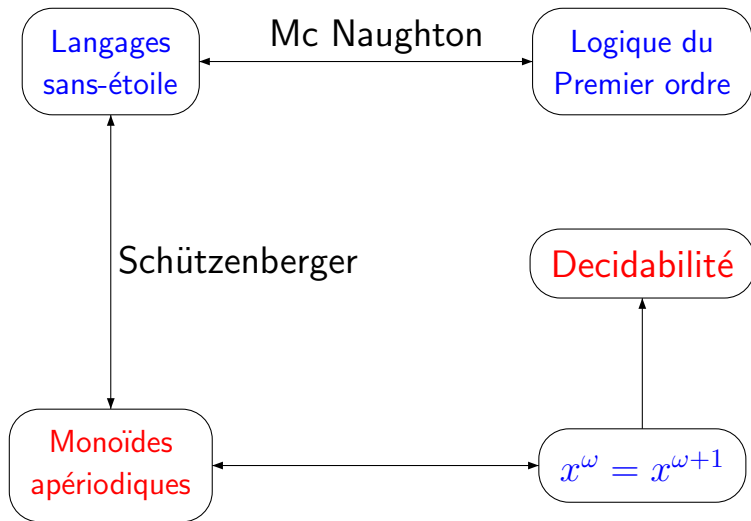


Sommaire

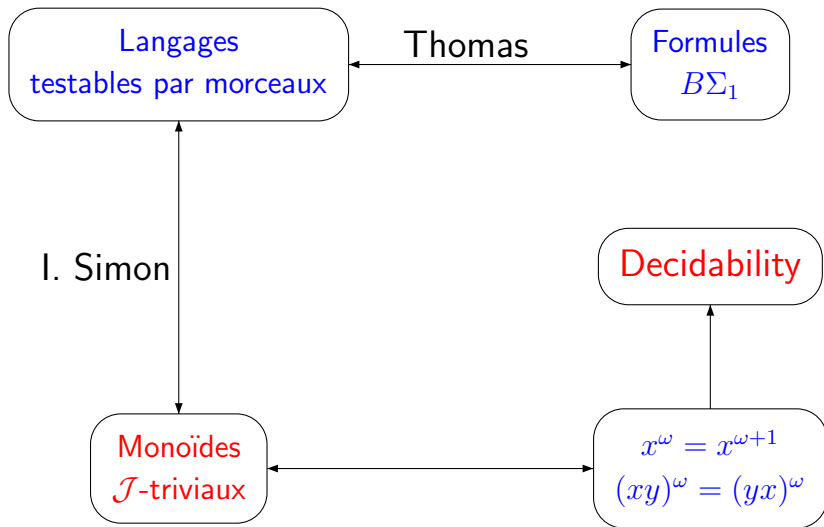
- (1) Treillis de langages
- (2) Le monde profini
- (3) Équations profinies



Un cercle vertueux



Un autre cercle vertueux



Première partie I

Treillis de langages



Parties reconnaissables

Un langage L de A^* est reconnu par un monoïde M s'il existe un morphisme surjectif de monoïdes $\varphi : A^* \rightarrow M$ qui **sature** L : si $u \in L$ et $\varphi(u) = \varphi(v)$, alors $v \in L$ [ou encore $L = \varphi^{-1}\varphi(L)$].

Un langage est **reconnaisable** s'il est reconnu par un monoïde fini. Un langage est reconnaissable ssi il est **rationnel** (Kleene).

Le **monoïde syntactique** d'un langage reconnaissable L de A^* est le **plus petit** monoïde qui le reconnaît.

Treillis de langages

On fixe un alphabet fini A . Un **treillis de langages** est un ensemble de langages reconnaissables de A^* contenant \emptyset et A^* et fermé par **union finie** et **intersection finie**.

Définition

Soient u et v des mots de A^* . Un langage L de A^* **satisfait l'équation** $u \rightarrow v$ si la condition $u \in L$ entraîne $v \in L$.

Treillis finis de langages

Soit E un ensemble d'équations de la forme $u \rightarrow v$. Alors l'ensemble des langages de A^* vérifiant les équations de E est un treillis de langages.

Proposition

Soient L, L_1, \dots, L_n des langages. Si L satisfait toutes les équations satisfaites par L_1, \dots, L_n , alors L appartient au treillis de langages engendré par L_1, \dots, L_n .

Corollaire

Un ensemble fini de langages de A^ est un **treillis de langages** ssi il peut être défini par un ensemble d'équations de la forme $u \rightarrow v$ avec $u, v \in A^*$.*

Il y a donc une théorie équationnelle pour les **treillis finis** de langages. Que peut-on dire des treillis infinis ?

Corollaire

Un ensemble fini de langages de A^ est un **treillis de langages** ssi il peut être défini par un ensemble d'équations de la forme $u \rightarrow v$ avec $u, v \in A^*$.*

Il y a donc une théorie équationnelle pour les **treillis finis** de langages. Que peut-on dire des treillis infinis ?

Il faut se plonger dans le monde profini...

Deuxième partie II

Le monde profini



Séparer deux mots

Un monoïde M sépare deux mots u et v de A^* s'il existe un morphisme de monoïde $\varphi : A^* \rightarrow M$ tel que $\varphi(u) \neq \varphi(v)$.

Par exemple, le morphisme qui envoie chaque mot sur sa longueur modulo 2 est un morphisme de $\{a, b\}^*$ sur $\mathbb{Z}/2\mathbb{Z}$ qui sépare $abaaba$ et $abaabab$.

Soit $M = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \right\}$ et soit $\varphi : \{a, b\}^* \rightarrow M$ défini par $\varphi(a) = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$ and $\varphi(b) = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$. Alors, pour chaque mot u , φ sépare ua et ub car $\varphi(ua) = \varphi(a)$ et $\varphi(ub) = \varphi(b)$.

La distance profinie

Soient u et v deux mots. Posons

$$r(u, v) = \min\{|M| \mid M \text{ est un monoïde fini} \\ \text{qui sépare } u \text{ et } v\}$$

$$d(u, v) = 2^{-r(u, v)}$$

Dans ce cas, d est une distance **ultramétrique**. Pour tout $x, y, z \in A^*$,

- (1) $d(x, x) = 0$,
- (2) $d(x, y) = d(y, x)$,
- (3) $d(x, z) \leq \max\{d(x, y), d(y, z)\}$

Propriétés de d

Intuitivement, deux mots sont proches pour d s'il faut un monoïde **de grande taille** pour les séparer.

Une suite de mots u_n est une **suite de Cauchy** ssi, pour tout morphisme φ de A^* vers un monoïde fini, la suite $\varphi(u_n)$ est ultimement constante.

Une suite de mots u_n **converge** vers un mot u ssi, pour tout morphisme φ de A^* sur un monoïde fini, la suite $\varphi(u_n)$ est ultimement égale à $\varphi(u)$.

Le monoïde profini libre

La complétion de l'espace métrique (A^*, d) est le **monoïde profini libre** sur A , noté $\widehat{A^*}$. Ses éléments sont appelés des **mots profinis**.

Le produit est **uniformément continu** sur A^* et s'étend donc par continuité à $\widehat{A^*}$. Si A est **fini**, $\widehat{A^*}$ est **compact**.

Tout morphisme $\varphi : A^* \rightarrow M$, où M est **fini** (discret) est **uniformément continu**. Puisque A^* est **dense** dans $\widehat{A^*}$, φ s'étend de façon unique en un morphisme uniformément continu $\widehat{\varphi} : \widehat{A^*} \rightarrow M$.

Le monoïde profini libre comme limite projective

Le monoïde profini peut être défini comme la **limite projective** du **système dirigé** formé par les morphismes surjectifs entre monoïdes finis.

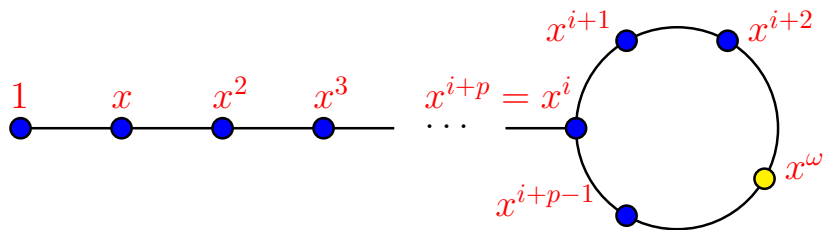
En particulier, un mot profini u est complètement déterminé par ses images $\widehat{\varphi}(u)$, où φ décrit la classe des morphismes de A^* sur un monoïde fini.

$$\text{Mot profini} = \{\widehat{\varphi}(u)\}_{\varphi:A^* \rightarrow M}$$



Un mot profini non fini

Pour chaque $u \in \widehat{A}^*$, la suite $u^{n!}$ est une suite de Cauchy et donc converge dans \widehat{A}^* vers une limite, notée u^ω . Si φ est un morphisme de A^* sur un monoïde fini M , $\widehat{\varphi}(u^\omega)$ est l'unique idempotent du sous-semigroupe de M engendré par $x = \varphi(u)$.



Une autre définition de x^ω

Dans un semigroupe **compact**, le plus petit sous-semigroupe **fermé** contenant un élément donné x contient exactement un **idempotent**, noté x^ω .

Troisième partie III

Théorie équationnelle des langages



Soit L un langage reconnaissable de A^* et soit $\eta : A^* \rightarrow M$ son morphisme syntactique. On sait que η se prolonge en un morphisme uniformément continu de $\widehat{A^*}$ vers M . On note \overline{L} la fermeture topologique de L pour la distance profinie.

Définition

Soient u et v des mots profinis de $\widehat{A^*}$. Alors L satisfait l'équation $u \rightarrow v$ si la condition $u \in \overline{L}$ entraîne $v \in \overline{L}$.

Une définition équivalente

Soit L un langage reconnaissable de A^* et soit $\eta : A^* \rightarrow M$ son monoïde syntactique.

Définition

Soient u et v des mots profinis de $\widehat{A^*}$. Alors L satisfait l'équation $u \rightarrow v$ si la condition $\widehat{\eta}(u) \in \eta(L)$ entraîne $\widehat{\eta}(v) \in \eta(L)$.

Caractérisation équationnelle des treillis

Soit E un ensemble d'équations de la forme $u \rightarrow v$.
Le sous-ensemble de $\text{Rec}(A^*)$ défini par E est
l'ensemble de tous les langages reconnaissables de
 A^* satisfaisant toutes les équations de E .

Théorème

Un ensemble de langages reconnaissables de A^ est un treillis de langages ssi il peut être défini par un ensemble d'équations de la forme $u \rightarrow v$, où $u, v \in \widehat{A}^*$.*

Treillis de langages fermés par quotient

Soient $x, y \in A^*$. Le **quotient** de L par x et y est le langage $x^{-1}Ly^{-1} = \{u \in A^* \mid xuy \in L\}$.

Un treillis de langages \mathcal{L} est **fermé par quotient** si pour tout $x, y \in A^*$, $L \in \mathcal{L}$ entraîne $x^{-1}Ly^{-1} \in \mathcal{L}$.

Treillis de langages fermés par quotient

On dit que L satisfait l'équation $u \leq v$ si, pour tout $x, y \in \widehat{A}^*$, il satisfait l'équation $xvy \rightarrow xuy$.

Proposition

Soit $L \in \text{Rec}(A^*)$, soit (M, \leq_L) son monoïde syntactique ordonné et soit $\eta : A^* \rightarrow M$ son morphisme syntactique. Alors L satisfait l'équation $u \leq v$ ssi $\widehat{\eta}(u) \leq_L \widehat{\eta}(v)$.

Treillis de langages fermés par quotient (2)

On dit que L satisfait l'équation $u \leq v$ si, pour tout $x, y \in \widehat{A}^*$, il satisfait l'équation $xvy \rightarrow xuy$.

Théorème

Un ensemble de langages reconnaissables de A^ est un treillis de langages fermé par quotient ssi il peut être défini par un ensemble d'équations de la forme $u \leq v$, où $u, v \in \widehat{A}^*$.*

Algèbres de Boole fermées par quotient

On note $u = v$ si $u \leq v$ et $v \leq u$.

Théorème

Un ensemble de langages reconnaissables de A^ est une algèbre de Boole fermée par quotient ssi il peut être défini par un ensemble d'équations de la forme $u = v$, où $u, v \in \widehat{A^*}$.*

On retrouve ainsi une conséquence des théorèmes d'Eilenberg et de Reiterman.

Fermeture par inverse de morphismes

Une classe de langages \mathcal{K} associe à chaque alphabet A un treillis $\mathcal{K}(A^*)$ de langages de A^* .

On dit que \mathcal{K} est fermée par inverse de morphismes si pour tout morphisme $\varphi : A^* \rightarrow B^*$, $L \in \mathcal{K}(B^*)$ entraîne $\varphi^{-1}(L) \in \mathcal{K}(A^*)$.

Une classe d'équations profinies E associe à chaque alphabet A un ensemble $E(A)$ d'équations profinies sur $\widehat{A^*}$.

Comment retrouver Eilenberg-Reiterman ?

Une classe E d'équations profinies est **fermée par morphisme** si pour tout morphisme $\varphi : A^* \rightarrow B^*$, $u \rightarrow v \in E(A)$ entraîne $\varphi(u) \rightarrow \varphi(v) \in E(B)$. On parle alors d'**identités** au lieu d'équations.

Théorème

*Chaque classe de langages reconnaissables est définie par une classe d'équations profinies. Elle est définie par une classe d'identités profinies ssi elle est fermée par **inverse de morphismes**.*

\mathcal{C} -morphismes

Soit \mathcal{C} une classe de morphismes fermée par composition et contenant les morphismes lettre-à-lettre.

Exemples de telles classes \mathcal{C} de morphismes :

- Tous
- Les lettre-à-lettre ($\varphi(A) \subseteq B$)
- Les non-effaçants ($\varphi(A) \subseteq B^+$)
- Les alphabétiques ($\varphi(A) \subseteq B \cup \{1\}$)
- Les uniformes ($\varphi(A) \subseteq B^k$)

Fermeture par inverse de \mathcal{C} -morphisms

Une classe de langages \mathcal{K} est fermée par inverse de \mathcal{C} -morphisms si pour tout \mathcal{C} -morphisme $\varphi : A^* \rightarrow B^*$, on a $L \in \mathcal{K}(B^*)$ entraîne $\varphi^{-1}(L) \in \mathcal{K}(A^*)$.

Comment retrouver Esik-Straubing-Kunc ?

Une classe E d'équations profinies est **fermée par \mathcal{C} -morphisme** si, pour tout \mathcal{C} -morphisme $\varphi : A^* \rightarrow B^*$, $u \rightarrow v \in E(A)$ entraîne $\varphi(u) \rightarrow \varphi(v) \in E(B)$. On parle alors de **\mathcal{C} -identités** au lieu d'équations.

Théorème

*Une classe de langages reconnaissables est définie par une classe de **\mathcal{C} -identités profinies** ssi elle est fermée par **inverse de \mathcal{C} -morphisme**.*

Interprétation des équations (Notation : $P = \eta(L)$)

Fermé par		Interprétation
\cup, \cap	$u \rightarrow v$	$\hat{\eta}(u) \in P \Rightarrow \hat{\eta}(v) \in P$
quotient	$u \leq v$	$xvy \rightarrow xuy$
complément	$u \leftrightarrow v$	$u \rightarrow v$ et $v \rightarrow u$
quotient et L^c	$u = v$	$xuy \leftrightarrow xvy$
φ^{-1}		Variables = mots
φ^{-1} lettre à lettre		Variables = lettres
φ^{-1} uniforme		Variables = mots de même longueur

Langages avec zéro

Un **langage avec zéro** est un langage dont le monoïde syntactique possède un zéro.

Proposition

*La classe des langages reconnaissables **avec zéro** est fermée pour les **opérations booléennes** et par **quotients**.*

Par conséquent, il peut être défini par un ensemble d'équations de la forme $u = v \dots$



Un mot profini

Fixons un ordre total sur l'alphabet A . Soit u_0, u_1, \dots la suite ordonnée de tous les mots de A^+ dans l'ordre des mots croisés (longueur d'abord).

$1, a, b, aa, ab, ba, bb, aaa, aab, aba, abb, baa, bab, \dots$

Reilly et Zhang (voir aussi Almeida-Volkov) ont prouvé que la suite $(v_n)_{n \geq 0}$ définie par

$$v_0 = u_0, \quad v_{n+1} = (v_n u_{n+1} v_n)^{(n+1)!}$$

converge vers un idempotent ρ_A de l'idéal minimal du monoïde profini libre sur A .



Proposition

*Un langage reconnaissable possède un **zéro** ssi il satisfait l'équation $x\rho_A = \rho_A = \rho_A x$ pour tout $x \in A^*$.*

Langages non denses

Un langage L de A^* est **dense** si, pour tout mot $u \in A^*$, $L \cap A^*uA^* \neq \emptyset$.

Les langages **non-denses** ou **pleins** sont fermés par **union** finie, **intersection** finie et **quotients**. Ils **ne sont pas fermés** par **inverse de morphismes**, même pour les morphismes strictement alphabétiques.

Théorème

Un langage de A^ est **non-dense** ou **plein** ssi il satisfait l'équation $x \leq 0$ pour tout $x \in A^*$.*

Langages minces (slender)

Un langage L est **mince** si sa densité $d_L(n) = |L \cap A^n|$ est $O(1)$.

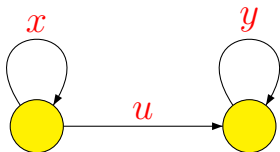
Théorème (Folklore)

*Un langage reconnaissable est **mince** ssi il est union finie de langages de la forme xu^*y .*

Ces langages sont fermés par **union** finie, **intersection** finie, **quotients** et **morphismes** mais **pas par inverse de morphismes**, même pour les morphismes strictement alphabétiques.

Théorème

Un langage reconnaissable est *mince* ssi son automate déterministe minimal ne contient *aucune paire connectée de cycles simples*.



Deux cycles connectés, où $x, y \in A^+$ et $u \in A^*$.

Équations définissant les langages minces

On note $i(u)$ l'initiale du mot u .

Théorème

Supposons que $|A| \geq 2$. Un langage reconnaissable de A^* est *mince ou plein* ssi il satisfait les équations $x \leq 0$ et $0 \leq x^\omega u y^\omega$ pour tout $x, y \in A^+$, $u \in A^*$ tels que $i(uy) \neq i(x)$.

Langages clairsemés (sparse)

Un langage L est **clairsemé** si sa densité est polynomiale.

Théorème (Folklore)

Un langage reconnaissable est clairsemé ssi il est union finie de langages de la forme $u_0 v_1^ u_1 \cdots v_n^* u_n$.*

Ces langages sont fermés par **union** finie, **intersection** finie, **produit de concatenation**, **quotients** et **morphismes** mais **pas par inverse de morphismes**, même pour les morphismes strictement alphabétiques.



Théorème

Supposons que $|A| \geq 2$. Un langage reconnaissable de A^* est *clairsemé ou plein* ssi il satisfait les équations $0 \leq (x^\omega y^\omega)^\omega$ pour tout $x, y \in A^+$, tels que $i(x) \neq i(y)$.

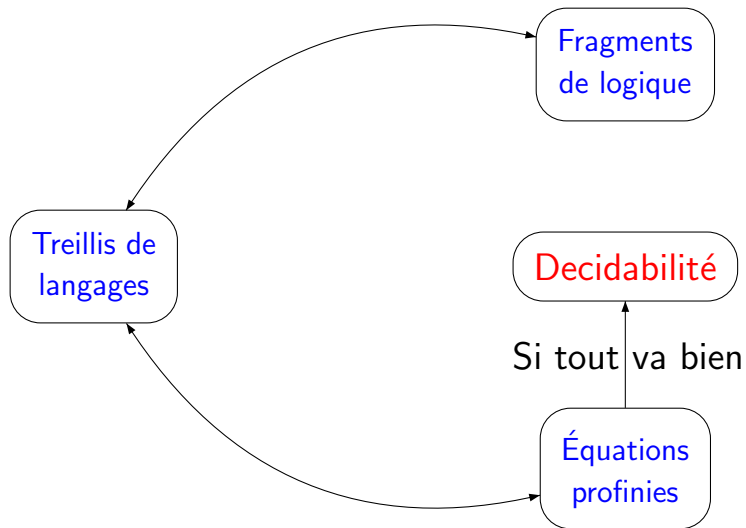
Résumé

On a montré que tout **treillis de langages** admet une description **équationnelle**, un résultat qui étend le **théorème des variétés d'Eilenberg** et ses **extensions**.

En particulier, toute classe de langages reconnaissables définie par un **fragment de logique** fermé par **conjonctions** et **disjonctions** (premier ordre, second ordre monadique, logique temporelle, etc.) admet une description **équationnelle**.



Retour sur les cercles vertueux



Difficultés

On a montré que tout **treillis de langages** admet une description **équationnelle**. Cependant. . .

- Le théorème ne donne pas d'algorithme pour trouver un ensemble d'équations.
- Les mots profinis sont parfois difficiles à manipuler, voire à définir.
- Même si on a une description récursive d'un ensemble d'équations profinies, il reste à trouver un algorithme pour tester si un langage reconnaissable donné les vérifie.

Extensions

Le résultat s'étend aux **mots infinis** (en utilisant des ω -semigroupes) et très certainement aux mots sur des **ordinaux** ou des **ordres linéaires** en utilisant les travaux de Bruyère-Carton-Colcombet.

Il devrait s'étendre également aux langages d'**arbres** (finis, infinis) et à d'autres structures.

On a une théorie équationnelle pour n'importe quel **treillis de langages** (non reconnaissables !), à condition d'utiliser des **équations procompactes**.

