Chapitre 2 Langages algébriques et BNF (définitions)

Comment suivre ce cours ?

Cours théorique avec beaucoup de concepts

- Examen final sur capacité à appliquer la théorie, mais pas sur capacité à prolonger ou même à restituer la théorie.
- Ceci dit, comprendre finement la théorie aide à l'appliquer.
- De plus, "*mathématiser la programmation*" aide à faire du code plus robuste, plus efficace, plus évolutif, etc.
- Beaucoup de théorie & d'exos ⇒ ça avance vite...

Prise de note

- Seul les exos marqués d'un † sont à savoir refaire pour l'examen.
- Sur les autres exos :
  Il est inutile de recopier les démonstrations (qui vont à toute vitesse...)
  Il est préférable de chercher à les comprendre (en posant des questions en cas de doute).

Idée du chapitre

Exo 2.1[†] Soit V un vocabulaire fini. Soient $A, B \subseteq V^*$. Quel est le plus petit ensemble $X \subseteq V^*$ tel que $X = A.X \cup B$?

$\Rightarrow$ On cherche conditions gales sur f pour définir langage L comme
"plus petit ensemble X qui vérifie $X = f(X)$"

- généraliser le lemme d'Arden (et les expressions régulières).
- une notion de "*définition récursive d'ensemble*".
- sous certaines conditions, $L =$ *limite* de suite infinie $\emptyset$, $f(\emptyset)$, $f(f(\emptyset))$, ..., $f^i(\emptyset)$, ...
  - permet de démontrer des propriétés par récurrence sur i.
  - limite éventuellement calculable.

Préliminaires sur la relation d'inclusion entre ensembles

Exo 2.2[†] Dessiner *diagramme de Hasse* de $\mathcal{P}(\{1, 2, 3, 4\})$ obtenu en reliant tte partie X aux Y minimaux tq $X \subsetneq Y$.

Exo 2.3 Soit E un ensemble. Soient $X, Y \subseteq E$ quelconques.

- Montrer que $\subseteq$ est une relation d'ordre sur $\mathcal{P}(E)$.
- Est-elle totale ? (a-t-on $X \subseteq Y$ ou $Y \subseteq X$?)
- Sur un ordre (partiel) $\leq$, on définit la notion de *borne sup* :

  $\sup(A, B)$ *est le plus petit X tq $A \leq X$ et $B \leq X$.*

  À quoi cela correspond sur un ordre total ? Et, sur $\subseteq$?
  (Attention, dans cas général, borne sup pas toujours définie).
- idem pour *borne inf*.

NB : ordre avec bornes inf/sup = *treillis* (anglais : *lattice*).

Chapitre 2 Langages algébriques et BNF (définitions)

Introduction à la notion de +petit point fixe

Defs Soit E un ensemble et f application de $\mathcal{P}(E) \to \mathcal{P}(E)$.
Un point fixe de f est un $X \subseteq E$ tq $X = f(X)$.
Un plus petit point fixe est un point fixe X tq tout point fixe Y vérifie aussi $X \subseteq Y$ (i.e. *unique* point fixe *minimal*).

Exo 2.4 Soit $\mathcal{V} \stackrel{def}{=} \{a, b\}$. Pour chacune des équations suivantes, quel est le +petit $X \subseteq \mathcal{V}^*$ qui la vérifie ?

- $X = \{a\}.X \cup \{b\}$

  ...

- $X = \{a\}.X \cup X$

  ...

- $X = \{a\}.X.\{b\} \cup \{\epsilon\}$?

  ...

Exo 2.5 M̂ question avec $X \subseteq \mathbb{N}$ pour $X = \{u + 2 \mid u \in X\} \cup \{0\}$

...

Inexistence ou non-unicité des points fixes minimaux

Soit E ensemble avec au moins deux éléments a et b distincts.

Exo 2.6 Quels sont les ensembles $X \subseteq E$ tq $X = E \setminus X$?

...

Exo 2.7 Quels sont les ensembles minimaux $X \subseteq E$ tq

$$X = \begin{cases} \{b\} & \text{si } b \in X \\ \{a\} & \text{sinon} \end{cases}$$

...

Solution pour éliminer ces contre-exemples

Garantir que "agrandir" le membre gauche de l'équation implique "agrandir" le membre droit de l'équation.

donc, se restreindre aux équations "$X = f(X)$" avec f croissante, c-à-d. $X \subseteq Y \Rightarrow f(X) \subseteq f(Y)$.

Théorème du point fixe de Knaster-Tarksi (1928)

Énoncé Si f application *croissante* de $\mathcal{P}(E) \rightarrow \mathcal{P}(E)$, alors f admet un +petit point fixe : $\bigcap\{X \in \mathcal{P}(E) \mid f(X) \subseteq X\}$.

Catalogue de fonctions croissantes

- pour A fixé, fonctions $X \mapsto X \cup A$ et $X \mapsto X \cap A$ croissantes.
- sur $\mathcal{V}^*$, $X \mapsto X.A$ et $X \mapsto A.X$ et $X \mapsto X^*$ croissantes.
- composée de fonctions croissantes est croissante.

Exemple : caractère croissant des membres droits de l'exo 2.4, en décomposant la vérification à l'aide des "briques" ci-dessus.

Avec ce thm, +petit point fixe "connu" mais pas "calculable".
Idée : f croissante, donc $\emptyset \subseteq f(\emptyset) \subseteq f(f(\emptyset)) \subseteq \ldots$
La suite des $(f^i(\emptyset))_{i \in \mathbb{N}}$ est croissante.
Et, s'il existe i tq $f^i(\emptyset) = f^{i+1}(\emptyset)$, alors point fixe atteint !

Vers le calcul des +petits points fixes

Exo 2.8† Soit $f : X \mapsto \{a\}.X.\{b\} \cup \{\epsilon\}$ (pour $X \subseteq \{a, b\}^*$).
Que vaut $f^i(\emptyset)$ pour $i \in \mathbb{N}$? ...
Que vaut $\bigcup_{i\in\mathbb{N}} f^i(\emptyset)$? ...

Notation Pour $(A_i)_{i\in\mathbb{N}}$ suite sur $\mathcal{P}(E)$, $\lim_{i\to+\infty} A_i \stackrel{def}{=} \bigcup_{i\in\mathbb{N}} A_i$.

Exo 2.9 Soit f application *croissante* de $\mathcal{P}(E) \to \mathcal{P}(E)$.
Montrer que :

1. pour tout i, $f^i(\emptyset) \subseteq f^{i+1}(\emptyset)$.
   En déduire, $\bigcup_{i\in[0,n]} f^i(\emptyset) = f^n(\emptyset)$.
2. tout point fixe de f contient $\lim_{i\to+\infty} f^i(\emptyset)$.
3. pour tout i, si $f^i(\emptyset)$ pas point-fixe, alors son cardinal $\geq i$.
4. si E est fini de cardinal n, alors $\lim_{i\to+\infty} f^i(\emptyset) = f^n(\emptyset)$ est un point fixe de f !

Quand $\lim_{i \to +\infty} f^i(\emptyset)$ n'est pas un point fixe...

On se place sur $E \stackrel{def}{=} \{a, b\}^*$. On pose

$$f(X) \stackrel{def}{=} \begin{cases} \{a\}.X \cup \{\epsilon\} & \text{si } X \text{ partie finie de } \{a\}^* \\ \{a\}^* \cup \{b\} & \text{sinon} \end{cases}$$

On a les propriétés suivantes (aisément vérifiables) :

- f croissante
- pour tout $i \in \mathbb{N}$, $f^i(\emptyset) = \{a^n \mid n \in \mathbb{N} \text{ et } n < i\}$
- $\lim_{i \to +\infty} f^i(\emptyset) = \{a\}^*$
- $f(\lim_{i \to +\infty} f^i(\emptyset)) = \{a\}^* \cup \{b\}$
  $\Rightarrow$ c'est lui le +petit point fixe !

Pb de "*discontinuité*" en l'infini !

Continuité (de Scott) & Point fixe de Kleene

Def Soient E_1, E_2 ensembles et f application de $\mathcal{P}(E_1) \to \mathcal{P}(E_2)$,
f est *continue* (au sens de Scott)
ssi pour toute suite de $(A_i)_{i\in\mathbb{N}}$ de $\mathcal{P}(E_1)$ tq $\forall i, A_i \subseteq A_{i+1}$,
on a $f(\lim_{i\to+\infty} A_i) = \lim_{i\to+\infty} f(A_i)$.

Exo 2.10 Montrer que :

- Une fonction continue est croissante.
- Toutes fonctions croissantes en exemple sur la diapo du thm "Knaster-Tarski" sont en fait continues.
- La composée de 2 fonctions continues est continue.

Thm de Kleene (1938) Si f application *continue* de $\mathcal{P}(E) \to \mathcal{P}(E)$, alors le +petit point fixe est $\lim_{i\to+\infty} f^i(\emptyset)$.

Application aux langages du TP

Soit $\mathbb{N}_1 \overset{def}{=} \mathbb{N} \backslash \{0\}$ et soit $V \overset{def}{=} \{-, \&, |, >, \mathtt{t}, \mathtt{f}\} \cup \mathbb{N}_1$.

Exo 2.11† Définir par plus petit point fixe, l'ensemble des mots de V^* qui correspondent à la notation *préfixe* d'une formule propositionnelle (cf. syntaxe du TP).
On doit trouver un f tq le langage recherché est $\lim_{h \to +\infty} f^h(\emptyset)$.

Exo 2.12 Calculer $f(\emptyset)$ et $f^2(\emptyset)$. Exprimer "$f^h(\emptyset)$" en fonction de la structure d'AST du TP.

Exo 2.13† Définir par plus petit point fixe, l'ensemble des mots de $(V \cup \{(,)\})^*$ qui correspondent à la notation *infixe* d'une formule propositionnelle (cf. syntaxe du TP).

Une technique centrale pour calculer image d'un pt fixe !

Lemme de commutation Si pour $k \in \{1, 2\}$, f_k applications de $\mathcal{P}(E_k) \to \mathcal{P}(E_k)$, et g application *continue* de $\mathcal{P}(E_1) \to \mathcal{P}(E_2)$ avec $g(\emptyset) = \emptyset$ et $g \circ f_1 = f_2 \circ g$, alors

$$g(\lim_{i \to +\infty} f_1^i(\emptyset)) = \lim_{i \to +\infty} f_2^i(\emptyset)$$

Exo 2.14†

Soient $A, B \subseteq V^*$, $f_1(X) \stackrel{def}{=} A.X \cup \{\epsilon\}$ et $f_2(Y) \stackrel{def}{=} A.Y \cup B$

Par définition $A^* = \lim_{i \to +\infty} f_1^i(\emptyset)$.

En appliquant le lemme de commutation, redémontrer

$$A^*.B = \lim_{i \to +\infty} f_2^i(\emptyset) \qquad \textbf{(lemme d'Arden)}.$$

Applications typiques du lemme de commutation

Pour f_1 et g fixés avec E_1 infini et E_2 de cardinal fini n.
Pour calculer $g(\lim_{i \to +\infty} f_1^i(\emptyset))$ **sans calculer** $\lim_{i \to +\infty} f_1^i(\emptyset)$

1. On trouve f_2 en exprimant $g(f_1(X))$ à partir de $g(X)$ sous la forme $g(f_1(X)) = f_2(g(X))$.
2. On se ramène au calcul de $\lim_{i \to +\infty} f_2^i(\emptyset) = f_2^n(\emptyset)$.

NB un tel f_2 n'existe pas forcément !
Auquel cas, méthode inapplicable.

Généralisation/application aux systèmes d'équations

Idée : système d'équations codée comme une unique équation.

Soit un *système d'équations* donné par fonction

$$f : \mathcal{P}(E_1) \times \ldots \times \mathcal{P}(E_n) \ \rightarrow \ \mathcal{P}(E_1) \times \ldots \times \mathcal{P}(E_n)$$

Comme $\mathcal{P}(E_1) \times \ldots \times \mathcal{P}(E_n) \simeq \mathcal{P}(\{1\} \times E_1 \cup \ldots \cup \{n\} \times E_n)$
on peut appliquer la théorie des $+$-petits points fixes à f

Pour $(X_1, \ldots, X_n)$ et $(Y_1, \ldots, Y_n)$ de $\mathcal{P}(E_1) \times \ldots \times \mathcal{P}(E_n)$
on a "$(X_1, \ldots, X_n) \subseteq (Y_1, \ldots, Y_n)$" ssi pr tt i, $X_i \subseteq Y_i$
et "$(X_1, \ldots, X_n) \cup (Y_1, \ldots, Y_n)$" $= (X_1 \cup Y_1, \ldots, X_n \cup Y_n)$

Exo 2.15† Soit le système suivant sur $\{a, b\}^* \times \{a, b\}^*$,

$$X_1 = \{b\} \cup X_2.X_2 \qquad X_2 = \{a\}.X_1$$

Calculer $f^4(\emptyset, \emptyset)$.

Chapitre 2 Langages algébriques et BNF (définitions)

Système d'équations algébriques sur V^*

Def Soit V ensemble dénombrable. Un *système d'équations algébriques sur* V^* est un ensemble d'équations (avec $(X_k)_{k\in[1,n]}$ suite de variables 2 à 2 distinctes)

$$\begin{array}{lcl} X_1 & = & f_1(X_1, \ldots, X_n) \\ \ldots & & \ldots \\ X_n & = & f_n(X_1, \ldots, X_n) \end{array}$$

où chaque $f_k(X_1, \ldots, X_n)$ est une *expression* constituée uniquement à partkr des varkables X_k et des "opérateurs" ensemblistes :

- $\{\epsilon\}$
- A, pour tout $A \subseteq V$ fixé (indépendant des X_k)
- union $\cup$
- concénation .

Thm Soit
$f \stackrel{def}{=} (X_1, \ldots X_n) \mapsto (f_1(X_1, \ldots, X_n), \ldots, f_n(X_1, \ldots, X_n))$.
On a f continue (avec $\lim_{i\to+\infty} f^i(\vec{\emptyset})$ comme +petit point fixe).

Langages algébriques

Définition Un langage L_1 est *algébrique* sur V^* ssi il existe $(L_k)_{i \in 2..n}$ tel que $(L_1, \ldots, L_n)$ est $+$petit point-fixe d'un système d'équations algébriques sur V^*.

Exo 2.16† Montrer que les langages définis dans le TP (`Prop`, `Nnf` en notations préfixes ou infixes) sont algébriques.

Les BNF "Backus-Naur Form" (années 1960)

BNF=notation pour définir des langages algébriques
(inventée pour syntaxe du 1er langage de prog structurée ALGOL).
Par ex, sur l'alphabet $V \stackrel{def}{=} \{0, 1, -, (,)\}$, la BNF

```
E  ::= L  |  E - E  |  ( E )
L  ::= 0  |  1  |  L L
```

définit E comme langage algébrique associé au système

$$
\begin{array}{rcl}
E & = & L \cup E.\{-\}.E \cup \{(\}.E.\{)\} \\
L & = & \{0\} \cup \{1\} \cup L.L
\end{array}
$$

Terminologie

- Les éléments de V s'appellent aussi "*symboles terminaux*".
- Les "variables" s'appellent aussi "*symboles non-terminaux*".
- Membre gauche de la 1ère équation s'appelle aussi "*axiome*".
- Membre droit d'une équation = union "*d'alternatives*".

Mini-exemples de langages algébriques non-réguliers

Exo 2.17† Pour $V \stackrel{def}{=} \{a, b, c\}$, donner une BNF pour chacun des langages suivants (dans chaque cas, justifier en calculant $f^n(\emptyset)$ où f est la fonction du point-fixe).

1. $\{a^n b^n \mid n \in \mathbb{N}\}$
2. $\{a^n b^p \mid n \geq p \geq 0\}$
3. $\{a^n b^p \mid n \neq p\}$
4. $\{a^n b^p \mid 2p \geq n \geq p\}$
5. $\{a^n b^p c^q \mid n + p = q\}$
6. $\{w \in \{a, b\}^* | w = \overline{w}\}$ où $\overline{w}$ est le renversé de w.
   Exemple de renversé : $\overline{a\ a\ b\ a} = a\ b\ a\ a$.
   NB : un mot égal à son renversé s'appelle un *palindrome*.
   Exemples de palindromes : "$a\ b\ a$" et "$a\ b\ b\ a$".

Autres exos sur les BNF

Exo 2.18[†] Donner une BNF sur $\{0, 1\}$ qui définit le langage des mots ayant un nombre pair de 0 et un nombre impair de 1.

Exo 2.19 Montrer que tout langage régulier peut être défini par une BNF. Que représente alors $f^n(\emptyset)$? Réciproquement, à quelles conditions (suffisantes), une BNF définit-elle un langage régulier ?

Exo 2.20 Définir la syntaxe des BNF comme un langage algébrique sur un alphabet formés de deux sous-ensembles disjoints : V (pour les symboles) et $\{::=, |, \backslash\mathtt{n}\}$. On autorisera l'alternative vide pour représenter ϵ. Par convention, les non-terminaux sont les symboles qui apparaissent en tant que membre gauche d'une équation.

Exemple d'algo : décider $\epsilon \in L$ avec L algébrique

On se ramène au calcul de $\mathcal{E}(L) \stackrel{def}{=} L \cap \{\epsilon\}$.

On définit $g(X_1, \ldots, X_n) \stackrel{def}{=} (\mathcal{E}(X_1), \ldots, \mathcal{E}(X_n))$ en bijection avec fct de $\mathcal{P}([1, n] \times V^*) \to \mathcal{P}([1, n] \times \{\epsilon\})$
pour appliquer le **lemme de commutation**.

Système à résoudre obtenu en transformant chaque équation "$X_k ::= e_k$" de la BNF de L en équation "$\mathcal{E}(X_k) = \mathcal{E}(e_k)$"
où "$\mathcal{E}(X_k)$" est vue comme une variable
et "$\mathcal{E}(e_k)$" calculé récursivement sur syntaxe de e_k pour s'exprimer en fonction de $\mathcal{E}(X_1), \ldots, \mathcal{E}(X_n)$:

- $\mathcal{E}(\epsilon) = \{\epsilon\}$ et pour tout terminal a, $\mathcal{E}(a) = \emptyset$
- $\mathcal{E}(\alpha.\beta) = \mathcal{E}(\alpha) \cap \mathcal{E}(\beta)$
- $\mathcal{E}(\alpha \mid \beta) = \mathcal{E}(\alpha) \cup \mathcal{E}(\beta)$

Calcul du +ptit pt fixe sur $E = [1, n] \times \{\epsilon\}$
ou par éliminations successives en exploitant la propriété suivante :
la +petite solution de $X = (X \cap \alpha) \cup \beta$ *vérifie aussi* $X = \beta$

Illustrations de cet algo

Exo 2.21† Appliquer cette méthode sur la BNF

```
S ::= A B C
A ::= B C | B a C | A C B
B ::= C b | A C
C ::= C c B A | ϵ
```

Exo 2.22† Idem en remplaçant l'équation de `B` ci-dessus par

```
B ::= C b | A C | C C
```

Une idée de la suite du cours

Problématique Soit L est un langage algébrique sur V^*.

1. Comment définir $T \in L \to D$ (pour un D donné) ?
   Problème : $T(1.|.2.\&.3) = T(1.|.(2.\&.3)) \neq T((1.|.2).\&.3)$
2. Algorithme efficace pour étant donné un mot $w \in V^*$, déterminer si $w \in L$, et si oui retourner $T(w)$?

Plan du cours :

- d'abord le cas où L est 1 notation préfixe (sans pb de parenthèsage).
  NB : notation préfixe = AST.
- ensuite, cas des autres langages algébriques.