Dans ce sujet, on s'intéresse aux différentes classes de langages étudiées en MP2I/MPI : langages réguliers, langages non-contextuels et langages décidables. Ce sujet se décompose en quatre problèmes : un problème introductif, puis un problème par classe de langages.

Le problème 1 traite d'une preuve alternative pour montrer qu'un langage régulier est non-contextuel.

Le problème 2 traite des langages réguliers, notamment sur une propriété intéressante de l'étoile de Kleene dans les langages unaires. Il est assez long.

Le problème 3 traite des langages décidables, toujours dans le cas des langages unaires, pour construire un langage indécidable unaire.

Le problème 4 traite des langages non-contextuels, où l'on montre que le langage des mots qui ne sont pas des carrés est non-contextuel.

Mon conseil est de ne pas procéder dans l'ordre mais de suivre un ordre topologique du graphe ci-dessous.

$$
\begin{tikzcd}
 & \text{Problème } 3 \arrow[dr] & \\
\text{Problème } 1 \arrow[ur] \arrow[dr] & & \text{Problème } 2 \\
 & \text{Problème } 4 \arrow[ur] &
\end{tikzcd}
$$

Table des matières

1 Introduction

Dans ce problème, on montre que les langages réguliers sont non-contextuels à l'aide d'une preuve différente de celle donnée mercredi.

Considérons L un langage régulier sur l'alphabet Γ, et $\mathscr{A} = (\Gamma, Q, I, F, \delta)$ un automate fini sans ε-transitions reconnaissant L.

On se donne $|Q|$ symboles non-terminaux (*i.e.* variables), que l'on notera $(\mathsf{X}_q)_{q \in Q}$.

Pour $q \in Q$, on note $\mathscr{A}_q := (\Gamma, Q, \{q\}, F, \delta)$ l'automate identique à $\mathscr{A}$ mais ayant q comme unique état initial.

Q1. Proposer un ensemble de règles de production Π tel que, pour tout $q \in Q$, on ait

$$\mathscr{L}(\mathscr{A}_q) = \underbrace{\{u \in \Gamma^\star \mid \mathsf{X}_q \Rightarrow^\star u\}}_{\mathscr{L}(\mathscr{G}, \mathsf{X}_q)}.$$

Indication.

On pourra procéder « à la manière » des fonctions récursives mutuelles où, ici, une fonction récursive correspond à un certain X_q.

Q2. Démontrer qu'avec l'ensemble de règles Π donné, on ait bien la propriété ci-dessus.

Indication.

On pourra procéder par récurrence sur n afin de montrer, pour tout état $q \in Q$, l'égalité $\mathscr{L}(\mathscr{A}_q) \cap \Gamma^n = \mathscr{L}(\mathscr{G}, \mathsf{X}_q) \cap \Gamma^n$.

Q3. Conclure.

Indication.

On pourra ajouter des règles supplémentaires à Π, ou alors ajouter une hypothèse supplémentaire sur $\mathscr{A}$.

2 Langages réguliers.

Dans ce problème, on se place sur l'alphabet $\Sigma = \{\mathtt{a}\}$. Un langage sur un tel alphabet est appelé *unaire*.

Q4. Donner un exemple de langage unaire X qui n'est pas régulier.

L'application

$$\begin{aligned} f : \wp(\mathbb{N}) &\longrightarrow \wp(\Sigma^\star) \\ A &\longmapsto \{\mathtt{a}^n \mid n \in \mathbb{N}\} \end{aligned}$$

est une bijection. Ainsi, il est équivalent de s'intéresser à des parties de $\mathbb{N}$ ou à des langages unaires.

Dans la suite de ce problème, on dira que $A \subseteq \mathbb{N}$ est décidable, régulier, indécidable, *etc* si le langage associé $f(A)$ l'est.

Pour $a, b \in \mathbb{N}$ donnés, on note $L(a, b) := \{a + bn \mid n \in \mathbb{N}\}$. On dit qu'un ensemble A est linéaire s'il existe un couple $(a, b) \in \mathbb{N}^2$ vérifiant que $A = L(a, b)$.

Un ensemble *semi-linéaire* est une union finie d'ensembles linéaires.

Q5. Montrer que tout ensemble linéaire est régulier. En déduire que tout ensemble semi-linéaire est aussi régulier.

On s'intéresse maintenant à montrer la réciproque : tout langage unaire régulier correspond à un ensemble semi-linéaire. On commence par une propriété plus simple.

Q6. Montrer qu'un ensemble A est semi-linéaire si et seulement si il est *ultimement périodique*. C'est-à-dire qu'il existe une borne $k \in \mathbb{N}$ et un entier $p \in \mathbb{N} \setminus \{0\}$ tel que, pour tout entier $n > k$,

$$n \in A \text{ si et seulement si } n + p \in A.$$

Indication.

On pourra commencer par montrer que tout ensemble linéaire est ultimement périodique, puis que l'union de deux ensembles ultimement périodiques est ultimement périodique.

Q7. Montrer qu'un ensemble $A \subseteq \mathbb{N}$ régulier est ultimement périodique.

Indication.

Quelles informations donne réellement le lemme de l'étoile dans le cas unaire ? Se rappeler du « dessin » de la preuve du lemme de l'étoile, et essayer de généraliser.

Dans la suite de cette section, on se propose de montrer la propriété suivante :

Si X est un langage unaire, alors $X^\star$ est régulier.

Q8. On note $g := f^{-1} : \wp(\Sigma^\star) \to \wp(\mathbb{N})$. Montrer que, pour tout langage unaire $X \subseteq \Sigma^\star$, on a

$$g(X^\star) = \{x_1 + x_2 + \cdots + x_n \mid n \in \mathbb{N}, x_1, \ldots, x_n \in g(X)\}.$$

Q9. Montrer que, pour $A \subseteq \mathbb{N}$ une partie quelconque, l'ensemble

$$A_\star := \{a_1 + a_2 + \cdots + a_n \mid n \in \mathbb{N}, a_1, \ldots, a_n \in A\}$$

est semi-linéaire. Pour cela, on procèdera ainsi :

- ▷ traiter le cas où $A = \emptyset$;
- ▷ puis, on choisit $a \in A$, et on décompose $A_\star$ en « classes de congruences » modulo a.

Indication.

On posera, par exemple, $A_r = \{x \in A_\star \mid x \equiv r \pmod a\}$. Lorsque $A_r \subseteq \mathbb{N}$ est non vide, il admet un minimum a_r. On pourra exprimer A_r à l'aide de a_r et a, et conclure que $A_\star$ est semi-linéaire.

Q10. Conclure que $X^\star$ est régulier pour tout $X \subseteq \{\mathtt{a}\}^\star$.

3 Langages décidables.

Dans ce problème, on continue sur les langages unaires, et on va construire un langage unaire indécidable.

On rappelle que, pour un langage L donné, on dit que L est *(in)décidable* si le problème MEMBERSHIP$_L$ est *(in)décidable*.

MEMBERSHIP$_L$	**Entrée.** Un mot $w \in \Sigma^\star$ **Sortie.** Est-ce que w appartient à L ?

On pose Σ un alphabet arbitraire (pas nécessairement avec une lettre).

Q11. Montrer que le problème FINITE est indécidable.

FINITE	**Entrée.** Une machine $\mathcal{M}$ **Sortie.** Est-ce que le langage $\mathcal{L}(\mathcal{M})$ est fini ?

Q12. Exprimer mathématiquement le langage des « instances positives » du problème FINITE, c'est-à-dire l'ensemble des entrées vérifiant la condition du problème. Par la suite, on le notera FINITE^+. Ce langage est-il décidable ?

Q13. Définir une fonction injective calculable $f : \Sigma^\star \to \{a\}^\star$.

Indication.

On pourra construire une fonction $g : \{0, 1\}^\star \to \{a\}^\star$, à l'aide de la fonction donnée au début du problème 2.

Q14. On pose $L := f(\text{FINITE}^+)$. Montrer que L est indécidable.

4 Langages non-contextuels.

Dans cet exercice, on pose $\Gamma := \{a, b\}$. On souhaite montrer que le langage

$$L := \Gamma^\star \setminus \{uu \mid u \in \Gamma^\star\}$$

est non-contextuel (*i.e.* que L est le langage d'une grammaire non-contextuelle).

Q15. Donner deux exemples de mots dans L, le premier de longueur 4, et le second de longueur 5.

Q16. Montrer que tout mot de longueur impaire est dans L, et que tout mot $w = w_1 \dots w_{2n}$ de longueur paire appartient à L si, et seulement si, il existe un indice $i \in [\![1, n]\!]$ tel que $w_i \neq w_{n+i}$.

Indication.

Procéder par contraposée : montrer que $w \in \Gamma^\star \setminus L$ si, et seulement si, [...à compléter...].

On considère la grammaire $\mathscr{G} := (\Gamma, \mathscr{V}, \Pi, \mathsf{S})$ où $\mathscr{V} := \{\mathsf{S}, \mathsf{A}, \mathsf{B}\}$ où

$$\begin{aligned}
\mathsf{S} &\to \mathsf{A} \mid \mathsf{B} \mid \mathsf{AB} \mid \mathsf{BA} \\
\mathsf{A} &\to a \mid a\mathsf{A}a \mid a\mathsf{A}b \mid b\mathsf{A}a \mid b\mathsf{A}b \\
\mathsf{B} &\to b \mid a\mathsf{B}a \mid a\mathsf{B}b \mid b\mathsf{B}a \mid b\mathsf{B}b.
\end{aligned}$$

On notera

$$\mathscr{L}(\mathscr{G}, \mathsf{A}) := \{w \in \Gamma^\star \mid \mathsf{A} \Rightarrow^\star w\}.$$

Ainsi le langage de $\mathscr{G}$, $\mathscr{L}(\mathscr{G})$, est $\mathscr{L}(\mathscr{G}, \mathsf{S})$ où S est le symbole initial de $\mathscr{G}$.

Q17. Décrire les langages $\mathscr{L}(\mathscr{G}, \mathsf{A})$ et $\mathscr{L}(\mathscr{G}, \mathsf{B})$.

Q18. Montrer que tout mot w de longueur impaire est dans $\mathscr{L}(\mathscr{G})$.

Indication.

Regarder la lettre centrale du mot considéré. On pourra construire la dérivation en se basant sur cette lettre, et à l'aide de la question précédente.

Q19. Montrer que tout mot $w \in L$ de longueur paire est dans $\mathscr{L}(\mathscr{G})$.

Indication.

Utiliser la question 16 en découpant w en deux mots.

Q20. Conclure que L est non-contextuel.

Indication.

Montrer que $L = \mathscr{L}(\mathscr{G})$ par double inclusion. Une des deux inclusions est déjà faite par les questions précédentes.