Fondements de l'informatique: Examen
Durée: 3h

Sujet proposé par Olivier Bournez
Version 3

L'énoncé comporte 4 parties (sections), certaines avec des sous-parties (sous-sections), chacune indépendante, qui pourront être traitées dans un ordre quelconque. En revanche, dans chaque partie, il peut être utile, dans la réponse à une question, d'utiliser les questions précédentes ! On pourra librement admettre le résultat d'une question pour passer aux questions suivantes. La difficulté des questions n'est pas une fonction linéaire ni croissante de leur numérotation.

Il est possible d'avoir la note maximale sans avoir répondu à toutes les questions. Les questions avec un barème plus élevé sont indiquées par le symbole $(*)$

On pourra utiliser les résultats et théorèmes démontrés en cours sans chercher à les redémontrer.

Dans tout l'énoncé, on demande des algorithmes et des solutions à un haut niveau : dans aucune des questions il n'est demandé de décrire complètement une machine de Turing, ni même d'en donner une description graphique ; on pourra se contenter pour décrire un algorithme de le décrire par exemple en français ou dans un langage de programmation classique comme JAVA, C ou CAML.

1 A propos de 007

Question 1. *Parmi les problèmes suivants, lesquels sont décidables, récursivement énumérables ou non-récursivement énumérables ? Justifier votre réponse.*

On considère des machines de Turing sur l'alphabet $\{0, 1, 2, \ldots, 9\}$.

1. *Déterminer si une machine de Turing M, partant sur le mot vide, est telle qu'à un certain instant son ruban contient quelque part le mot* 007.
2. *Déterminer si le langage accepté par une machine de Turing M contient le mot* 007.
3. *Déterminer si une machine de Turing M partant sur le mot* 007 *est telle que sa tête de lecture visite au plus* 700 *cases du ruban.*

2 Gérer le catalogue des cours est difficile

On se place dans une école que l'on appellera Y.

Chaque semestre, on propose aux étudiants de Y un catalogue de cours. Chaque cours est constitué d'un nombre fini de séances. Chaque séance se déroule selon un horaire (intervalle de temps) précis.

On précise qu'il est impossible dans l'école Y de déplacer une séance de cours.

Dans tout ce qui suit, on suppose que chaque date est codée par un entier.

2.1 Le point de vue de l'étudiant

Pour valider un cours, chaque élève doit assister à toutes les séances du cours. L'objectif d'un étudiant de l'école Y est de valider le plus grand nombre de cours dans le semestre.

Le problème PLANNING que l'étudiant a à résoudre peut donc se modéliser de la façon suivante :

— **Donnée**: Un ensemble $\mathcal{S}$ de cours, chaque cours correspondant à un ensemble fini d'intervalles, chaque intervalle correspondant à une date de début et de fin d'une séance. Un entier k.

— **Réponse**: Décider s'il existe un sous-ensemble $S' \subseteq S$ de cardinal $\geq k$ tel qu'il est possible de suivre tous les séances des cours de S' (c'est-à-dire tel qu'aucune séance des cours de S' ne se chevauche)

Question 2. $(*)$ *Prouver que le problème PLANNING est NP-complet.*

On admettra la NP-complétude du problème STABLE :

— ***Donnée**: Un graphe $G = (V, E)$ non-orienté et un entier k.*

— ***Réponse**: Décider s'il existe $V' \subset V$, avec $|V'| = k$, tel que $u, v \in V' \Rightarrow (u, v) \notin E$.*

2.2 Le point de vue de la scolarité

La personne qui gère l'affectation des salles de l'école Y possède de son côté s salles, et doit affecter à chaque séance une salle. Son problème est de comprendre si le nombre de salles est suffisant.

Elle a donc à résoudre le problème de décision AFFECTATION suivant (on suppose que tous les cours ont au moins un étudiant, et que toute séance peut avoir lieu dans n'importe quelle salle) :

— **Donnée**: Un ensemble $\mathcal{S}$ de cours, chaque cours correspondant à un ensemble fini d'intervalles, chaque intervalle correspondant à une date de début et de fin d'une séance ; Un entier s.

— **Réponse**: Décider s'il existe une façon de placer chacune des séances dans une des s salles de telle sorte qu'au plus une séance est affectée à chaque salle à tout moment.

Question 3. $(*)$ *Le problème AFFECTATION est-il NP-complet ? Si non, proposez un algorithme polynomial. Justifier.*

Indication : on pourra considérer les séances dans l'ordre de leur date de début.

3 Problèmes complets et difficiles pour RE

On dit qu'un problème A est *RE-complet* s'il est récursivement énumérable, et si tout problème récursivement énumérable B est tel que $B \leq_m A$. On dit qu'un problème A est *RE-difficile* si la seconde propriété est vraie : si tout problème récursivement énumérable B est tel que $B \leq_m A$.

Question 4. *Prouver que le problème* HALTING − PROBLEM *est* RE*-complet.*

Question 5. *Un problème* RE*-difficile peut-il être décidable ?*

Question 6. *Prouver la version suivante étendue du théorème de Rice :*

Soit une propriété P des langages semi-décidables non triviale, c'est-à-dire telle qu'il y a au moins une machine de Turing M telle que $L(M)$ satisfait P et une machine de Turing M' telle que $L(M')$ ne satisfait pas P.

Alors le problème de décision L_P :

— ***Donnée**: Le codage $\langle M \rangle$ d'une machine de Turing M ;*

— ***Réponse**: Décider si $L(M)$ vérifie la propriété P ;*

est RE-difficile ou son complémentaire est RE-difficile.

Question 7. *Démontrer que le problème suivant*
— ***Donnée**: Une machine de Turing $\langle M \rangle$*
— ***Réponse**: Décider si M accepte le mot $\langle M \rangle$*
est RE-*complet.*

Question 8. *Y a-t-il un problème récursivement énumérable considéré en cours ou en TD (PC) que l'on a prouvé indécidable qui ne soit pas RE-complet. Justifier la réponse.*

4 Définissabilité

L'objectif de cette partie est de prouver le résultat suivant, que l'on nommera résultat @ : *il n'existe pas de formule du premier ordre (i.e. du calcul des prédicats vu en cours) qui caractérise les graphes finis connexes.* Dit autrement : il n'existe pas de formule close du premier ordre (i.e. du calcul des prédicats vu en cours) qui est satisfaite par tout graphe fini connexe, et qui n'est pas satisfaite par aucun graphe fini non-connexe.

On considère pour cela une signature Σ avec le symbole de relation R binaire, et le symbole d'égalité $=$ binaire, et aucun symbole de constante ou de fonction. On se restreint dans tout cet énoncé aux modèles égalitaires, c'est-à-dire aux modèles où l'interprétation de $=$ correspond à l'égalité. Les modèles (structures) pour cette signature Σ correspondent donc à des graphes orientés.

On confondra souvent dans la suite par conséquent graphes et structures sur cette signature. On notera souvent $G = (V, E)$ pour un graphe, où V est l'ensemble des sommets, et E des arêtes orientées (on appelle aussi parfois cela des arcs). On notera $E(x, y)$ ou $(x, y) \in E$ pour le fait qu'il y a une arête orientée de x vers y.

On rappelle qu'un *chemin* entre les sommets x et y d'un graphe $G = (V, E)$ est une suite finie $x_0 = x, x_1, x_2, \ldots, x_k = y$ de sommets avec pour tout i, $(x_i, x_{i+1}) \in E$. Un tel chemin est dit de longueur k.

On rappelle qu'un graphe $G = (V, E)$ est *connexe* si pour toute paire x et y de sommets de G, il y a un chemin entre x et y.

4.1 Un échauffement

Question 9. *On considère la signature Σ' obtenue en ajoutant à la signature Σ deux symboles de constantes c_1 et c_2. Un modèle sur la signature Σ' correspond donc à un graphe orienté, avec deux sommets distingués s_1 et s_2 : s_1 est donné par l'interprétation de c_1 et s_2 par l'interprétation de c_2.*

On fixe un entier d.

Produire une formule sur la signature Σ' qui est satisfaite si et seulement s'il y a un chemin de longueur d entre les sommets distingués s_1 et s_2.

4.2 Une variante du résultat @

Si l'on ne fait pas l'hypothèse que les graphes sont finis[1], le résultat @ peut se prouver directement (via le théorème de compacité) :

Question 10. $(*)$ *Prouver qu'il n'existe pas de formule du premier ordre (i.e. du calcul des prédicats vu en cours) qui caractérise les graphes connexes : dit autrement, démontrer qu'il n'existe pas de formule close du premier ordre (i.e. du calcul des prédicats vu en cours) qui est satisfaite par tout graphe[2] connexe, et qui n'est pas satisfaite par aucun graphe[3] non-connexe.*

1. Un graphe est dit fini si son nombre de sommets est fini.
2. fini ou infini
3. fini ou infini

Indication : on pourra raisonner par l'absurde, et chercher à exprimer l'inexistence d'un chemin entre deux sommets distingués s_1 et s_2 et utiliser le théorème de compacité du calcul des prédicats.

Observons que cela n'implique pas le résultat @ : le fait qu'il n'existe pas de formule du premier ordre qui caractérise la propriété d'être connexe pour un graphe (fini ou infini) n'implique pas qu'il n'y a pas de formule du premier ordre qui caractérise la propriété d'être connexe pour un graphe **fini**. Aussi, on va développer plusieurs résultats dans la suite pour prouver le résultat @.

4.3 Une première remarque

Question 11. *Soit ϕ une formule sur la signature Σ.*

Expliquer pourquoi toute formule atomique qui apparait dans ϕ (i.e. toute sous-formule de ϕ) est de l'une des deux formes suivantes :

1. *$x = y$ où x, y sont des variables ;*
2. *$R(x, y)$, où x, y sont des variables.*

4.4 Rang de quantification

On dit que deux structures $\mathcal{S}_1$ et $\mathcal{S}_2$ sur une même signature sont élémentairement équivalentes, noté $\mathcal{S}_1 \equiv \mathcal{S}_2$ si pour toute formule close ϕ, ϕ est satisfaite dans $\mathcal{S}_1$ si et seulement si ϕ est satisfaite dans $\mathcal{S}_2$.

Le *rang* $rg(\phi)$ d'une formule ϕ (du calcul des prédicats) se définit inductivement comme suit (c'est le nombre maximal de quantificateurs sur un chemin si l'on voit la formule comme un arbre) :

— $rg(\phi) = 0$ si ϕ est une formule atomique

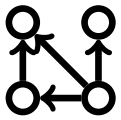
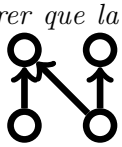
— $rg(\phi \wedge \psi) = rg(\phi \vee \psi) = rg(\phi \Rightarrow \psi) = rg(\phi \Leftrightarrow \psi) = \max(rg(\phi), rg(\psi))$

— $rg(\exists x\ \phi) = rg(\forall x\ \phi) = 1 + rg(\phi)$

Une formule sans quantificateur est en particulier par définition une formule ϕ avec $rg(\phi) = 0$.

Soit r un entier. Deux structures $\mathcal{S}_1$ et $\mathcal{S}_2$ seront dites *r-indistinguables* si pour toute formule close ϕ telle que $rg(\phi) \leq r$, on a que ϕ est satisfaite dans $\mathcal{S}_1$ si et seulement si ϕ est satisfaite dans $\mathcal{S}_2$. On dit qu'elles sont *r-distinguables* dans le cas contraire : il y a donc une formule close ϕ, avec $rg(\phi) \leq r$, qui est satisfaite dans l'une et pas dans l'autre.

Observons que par définition, $\mathcal{S}_1 \equiv \mathcal{S}_2$ si et seulement si $\mathcal{S}_1$ et $\mathcal{S}_2$ sont *r-indistinguables* pour tout entier r.

Question 12. *Montrer que la structure correspondant au graphe*  *et la structure correspondant au graphe*  *sur la signature Σ sont 2-distinguables.*

4.5 Isomorphismes partiels et 0-distinguabilité

Un *isomorphisme partiel* h d'un graphe orienté $G_1 = (V_1, E_1)$ dans un graphe orienté $G_2 = (V_2, E_2)$ est une fonction partielle injective h de V_1 dans V_2 telle que pour tout x, y dans le domaine de h,

$$E_1(x, y) \text{ si et seulement si } E_2(h(x), h(y)).$$

Lorsque ϕ est une formule de variables libres $x_1, \ldots, x_n$, et G est une structure (i.e. dans notre contexte un graphe orienté), $a_1, a_2, \ldots, a_n$ sont des éléments du domaine de la structure

(i.e. dans notre contexte des sommets), on écrira $G, \{x_1 \mapsto a_1, \ldots, x_n \mapsto a_n\} \models \phi$ lorsque la formule ϕ est satisfaite sur G pour la valuation qui envoie x_i sur a_i.

Les isomorphismes partiels sont suffisants pour caractériser la 0-distinguabilité :

Question 13. *Soient $a_1, \ldots, a_n \in V_1$ et $b_1, \ldots, b_n \in V_2$ où $G_1 = (V_1, E_1)$ et $G_2 = (V_2, E_2)$ sont deux graphes orientés.*

On considère la fonction partielle h définie par $h(a_i) = b_i$ pour $i = 1, \ldots, n$ (et de domaine exactement $\{a_1, \ldots, a_n\}$).

*Montrer que h est un isomorphisme partiel de G_1 dans G_2 si et seulement si pour toute formule **sans quantificateur** ϕ de variables libres $x_1, \ldots, x_n$*

$$G_1, \{x_1 \mapsto a_1, \ldots, x_n \mapsto a_n\} \models \phi \text{ si et seulement si } G_2, \{x_1 \mapsto b_1, \ldots, x_n \mapsto b_n\} \models \phi$$

4.6 Jeux de Ehrenfeucht-Fraïssé

Un jeu (de EF) est un jeu qui se joue avec deux joueurs, l'un appelé S (spoiler) et l'autre D (duplicator). Il se joue sur deux graphes orientés $G_1 = (V_1, E_1)$ et $G_2 = (V_2, E_2)$.

On fixe un entier r, que l'on appelle *nombre de tours de jeu* (on dit aussi *nombre de coups*). Pour $i = 1$ à r,

- S choisit un élément dans le domaine de l'une des deux structures : c'est-à-dire, soit il décide de choisir un élément dans V_1, soit il décide de choisir un élément dans V_2.
  Dans le premier cas, on appellera cet élément a_i, et dans le second b_i.
- D choisit ensuite un élément dans le domaine de l'autre structure : si S avait choisit a_i dans V_1, D choisit un élément que l'on appelle b_i dans V_2. Si S avait choisit b_i dans V_2, D choisit un élément que l'on appelle a_i dans V_1.

Pour déterminer qui a gagné, on fait ainsi : on considère la fonction partielle h définie par $h(a_i) = b_i$ (et pas définie ailleurs, i.e. le domaine de h est $\{a_1, \ldots, a_r\}$). Le joueur D gagne si h est un isomorphisme partiel de G_1 dans G_2.

Une *stratégie gagnante* (pour le jeu à r tours de jeu) pour le joueur D est une manière systématique de jouer qui lui permet de gagner, quelle que soit la manière de jouer du joueur S.

Question 14. *On considère les deux structures de la question 12. Montrer que S peut gagner le jeu à 2 tours de jeu (= D n'a pas de stratégie gagnante pour $r = 2$). Montrer que par contre, S ne gagne jamais le jeu à 1 tour de jeu (= D possède bien une stratégie gagnante pour $r = 1$).*

Question 15. $(*)$ *Soient n et r deux entiers.*

Soient $a_1, \ldots, a_n \in V_1$ et $b_1, \ldots, b_n \in V_2$ où $G_1 = (V_1, E_1)$ et $G_2 = (V_2, E_2)$ sont deux graphes orientés.

On suppose que le joueur D possède une stratégie gagnante pour le jeu à $n + r$ tours de jeu sur G_1 et G_2 (c'est-à-dire, plus formellement : en supposant les choix fixés pour les n premiers tours de jeu par les $a_1, \ldots, a_n \in V_1$ et $b_1, \ldots, b_n \in V_2$, D est toujours capable de jouer pour les r tours de jeu restants de façon à gagner quels que soient les choix de S).

Prouver par induction que pour toute formule ϕ sans quantificateur universel, de rang $\leq r$, dont les variables libres sont $x_1, \ldots, x_n$ on a

$$G_1, \{x_1 \mapsto a_1, \ldots, x_n \mapsto a_n\} \models \phi \text{ si et seulement si } G_2, \{x_1 \mapsto h(a_1), \ldots, x_n \mapsto h(a_n)\} \models \phi.$$

Le résultat reste vrai en supprimant l'hypothèse "sans quantificateur universel", puisque $\forall x \phi$ est logiquement équivalent à $\neg\exists\neg\phi$ qui est de même rang.

Question 16. *Prouver que si le joueur D possède une stratégie gagnante pour le jeu à r tours de jeux, alors G_1 et G_2 sont r-indistinguables.*

4.7 La connexité n'est pas définissable

Question 17. *On considère un graphe G_1 qui correspond au cycle avec 5 sommets : c'est-à-dire le graphe $G_1 = (V_1, E_1)$ dont les sommets sont $V_1 = \{s_1, \ldots, s_5\}$, et dont l'ensemble des arêtes est $E_1 = \{(s_1, s_2), (s_2, s_3), (s_3, s_4), (s_4, s_5), (s_5, s_1)\}$.*

On considère le graphe G_2 avec 10 sommets constitué de deux copies disjointes du graphe précédent, c'est-à-dire de 2 cycles disjoints à 5 sommets.

Montrer que, quels que soient les deux sommets distincts b_1 et b_2 choisis dans G_2, on peut trouver deux sommets a_1 et a_2 dans G_1 tels que la fonction partielle définie par $h(a_1) = b_1$, $h(a_2) = b_2$ (et de domaine $\{a_1, a_2\}$) soit un isomorphisme partiel de G_1 dans G_2.

Qu'en est-il si on avait considéré le cycle à 3 sommets au lieu de 5 (G_2 étant toujours 2 copies disjointes, i.e. un graphe à 6 sommets) ?

Question 18. $(*)$ *On fixe un entier r. On considère un entier $r' > r2^{r+1}$. On considère le cycle G_1 avec r' sommets : c'est-à-dire le graphe dont les sommets sont $s_1, \ldots, s_{r'}$, et dont l'ensemble des arêtes est $\{(s_i, s_{i+1}) | 1 \leq i \leq r' - 1\} \cup \{(s_{r'}, s_1)\}$.*

On considère le graphe G_2 avec $2r'$ sommets constitué de deux copies disjointes du graphe précédent, i.e ; de 2 cyles disjoints à r' sommets.

Expliquer pourquoi D possède une stratégie gagnante pour le jeu à r tours de jeu sur les graphes G_1 et G_2.

Question 19. *En déduire le résultat @.*