- (1) Demostreu que també és indecidible el problema de, donades dues gramàtiques, saber si tenen com a intersecció dos o més mots de longitud parell. (anomenem *2-intersecció parella* a aquest nou problema.)

$$\{\langle G_1, G_2\rangle \mid \exists w_1, w_2 : (w_1 \neq w_2 \wedge |w_1|, |w_2| \in \dot{2} \wedge w_1, w_2 \in (\mathtt{L}(G_1) \cap \mathtt{L}(G_2)))\}$$

Resposta:

- **Donada la entrada $\langle G_1, G_2\rangle$ d'intersecció no buida, construïm una nova entrada $\langle G_1', G_2'\rangle$ per a intersecció parella com segueix. Ens inventem un símbol nou #. Si S_1 és el símbol inicial de G_1, llavors G_1' té una nova variable S_1' com a símbol inicial, i les seves regles són les de G_1 més $S_1' \to S_1\#\#S_1$. Si S_2 és el símbol inicial de G_2, llavors G_2' té una nova variable S_2' com a símbol inicial, i les seves regles són les de G_2 més $S_2' \to S_2\#\#S_2$.**

  Si G_1 i G_2 generen una paraula comuna w, llavors G_1' i G_2' generen la paraula comuna $w\#\#w$ de longitud parella. Per a la direcció contrària, si G_1' i G_2' generen una paraula comuna u de longitud parella, llavors, per la forma d'aquestes gramàtiques, u és de la forma $u_1\#\#u_2$, on tant u_1 com u_2 són generables tant amb G_1 com amb G_2. Per tant, G_1 i G_2 generen alguna paraula comuna, i això conclou la prova.

- **Reduïm desde el llenguatge de l'apartat anterior. Donada la entrada $\langle G_1', G_2'\rangle$ d'intersecció parella, construïm una nova entrada $\langle G_1'', G_2''\rangle$ per a 2-intersecció parella com segueix. Ens inventem un símbol nou \$. Si S_1' és el símbol inicial de G_1', llavors G_1'' té una nova variable S_1'' com a símbol inicial, i les seves regles són les de G_1' més $S_1'' \to S_1'|\$\$$. Si S_2' és el símbol inicial de G_2', llavors G_2'' té una nova variable S_2'' com a símbol inicial, i les seves regles són les de G_2' més $S_2'' \to S_2'|\$\$$.**

  Si G_1' i G_2' generen una paraula comuna de longitud parella, llavors G_1'' i G_2'' també la generen, i ademés totes dues generen també \$\$, que és de longitud parella. Per a la direcció contrària, si G_1'' i G_2'' generen dues paraules comunes de longitud parella, llavors almenys una no és \$\$, i ademés és generable desde G_1' i G_2'. Això conclou la prova.

(Examen de Juny-2008)

1. (2 punts en total, temps màxim de resolució estimat: 40 minuts) Contesteu amb una, dues o tres frases, i sobre el propi full de l'enunciat, les qüestions següents. (que no requereixen justificació, tret que la pregunta indiqui el contrari).

   (a) (0.25) Doneu un exemple de llenguatge que sigui decidible però no incontextual.

   Resposta: $\{a^n b^n c^n \mid n \geqslant 0\}$.

   (b) (0.25) Doneu un exemple de llenguatge que sigui incontextual i inherentment ambigu.

   Resposta: $\{a^n b^n c^m \mid n, m \geqslant 0\} \cup \{a^m b^n c^n \mid n, m \geqslant 0\}$

(c) (0.5) Acabeu la següent demostració del fet que $\{a^n b^n | n \geqslant 0\}$ és no-regular: **suposem que $\{a^n b^n | n \geqslant 0\}$ és regular. Llavors existeix un autòmat A que el reconeix. Sigui N el nombre d'estats de A, i q_0 l'estat inicial de A. L'autòmat A, amb entrada $a^N b^N$ passa almenys pels estats $q_0, q_0a, q_0aa, \ldots, q_0a^N$. Com que n'hi ha $N+1$, algun està repetit.** ...

Resposta: Per tant existeixen $0 \leqslant i < j \leqslant N$ tals que $q_0a^i = q_0a^j$. Així doncs, $q_0(a^ib^i) = (q_0a^i)b^i = (q_0a^j)b^i = q_0(a^jb^i)$. Però llavors aquest estat hauria de ser acceptador i rebutjador al mateix temps, contradicció.

(d) (0.25) Donats dos autòmats $A_1 = \langle \Sigma, Q_1, \delta_1, q_{i1}, F_1 \rangle$ i $A_2 = \langle \Sigma, Q_2, \delta_2, q_{i2}, F_2 \rangle$, reconeixent llenguatges L_1 i L_2, un autòmat que reconeix $L = L_1 \cap L_2$ s'obté definint com a estats totes les parelles d'estats $Q = Q_1 \times Q_2$, definint les transicions com $\delta(\langle q_1, q_2 \rangle, a) = \langle \delta_1(q_1, a), \delta_2(q_2, a) \rangle$, i escollint com a estats inicial i acceptadors ...

Resposta: $\langle q_{i1}, q_{i2} \rangle$ com a inicial i $F_1 \times F_2$ com a acceptadors.

(e) (0.25) Si el llenguatge L és semi-decidible i el seu complementari $\overline{L}$ també, què en podem deduir?

Resposta: Que L és decidible.

(f) (0.5) Suposem que A és un DFA mínim que reconeix un cert llenguatge L. El DFA A' obtingut intercanviant estats acceptadors per no acceptadors de A reconeix el complementari $\overline{L}$. A continuació es comença a justificar que A' també resulta ser un DFA mínim. Acabeu la justificació. **Suposem que A' no és el DFA mínim reconeixent $\overline{L}$. Llavors, existeix un DFA més petit A'' que també reconeix $\overline{L}$. Si intercambiem estats acceptadors per no acceptadors en A''** ...

Resposta: Obtenim un nou DFA A''' que accepta L i amb menys estats que A, cosa que contradiu el fet que A fos mínim.

2. (3.5 punts en total, temps màxim de resolució estimat: 60 minuts) Diem que un llenguatge és un *singletó* si conté exàctament un mot, i.e. és de la forma $\{w\}$. A continuació definim el concepte de gramàtica singletó, que és una gramàtica pensada per tal que el seu llenguatge generat sigui singletó.

   Una gramàtica singletó té els símbols no terminals (variables) indexats de la forma $X_1, \ldots, X_n$. Cada variable X_i apareix com a part esquerra de regla en exàctament una regla, que pot ser de la forma $X_i \to c$ per a un cert símbol terminal c, o de la forma $X_i \to X_j X_k$ per a $j, k < i$. El símbol inicial de la gramàtica és X_n.

   Per exemple, la següent gramàtica singletó genera $\{aabaab\}$.

$$\begin{array}{lcl} X_5 & \to & X_4X_4 \\ X_4 & \to & X_2X_3 \\ X_3 & \to & b \\ X_2 & \to & X_1X_1 \\ X_1 & \to & a \end{array}$$

   Donada una gramàtica singletó G i una de les seves variables X_i, denotem amb $\texttt{mot}(G, X_i)$ al mot generat des de X_i. Observeu que si G conté una regla de la forma $X_i \to c$, llavors $\texttt{mot}(G, X_i) = c$, i que si G conté una regla de la forma $X_i \to X_j X_k$, llavors $\texttt{mot}(G, X_i) = \texttt{mot}(G, X_j)\texttt{mot}(G, X_k)$.

   La gràcia de les gramàtiques singletó és que permeten descriure mots "llargs" utilitzant una representació "curta", i a més, permeten calcular ràpidament propietats sobre el mot sense haver de generar-lo (cosa que seria molt costosa si el mot és "llarg").

   (a) (0.5) Escriviu una gramàtica singletó per a *ababab* utilitzant només 5 variables.

   (b) (1) Descriviu una gramàtica singletó per a $a^{(2^n)}$ utilitzant només $n + 1$ variables.

   (c) (1) L'exemple de l'apartat anterior mostra que les gramàtiques singletó poden comprimir molt. Però aquesta compressió té un límit. Volem justificar que una gramàtica singletó amb n variables $X_1, \ldots, X_n$ pot arribar a definir un mot de com a molt mida 2^{n-1}. Feu-ho per inducció sobre la definició de la gramàtica i la forma de les regles, veient que, per tot i entre 1 i n, tenim que $|\texttt{mot}(G, X_i)| \leqslant 2^{i-1}$.

   (d) (1) Ja hem vist que podem comprimir molt, però fins a un cert límit. Ara veiem que podem fer un càlcul sobre el mot generat, però sense necessitat de generar-lo. Donat un autòmat A i una gramàtica singletó G amb variables $X_1, \ldots, X_n$, volem comprobar si A accepta $\texttt{mot}(G, X_n)$, però sense generar-lo, ja que podria tenir mida 2^n i això implicaria cost exponencial. Doneu un algorisme de cost polinòmic que calculi a quin estat arriba A amb entrada $\texttt{mot}(G, X_n)$. (Pista: necessitareu calcular certa informació per cada $\texttt{mot}(G, X_i)$. Utilitzeu la definició de la gramàtica i la forma de les regles.)

Resposta:

(a)

$$\begin{array}{lcl} X_5 & \to & X_3X_4 \\ X_4 & \to & X_3X_3 \\ X_3 & \to & X_1X_2 \\ X_2 & \to & b \\ X_1 & \to & a \end{array}$$

(b)

$$\begin{array}{lcl} X_{n+1} & \to & X_nX_n \\ X_n & \to & X_{n-1}X_{n-1} \\ \vdots & & \vdots \\ X_2 & \to & X_1X_1 \\ X_1 & \to & a \end{array}$$

(c) Demostrem per inducció sobre i que cada X_i cumpleix $|\mathtt{mot}(G, X_i)| \leqslant 2^{i-1}$.

Sigui un i qualsevol entre 1 i n. Assumim inductivament que $|\mathtt{mot}(G, X_j)| \leqslant 2^{j-1}$ per a qualsevol $1 \leqslant j < i$. Distingim dos casos. Si la regla de X_i és de la forma $X_i \to c$, llavors $\mathtt{mot}(G, X_i) = c$, i per tant $|\mathtt{mot}(G, X_i)| = 1 \leqslant 2^{i-1}$. En canvi, si la regla de X_i és de la forma $X_i \to X_j X_k$, llavors $\mathtt{mot}(G, X_i) = \mathtt{mot}(G, X_j)\mathtt{mot}(G, X_k)$ i per tant $|\mathtt{mot}(G, X_i)| = |\mathtt{mot}(G, X_j)| + |\mathtt{mot}(G, X_k)| \leqslant 2^{j-1} + 2^{k-1} \leqslant 2^{i-2} + 2^{i-2} = 2^{i-1}$.

(d) Sigui δ la funció de transició de A. El que farem és calcular $\delta(q, \mathtt{mot}(X_i))$ per a tots els estats q de A i tots els X_i. Si som capaços de fer això ja tindrem la resposta, doncs n'hi haurà prou amb comprobar que $\delta(q_0, \mathtt{mot}(X_i))$ és acceptador, on q_0 és l'estat inicial de A. Aquest càlcul el fem per ordre, desde i igual a 1 fins a n. Per a cada X_i, assumim ja calculats els $\delta(q, \mathtt{mot}(X_j))$ per als $j < i$ i tots els estats q. Així doncs, n'hi ha prou amb veure que cada $\delta(q, \mathtt{mot}(X_i))$ es pot calcular eficientment sota aquestes assumcions. Distingim dos casos. Si la regla de X_i és de la forma $X_i \to c$, llavors $\mathtt{mot}(G, X_i) = c$, i per tant $\delta(q, \mathtt{mot}(X_i)) = \delta(q, c)$. Donat que $\delta(q, c)$ s'obté diréctament de l'autòmat d'entrada, aquest càlcul es pot fer eficientment. En canvi, si la regla de X_i és de la forma $X_i \to X_j X_k$, llavors $\mathtt{mot}(G, X_i) = \mathtt{mot}(G, X_j)\mathtt{mot}(G, X_k)$, i per tant $\delta(q, \mathtt{mot}(X_i)) = \delta(q, \mathtt{mot}(G, X_j)\mathtt{mot}(G, X_k)) = \delta(\delta(q, \mathtt{mot}(G, X_j)), \mathtt{mot}(G, X_k))$. Sigui q' l'estat $\delta(q, \mathtt{mot}(G, X_j))$. Per les nostres hipòtesis, q' ja està precalculat, i també $q'' = \delta(q', \mathtt{mot}(G, X_k))$, que és el que voliem obtenir.

3. (2 punts, temps màxim de resolució estimat: 30 minuts) Classifiqueu com a decidible, semi-decidible però no decidible, o no semi-decidible, el problema següent: $C = \{\langle x, y \rangle \mid \exists z : (M_x(z) \downarrow \ \wedge M_y(z) \downarrow \ \wedge M_x(z) < M_y(z))\}$

Resposta:

Demostrem que és indecidible reduïnt desde $\mathtt{K}$. La reducció consisteix a, per cada x d'entrada, generar el parell $\langle p(x), q \rangle$, on $p(x)$ és número que codifica el programa

```
entrada y
   Simular M_x(x)
   sortida 1
```

i q és el número que codifica el programa

```
entrada y
   sortida 2
```

Si x pertany a $\mathtt{K}$, llavors $\mathtt{M}_x(x) \downarrow$, i per tant el programa codificat per $p(x)$ dona sempre sortida 1. Com que el programa codificat per q dona sempre sortida 2, resulta que el parell $\langle p(x), q \rangle$ és del llenguatge C.

Si x no pertany a $\mathtt{K}$, llavors $\mathtt{M}_x(x) \uparrow$, i per tant el programa codificat per $p(x)$ no s'atura per a cap entrada. Així doncs, el parell $\langle p(x), q \rangle$ no és del llenguatge C.

Ara justificarem que el llenguatge anterior és semi-decidible. N'hi ha prou amb mostrar un programa que semi-decideix el conjunt.

```
entrada ⟨x,y⟩
   t:=1
```

```
Executar indefinidament:
   Per a cada i entre 0 i t fer:
      Si Mx(i) i My(i) s'aturen en t passos i Mx(i) < My(i) llavors acceptar
   t:=t+1
```

4. (2.5 punts en total, temps màxim de resolució estimat: 50 minuts) Durant el curs s'ha vist que els següents dos problemes són indecidibles:

- (intersecció no buida) Donades dues gramàtiques, saber si tenen intersecció no buida.

$$\{\langle G_1, G_2\rangle \mid \mathtt{L}(G_1) \cap \mathtt{L}(G_2) \neq \emptyset\}$$

- (accessibilitat de mots) Donada una llista finita R de regles de reescriptura (un conjunt finit de parells de mots (u_i, v_i)) i dos mots u (inicial), v (final), determinar si és possible passar de u a v utilitzant les regles de reescriptura (un seguit de substitucions d'algun submot u_i pel seu corresponent v_i).

$$\{\langle R, u, v\rangle \ \mid \ u \to_R^* v\}.$$

Volem donar reduccions entre aquests dos problemes.

(a) Primer volem veure que es pot reduïr fàcilment el primer problema (intersecció no buida) al segon (accessibilitat de mots). Per a fer-ho, us fem una proposta de com començar la reducció, i hi heu d'afegir el que hi falta.

Reducció: Sigui $\langle G_1, G_2\rangle$ el parell de gramàtiques que són instància d'intersecció no buida. Per simplificar, suposarem que els noms de variables que apareixen a G_1 són disjunts dels noms de variables que apareixen a G_2 (si calgués, es poden reanomenar). El nostre objectiu és generar una entrada d'accessibilitat de mots adequada.

Sigui S_1 el símbol inicial de G_1, i S_2 el símbol inicial de G_2. El que farem és generar una entrada de la forma $\langle R, S_1, S_2\rangle$, és a dir, un sistema de reescriptura R, i el parell de mots S_1 i S_2. Queda per determinar R per tal que la reducció sigui correcta, i això és el que heu de fer. És a dir, descriviu com s'obté R a partir de les regles de G_1 i G_2, i justifiqueu que $S_1 \to_R^* S_2$ si i només si G_1 i G_2 tenen intersecció no buida.

(descriure R encertadament val 1 punt, i la justificació val 0.5 punts.)

Resposta: Definim la llista de regles R com la llista de regles/produccions de G_1, més la llista de regles/produccions de G_2 invertides. Formalment, si R_1 són les produccions de G_1 i R_2 són les produccions de R_2, llavors definim $R = \{X \to \alpha | X \to \alpha \in R_1\} \cup \{\alpha \to X | X \to \alpha \in R_2\}$. En altres paraules, $R = R_1 \cup R_2^{-1}$ amb una definició natural d'invers adequada.

Per a veure que la reducció és correcta, mirem les dues direccions. Suposem que G_1 i G_2 tenen intersecció no buida. Llavors existeix un mot comú w generat per totes dues gramàtiques. Per tant, $S_1 \to_{R_1}^* w$ i $S_2 \to_{R_2}^* w$. Això implica $S_1 \to_{R_1}^* w \to_{R_2^{-1}}^* S_2$, de manera que concloem $S_1 \to_R^* S_2$. Per a la direcció contrària, suposem $S_1 \to_R^* S_2$. El conjunt R consta de les regles de R_1 i de les regles de R_2^{-1}. El primer que fem és observar que una aplicació d'una regla de R_2^{-1} seguida d'una aplicació d'una regla de R_1 es poden commutar. És a dir, que si tenim $u \to_{R_2^{-1}} v \to_{R_1} w$, llavors existeix v' tal que $u \to_{R_1} v' \to_{R_2^{-1}} w$. Per a veure-ho, siguin $\alpha \to X$ i $Y \to \beta$ les corresponents regles de R_2^{-1} i R_1 aplicades. Llavors la derivació anterior és de la forma o bé $u = u_1\alpha u_2 Y u_3 \to_{R_2^{-1}} u_1 X u_2 Y u_3 \to_{R_1} u_1 X u_2 \beta u_3 = w$, o bé $u = u_1 Y u_2 \alpha u_3 \to_{R_2^{-1}} u_1 Y u_2 X u_3 \to_{R_1} u_1 \beta u_2 X u_3 = w$. Podem construir la derivació que volem, així: o bé $u = u_1 \alpha u_2 Y u_3 \to_{R_1} u_1 \alpha u_2 \beta u_3 \to_{R_2^{-1}} u_1 X u_2 \beta u_3 = w$, o bé $u = u_1 Y u_2 \alpha u_3 \to_{R_1} u_1 \beta u_2 \alpha u_3 \to_{R_2^{-1}} u_1 \beta u_2 X u_3 = w$. Un cop que hem vist que podem commutar els passos de reescriptura, concloem que, a partir de la

derivació $S_1 \to_R^* S_2$ en podem construir una de la forma $S_1 \to_{R_1}^* w \to_{R_2^{-1}}^* S_2$. La paraula w no pot tenir variables de G_1 perque G_1 i G_2 no comparteixen variables, i per tant les regles de R_2^{-1} no les podrien eliminar, de manera que no podriem arribar a S_2. Anàlogament es pot veure que w no pot contenir variables de G_2. Per tant, w és un mot terminal arribable desde S_1 amb R_1 i desde S_2 amb R_2. Això implica que G_1 i G_2 generen el mot comú w, i això conclou la prova.

(b) Ara volem donar una reducció del segon problema (accessibilitat de mots) al primer (intersecció no buida). Això és el que es fa a continuació.

Sigui $\langle R = \langle u_1 \to v_1, \ldots, u_n \to v_n \rangle, u, v \rangle$ una entrada d'accessibilitat de mots. Per simplificar suposem que R conté la regla $u \to u$. Si no hi és, la podem afegir sense problemes, ja que no modifica que v sigui accessible des de u. Gràcies a aquesta regla comodí podrem afirmar tranquilament que, si v és accessible des de u, llavors existeix una derivació de u a v amb almenys un pas de reescriptura.

La idea de la reducció que intentarem és la següent. A partir de R, u i v, obtindrem dues gramàtiques G_1, G_2 tals que, donada una derivació per reescriptura $u = s_1 \to_R s_2 \to_R \ldots \to s_m = v$ amb $m \geqslant 2$, G_1 i G_2 podran generar el mot comú $s_1\#s_2\#s_3\#\ldots s_m\#s_m^r\ldots\#s_3^r\#s_2^r\#s_1^r$, on $\#$ és un símbol nou, i l'exponent r significa el revessat. De fet, G_1 generarà mots qualssevols de la forma $w_1\#w_2\#w_3\#\ldots\#w_k\#w_k^r\#\ldots\#w_3^r\#w_2^r\#w_1^r$ amb $k \geqslant 2$ complint que w_1 és u, i w_k és v. Per altra banda, G_2 generarà mots qualssevols de la forma $w_1\#w_2\#w_3\#\ldots\#w_k\#w_k'\#\ldots\#w_3'\#w_2'\#w_1'$ amb $k \geqslant 2$ complint que amb un pas de reescriptura sobre cada w_i amb $i < k$ es pot obtenir ${w_{i+1}'}^r$. Tot plegat garantitzarà que tindran intersecció no buida si i només si v és accessible des de u. Per simplificar suposem que l'alfabet és $\{a, b\}$. Llavors G_1 és:

$S \to u\#X\#u^r$
$X \to v\#v^r \mid Y$
$Y \to aYa \mid bYb \mid \#X\#$

Només falta per obtenir la descripció d'una gramàtica G_2 adequada, i això és el que heu de fer (no cal justificació i val 1 punt).

Resposta:

$S \to S'\#X$
$X \to aX \mid bX \mid \lambda$
$S' \to aS'a \mid bS'b \mid u_1S''v_1^r \mid \ldots \mid u_nS''v_n^r$
$S'' \to aS''a \mid bS''b \mid \#X\# \mid\#S'\#$