Resposta alternativa a l'últim apartat: Podriem construïr un autòmat que reconegui aquest llenguatge a base de recordar, en cada estat, en quines de les últimes $2^n + 1$ lectures hi hem vist una 'a'; i acceptant en el cas en que el de la posició $2^n + 1$ començant pel final contingui una a. Això requereix 2^{2^n+1} estats.

Però aquesta idea ja dona l'autòmat mínim, com justifiquem a continuació. En un exercici del curs haviem vist que un llenguatge de la forma $\{a, b\}^* a\{a, b\}^t$ requereix almenys de 2^{t+1} estats. Com que en el nostre cas t és 2^n, resulta que necessitem 2^{2^n+1} estats.

Resposta alternativa a l'últim apartat: Podriem construïr un autòmat que reconegui aquest llenguatge a base de recordar, en cada estat, en quines de les últimes $2^n + 1$ lectures hi hem vist una 'a'; i acceptant en el cas en que el de la posició $2^n + 1$ començant pel final contingui una a. Això requereix 2^{2^n+1} estats.

Però aquesta idea ja dona l'autòmat mínim, com justifiquem a continuació. Suposem que existeix un autòmat A que reconeix el llenguatge i amb menys estats que 2^{2^n+1}, i arribarem a una contradicció. Considerem totes les paraules diferents de mida $2^n + 1$. D'aquestes n'hi ha 2^{2^n+1}. Com que n'hi ha més que estats, dues ens han de portar al mateix estat començant l'execució desde l'estat inicial. Siguin x i y dues paraules diferents de mida $2^n + 1$ que porten al mateix estat q. Sigui z el prefix comú més llarg de x i y. Donat que x i y són diferents, llavors són de la forma, o bé $x = zax'$ i $y = zby'$, o bé $x = zbx'$ i $y = zay'$. Sense pèrdua de generalitat considerem només el primer cas (el segon es tracta de manera anàloga), i per tant $x = zax'$ i $y = zby'$. Donat que aquestes dues paraules ens porten al mateix estat, les paraules $zax'a^{2^n-|x'|}$ i $zby'a^{2^n-|x'|}$ (noteu que $|x'| = |y'|$) també ens porten al mateix estat, que serà acceptador o no. Per tant, o bé s'accepta una paraula com $zby'a^{2^n-|x'|}$ que no conté una a en la posició $2^n + 1$ abans del final, o bé es rebutja una paraula com $zax'a^{2^n-|x'|}$ que sí conté una a en la posició $2^n + 1$ abans del final: contradicció amb el fet que l'autòmat A reconeix el llenguatge.

(Examen de Gener-2007)

1. (7 punts en total) Considereu els llenguatges següents:

   - $L_1 = \{0^{n_1}10^{n_2}1 \ldots 0^{n_{k-1}}10^{n_k} \mid (\forall i : (1 \leqslant i \leqslant k \Rightarrow n_i \geqslant 0)) \wedge (\exists i, j : (1 \leqslant i < j \leqslant k \ \wedge\ i + 2 \neq j \ \wedge\ n_i = n_j))\}$
   - $L_2 = \{0^i 1 0^j 1 0^k \mid i = j \vee j = k\}$
   - $L_3 = \{w_1 1 w_2 1 w_3 \mid w_1, w_2, w_3 \in \{a, b, c\}^* \wedge (|w_1| = |w_2| \vee |w_2| = |w_3|)\}$
   - $L_4 = \{a^i 1 b^j 1 c^k \mid i = j \vee j = k\}$

   Responeu raonadament a les qüestions següents. Tingueu en compte que, tot i estar relacionades, es poden resoldre de manera independent. Quan us convingui, podeu suposar que els apartats anteriors han estat ja demostrats. De manera indicativa, la segona i penúltima qüestions són les més difícils.

   (a) (0.5 punts) Quins dels mots següents són de L_1? 0, 1, 010, 01010, 010010, 010010100 (en aquest apartat no cal que justifiqueu la resposta).

   (b) (2 punts) Doneu una gramàtica que generi L_1.

   (c) (0.75 punts) Definiu formalment un llenguatge regular L_2' tal que $L_1 \cap L_2' = L_2$, i trobeu l'autòmat mínim que el reconeix.

   (d) (0.75 punts) Definiu un morfisme σ tal que $\sigma^{-1}(L_2) = L_3$. Amb el σ que heu trobat, quines inclusions es compleixen d'entre $\sigma(L_4) \subseteq L_2$ i $L_2 \subseteq \sigma(L_4)$?

(e) (0.5 punts) Definiu, amb una expressió regular, un llenguatge regular L'_4 tal que $L_3 \cap L'_4 = L_4$ (no cal que en trobeu cap autòmat ni que justifiqueu aquest apartat).

(f) (0.75 punts) Als exercicis del curs vàrem veure que el llenguatge $\{a^i b^j c^k \mid i = j \vee j = k\}$, de definició molt semblant a L_4, era inherentment ambigu. La demostració es pot adaptar molt fàcilment per veure que el propi L_4 també és inherentment ambigu, i per tant, suposem aquest fet com a cert. Utilitzeu això, i el fet que els llenguatges incontextuals no inherentment ambigus són tancats per intersecció amb regulars i per morfisme invers, per justificar que L_1 també és inherentment ambigu.

(g) (1 punts) Justifiqueu que els llenguatges regulars són no ambigus (en aquesta pregunta no cal una justificació massa detallada; n'hi ha prou amb què doneu una idea de com es podria demostrar).

(h) (0.75 punts) Demostreu que el llenguatge L_1 no és regular.

Resposta:

(a) Els mots 1, 010, 01010 són de L_1, mentre que els mots 0, 010010, 010010100 no ho són.

(b) Intuïtivament, els mots de L_1 estan formats per grups de 0's separats per 1's. La condició $(\exists i,j : (1 \leqslant i < j \leqslant k \ \wedge\ i+2 \neq j \ \wedge\ n_i = n_j))$ ens demana que hi hagi dos grups de 0's que coincideixin en nombre de 0's, i que no quedi enmig d'ells exactament un altre grup de 0's.

Així doncs, qualsevol mot de L_1 és de la forma $x0^k y 0^k z$, on x, y i z satisfan les condicions següents. El mot x és o bé λ o bé qualsevol mot acabat en 1. El mot z és o bé λ o bé qualsevol mot que comença per 1. El mot y és o bé 1 o bé comença i acaba per 1 i té almenys tres 1's. Així garantitzem que els dos 0^k no estan separats per exactament un grup de 0's.

Per tant, una possible gramàtica per a L_1 és la següent:

$$\begin{array}{lcccccccc} S & \to & Y & \mid & A1Y & \mid & Y1A & \mid & A1Y1A \\ Y & \to & 0Y0 & \mid & 1 & \mid & 1A1A1 & & \\ A & \to & 0A & \mid & 1A & \mid & \lambda & & \end{array}$$

Com a resposta alternativa podem obtenir una gramàtica lineal:

$$\begin{array}{lcccccc} S & \to & S_2 & \mid & X & & \\ X & \to & 0X & \mid & 1X & \mid & 1S_2 \\ & & & & & & \\ S_2 & \to & S_3 & \mid & Z & & \\ Z & \to & Z0 & \mid & Z1 & \mid & S_3 1 \\ & & & & & & \\ S_3 & \to & 0S'_3 & \mid & 1 & \mid & 1S_4 \\ S'_3 & \to & S_3 0 & & & & \\ & & & & & & \\ S_4 & \to & 0S_4 & \mid & 1S_5 & & \\ S_5 & \to & 0S_5 & \mid & 1S_5 & \mid & 1 \end{array}$$

(c) El llenguatge $L'_2 = 0^*10^*10^*$ intersecat amb L_1 ens escull de L_1 els mots amb exactament tres grups de 0's. La resta de condicions de L_1 garanteixen que almenys dos grups de 0's consecutius coincidiran en nombre de 0's. Per tant $L_1 \cap L'_2 = L_2$. Una definició alternativa en notació clàssica de conjunts és $L'_2 = \{0^i 1 0^j 1 0^k\}$.

L'autòmat mínim que reconeix aquest llenguatge és prou senzill com per a escriu-re'l directament:

	0	1
q_0	q_0	q_1
q_1	q_1	q_2
$\dagger q_2$	q_2	q_r
q_r	q_r	q_r

És mínim perque el mot λ distingeix a q_2 de tots els altres estats, el mot 1 distingeix a q_0 de q_1, i a q_1 de q_r, i el mot 11 distingeix a q_0 de q_r.

(d) **El morfisme σ definit per $\sigma(a) = \sigma(b) = \sigma(c) = 0$ i $\sigma(1) = 1$ satisfà $\sigma^{-1}(L_2) = L_3$ com justifiquem a continuació.**

Tot mot de L_3 és de la forma $w_1 1 w_2 1 w_3$ amb $w_1, w_2, w_3 \in \{a, b, c\}^*$, i té com a imatge $0^{|w_1|}10^{|w_2|}10^{|w_3|}$ per σ. A més, cumpleix la condició $(|w_1| = |w_2| \vee |w_2| = |w_3|)\}$, que és precisament la que es requereix per tal que aquesta imatge sigui de L_2; així doncs, $L_3 \subseteq \sigma^{-1}(L_2)$.

Per altra banda, qualsevol mot w de $\sigma^{-1}(L_2)$ té com a imatge un mot $0^i10^j10^k$ de L_2 cumplint en particular $i = j \vee j = k$. Donat que σ preserva els 1's i no n'afegeix de nous, w ha de ser de la forma $w_1 1 w_2 1 w_3$ on $\sigma(w_1) = 0^i$, $\sigma(w_2) = 0^j$ i $\sigma(w_3) = 0^k$. Per la definició de σ, $w_1, w_2, w_3 \in \{a, b, c\}^*$, i a més $(|w_1| = |w_2| \vee |w_2| = |w_3|)\}$, de manera que $w \in L_3$.

Finalment, qualsevol paraula de la forma $a^i1b^j1c^k$ es transforma en $0^i10^j10^k$. Com que a més a més tenim la mateixa condició $i = j \vee j = k$ tant en L_2 com en L_4, resulta que $\sigma(L_4) = L_2$.

(e) **El llenguatge $L_4' = a^*1b^*1c^*$ intersecat amb L_3 ens escull de L_3 les paraules de L_4.**

(f) **Ho fem per reducció a l'absurd. Suposem que L_1 no és inherentment ambigu. Per l'apartat 1b sabem que L_1 és incontextual. Donat que els incontextuals no inherentment ambigus són tancats per intersecció amb regulars, per l'apartat 1c resulta que L_2 és incontextual i no inherentment ambigu. Donat que els incontextuals no inherentment ambigus són tancats per morfisme invers, per l'apartat 1d resulta que L_3 és incontextual i no inherentment ambigu. Donat que els incontextuals no inherentment ambigus són tancats per intersecció amb regulars, per l'apartat 1e resulta que L_4 és incontextual i no inherentment ambigu. Això últim contradiu el fet que, com es suposava en aquest apartat, L_4 és inherentment ambigu.**

(g) **Els llenguatges regulars es poden reconèixer amb autòmats deterministes. Els autòmats deterministes poden ser simulats per una gramàtica lineal. La transformació usual ja assegura que a tota execució amb una paraula d'entrada acceptada per l'autòmat li correspon una i només una derivació amb la gramàtica. També assegura que la gramàtica no genera més paraules que les reconegudes per l'autòmat. Tot plegat implica que la corresponent gramàtica lineal és no ambigua. (Per a més informació consulteu la pàgina 143 dels apunts de llenguatges incontextuals i regulars de la assignatura).**

(h) **Per l'apartat 1f tenim que L_1 és inherentment ambigu. Per l'apartat 1g sabem que tot llenguatge regular és no inherentment ambigu. Tot plegat implica que L_1 no és regular.**

Com a resposta alternativa ho podem fer pel lema de bombament. Fixat N, escollim la paraula 0^N10^N que és de L_1. Qualsevol subparaula xyz cumplint $|xy| \leqslant N$ i $|y| \geqslant 1$ satisfà que x és de la forma 0^i, que y és de la forma 0^j amb $j \geqslant 1$, i que z és de la forma $0^{N-i-j}10^N$. Llavors, la paraula xy^2z no és de L_1.

2. (1.5 punts) Classifiqueu com a decidible, indecidible però semi-decidible, o no semi-decidible, el llenguatge $L = \{x \mid \forall y : (\mathsf{M}_x(y)\downarrow \Rightarrow \mathsf{M}_x(y) = 1)\}$.

Resposta:

Pel teorema de Rice, donat que L és un conjunt d'índexos, que no és ni el total ni el buit, i que conté els nombres que representen la funció totalment indefinida, resulta que L no és ni semi-decidible.

Més en detall, L és un conjunt d'indexos perque en la condició, sempre que apareix x, ho fa amb el contexte M_x, és a dir, és una condició només sobre el comportament del programa M_x, independentment de la seva implementació. L no és buit perque conté, per exemple, a qualsevol nombre de programa que sempre dona sortida 1. L no és el total perque no conté per exemple als nombres de programes que donen sortida 2 per a alguna entrada.

Com a resposta alternativa podem fer la següent reducció des de $\overline{\mathtt{K}}$. Si x és l'element d'entrada, donem com a sortida el programa p següent:

```
entrada y
   Simular Mx(x)
   sortida 2
```

Si x pertany a $\overline{\mathtt{K}}$, llavors $\mathrm{M}_x(x)\uparrow$, de manera que el programa p no s'atura, i llavors la implicació ($\mathsf{M}_x(y)\downarrow\Rightarrow \mathsf{M}_x(y)=1$) és certa per tot y perque l'antecedent és fals per tot y, de manera que p pertany a L.

En canvi, si x no pertany a $\overline{\mathtt{K}}$, llavors $\mathrm{M}_x(x)\downarrow$, de manera que el programa p s'atura amb totes les entrades i dóna sortida 2, així que la implicació ($\mathsf{M}_x(y)\downarrow\Rightarrow \mathsf{M}_x(y)=1$) és falsa per tot y i, per tant, p no pertany a L.

3. (1.5 punts) L'objectiu d'aquest problema és demostrar que el llenguatge que descrivim a continuació és indecidible. Donada una llista finita R de regles de reescriptura (un conjunt finit de parells de mots (u_i, v_i)) i un mot u (inicial i final), determinar si és possible passar de u a u aplicant una o més regles de reescriptura (un seguit d'una o més substitucions d'algun submot u_i pel seu corresponent v_i). Formalment, el llenguatge associat a aquest problema és el següent:

$$\{\langle R,u\rangle \ \mid\ \exists k\geqslant 2\ \exists w_1,\dots w_k : (u=w_1 \underset{R}{\to} w_2 \underset{R}{\to} w_3 \dots \underset{R}{\to} w_k = u)\}$$

   (a) (0.75) Justifiqueu primer que la reducció $\langle R,u,v\rangle \ \mapsto\ \langle R\cup\{v\to u\},u\rangle$ des del problema de `Mots` no funciona, mitjançant un contraexemple. (És a dir, que no és cert que una entrada del problema dels mots té resposta afirmativa si i només si la seva transformada en una entrada del problema que hem enunciat també en té).

   (b) (0.75) Doneu una nova reducció des del problema de `Mots` i justifiqueu que és correcta.

Resposta:

(a) Hi ha moltes maneres de respondre a aquest apartat. Per exemple, la transformació anterior aplicada sobre $\langle\{a\to aaa\},a,aa\rangle$ donaria lloc a $\langle\{a\to aaa, aa\to a\},a\rangle$. El primer té resposta negativa per al problema dels mots, mentre que el segon té resposta afirmativa per al llenguatge que s'ha descrit aquí.

O també, i més senzill, sobre $\langle\{a\to a\},a,aa\rangle$ s'obtindria $\langle\{a\to a, aa\to a\},a\rangle$ obtenint resposta negativa en el primer cas i positiva en el segon.

(b) Una possible reducció seria: $\langle R,u,v\rangle \ \mapsto\ \langle R',\#\rangle$, amb R' definit com $R\cup\{\#\to \$u\$\ ,\ \$v\$\to\#\}$, on $\#$ i $\$$ són símbols nous (que no apareixen a R ni u ni v).
Si $\langle R,u,v\rangle$ és del problema de `Mots`, llavors $u \xrightarrow[R]{*} v$, i també $\# \xrightarrow[\#\to\$u\$]{} \$u\$ \xrightarrow[R]{*} \$v\$ \xrightarrow[\$v\$\to\#]{} \#$, de manera que $\langle R',\#\rangle$ satisfà la condició de l'enunciat.

Per altra banda, si $\langle R', \# \rangle$ satisfà la condició de l'enunciat, llavors $\# \xrightarrow[R']{+} \#$. Escollim aquí la derivació més curta de la forma $\# \xrightarrow[R']{+} \#$, de manera que cap paraula intermedia és $\#$. Per les regles que tenim al sistema de rescriptura, aquesta derivació ha de ser de la forma $\# \xrightarrow[\# \to \$u\$]{} \$u\$ \xrightarrow[R']{*} \$v\$ \xrightarrow[\$v\$ \to \#]{} \#$. Es pot veure inductivament que en la derivació intermedia $\$u\$ \xrightarrow[R']{*} \$v\$$ totes les paraules són de la forma $\$w\$$, on w no conté $\#$ ni $\$$: si en un cert punt abans del final es cumpleix la condició, llavors no es pot fer servir $\# \to \$u\$$ per a obtenir la següent paraula, però tampoc $\$v\$ \to \#$, ja que si no $\#$ seria paraula intermedia. De l'anterior també es dedueix que totes les regles que s'apliquen en aquesta subderivació són de R. Per tant $\$u\$ \xrightarrow[R]{*} \$v\$$, i com que les regles de R no contenen $\$$, tambe tenim que $u \xrightarrow[R]{*} v$, de manera que $\langle R, u, v \rangle$ és del problema de `Mots`.