Сега ще посочим един алгоритъм за намиране на регулярен израз, който съответствува на езика разпознаван от даден автомат (независимо какъв е той КДА или НДА). Преди това ще определим в каква форма да се намира съответния автомат. Тази форма ще наричаме за сега „специална" форма. Без ограничение на общността можем да считаме, че този автомат M е НДА.

M се намира в „специална форма", ако:

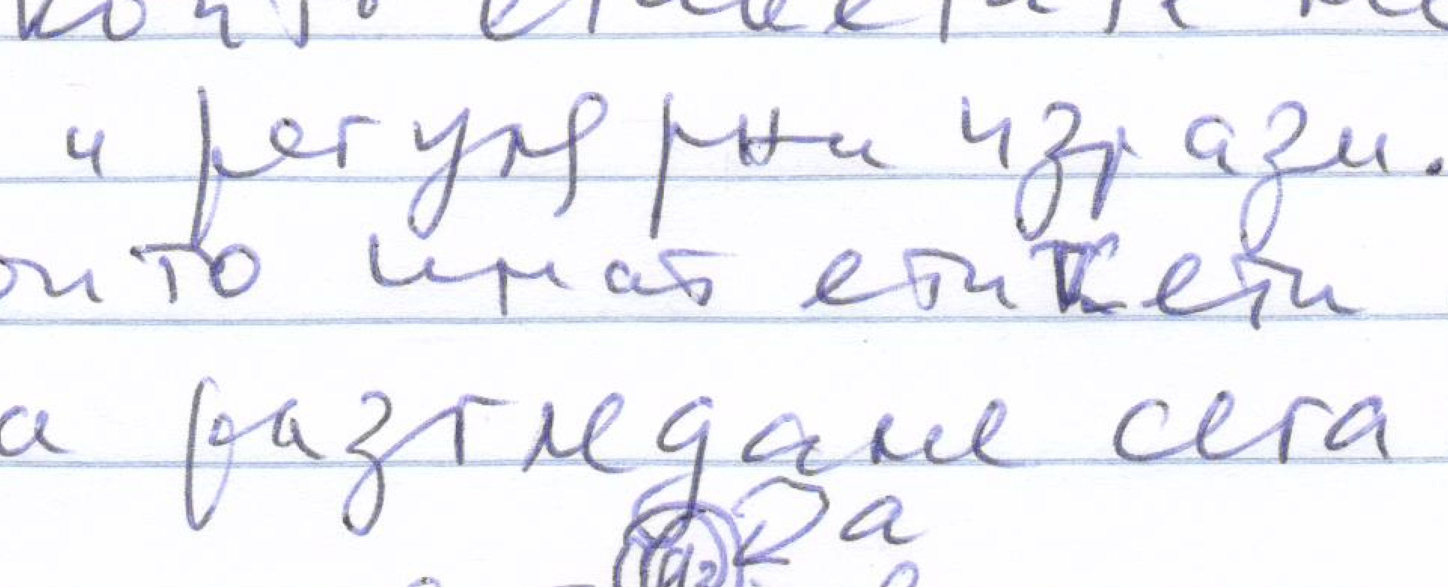
а) M има единствено заключително състояние, т.е. $F=\{f\}$;

б) Ако $(q,u,p)\in\Delta$, то $q\neq f$ и $p\neq s$, т.е. няма стрелка, която влиза в нач. състояние s и няма стрелки, която излиза от заключ. състояние f;

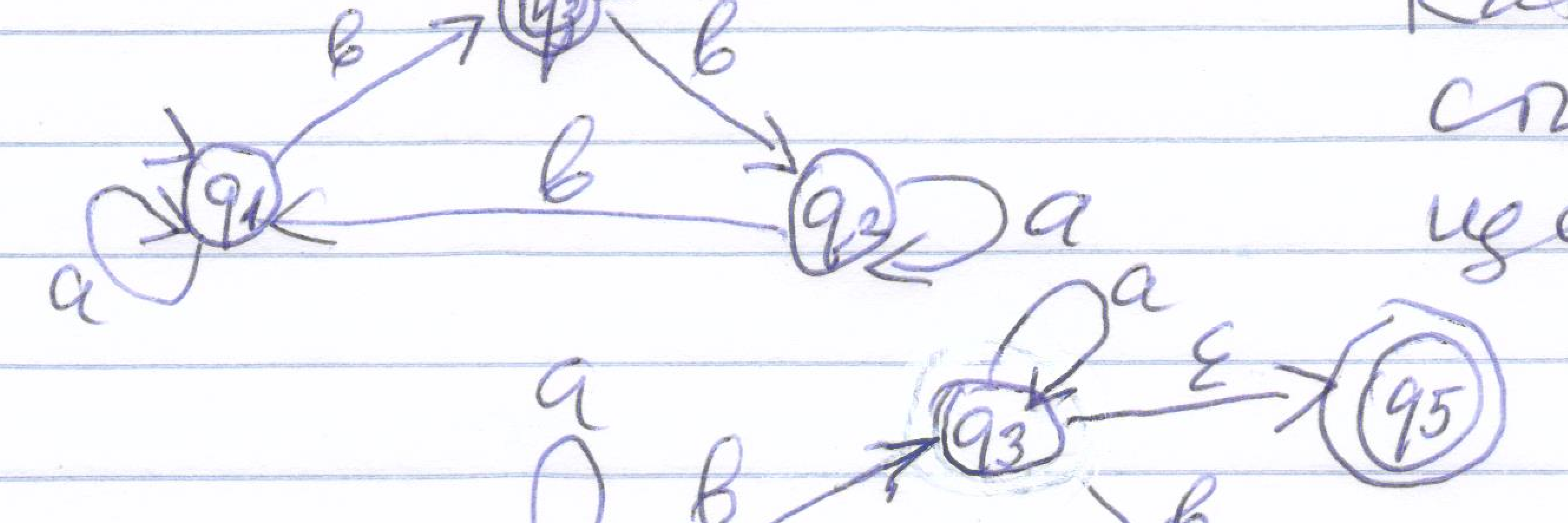
в) За удобство освен това ще считаме, че ако $K=\{q_1,\dots,q_n\}$, то $s=q_{n-1}$ и $f=q_n$.

След малко ще видим, че всеки автомат може да се сведе много лесно до такъв в специална форма. Най-напред ще пресметнем $R(i,j,0)$, след това $R(i,j,1)$ и т.н. докато най-накрая получим $R(n-1,n,n)$, който ще е искания от нас регулярен израз, чийто език се разпознава от разглеждания автомат. На всеки етап от построяването на изразите $R(i,j,k)$ ще бъде етикет на ребро в така наречения „обобщен краен автомат" в който етикетите могат да бъдат не само $\Sigma\cup\{\varepsilon\}$ но и регулярни изрази. Ще изпускаме стрелки, които имат етикети ϕ или $\{\varepsilon\}$.

Да разгледаме сега примера на КДА:

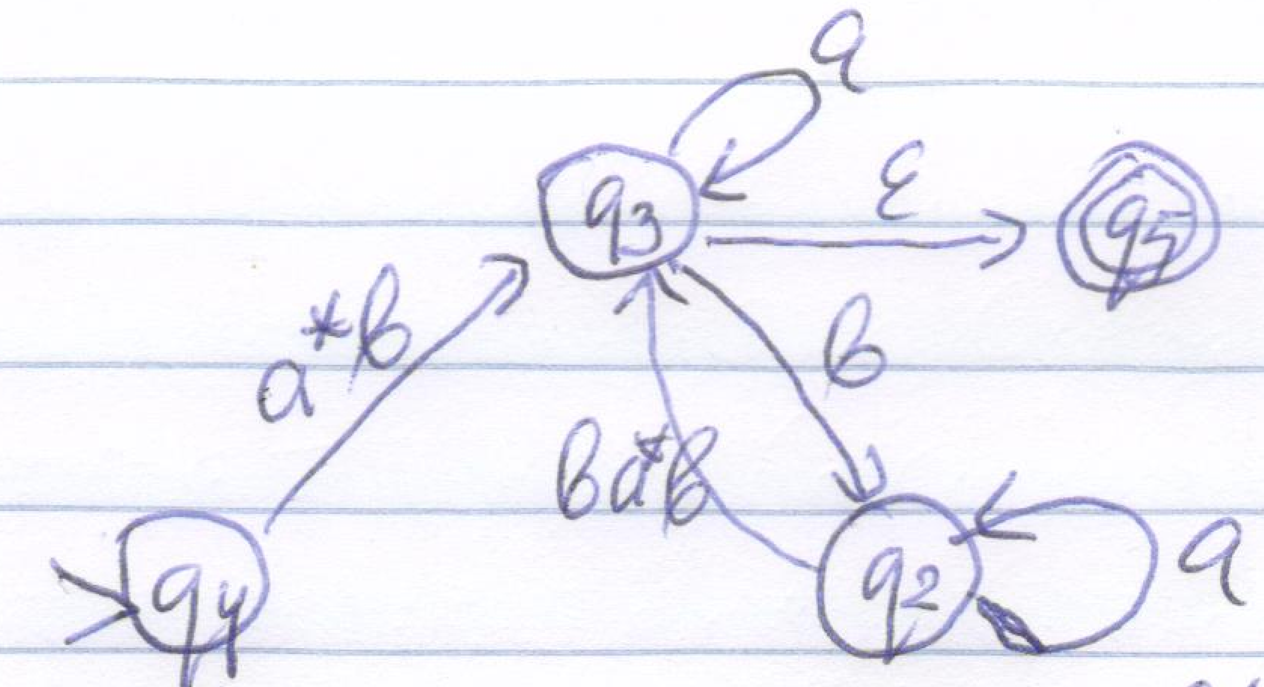


Като го приведем в специална форма, той ще изглежда така:

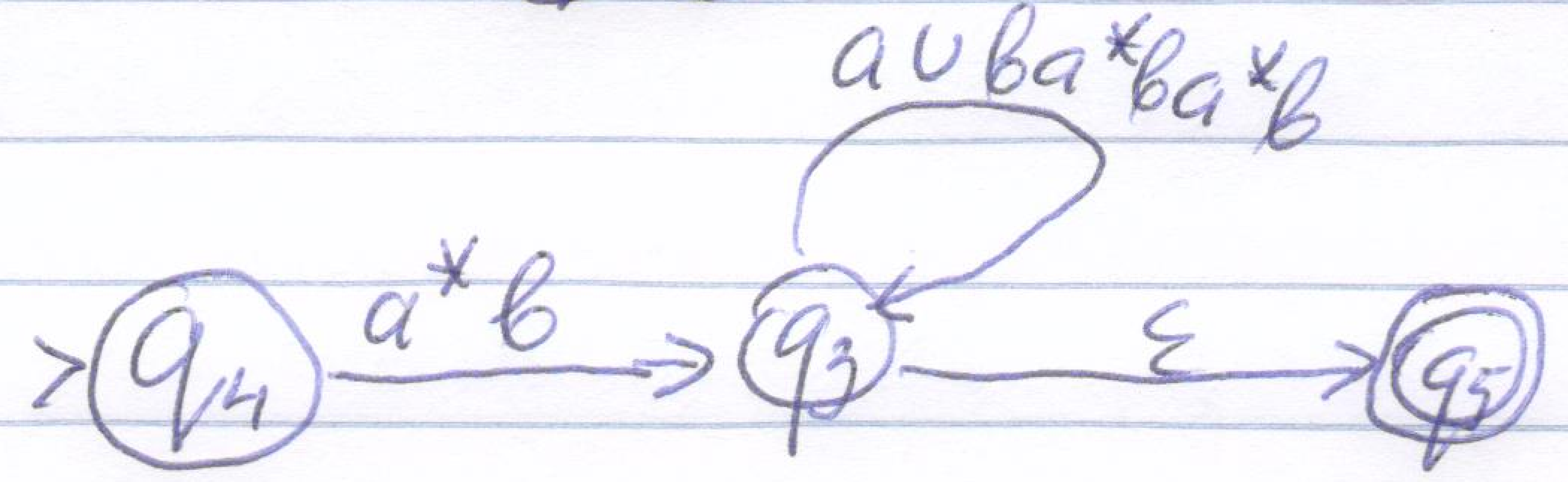


I ст.

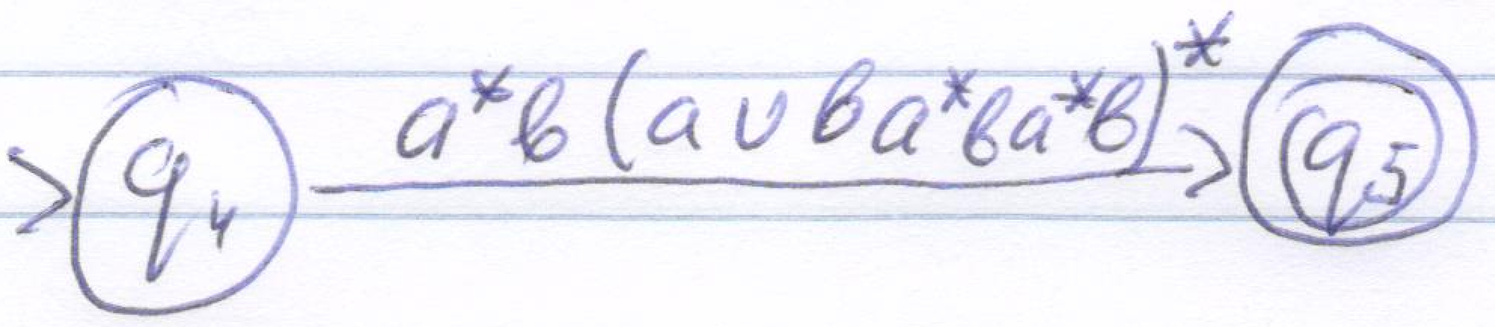
II ст. (елиминиране) на q_1



III ст. (ел. на q_2)

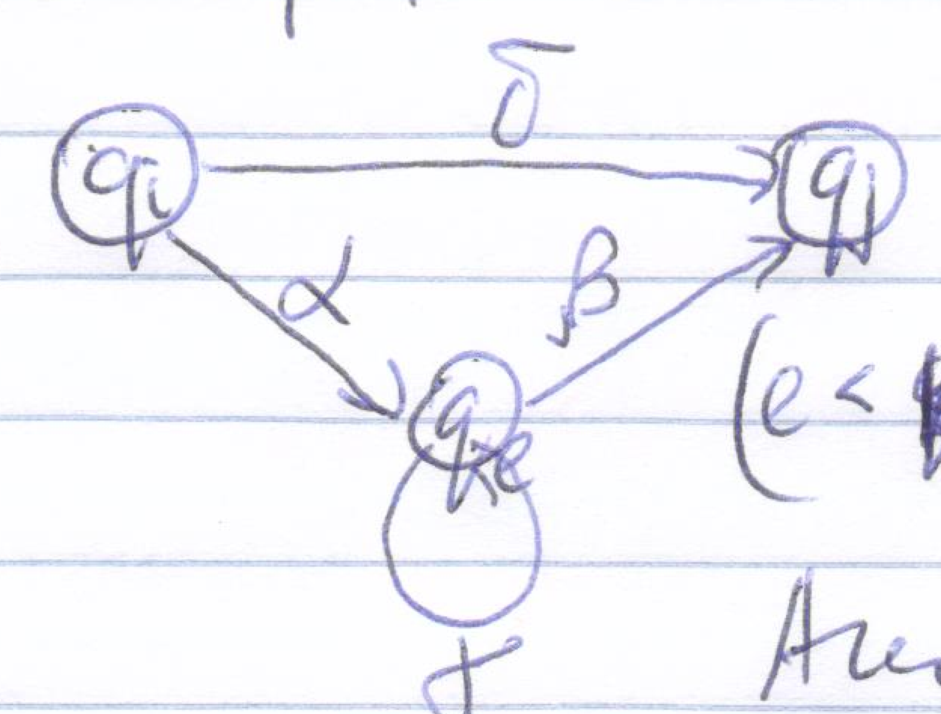


IV ст. (ел. на q_3)

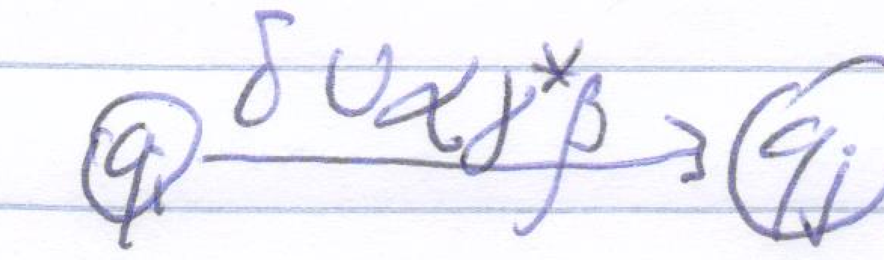


$R(4,5,5) = a^*b(a \cup ba^*ba^*b)^* = L(M)$

Да обърнем внимание, че там където имаме



($l < k$, q_l се елиминира)



Ако γ го няма, то $\gamma = \emptyset$ и $\gamma^* = \{\varepsilon\}$ и $\alpha\gamma^*\beta = \alpha\beta$.

Така за всеки автомат M можем да намерим регулярен израз α, такъв, че $\mathcal{L}[\alpha] = L(M)$.

Лема за Разрастването за КДА

Теорема (Лема за разрастването) Нека L е регулярен език. Тогава съществува такова естествено число $n \geq 1$, че за всяка дума $w \in L$, такава, че $|w| \geq n$ съществуват думи x, y, z, такива, че $w = xyz$, $y \neq \varepsilon$, $|xy| \leq n$ и за всяко $i \in \mathbb{N}$ е изпълнено, че $xy^iz \in L$.

Д-во. Тъй като L е регулярен, то съществува КДА $M = (K, \Sigma, \delta, s, F)$, който разпознава L. Да положим $n = |K|$ и нека $w \in L$, $|w| \geq n$. Следователно, ако $q_0 = s$ и $|w| = m \geq n$ то съществува извод $(q_0, w) \vdash_M (q_1, w_1) \vdash_M \dots \vdash (q_m, \varepsilon)$, $q_m \in F$. и $w = a_1 \dots a_m$, $a_i \in \Sigma$. Тогава състоянията $q_0, \dots, q_m$ са $n+1$ а $|K| = n$. Следователно съществуват поне две различни i, j, такива че $q_i = q_j$. Да си мислим, че $i < j$, $1 \leq i < j \leq n$.

Да означим с $x = a_1 \dots a_i$, $y = a_{i+1} \dots a_j$, $z = a_{j+1} \dots a_m$. Очевидно е, че $y \neq \varepsilon$ $(i<j)$, $|xy| = |a_1 \dots a_j| = j \leq n$. Остава да покажем, че $xy^k z \in L$ за произволно естествено k. Ще го докажем с индукция относно k. Да разпишем по-подробно стъпките на извода $(q_0, w) \vdash_M^* (q_m, \varepsilon)$, а именно
$(q_0, w) \vdash_M (q_i, a_{i+1} \dots a_m) \vdash_M^* (q_j, a_{j+1} \dots a_m) \vdash_M^* (q_m, \varepsilon)$, $q_m \in F$ и $q_i = q_j$. С други думи

$$(q_0, xz) = (q_0, a_1 \dots a_i a_{j+1} \dots a_m) \vdash_M^* (\overset{q_j}{\overset{\shortparallel}{q_i}}, a_{j+1} \dots a_m) \vdash_M^* (q_m, \varepsilon)$$

т.е. $xz = xy^0 z \in L$.
Допускаме, че тв. е вярно за някое k. Ще го докажем за $k+1$. Наистина $(q_0, xy^{k+1}z) \vdash_M^* (q_i, y^{k+1}z) \vdash_M^* (q_i, y^k z) \vdash_M^* (q_i, y^{k-1}z) \vdash_M^* \dots \vdash_M^* (q_m, \varepsilon)$, т.е. $xy^{k+1}z \in L$, защото $q_m \in F$.
С това лемата за разрастването е доказана.
След като доказахме теоремата (лемата за разрастването) ще обърна внимание на сложността на лемата за разрастването (формулировката). Обикновенно, за логическата сложност на едно твърдение се съди по броя на смяната на кванторите в него. Същото разбира се отнася и за дефинициите. В лемата за разрастването имаме 5 алтернативно сменящи се квантори:

За всеки регулярен език L,
Съществува естествено число $n \geq 1$, такова че
За всяка дума $w \in L$, такава, че $|w| \geq n$
Съществуват думи x, y, z: $w = xyz$, $y \neq \varepsilon$, $|xy| \leq n$
такива, че За всяко естествено число k, $xy^k z \in L$.

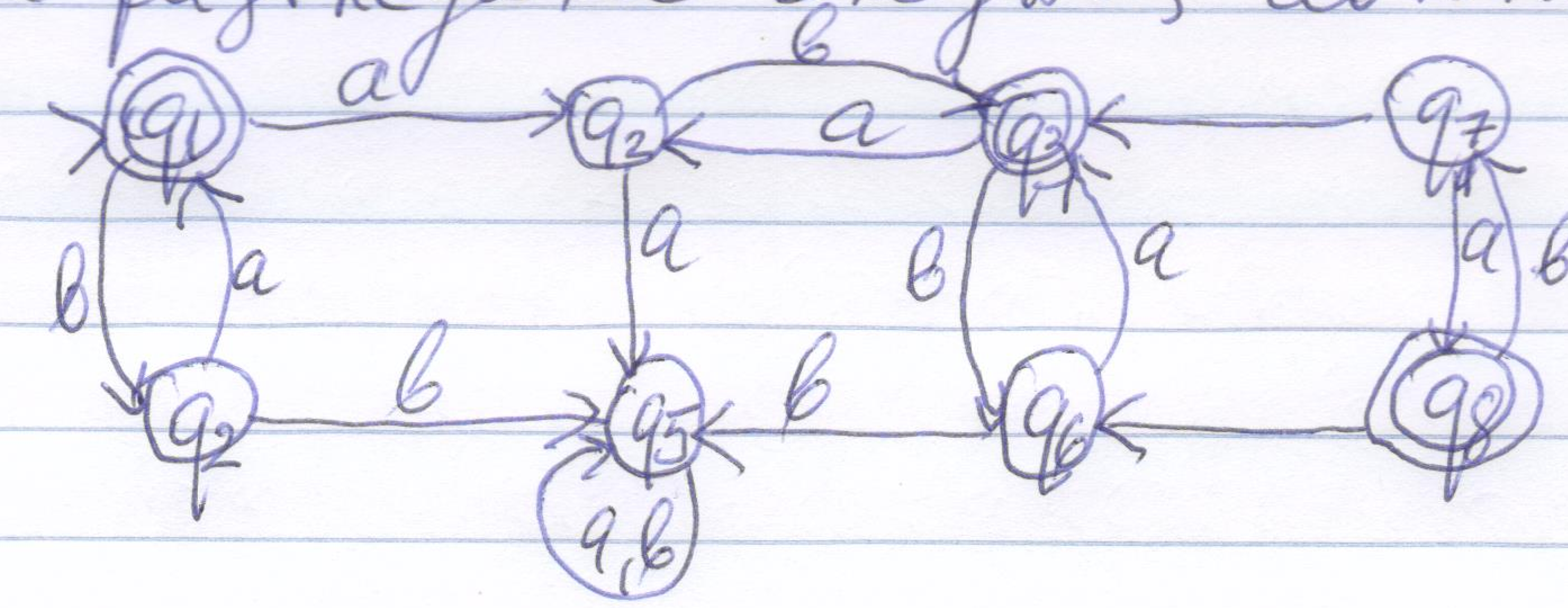
Досега най-сложните дефиниции, които сте срещали са при $\lim_{n \to \infty} a_n = l$, $\lim_{x \to x_0} f(x) = l$, където имаме 3 алтернативно сменящи се квантори.
Обръщам специално внимание на това, защото всяка промяна или неточност в кванторите променя смисъла и го превръща в неточна (грешна) формулировка или дефиниция.
Сега да се върнем на смисъла на лемата за разрастването за прилагането ѝ. Знаем, че ако искаме да проверим, че

един език е регулярен (ако очакваме, че той е такъв) ние се стремим да построим автомат, който да разпознава този език. Какво ще направим ако не можем да построим такъв автомат? Ще се опитаме да докажем, че този език не се разпознава от автомат. Как? Ами лемата за разрастването ни дава това средство.
Да разгледаме езикът $L = \{a^i b^i \mid i \in \mathbb{N}\}$, L език в азбуката Σ, която съдържа буквите a, b. Ще докажем, че L не е регулярен език. Допускаме, че той е регулярен. Тогава по лемата за разрастването същ. естествено число n, такова, че за всяко $w \in L$, $|w| \geq n$ съществуват $x, y, z \in \Sigma^*$, такива, че $|xy| \leq n$, $y \neq \varepsilon$, $w = xyz$ и за всяко ест. i е изп. $xy^i z \in L$. За точно това n разглеждаме думата $w = a^n b^n$. Тогава $|w| = 2n > n$, $w \in L$ и съгласно лемата същ. x, y, z, със изредените свойства за точно тази дума w. Тогава $x = a^k$, $y = a^l$, $z = a^m b^n$, $k \geq 0$, $l > 0$, $m \geq 0$.
Тогава за $i = 0$ имаме $xz = a^k . a^m b^n = a^{n-l} . b^n \in L$, което е невъзможно, защото $l > 0$ и $n - l \neq n$. Стигнахме до противоречие заради допускането, че L е регулярен. Следователно, L не е регулярен. На упражнения ще разгледате още примери.

Минимизация на състоянията

Задачата, която ще си поставим, е да можем по всеки КДА да намерим еквивалентен на него КДА с най-малко състояния, или, както се казва за краткост да решим задачата за минимизация на КДА.
Това няма да го направим директно, а междувременно ще докажем и някои важни теореми, като теоремата на Майхил-Нероуд (М.-Н.) и следствията ѝ, които ще дадат още един способ за разпознаване дали един език е регулярен или не.
Най-напред да започнем с минимизацията на КДА изобщо и кои са първите стъпки за това

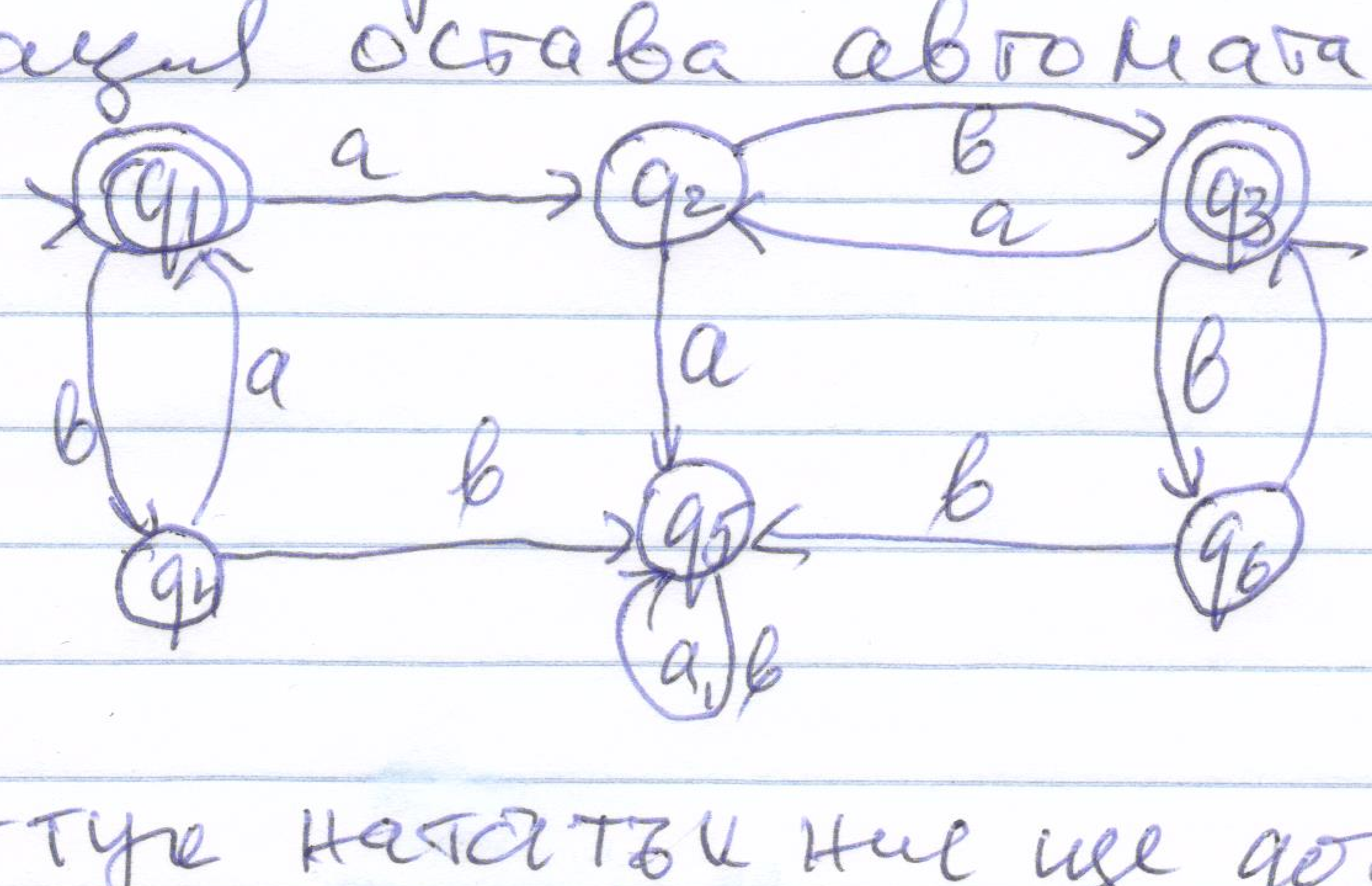
Да разгледаме следния автомат:



Както ще видим, този автомат разпознава езика $L = (ав \cup ва)^*$. Първата стъпка за минимизация е да видим дали всички състояния „се използват" или още са достижими. Едно състояние q е достижимо за даден автомат $M = \langle K, \Sigma, \Delta, s, F \rangle$, ако съществува дума w, такава, че $(s, w) \vdash_M^* (q, \varepsilon)$, т.е. ако има поредица от стрелки, започваща от s и стигаща до q. Това, което можем да забележим без особена трудност, е, че q_7 и q_8 не са достижими.
В същност множеството S от достижимите състояния можем да го дефинираме така:

$S := \{s\}$
while същ. състояние $p \in S$ и $a \in \Sigma$ такова, че
$\delta(p, a) \notin S$ do добави $\delta(p, a)$ към S.

Или да определим релацията $R = \{(p, q) \mid \delta(p, a) = q$ за някое $a \in \Sigma\}$. Тогава S е затварянето на множ. $\{s\}$ относно релацията R.
И така ние свършихме грубата работа да отхвърлим недостижимите състояния. Така за минимизация остава автомата:



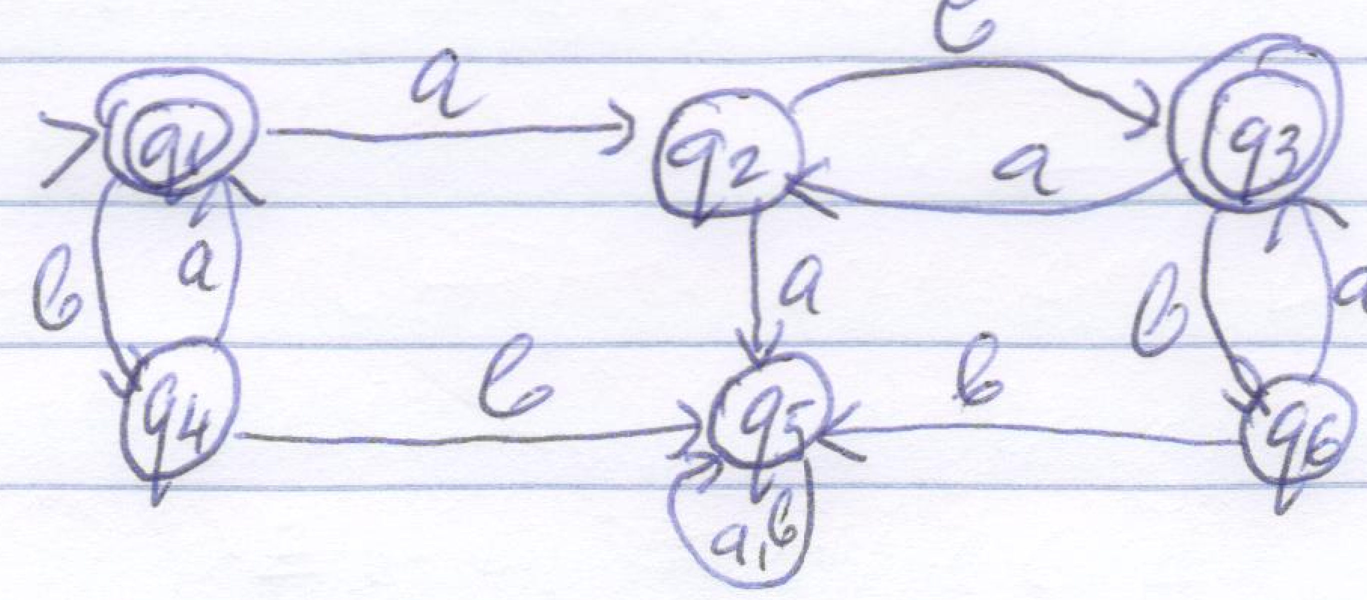
В този автомат сме оставили само достижимите състояния и с това сме завършили първия етап от минимиз

Оттук нататък ние ще докажем някои допълнителни

твърдения, необходими за алгоритъма за минимизация.
Ще започнем със свързаната дефиниция:
Д. Нека L е език в азбуката Σ и $x, y \in \Sigma^*$. Казваме, че x и y са еквивалентни относно L (пишем $x \approx_L y$), ако за всяка дума $z \in \Sigma^*$ е изпълнена еквивалентността:
$$xz \in L \Leftrightarrow yz \in L.$$
Релацията $\approx_L$ е релация на еквивалентност в Σ^*.
Наистина, очевидно е, че $x \approx_L x$ (рефлексивност) и ако $x \approx_L y$, то $y \approx_L x$ (симетричност).
Нека $x \approx_L y$ и $y \approx_L z$. Тогава за всяко $z_1 \in \Sigma^*$ е изп. $xz_1 \in L \Leftrightarrow yz_1 \in L \Leftrightarrow zz_1 \in L$, т.е. $x \approx_L z$.
Пример 1. Нека $L = (ab \cup ba)^*$. Да намерим класовете на еквивалентност отн. $\approx_L$. Това са класовете:
1) $[\varepsilon] = L$; 2) $[a] = La$; 3) $[b] = Lb$; 4) $[aa] = [bb] = L(aa \cup bb)\Sigma^*$
Да обърнем внимание, че $L = \{ab, ba\}^*$ и че $\varepsilon \in L$.
Да обосновем 1) $x \in [\varepsilon]$ т.т.к. за всяко $z \in \Sigma^*$ е изп. $z \in L \Leftrightarrow xz \in L$. В частност за $z = \varepsilon$ имаме, че $\varepsilon \in L \Leftrightarrow x \in L$, т.е. $x \in L$. С това $[\varepsilon] = L$. 2) $x \in [a]$ т.т.т. когато можем да допълним x отдясно така, че a и x едновременно да принадлежат на L. Това е възможно само ако отдясно допълним с дума от вида bL; Аналогично за 3), 4).
Да разгледаме сега още една дефиниция, която се отнася до КДА.
Д. Нека $M = (K, \Sigma, \delta, s, F)$ е КДА. Казваме, че думите x и y са еквивалентни относно M (пишем $x \sim_M y$) т.т.к съществува състояние $q \in K$, такова, че $(s, x) \vdash_M^* (q, \varepsilon)$ и $(s, y) \vdash_M^* (q, \varepsilon)$.
Очевидно е, че релацията $\sim_M$ е релация на еквивалентност. Какво представляват класовете на еквивалентност отн. $\sim_M$? Те зависят от състоянието q. С други думи всеки клас на еквивалентност е $E_q = \{x \mid (s, x) \vdash_M^* (q, \varepsilon)\}$.
Да видим кои са класовете на еквивалентност за автомата, който разгледахме, когато започнахме минимизацията.

$E_{q_1}=(ba)^*$, $E_{q_2}=La$, $E_{q_3}=abL$, $E_{q_4}=b(ab)^*$, $E_{q_5}=L(aa\cup bb)\Sigma^*$

$E_{q_6}=abLb$



Сега ще докажем едно твърдение

Тв. За всеки КДА $M=(K,\Sigma,\delta,s,F)$ и за произволни думи $x,y\in\Sigma$ е изпълнено: ако $x\sim_M y$, то $x\approx_{L(M)}y$.

Д-во. Нека $x\sim_M y$. За произволно z нека с $q(z)$ означим единственото състояние q, такова, че $(s,z)\vdash^*_M(q,\varepsilon)$. Тогава за x,y е изпълнено $q(x)=q(y)$. Трябва да докажем, че за произволно $z\in\Sigma^*$ е изп. екв. $xz\in L(M)\Leftrightarrow yz\in L(M)$.

Наистина $xz\in L(M)\Leftrightarrow(s,xz)\vdash^*_M(f,\varepsilon)$ за $f\in F\Leftrightarrow$ $(q(x),z)\vdash^*_M(f,\varepsilon)$ за $f\in F\Leftrightarrow(q(y),z)\vdash^*_M(f,\varepsilon)$ за $f\in F\Leftrightarrow$ $(s,yz)\vdash^*_M(f,\varepsilon)$ за $f\in F\Leftrightarrow yz\in L(M)$.

С това твърдението е доказано

Така в Σ^* имаме две релации на еквивалентност за даден КДА М – $\approx_{L(M)}$ и $\sim_M$. Това, че класовете на еквивалентност относно $\sim_M$ се съдържат в класовете на еквивалентност относно $\approx_{L(M)}$ означава, че релацията $\sim_M$ има повече класове на еквивалентност отколкото втората $\approx_{L(M)}$ и още казване, че $\approx_{L(M)}$ е по-груба отколкото $\sim_M$, а $\sim_M$ е по-финна отколкото $\approx_{L(M)}$.

За примера, който разглеждаме $[\varepsilon]_L=E_{q_1}\cup E_{q_3}$, $[a]_L=E_{q_2}$, $[b]_L=E_{q_4}\cup E_{q_6}$ и $[aa]_L=E_{q_5}$.

Сега ще докажем следната

Теорема (Майхил-Нероуд). Нека L е регулярен език в азбуката Σ. Тогава съществува КДА M, който разпознава езика L и има точно толкова състояния, колкото са класовете на еквивалентност относно $\approx_L$

Д-во. Нека с $[x]$ да означаване класът на еквивалентност относно $\approx_L$, който съдържа думата x, за произволно $x\in\Sigma^*$.

По дадения език L построяваме КДА $M=(K,\Sigma,\delta,s,F)$

както следва:

$K = \{[x] \mid x \in \Sigma^*\}$, $s = [\varepsilon]$, $F = \{[x] \mid x \in L\}$ и

$\delta([x], a) = [xa]$, за $x \in \Sigma^*$ и $a \in \Sigma$.

Така построеният автомат се нарича стандартен автомат за езика L.

Първият въпрос, на който трябва да си отговорим е дали K е крайно множество, защото говорим за КДА. Това е така защото L е регулярен и, следователно, той се разпознава от някой КДА и класовете на еквивалентност относно $\approx_L$ по брой са по-малко отколкото броя на състоянията на този КДА, съгласно предното твърдение. Така, че K е крайно множество.

След това трябва да отбележим, че функцията δ е добре дефинирана и не зависи от представителя, който избираме от класа $[x]$. Наистина, ако $y \in [x]$, то за всяко $z \in \Sigma^*$ е изп., че $yz \in L \Leftrightarrow xz \in L$, т.е. ако вземем произволна буква $a \in \Sigma$ и $z \in \Sigma^*$, то $xaz \in L \Leftrightarrow yaz \in L$. Следователно $xa \approx_L ya$, т.е. $\delta([x], a) = \delta([y], a)$, ако $x \approx_L y$, т.е. дефиницията е коректна.

Следващата стъпка е да покажем, че $L = L(M)$.

За тази цел ще покажем следното помощно твърдение: За произволни $x, y \in \Sigma^*$ е изпълнено

$$(*) \qquad ([x], y) \vdash_M^* ([xy], \varepsilon)$$

Това ще го направим с индукция относно д-ната $|y|$ на думата y.

1) Нека $|y| = 0$, т.е. $y = \varepsilon$. Тогава очевидно $(*)$ е налице.

2) Нека $(*)$ е изпълнено за всички думи $y \in \Sigma^*$ с д-на n и нека $|y| = n+1$, $y \in \Sigma^*$. Тогава $y = y'a$, за някоя y', $|y'| = n$ и $a \in \Sigma$. Следователно, $([x], y') \vdash_M^* ([xy'], \varepsilon)$ и $([x], y'a) \vdash_M^* ([xy'], a) \vdash_M ([xy'a], \varepsilon)$, т.е. $(*)$ е налице и за y, с което помощното тв. е доказано.

Сега за произволно $x \in \Sigma^*$ имаме $x \in L(M) \Leftrightarrow ([\varepsilon], x) \vdash_M^* ([x], \varepsilon)$ и $[x] \in F \Leftrightarrow x \in L$. С това Теоремата на М.-Н. е доказана

Доказаната теорема на М.-Н. ни дава минимален КДА, но не ни дава ефективен начин и алгоритъм за намиране на класовете на еквивалентност, както и техния брой. Въпреки това, тя ни дава възможност да разпознаваме дали един език е регулярен или не — втора такава възможност. Това ще видим след като докажем следствието от Т-мата на М.-Н..

Следствие. Един език L е регулярен т.т.к. релацията на еквивалентност $\approx_L$ има крайно много класове на еквивалентност, т.е. т.т.к. когато $\approx_L$ има краен индекс.

Д-во. Нека L е регулярен език. Тогава той се разпознава от краен автомат. Съгласно доказателството на т-мата на М.-Н. класовете на еквивалентност на $\approx_L$ са по-малко на брой от броя на състоянията на този КА, т.е. класове на екв. на $\approx_L$ са крайно много.

Обратно, нека $\approx_L$ има краен брой класове на еквивалентност. Тогава стандартния автомат за L ще разпознава този език L, откъдето L е регулярен.

Да видим как се прилага следствието.

Нека отново разгледаме езика $L=\{a^n b^n \mid n\in N\}$. Да видим, че $\approx_L$ има безкрайно много класове на еквивал. Да разгледаме редицата $[\varepsilon], [a], [aa], [aaa], \dots$

Да обясним, че $a^i \not\approx_L a^j$ за $i\neq j$. Наистина, нека да определеност $i<j$. Тогава $a^i \not\approx_L a^j$ защото $a^i b^i \in L$ но $a^j b^i \notin L$. Това показва, че класовете на еквивалентност отн. $\approx_L$ не са краен брой и L не е регулярен