TD 6: Residues and minimization

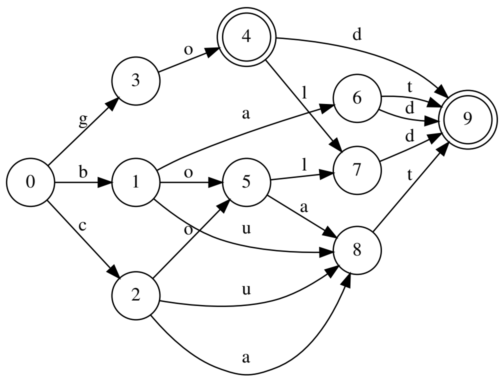
Exercice 1: Let L be the following language

bad	bat	boat	bold	but	cat
coat	cold	cut	go	god	gold

1. Write all the residues of the language L.
2. Draw the smallest deterministic automaton accepting L.

Solutions:

1. The residuals are:
   - $\varepsilon^{-1}L = \{bad, bat, boat, bold, but, cat, coat, cold, cut, go, god, gold\}$
   - $b^{-1}L = \{ad, at, oat, old, ut\}$
   - $c^{-1}L = \{at, oat, old, ut\}$
   - $g^{-1}L = \{o, od, old\}$
   - $(go)^{-1}L = \{\varepsilon, d, ld\}$
   - $(bo)^{-1}L = (co)^{-1}L = \{at, ld\}$
   - $(ba)^{-1}L = \{d, t\}$
   - $(bol)^{-1}L = (col)^{-1}L = (gol)^{-1}L = \{d\}$
   - $(boa)^{-1}L = (bu)^{-1}L = (ca)^{-1}L = (coa)^{-1}L = (cu)^{-1}L = \{t\}$
   - $(bad)^{-1}L = (bat)^{-1}L = (boat)^{-1}L = (bold)^{-1}L = (but)^{-1}L = (cat)^{-1}L = (coat)^{-1}L = (cold)^{-1}L = (cut)^{-1}L = (go)^{-1}L = (god)^{-1}L = (gold)^{-1}L = \{\varepsilon\}$
2. The automaton is:



Exercise 2: Let L be the language $\{a^n b^n \mid n \in \mathbb{N}\}$.

1. Given $k \in \mathbb{N}$ calculate the residue $(a^k)^{-1}L$.
2. Deduce that L is not regular.

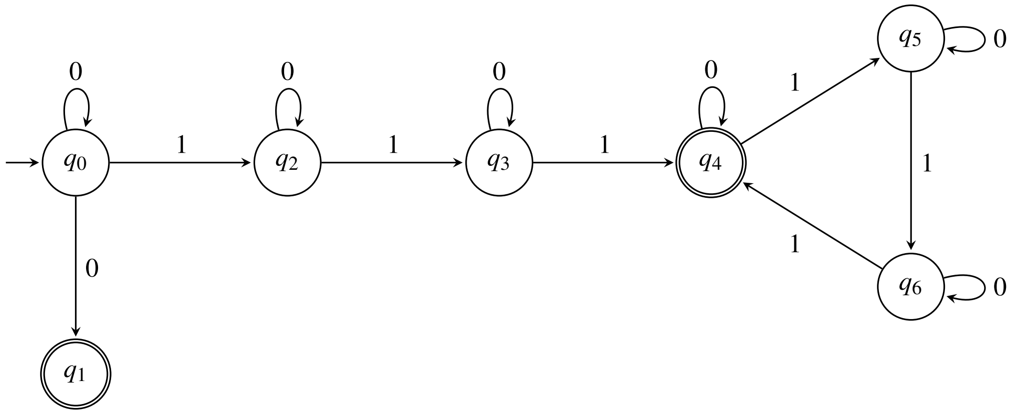
Solutions:

1. $(a^k)^{-1}L = \{a^{n-k}b^n \mid n \geq k\}$
2. By Myhill-Nerode's theorem, L is regular *iff* it has a finite number of residues. To prove that L is not regular, we show that it has infinitely many residues. To do that, it is sufficient to show that all the languages $((a^k)^{-1}L)_{k\in\mathbb{N}}$ are distinct.

   Let $k, \ell \in \mathbb{N}$ with $k \neq \ell$, we want to show that $(a^k)^{-1}L \neq (a^\ell)^{-1}L$.

   The word b^k belongs to $(a^k)^{-1}L$ (because for $n = k$, $a^{n-k}b^n = a^0 b^k = b^k$). But it does not belong to $(a^\ell)^{-1}L$, because the only word of $(a^\ell)^{-1}L$ without a's is $b^\ell \neq b^k$. So, the two languages are distinct.

Exercice 3: Let L be the language over the alphabet $\{0,1\}$ recognized by the following automaton M.



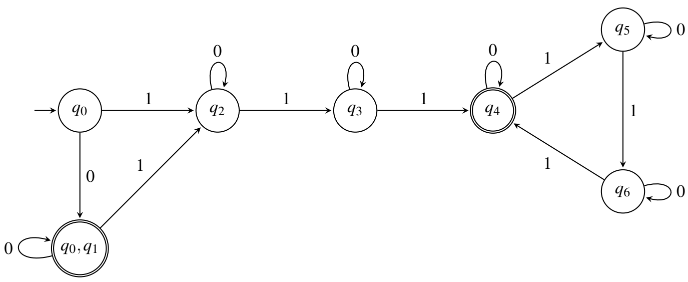
1. Compute the automaton $\det(M)$.
2. Find the minimal deterministic automaton recognizing L using Moore's algorithm.

3a. What is the language recognized by $\mathrm{tr}(M)$?

3b. Find the minimal deterministic automaton recognizing L using Brzozowski's algorithm.

Solutions:

1. Using the powerset algorithm, we get the following automaton $\det(M)$:



2. **Remember that Moore's algorithm only works when we start from a <u>deterministic</u> and <u>reachable</u> automaton.** So, we apply the algorithm to $\det(M)$.

   For simplicity, we rename the state $\{q_0, q_1\}$ of $\det(M)$ and just call it q'_1.

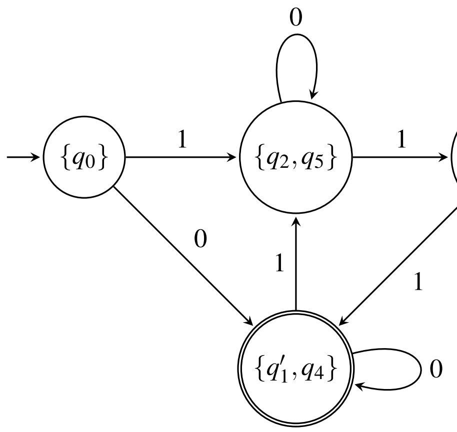
   Step 1: We compute successively the equivalence classes $\sim_n$.

   $\sim_0$: The two equivalence classes of $\sim_0$ are $\{q_0, q_2, q_3, q_5, q_6\}$ and $\{q'_1, q_4\}$.

   $\sim_1$: The equivalence classes of $\sim_1$ are $\{q_0\}$ and $\{q_2, q_5\}$ and $\{q_3, q_6\}$ and $\{q'_1, q_4\}$.

   $\sim_2$: $\sim_2 = \sim_1$, so we stop here.

   Step 2: We obtain the following automaton, whose states are the equivalence classes of $\sim_1$. The initial state is the class containing q_0, i.e., $\{q_0\}$. The final states are the ones containing final states, i.e., $\{q'_1, q_4\}$.



3a. In general, we know that $\mathrm{tr}(M)$ recognizes the mirror language of $[\![M]\!]$.
   But in this particular case, $\mathrm{mirror}([\![M]\!]) = [\![M]\!]$, because $[\![M]\!]$ is the language of non-empty words whose number of 1's is divisible by 3, and this property is still true if we reverse the word.

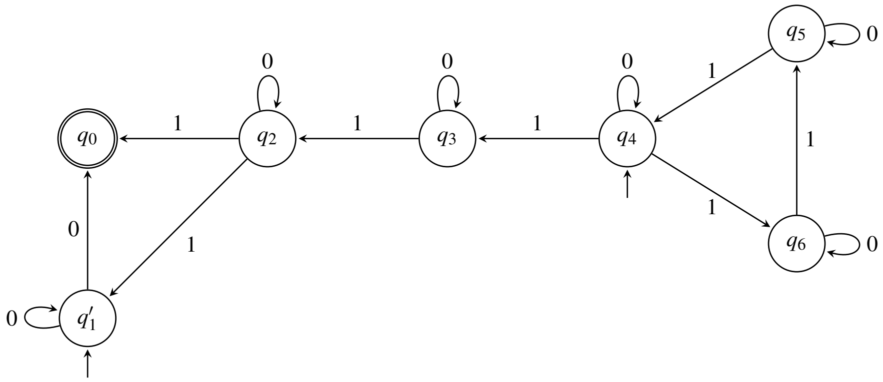
   So, $[\![\mathrm{tr}(M)]\!] = [\![M]\!]$.

3b. Normally, Brzozowski's algorithm is to compute $\det(\mathrm{tr}(\det(\mathrm{tr}(M))))$, which is the minimal deterministic automaton of M.

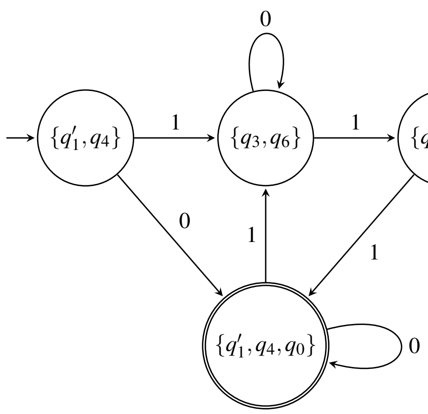
   But in this particular case, we can take a shortcut, thanks to question 3a. Indeed, remember that Brzozowski's algorithm works because whenever M is deterministic and reachable, $\det(\mathrm{tr}(M))$ is the minimal deterministic automaton recognizing the language $\mathrm{mirror}([\![M]\!])$.

   Since we already computer in question 1 the automaton $M' = \det(M)$ which is deterministic and reachable, we just have to do $\det(\mathrm{tr}(M'))$, which is the minimal automaton of $\mathrm{mirror}([\![M]\!]) = [\![M]\!]$.

- $\mathrm{tr}(M)$ is the following automaton:



- $\det(\mathrm{tr}(M))$ is the following automaton:



Exercice 4: Let L be the Dyck language on the alphabet $\{a, b\}$, in other words a and b are put for the opening and the closing parentheses respectively.

1. Given $k \in \mathbb{N}$ calculate the residue $(a^k)^{-1}L$.
2. Deduce that L is not regular.

Solutions:

1. We use the following characterization of the Dyck language: $w \in L$ *iff* $|w|_a = |w|_b$ and for every prefix x of w, $|x|_a \geq |x|_b$. We get: $(a^k)^{-1}L = \{u_0\, b\, u_1\, b \cdots b\, u_k \mid u_0\, u_1 \cdots u_k \in L\}$.
2. Same reasoning as in Exercise 2: for $k \neq \ell$, we have $b^k \in (a^k)^{-1}L$ but $b^k \notin (a^\ell)^{-1}L$, so all these residues are distinct, and since there is an infinity of them, L is not regular.

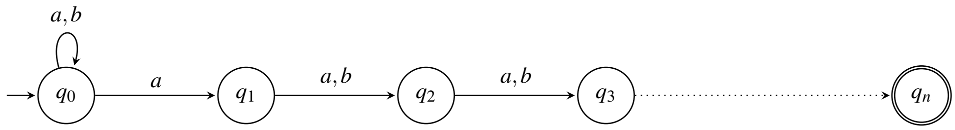
Exercise 5: Let the alphabet Σ be $\{a, b\}$ and $n \in \mathbb{N} \setminus \{0\}$ and consider the language

$$L \quad = \quad \{\gamma a \delta \mid \gamma \in \Sigma^*;\ \delta \in \Sigma^{n-1}\}\ .$$

1. Draw a *nondeterministic* finite automaton M with $n+1$ states and such that $[\![M]\!] = L$.

2. Find all the residues of the language L.
3. Deduce that any *deterministic* finite automaton accepting L has at least 2^n states.

Solutions:

1. 

2. Let us compute a few residues of L:
   - $\varepsilon^{-1}L = L$
   - $a^{-1}L = L \cup \Sigma^{n-1}$
   - $b^{-1}L = L$
   - $(aa)^{-1}L = L \cup \Sigma^{n-1} \cup \Sigma^{n-2}$
   - $(ab)^{-1}L = L \cup \Sigma^{n-2}$
   - $(aaa)^{-1}L = L \cup \Sigma^{n-1} \cup \Sigma^{n-2} \cup \Sigma^{n-3}$
   - $(aab)^{-1}L = L \cup \Sigma^{n-2} \cup \Sigma^{n-3}$
   - $(aba)^{-1}L = L \cup \Sigma^{n-1} \cup \Sigma^{n-3}$
   - $(abb)^{-1}L = L \cup \Sigma^{n-3}$
   - $\ldots$

   Notice that, except for $b^{-1}L$, all these residues are distinct.
   We can see a pattern emerge and we conjecture that the set $\mathscr{R}$ of residues of L is:

   $$\mathscr{R} = \left\{ L \,\cup\, \bigcup_{i \in I} \Sigma^i \;\middle|\; I \subseteq \{0, \ldots, n-1\} \right\}$$

   (Note in particular that for $I = \varnothing$, this formula gives $L \,\cup\, \bigcup_{i \in \varnothing} \Sigma^i = L = \varepsilon^{-1}L$.)

   To prove this equality we must show that: (1) all the residues of L are of this form; and (2) all the sets of this form are residues of L.

   Keeping in mind the iterative process in the examples that we computed by hand above, we will actually prove the following fact by induction on k:

   For all $k \in \{0, \ldots, n\}$, the set $\mathscr{R}_{\leq k}$ of residues of the form $w^{-1}L$ for $|w| \leq k$ is

   $$\mathscr{R}_{\leq k} = \left\{ L \,\cup\, \bigcup_{i \in I} \Sigma^i \;\middle|\; I \subseteq \{n-k, \ldots, n-1\} \right\}$$

   - For $k = 0$ (or $k = 1$), it works (see examples).
   - Assume this is true for k.
     We prove the equality for $\mathscr{R}_{\leq k+1}$ by double inclusion:

- $\subseteq$: Take an element of $\mathscr{R}_{\leq k+1}$, i.e., a residue $w^{-1}L$ with $|w| \leq k+1$.
  We want to show that $w^{-1}L = L \cup \bigcup_{i\in I} \Sigma^i$, for some set $I \subseteq \{n-(k+1), \ldots, n-1\}$.

  If $|w| \leq k$, we know by induction hypothesis that $w^{-1}L = L \cup \bigcup_{i\in I} \Sigma^i$, for some set $I \subseteq \{n-k, \ldots, n-1\}$, so in particular $I \subseteq \{n-(k+1), \ldots, n-1\}$ so we are done.

  If $|w| = k+1$, then w is either of the form ua or ub, with $|u| \leq k$. Moreover, we have the formula $(ua)^{-1}L = a^{-1}(u^{-1}L)$, and we know by induction hypothesis that $u^{-1}L = L \cup \bigcup_{i\in I} \Sigma^i$ for some $I \subseteq \{n-k, \ldots, n-1\}$.
  Then, $w^{-1}L = a^{-1}(L \cup \bigcup_{i\in I} \Sigma^i) = L \cup \Sigma^{n-1} \cup \bigcup_{i\in I} \Sigma^{i-1} = L \cup \bigcup_{i\in J} \Sigma^i$, where the union is now indexed over the set $J = \{n-1\} \cup \{i-1 \mid i \in I\} \subseteq \{n-k-1, \ldots, n-1\}$.
  We can do the same if $w = ub$, and we get $w^{-1}L = b^{-1}(L \cup \bigcup_{i\in I} \Sigma^i) = L \cup \bigcup_{i\in J} \Sigma^i$, where $J = \{i-1 \mid i \in I\} \subseteq \{n-k-1, \ldots, n-1\}$.

- $\supseteq$: Take $I \subseteq \{n-k-1, \ldots, n-1\}$. We want to show that there exists some word w such that $|w| \leq k+1$ and $w^{-1}L = L \cup \bigcup_{i\in I} \Sigma^i$.

  Either $n-k-1 \notin I$. In which case, we have $I \subseteq \{n-k, \ldots, n-1\}$, and we know that there is a word w that works.

  Or $n-k-1 \in I$. Then, there are two cases, depending whether $n-1 \in I$ or $n-1 \notin I$.
  $\rightarrow$ Case $n-1 \in I$: we take $J = \{i+1 \mid i \in I\} \setminus \{n\}$. We have $J \subseteq \{n-k, \ldots, n-1\}$, so by induction hypothesis there exists some word w that works. By doing the same calculations as above, we see that the word wa works for I.
  $\rightarrow$ Case $n-1 \notin I$: we take $J = \{i+1 \mid i \in I\}$. By induction hypothesis there exists some word w that works. By doing the same calculations as above, we see that the word wb works for I.

To complete the proof, we just have to remark that $\mathscr{R} = \mathscr{R}_{\leq n}$, which proves the initial conjecture about $\mathscr{R}$.

3. Since the residue automaton is the minimal deterministic automaton recognizing L, and there are 2^n residues (because there are 2^n subsets of $\{0, \ldots, n-1\}$), every deterministic automaton must have at least 2^n states.

Exercise 6: Let L be a language over an alphabet with a single element (i.e. $\Sigma = \{a\}$ and $L \subseteq \Sigma^*$). Describe the residues of L^*.

Solution: Remark that since Σ has only one letter, every word $w \in \Sigma^*$ is of the form $w = a^n$ for some $n \in \mathbb{N}$. So, we have a "natural" bijection between Σ^* and $\mathbb{N}$, and for ease of notation, we will consider $L \subseteq \Sigma^*$ as a subset of $\mathbb{N}$. Moreover, since $a^n a^m = a^{n+m}$, the analogue in $\mathbb{N}$ of concatenation is going to be addition. In short, we have we have an isomorphism of monoids between $(\Sigma^*, .)$ and $(\mathbb{N}, +)$.

So, given $L \subseteq \mathbb{N}$, the set L^* corresponds to $L^* = \{n_1 + \ldots + n_\ell \mid n_i \in L\}$.
A residue of L^* is $(a^k)^{-1}L^* = \{n_1 + \ldots + n_\ell - k \mid n_i \in L \text{ and } \Sigma_i n_i \geq k\}$, for any $k \in \mathbb{N}$.

Extra question*: Show that L^* has finitely many residues.* Remember Exercise 4 of Tutorial 4; the goal was to prove that, when the alphabet has only one letter, then L^* is always regular, even if L is not. The goal here will be to prove this result again, but this time using residues: we want to show that L^* has only finitely many residues, which implies, by Myhill-Nerode theorem, that L^* is regular.

So, let us look at the smallest "non-empty word" of L, i.e., the integer $m = \min(L \setminus \{0\})$. Then, we can also define for every $i \in \{1, \ldots, m-1\}$ the integer $s_i \in L$ which is the smallest element of L such that $s_i \equiv i \bmod m$. We take $k = m \times (\Sigma_i s_i)$ and $k' = k - m$, and we are going to show that $(a^k)^{-1}L^* = (a^{k'})^{-1}L^*$.

$\supseteq$: Assume $N \in (a^{k'})^{-1}L^*$, i.e., $N = n_1 + \ldots + n_\ell - k'$ with $\Sigma_i n_i \geq k'$. Then if we pick $n_0 = m$, we have $N = n_0 + n_1 + \ldots + n_\ell - k$ and the corresponding inequality, so $N \in (a^k)^{-1}L^*$.

$\subseteq$: Assume $N \in (a^k)^{-1}L^*$, i.e., $N = n_1 + \ldots + n_\ell - k$ with $\Sigma_i n_i \geq k$. This direction is a bit more tricky: now, we want to subtract m from the sum $n_1 + \ldots + n_\ell$, but m might not appear in this sum.

- if m appears in the sum, we can remove it and we are done.
- otherwise, if there is a term n_i in the sum which is not one of the s_j's, then we look at its value modulo m (we call it j), then it has to be of the form $n_i = s_j + qm$ for some $q \geq 1$, because by definition s_j is the smallest with this modulo. So we can remove n_i from the sum and replace it by $s_j + m + m + \ldots + m$ with $q-1$ occurrences of m.
- otherwise, all the terms of the sum are among the s_j's. But then, one of them has to appear at least m times, otherwise we could not have $\Sigma_i n_i \geq k = m \times (\Sigma_j s_j)$. Say s_j appears m times, then we replace m occurrences of s_j by $(s_j - 1)$ occurrences of m.

So, we found two residues, $(a^k)^{-1}L^*$ and $(a^{k'})^{-1}L^*$, which are equal. This actually implies (because Σ has only one letter) that there is only a finite number of residues in total. Indeed, for any $n \in \mathbb{N}$, we get $(a^{k+n})^{-1}L^* = (a^n)^{-1}((a^k)^{-1}L^*) = (a^n)^{-1}((a^{k'})^{-1}L^*) = (a^{k'+n})^{-1}L^*$. If we reformulate this by renaming the variables (and remember that $k = k' + m$), it says that for every $z \geq k'$, $(a^{z+m})^{-1}L^* = (a^z)^{-1}L^*$. So the sequence of residues is periodic after rang k': there are at most $m \times (\Sigma_i s_i)$ residues.

Exercise 7: Let L_ϕ be the language $\{a^{\phi(n)}b^n \mid n \in \mathbb{N}\}$ for some given mapping $\phi : \mathbb{N} \to \mathbb{N}$.

1. Suppose that $\phi(\mathbb{N})$ is infinite and let $\{k_0 < k_1 < \cdots < k_i < \cdots\}$ for $i \in \mathbb{N}$ be a strictly increasing enumeration of $\phi(\mathbb{N})$.
   1a. For all $k \in \mathbb{N}$, calculate the residue $(a^k)^{-1}L_\phi$.
   1b. Explain why L_ϕ is not regular.
2. Find a function ϕ such that L_ϕ is regular.
3. Find a function ϕ such that $\phi(\mathbb{N})$ is finite and L_ϕ is not regular.

Solutions:

1a. $(a^k)^{-1}L_\phi = \{a^{\phi(n)-k}b^n \mid \phi(n) \geq k\}$

1b. We show that for $i \neq j$, $(a^{k_i})^{-1}L \neq (a^{k_j})^{-1}L$. By definition k_i is in the image of ϕ, so there is $n_i \in \mathbb{N}$ such that $\phi(n_i) = k_i$. So, $b^{n_i} \in (a^{k_i})^{-1}L$.

On the other hand, we have $k_j \neq k_i$, so $b^{n_i} \notin (a^{k_j})^{-1}L$

Thus, we have infinitely many residues: at least one for each k_i in the image of ϕ, and we supposed that $\phi(\mathbb{N})$ is infinite.

2. $\phi(n) = 42$

3. Take $\phi(n) = \begin{cases} 1 & \text{if } n \text{ is prime} \\ 0 & \text{otherwise} \end{cases}$

Then $a^{-1}L_\phi = \{b^n \mid n \text{ is prime}\}$, which we know is not regular (cf. lesson on pumping lemma). But if L_ϕ was regular, its residue $a^{-1}L_\phi$ would be regular too, so L_ϕ is not regular.