5 Minimization of Finite Automata

As we saw when translating regular expressions into NFAs, the resulting automaton is not necessarily the smallest possible one. Similarly, when employing the subset construction to translate an NFA into a DFA, the result is not always the smallest possible DFA. It is often desirable to make automatons as small as possible. For example, if we wish to implement an automaton, the implementation will be more efficient the smaller the automaton is.

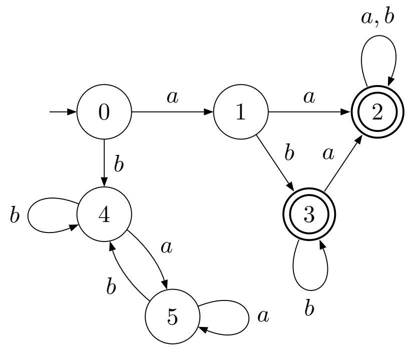
Given an automaton, the question, then, is how to construct an *equivalent* but smaller automaton. Recall that two automatons are equivalent of they accept the same language. In the following we will study a method for minimizing DFAs: the table-filling algorithm.

Another interesting question is if there, in general, is one unique automaton that is the smallest equivalent one, or if there can be many distinct equivalent automatons, none of which can be made any smaller. It turns out the answer is that the minimal equivalent DFA is unique up to naming of the states. This, in turn, means that we have obtained a mechanical decision procedure for determining whether two regular languages are equal: simply convert their respective representation (be it a DFA, an NFA, or a regular expression) to DFAs and minimize them. Because the minimal DFAs are unique, the languages are equal if and only if the minimal DFAs are equal.

5.1 The table-filling algorithm

For a DFA $(Q, \Sigma, \delta, q_0, F)$, $p, q \in Q$ are *equivalent* states if and only if, for all $w \in \Sigma^*$, $\hat{\delta}(p, w) \in F \Leftrightarrow \hat{\delta}(q, w) \in F$. If two states are *not* equivalent, then they are *distinguishable*.

Consider the following DFA, where $\Sigma = \{a, b\}$, $Q = \{0, 1, 2, 3, 4, 5\}$, $F = \{2, 3\}$:



The states 1 and 2 are distinguishable on ϵ because $\hat{\delta}(1, \epsilon) = 1 \notin F$ while $\hat{\delta}(2, \epsilon) = 2 \in F$. Similarly, 0 and 1 are distinguishable on e.g. b because $\hat{\delta}(0, b) = 4 \notin F$ while $\hat{\delta}(1, b) = 3 \in F$. On the other hand, in this case, we can easily see that 4 and 5 are not distinguishable on any word because it is not possible to reach any accepting (final) state from either 4 or 5.

The *Table-Filling Algorithm* recursively constructs the set of distinguishable pairs of states for a DFA. When all distinguishable state pairs have been identified, any remaining pairs of states must be equivalent. Such states can be merged, thereby minimizing the automaton. Assume a DFA $(Q, \Sigma, \delta, q_0, F)$:

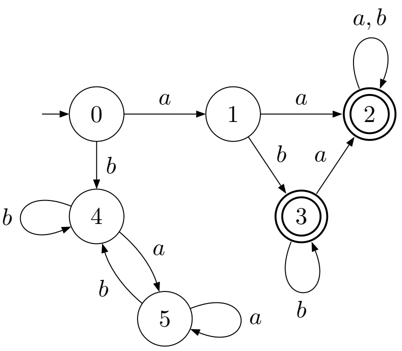
BASIS For $p, q \in Q$, if $(p \in F \wedge q \notin F) \vee (p \notin F \wedge qinF)$, then (p, q) is a distinguishable pair of states. (The states p and q are distinguishable on ϵ.)

INDUCTION For $p, q, r, s \in Q$, $a \in \Sigma$, if $(r, s) = (\delta(p, a), \delta(q, a))$ is a distinguishable pair of states, then (p, q) is also distinguishable. (If the states r and s are distinguishable on a word w, then p and q are distinguishable on aw.)

Theorem 5.1 *If two states are* not *distinguishable by the table-filling algorithm, then they are* equivalent*.*

5.2 Example of DFA minimization using the table-filling algorithm

This section illustrates how the table-filling algorithm can be used to minimize a DFA when working by hand through a fully worked example. We will minimize the following DFA, where $\Sigma = \{a, b\}$, $Q = \{0, 1, 2, 3, 4, 5\}$, $F = \{2, 3\}$:



First construct a table over all pairs of distinct states. That is, we do not consider pairs $(p \in Q, p \in Q)$ because a state obviously cannot be distinguishable from itself. An easy way of doing constructing the table is to order the states (e.g. numerically or alphabetically), and then list all states except the last one in *ascending* order along the top, and all states except the first one in *descending* order down the left-hand side of the table. The resulting table for our 6-state DFA looks like this:

	0	1	2	3	4
5					
4					
3					
2					
1					

Then mark the state pairs that are distinguishable according to the basis of the table-filling algorithm; i.e., the pairs where one state is accepting, and one is not. The accepting states of our DFA are 2 and 3. Thus the state pairs $(0,2)$, $(0,3)$, $(1,2)$, $(1,3)$, $(2,4)$, $(2,5)$, $(3,4)$, and $(3,5)$ have to be marked. List all remaining state pairs to the right: these are the potentially equivalent states that we now have to investigate further:

	0	1	2	3	4
5			x	x	
4			x	x	
3	x	x			
2	x	x			
1					

$(0,1)$	$(0,4)$	$(0,5)$	$(1,4)$

$(1,5)$	$(2,3)$	$(4,5)$

Recall that the induction step of the table-filling algorithm says that for $p,q,r,s \in Q$ and $a \in \Sigma$, if $(r,s) = (\delta(p,a), \delta(q,a)$ is distinguishable (on some word w), then (p,q) is (on the word aw). If, during systematic investigation of all state combinations, we, from a state pair (p,q) on some input symbol a, reach a state pair (r,s) for which it is not yet known whether it is distinguishable or not, we record (p,q) under the heading for (r,s). If it later becomes clear that (r,s) is distinguishable, that means that (p,q) also is distinguishable, and recording this implication allows us to carry out the deferred marking at that point.

Investigate all potentially equivalent state pairs on all input symbols (unless we find that a pair is distinguishable, which means we can stop):

$(0,1)$:	$(\delta(0,a), \delta(1,a)) = (1,2)$	Distinguishable! Mark in table.
$(0,4)$:	$(\delta(0,a), \delta(4,a)) = (1,5)$	Unknown as yet. Add $(0,1)$ under $(1,5)$.
	$(\delta(0,b), \delta(4,b)) = (4,4)$	Same state, no info.

Our table now looks as follows (we strike a line across the pairs we have considered):

	0	1	2	3	4
5			x	x	
4			x	x	
3	x	x			
2	x	x			
1	x				

~~$(0,1)$~~	~~$(0,4)$~~	$(0,5)$	$(1,4)$

$(1,5)$	$(2,3)$	$(4,5)$
$(0,4)$		

We continue:

$(0,5)$:	$(\delta(0,a), \delta(5,a)) = (1,5)$	Unknown as yet. Add $(0,5)$ under $(1,5)$.
	$(\delta(0,b), \delta(5,b)) = (4,4)$	Same state, no info.
$(1,4)$:	$(\delta(1,a), \delta(4,a)) = (2,5)$	Distinguishable! Mark in table.

Table:

	0	1	2	3	4
5			x	x	
4		x	x	x	
3	x	x			
2	x	x			
1	x				

~~$(0,1)$~~	~~$(0,4)$~~	~~$(0,5)$~~	~~$(1,4)$~~

$(1,5)$	$(2,3)$	$(4,5)$
$(0,4)$		
$(0,5)$		

Now we have come to the state pair $(1,5)$. If we can determine that $(1,5)$ is a distinguishable pair, then we also know that the pairs $(0,4)$ and $(0,5)$ are distinguishable:

$$(1,5){:}\quad (\delta(1,a),\delta(5,a)) = (2,5) \quad \text{Distinguishable!}$$

Thus we should mark $(1,5)$ along with and $(0,4)$ and $(0,5)$:

	0	1	2	3	4
5	x	x	x	x	
4	x	x	x	x	
3	x	x			
2	x	x			
1	x				

$\cancel{(0,1)}$ $\cancel{(0,4)}$ $\cancel{(0,5)}$ $\cancel{(1,4)}$

$\cancel{(1,5)}$	$(2,3)$	$(4,5)$
$\cancel{(0,4)}$		
$\cancel{(0,5)}$		

It remains to check the pairs $(2,3)$ and $(4,5)$:

$$\begin{array}{lll}
(2,3){:} & (\delta(2,a),\delta(3,a)) = (2,2) & \text{Same state, no info.}\\
 & (\delta(2,b),\delta(3,b)) = (2,3) & \text{No point in adding } (2,3) \text{ below } (2,3).\\
(4,5){:} & (\delta(4,a),\delta(5,a)) = (5,5) & \text{Same state, no info.}\\
 & (\delta(4,b),\delta(5,b)) = (4,4) & \text{Same state, no info.}
\end{array}$$

We have now systematically checked all potentially equivalent state pairs. Two pairs remain unmarked; i.e., we have not been able to show that they are distinguishable: $(2,3)$ and $(4,5)$. We can therefore conclude that these states are pairwise equivalent: $2 \equiv 3$ and $4 \equiv 5$. We thus proceed to merge these states by, informally, placing them "on top" of each other and "dragging along" the edges. The result is the following minimal DFA (where the merged states have been given the names 2 and 4):