Computer Science 313

Equivalence of Deterministic Finite Automata and Nondeterministic Finite Automata

Instructor: Wayne Eberly

Department of Computer Science
University of Calgary

Lecture #6

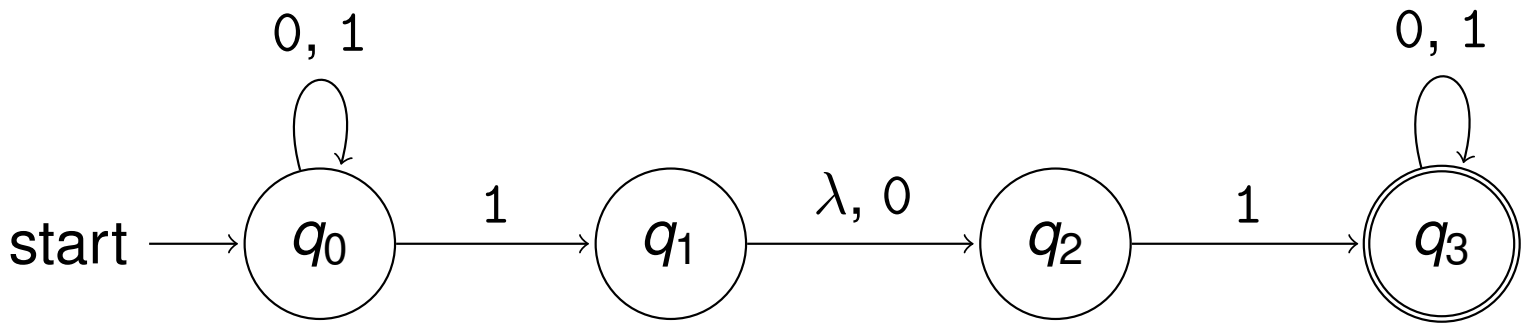
Goal for Today

- *Prove* that every language $L \subseteq \Sigma^\star$ is the language of an NFA *if and only if* it is the language of a DFA.
- A construction to *convert* an NFA into a DFA with the same language will be provided in order to complete the "hard part" of this proof.

Note: The proof of today's class is *similar to,* but not exactly the same as, the proof in Section 1.2 of *Introduction to the Theory of Computation:* A slightly different "subset construction" is described in these notes.

Ongoing Example

A DFA will be designed that has the same language (a language in $\Sigma^\star$, for $\Sigma = \{0, 1\}$) as the first NFA used as an example in the previous lecture:



Ongoing Example

Note: For every string $\omega = \Sigma^\star$...

- $q_0 \in \delta^\star(q_0, \omega)$.
- $q_1 \in \delta^\star(q_0, \omega)$ if and only if ω ends with a `1`.
- $q_2 \in \delta^\star(q_0, \omega)$ if and only if ω ends with either `1` or `10`.
- $q_3 \in \delta^\star(q_0, \omega)$ if and only if either `11` or `101` is a substring of ω — that is, if and only if either $\omega = \mu\texttt{11}\nu$ or $\omega = \mu\texttt{101}\nu$ for strings $\mu, \nu \in \Sigma^\star$ (or both).

Since $F = \{q_3\}$, it follows that the language of this NFA is the set of strings $\omega \in \Sigma^\star$ such that either `11` or `101` (or both) is a substring of ω.

The Easy Part of the Proof

Suppose you are given a DFA $M = (Q, \Sigma, \delta, q_0, F)$ for a language $L \subseteq \Sigma^\star$.

Consider the NFA $\widehat{M} = (\widehat{Q}, \Sigma, \widehat{\delta}, \widehat{q}_0, \widehat{F})$ (which also has some language of strings in $\Sigma^\star$) such that

- $\widehat{Q} = Q$ — that is, M and $\widehat{M}$ have the same ***states***;
- $\widehat{q}_0 = q_0$ — that is, M and $\widehat{M}$ have the same ***start state***;
- $\widehat{F} = F$ — that is, M and $\widehat{M}$ have the same set of ***accepting states***;
- For every state $q \in Q$ and for every symbol $\sigma \in \Sigma$

$$\widehat{\delta}(q, \sigma) = \{\delta(q, \sigma)\}$$

—- that is, $\widehat{\delta}(q, \sigma)$ is the *set*, with size one, including $\delta(q, \sigma)$;

- For every state $q \in Q$, $\widehat{\delta}(q, \lambda) = \emptyset$ — that is, there are no λ-transitions in $\widehat{M}$.

The Easy Part of the Proof

- The *pictures* of M and $\widehat{M}$ will be identical!
- Since there are no λ-transitions, it is easy to argue that $Cl_\lambda(q) = \{q\}$ for every state $q \in Q$.
- *This* makes it easy to show that $\widehat{\delta^\star}(q, \omega) = \{\delta^\star(q, \omega)\}$ for every state $q \in Q$ and string $\omega \in \Sigma^\star$ — that is, $\widehat{\delta^\star}(q, \omega)$ is a *set*, with size one, that contains the state $\delta^\star(q, \omega)$.
- Finally, notice that $\{q\} \cap F \neq \emptyset$ if and only if $q \in F$ for every state $q \in Q$.
- This makes it easy to argue that $L(\widehat{M}) = L(M)$.

Conclusion: For every language $L \subseteq \Sigma^\star$, *if* L is the language of a DFA *then* L is also the language of an NFA (with the same number of states).

The Hard Part — the "Subset Construction"

Suppose now that $L \subseteq \Sigma^\star$ (for some alphabet Σ) and that $L = L(M)$ for some ***nondeterministic*** finite automaton

$$M = (Q, \Sigma, \delta, q_0, F).$$

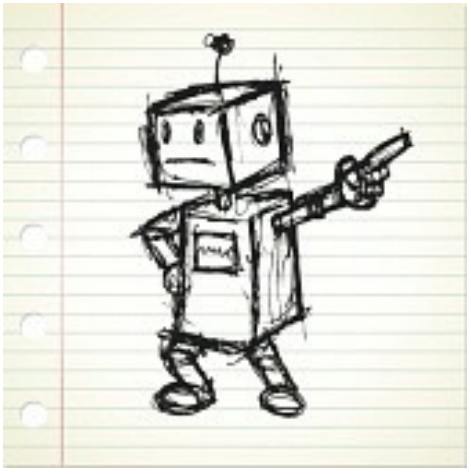
Suppose, now, that we would like to ***design*** a ***deterministic*** finite automaton

$$\widehat{M} = (\widehat{Q}, \Sigma, \widehat{\delta}, \widehat{q}_0, \widehat{F})$$

such that $L(\widehat{M}) = L = L(M)$.

The Hard Part — the "Subset Construction"

Recall that in order to do this we should try to...



Be the DFA!

The Hard Part — the "Subset Construction"

In order to this we have to try to answer the following:

- ***Question:*** What do you need to remember about the part μ of the input string you have seen so far?
- ***Answer (Which Might Be Surprising):*** $\delta^\star(q_0, \mu)$ — that is, the subset of Q including states that can be reached from q_0 by processing μ!

The Hard Part — the "Subset Construction"

As described in *Introduction to the Theory of Computation*, we should therefore introduce a state in our DFA $\widehat{M}$ for each subset of Q.

Since our example NFA, has four states in Q, we would now include $2^4 = 16$ states in our DFA. That seems like an awful lot — especially if you are being asked to find a DFA that is equivalent to a given NFA on a test!

The Hard Part — the "Subset Construction"

More generally, if your NFA had n states then you would include 2^n states in your DFA. Here are a few values for n and 2^n:

n	2^n
10	1,024
20	1,048,576
30	1,073,741,824
40	1,099,511,627,776

The Hard Part — the "Subset Construction"

The version of the "Subset Construction" described next is — admittedly — not quite a simple as the version in *Introduction to the Theory of Computation.*

The Hard Part — the "Subset Construction"

However, it avoids this exponential increase in the number of states you must work with... at least, *some* of the time.

In particular, the DFA we will design will include states corresponding to only *some* of the subsets of Q — namely, only those that are actually "reachable" from the start state.

The Hard Part — and the "Subset Construction"

- Suppose V is a subset of Q such that $V = \delta^\star(q_0, \mu)$ for some string $\mu \in \Sigma^\star$.
- Let $\sigma \in \Sigma$. Then — as noted in the previous lecture — the state that should be reached, from the one corresponding to V, by processing σ, corresponds to the set

$$W = \delta^\star(q_0, \mu \cdot \sigma) = \bigcup_{r \in \delta^\star(q_0, \mu)} \left(\bigcup_{s \in \delta(r, \sigma)} Cl_\lambda(s) \right)$$

$$= \bigcup_{r \in V} \left(\bigcup_{s \in \delta(r, \sigma)} Cl_\lambda(s) \right).$$

- Note as well that the start state for the DFA being designed should correspond to the subset $Cl_\lambda(q_0)$.

The Hard Part — and the "Subset Construction"

The algorithm to be described next will generate the subsets of Q that should correspond to states in the DFA being constructed by keeping track of *two* sets:

- $\widehat{Q}$ — subsets of Q that will correspond to states, such that the transitions out of these states (in the DFA) *have* been defined. This will initially be the empty set.
- $\widehat{R}$ — subsets of Q that will correspond to states, such that the transitions out of these states in the DFA *have not* yet been defined. This will initially be $\{Cl_\lambda(q_0)\}$.

New elements will be added to $\widehat{R}$ as the transition function of the DFA is being defined. Elements will move from $\widehat{R}$ to $\widehat{Q}$ as transitions for them are defined.

The Hard Part — and the "Subset Construction"

Pseudocode to compute a set $\widehat{Q} \subseteq \mathcal{P}(Q)$ of states for our DFA is as follows.

1. $\widehat{R} := \{Cl_\lambda(q_0)\}$; $\widehat{Q} := \emptyset$
2. `while` $(\widehat{R} \neq \emptyset)$ `do`
3. Choose an element V from $\widehat{R}$
4. `for each symbol` $\sigma \in \Sigma$ `do`
5. $W := \bigcup_{r \in V} \left(\bigcup_{s \in \delta(r,\sigma)} Cl_\lambda(s) \right)$ // $\widehat{\delta}(V, \sigma)$ will be set to be W
6. `if` $(W \notin \widehat{Q} \cup \widehat{R})$ `then` $\widehat{R} := \widehat{R} \cup \{W\}$ `end if`

   `end for`
7. $\widehat{R} := \widehat{R} \setminus \{V\}$; $\widehat{Q} := \widehat{Q} \cup \{V\}$

   `end while`
8. `return` $\widehat{Q}$

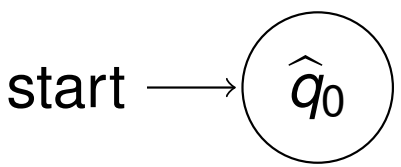
Subset Construction: Application to the Example

The initial state of the NFA M is the state q_0, so the initial state of the corresponding DFA (which we call $\widehat{q}_0$) is

$$\widehat{q}_0 = Cl_\lambda(q_0) = \{q_0\} \in \mathcal{P}(Q).$$

Consequently, the algorithm sets $\widehat{R}$ to be $\{\widehat{q}_0\}$, for $\widehat{q}_0$ as above, and sets $\widehat{Q}$ to be $\emptyset$.

Our DFA so far:



Subset Construction: Application to the Example

- In the first execution of the body of the `while` loop V must be set to be $\widehat{q}_0$ because this is the only member of $\widehat{R}$.
- Since $\Sigma = \{0, 1\}$, σ must be set to each of `0` and `1` during the execution of the inner `for` loop.
- Suppose that during this execution — and all other executions of the `for` loop — we set σ to be `0` during the first execution of the body of the `for` loop, and we set σ to be `1` during the second execution of the body of the `for` loop.

Subset Construction: Application to the Example

Then, during the first execution of the body of the `for` loop (included in the first execution of the body of the outer `while` loop) $V = \{q_0\}$ and $\sigma = 0$, so that

$$\begin{aligned} W &= \bigcup_{r \in \widehat{q}_0} \left(\bigcup_{s \in \delta(r,0)} Cl_\lambda(s) \right) \\ &= \bigcup_{s \in \delta(q_0,0)} Cl_\lambda(s) \\ &= Cl_\lambda(q_0) \\ &= \{q_0\} = \widehat{q}_0. \end{aligned}$$

Thus $\widehat{\delta}(\widehat{q}_0, 0) = \widehat{q}_0$. Since $\widehat{q}_0$ already belongs to $\widehat{Q} \cup \widehat{R}$, $\widehat{R}$ is not changed by this execution of the body of the `for` loop.

Subset Construction: Application to the Example

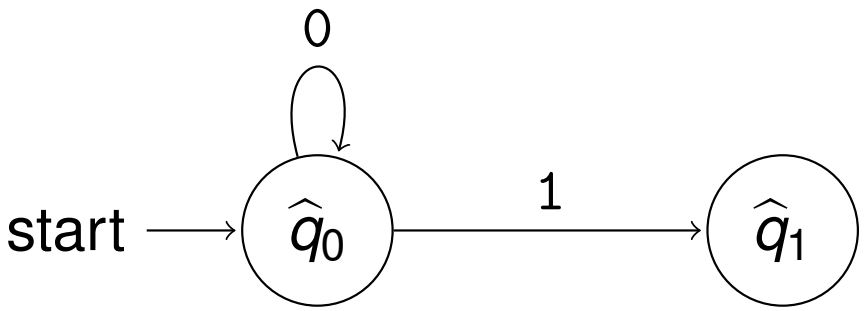
During the second execution of the body of the `for` loop (included in the first execution of the body of the outer `while` loop) $V = \{q_0\}$ and $\sigma = 1$, so that

$$\begin{aligned} W &= \bigcup_{r \in \widehat{q}_0} \left(\bigcup_{s \in \delta(r,1)} Cl_\lambda(s) \right) \\ &= \bigcup_{s \in \delta(q_0,1)} Cl_\lambda(s) \\ &= Cl_\lambda(q_0) \cup Cl_\lambda(q_1) \\ &= \{q_0\} \cup \{q_1, q_2\} \\ &= \{q_0, q_1, q_2\}. \end{aligned}$$

This set (which will now be called $\widehat{q}_1$) does not belong to $\widehat{Q} \cup \widehat{R}$, so it will now be added to $\widehat{R}$.

Subset Construction: Application to the Example

At the end of the first execution of body of the `while` loop $\widehat{q}_0$ is moved from $\widehat{R}$ to $\widehat{Q}$, so that $\widehat{Q} = \{\widehat{q}_0\}$ and $\widehat{R} = \{\widehat{q}_1\}$.The DFA, so far, is as follows.



Subset Construction: Application to the Example

In the second execution of the body of the `while` loop V must be set to be $\widehat{q}_1 = \{q_0, q_1, q_2\}$, since this is the only element of $\widehat{R}$. Since $\sigma = 0$ at the beginning of the first execution of the body of the `for` loop,

$$
\begin{aligned}
W &= \bigcup_{r \in \widehat{q}_1} \left(\bigcup_{s \in \delta(r,0)} Cl_\lambda(s) \right) \\
&= \bigcup_{s \in \delta(q_0,0)} Cl_\lambda(s) \;\cup\; \bigcup_{s \in \delta(q_1,0)} Cl_\lambda(s) \;\cup\; \bigcup_{s \in \delta(q_2,0)} Cl_\lambda(s) \\
&= Cl_\lambda(q_0) \cup Cl_\lambda(q_2) \cup \emptyset \\
&= \{q_0\} \cup \{q_2\} \cup \emptyset \\
&= \{q_0, q_2\}.
\end{aligned}
$$

This set (which will now be called $\widehat{q}_2$) does not belong to $\widehat{Q} \cup \widehat{R}$, so it will now be added to $\widehat{R}$.

Subset Construction: Application to the Example

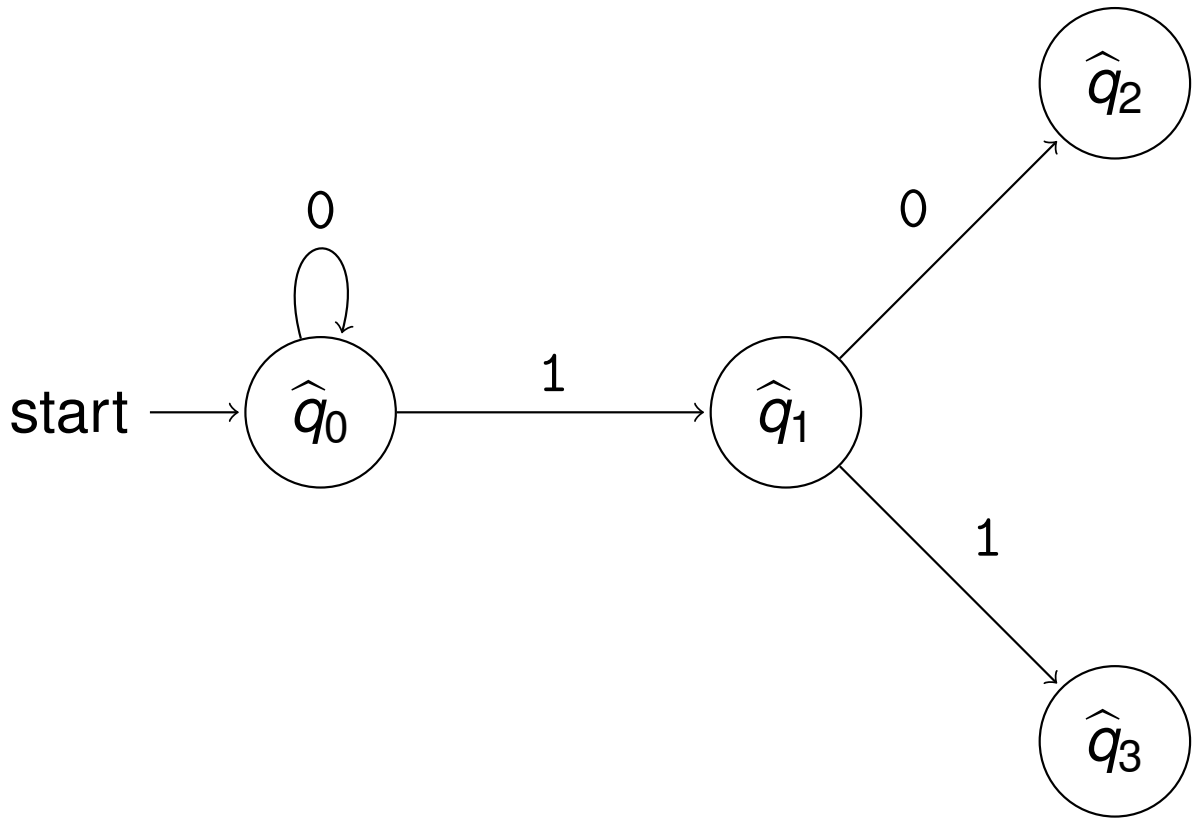
During the second execution of the body of the `for` loop (included in the second execution of the body of the outer `while` loop) $V = \widehat{q}_1 = \{q_0, q_1, q_2\}$ and $\sigma = 1$, so that

$$
\begin{aligned}
W &= \bigcup_{r \in \widehat{q}_1} \left(\bigcup_{s \in \delta(r,1)} Cl_\lambda(s) \right) \\
&= \bigcup_{s \in \delta(q_0,1)} Cl_\lambda(s) \;\cup\; \bigcup_{s \in \delta(q_1,1)} Cl_\lambda(s) \;\cup\; \bigcup_{s \in \delta(q_2,1)} Cl_\lambda(s) \\
&= (Cl_\lambda(q_0) \cup Cl_\lambda(q_1)) \cup \emptyset \cup Cl_\lambda(q_3) \\
&= Cl_\lambda(q_0) \cup Cl_\lambda(q_1) \cup Cl_\lambda(q_3) \\
&= \{q_0\} \cup \{q_1, q_2\} \cup \{q_3\} \\
&= \{q_0, q_1, q_2, q_3\}.
\end{aligned}
$$

This set (which will now be called $\widehat{q}_3$) does not belong to $\widehat{Q} \cup \widehat{R}$, so it will now be added to $\widehat{R}$.

Subset Construction: Application to the Example

At the end of the second execution of body of the `while` loop $\widehat{q}_1$ is moved from $\widehat{R}$ to $\widehat{Q}$, so that $\widehat{Q} = \{\widehat{q}_0, \widehat{q}_1\}$ and $\widehat{R} = \{\widehat{q}_2, \widehat{q}_3\}$.The DFA, so far, is as follows.



Subset Construction: Application to the Example

- Since $\widehat{R} = \{q_2, q_3\}$ we might continue by setting V (during the third execution of the body of the `while` loop to be *either* q_2 or q_3. States will be created in a different order for each choice (but it will turn out that the same set of states is ultimately created).
- We will continue, for this example, by setting V to be $\widehat{q}_2 = \{q_0, q_2\}$.

Subset Construction: Application to the Example

In the third execution of the body of the `while` loop V is set to be $\widehat{q}_2 = \{q_0, q_2\}$. Since $\sigma = 0$ at the beginning of the first execution of the body of the `for` loop,

$$\begin{aligned} W &= \bigcup_{r \in \widehat{q}_2} \left(\bigcup_{s \in \delta(r,0)} Cl_\lambda(s) \right) \\ &= \bigcup_{s \in \delta(q_0,0)} Cl_\lambda(s) \quad \cup \quad \bigcup_{s \in \delta(q_2,0)} Cl_\lambda(s) \\ &= Cl_\lambda(q_0) \cup \emptyset \\ &= \{q_0\} \cup \emptyset \\ &= \{q_0\} = \widehat{q}_0. \end{aligned}$$

Since $\widehat{q}_0$ already belongs to $\widehat{Q} \cup \widehat{R}$, neither $\widehat{Q}$ nor $\widehat{R}$ is changed.

Subset Construction: Application to the Example

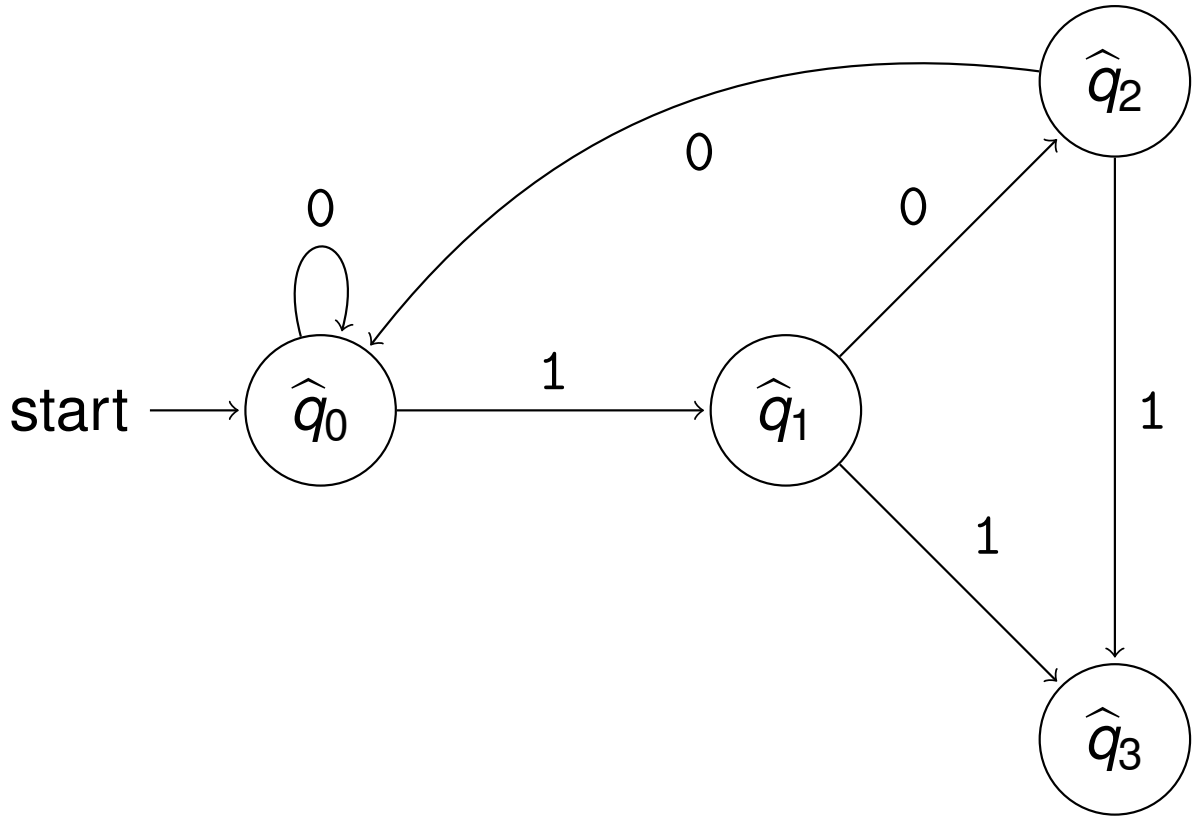
During the second execution of the body of the `for` loop (included in the third execution of the body of the outer `while` loop) $V = \widehat{q}_2 = \{q_0, q_2\}$ and $\sigma = 1$, so that

$$
\begin{aligned}
W &= \bigcup_{r \in \widehat{q}_2} \left(\bigcup_{s \in \delta(r,1)} Cl_\lambda(s) \right) \\
&= \bigcup_{s \in \delta(q_0,1)} Cl_\lambda(s) \quad \cup \quad \bigcup_{s \in \delta(q_2,1)} Cl_\lambda(s) \\
&= (Cl_\lambda(q_0) \cup Cl_\lambda(q_1)) \cup Cl_\lambda(q_3) \\
&= Cl_\lambda(q_0) \cup Cl_\lambda(q_1) \cup Cl_\lambda(q_3) \\
&= \{q_0\} \cup \{q_1, q_2\} \cup \{q_3\} \\
&= \{q_0, q_1, q_2, q_3\} = \widehat{q}_3
\end{aligned}
$$

Since $\widehat{q}_3$ already belongs to $\widehat{Q} \cup \widehat{R}$, neither $\widehat{Q}$ nor $\widehat{R}$ is changed.

Subset Construction: Application to the Example

At the end of the third execution of body of the `while` loop $\widehat{q}_2$ is moved from $\widehat{R}$ to $\widehat{Q}$, so that $\widehat{Q} = \{\widehat{q}_0, \widehat{q}_1, \widehat{q}_2\}$ and $\widehat{R} = \{\widehat{q}_3\}$.The DFA, so far, is as follows.



Subset Construction: Application to the Example

In the fourth execution of the body of the `while` loop V is set to be $\widehat{q}_3 = \{q_0, q_1, q_2, q_3\}$, since this is the only element of $\widehat{R}$. Since $\sigma = 0$ at the beginning of the first execution of the body of the `for` loop,

$$W = \bigcup_{r \in \widehat{q}_3} \left(\bigcup_{s \in \delta(r,0)} Cl_\lambda(s) \right)$$

$$= \bigcup_{s \in \delta(q_0,0)} Cl_\lambda(s) \cup \bigcup_{s \in \delta(q_1,0)} Cl_\lambda(s) \cup \bigcup_{s \in \delta(q_2,0)} Cl_\lambda(s) \cup \bigcup_{s \in \delta(q_3,0)} Cl_\lambda(s)$$

$$= Cl_\lambda(q_0) \cup Cl_\lambda(q_2) \cup \emptyset \cup Cl_\lambda(q_3)$$

$$= \{q_0\} \cup \{q_2\} \cup \emptyset \cup \{q_3\}$$

$$= \{q_0, q_2, q_3\}.$$

This set (which will now be called $\widehat{q}_4$) does not belong to $\widehat{Q} \cup \widehat{R}$, so it will now be added to $\widehat{R}$.

Subset Construction: Application to the Example

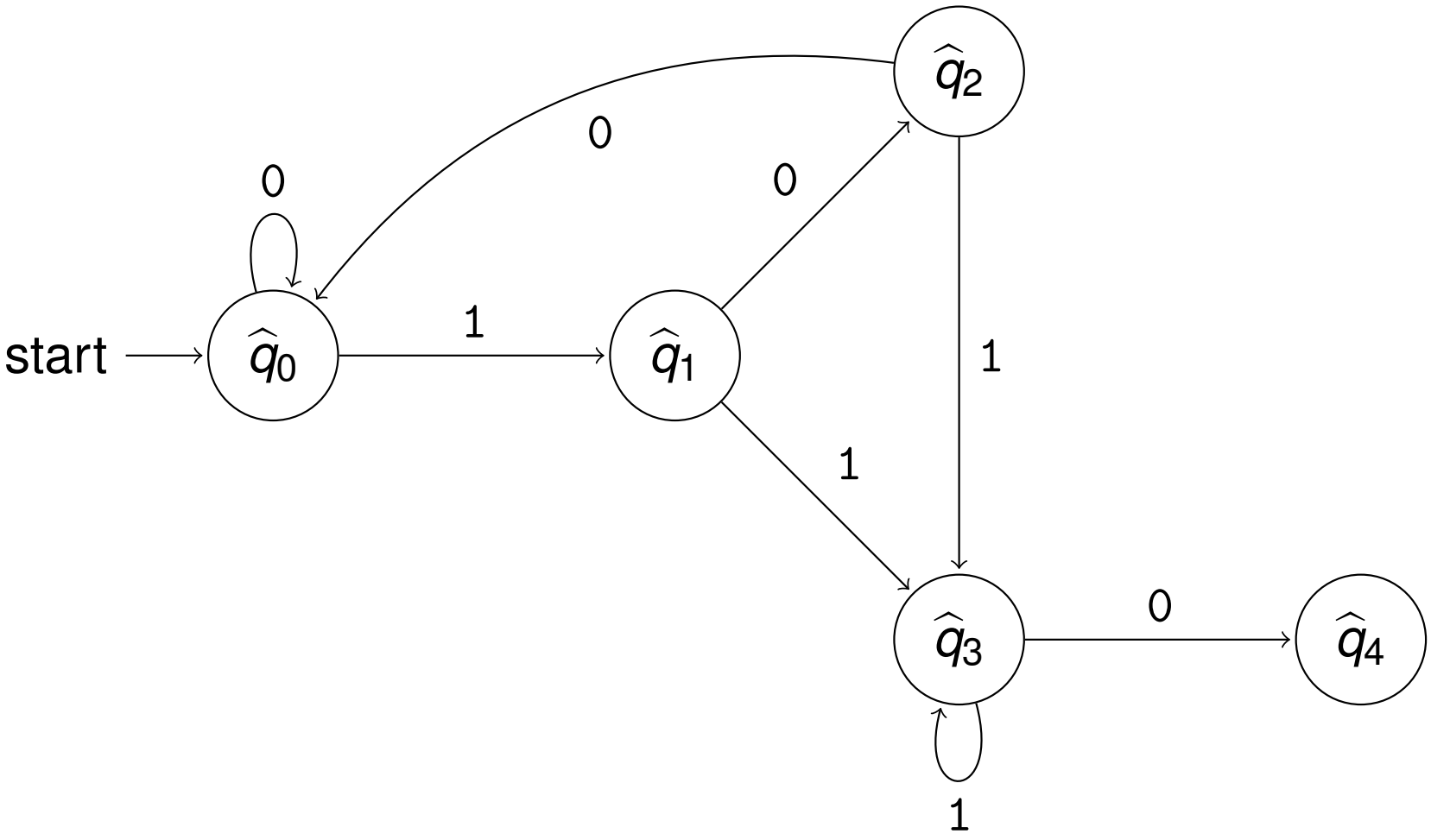
During the second execution of the body of the `for` loop (included in the fourth execution of the body of the outer `while` loop) $V = \widehat{q}_3 = \{q_0, q_1, q_2, q_3\}$ and $\sigma = 1$, so that

$$
\begin{aligned}
W &= \bigcup_{r \in \widehat{q}_3} \left(\bigcup_{s \in \delta(r,1)} Cl_\lambda(s) \right) \\
&= \bigcup_{s \in \delta(q_0,1)} Cl_\lambda(s) \cup \bigcup_{s \in \delta(q_1,1)} Cl_\lambda(s) \cup \bigcup_{s \in \delta(q_2,1)} Cl_\lambda(s) \\
&\qquad \cup \bigcup_{s \in \delta(q_3,1)} Cl_\lambda(s) \\
&= (Cl_\lambda(q_0) \cup Cl_\lambda(q_1)) \cup \emptyset \cup Cl_\lambda(q_3) \cup Cl_\lambda(q_3) \\
&= (\{q_0\} \cup \{q_1, q_2\}) \cup \emptyset \cup \{q_3\} \cup \{q_3\} \\
&= \{q_0, q_1, q_2, q_3\} = \widehat{q}_3.
\end{aligned}
$$

Since $\widehat{q}_3$ already belongs to $\widehat{Q} \cup \widehat{R}$, neither $\widehat{Q}$ nor $\widehat{R}$ is changed.

Subset Construction: Application to the Example

At the end of the fourth execution of body of the `while` loop $\widehat{q}_3$ is moved from $\widehat{R}$ to $\widehat{Q}$, so that $\widehat{Q} = \{\widehat{q}_0, \widehat{q}_1, \widehat{q}_2, \widehat{q}_3\}$ and $\widehat{R} = \{\widehat{q}_4\}$.The DFA, so far, is as follows.



Subset Construction: Application to the Example

In the fifth execution of the body of the `while` loop V is set to be $\widehat{q}_4 = \{q_0, q_2, q_3\}$, since this is the only element of $\widehat{R}$. Since $\sigma = 0$ at the beginning of the first execution of the body of the `for` loop,

$$
\begin{aligned}
W &= \bigcup_{r \in \widehat{q}_4} \left(\bigcup_{s \in \delta(r,0)} Cl_\lambda(s) \right) \\
&= \bigcup_{s \in \delta(q_0,0)} Cl_\lambda(s) \quad \cup \quad \bigcup_{s \in \delta(q_2,0)} Cl_\lambda(s) \quad \cup \quad \bigcup_{s \in \delta(q_3,0)} Cl_\lambda(s) \\
&= Cl_\lambda(q_0) \cup \emptyset \cup Cl_\lambda(q_3) \\
&= \{q_0\} \cup \emptyset \cup \{q_3\} \\
&= \{q_0, q_3\}.
\end{aligned}
$$

This set (which will now be called $\widehat{q}_5$) does not belong to $\widehat{Q} \cup \widehat{R}$, so it will now be added to $\widehat{R}$.

Subset Construction: Application to the Example

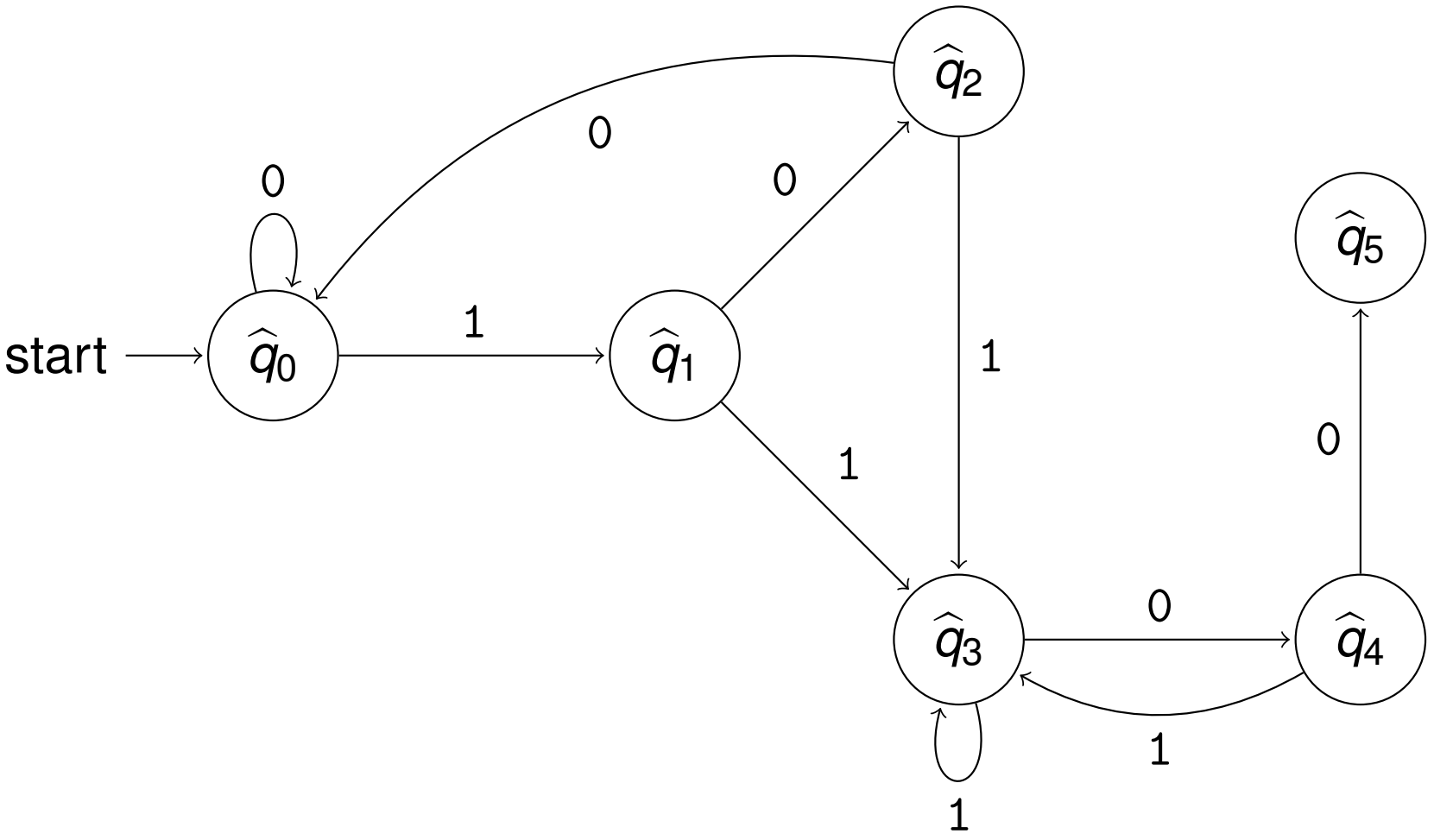
During the second execution of the body of the `for` loop (included in the fifth execution of the body of the outer `while` loop) $V = \widehat{q}_4 = \{q_0, q_2, q_3\}$ and $\sigma = 1$, so that

$$
\begin{aligned}
W &= \bigcup_{r \in \widehat{q}_4} \left(\bigcup_{s \in \delta(r,1)} Cl_\lambda(s) \right) \\
&= \bigcup_{s \in \delta(q_0,1)} Cl_\lambda(s) \quad \cup \quad \bigcup_{s \in \delta(q_2,1)} Cl_\lambda(s) \quad \cup \quad \bigcup_{s \in \delta(q_3,1)} Cl_\lambda(s) \\
&= (Cl_\lambda(q_0) \cup Cl_\lambda(q_1)) \cup Cl_\lambda(q_3) \cup Cl_\lambda(q_3) \\
&= (\{q_0\} \cup \{q_1, q_2\}) \cup \{q_3\} \cup \{q_3\} \\
&= \{q_0, q_1, q_2, q_3\} = \widehat{q}_3.
\end{aligned}
$$

Since $\widehat{q}_3$ already belongs to $\widehat{Q} \cup \widehat{R}$, neither $\widehat{Q}$ nor $\widehat{R}$ is changed.

Subset Construction: Application to the Example

At the end of the fifth execution of body of the `while` loop $\widehat{q}_4$ is moved from $\widehat{R}$ to $\widehat{Q}$, so that $\widehat{Q} = \{\widehat{q}_0, \widehat{q}_1, \widehat{q}_2, \widehat{q}_3, \widehat{q}_4\}$ and $\widehat{R} = \{\widehat{q}_5\}$.The DFA, so far, is as follows.



Subset Construction: Application to the Example

In the sixth execution of the body of the `while` loop V is set to be $\widehat{q}_5 = \{q_0, q_3\}$, since this is the only element of $\widehat{R}$. Since $\sigma = \mathtt{0}$ at the beginning of the first execution of the body of the `for` loop,

$$
\begin{aligned}
W &= \bigcup_{r \in \widehat{q}_5} \left(\bigcup_{s \in \delta(r,\mathtt{0})} Cl_\lambda(s) \right) \\
&= \bigcup_{s \in \delta(q_0,\mathtt{0})} Cl_\lambda(s) \quad \cup \quad \bigcup_{s \in \delta(q_3,\mathtt{0})} Cl_\lambda(s) \\
&= Cl_\lambda(q_0) \cup Cl_\lambda(q_3) \\
&= \{q_0\} \cup \{q_3\} \\
&= \{q_0, q_3\} = \widehat{q}_5.
\end{aligned}
$$

Since $\widehat{q}_5$ already belongs to $\widehat{Q} \cup \widehat{R}$, neither $\widehat{Q}$ nor $\widehat{R}$ is changed.

Subset Construction: Application to the Example

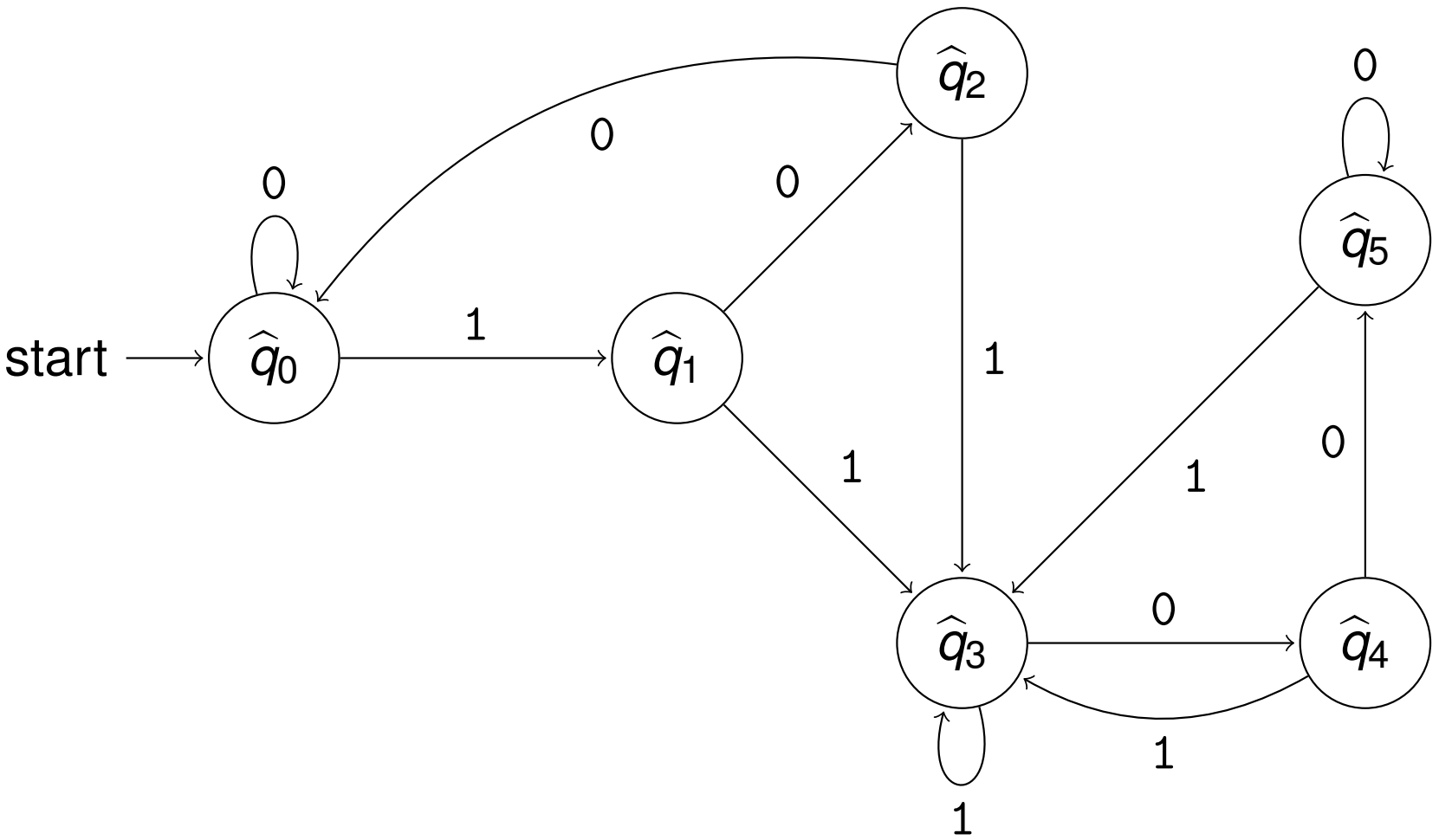
During the second execution of the body of the `for` loop (included in the sixth execution of the body of the outer `while` loop) $V = \widehat{q}_5 = \{q_0, q_3\}$ and $\sigma = 1$, so that

$$
\begin{aligned}
W &= \bigcup_{r \in \widehat{q}_5} \left(\bigcup_{s \in \delta(r,1)} Cl_\lambda(s) \right) \\
&= \bigcup_{s \in \delta(q_0,1)} Cl_\lambda(s) \quad \cup \quad \bigcup_{s \in \delta(q_3,1)} Cl_\lambda(s) \\
&= (Cl_\lambda(q_0) \cup Cl_\lambda(q_1)) \cup Cl_\lambda(q_3) \\
&= (\{q_0\} \cup \{q_1, q_2\}) \cup \{q_3\} \\
&= \{q_0, q_1, q_2, q_3\} = \widehat{q}_3.
\end{aligned}
$$

Since $\widehat{q}_3$ already belongs to $\widehat{Q} \cup \widehat{R}$, neither $\widehat{Q}$ nor $\widehat{R}$ is changed.

Subset Construction: Application to the Example

At the end of the sixth execution of body of the `while` loop $\widehat{q}_5$ is moved from $\widehat{R}$ to $\widehat{Q}$, so that $\widehat{Q} = \{\widehat{q}_0, \widehat{q}_1, \widehat{q}_2, \widehat{q}_3, \widehat{q}_4, \widehat{q}_5\}$ and $\widehat{R} = \emptyset$.The DFA, so far, is as follows.



The "Subset Construction:" Why This Works

It can be proved (using induction on the number of executions of the body of the `while` loop) that the following properties are satisfied at the beginning and end of ***every*** execution of the body of this loop:

1. $\delta^\star(q_0, \lambda) = Cl_\lambda(q_0)$ belongs to either $\widehat{Q}$ or $\widehat{R}$.
2. For every subset V of Q such that $V \in \widehat{Q} \cup \widehat{R}$, there is a string $\mu \in \Sigma^\star$ such that $V = \delta^\star(q_0, \mu)$.
3. $\widehat{Q} \cap \widehat{R} = \emptyset$.
4. For every subset V of Q such that $V \in \widehat{Q}$ and for every symbol $\sigma \in \Sigma$, the set

$$W = \bigcup_{r \in V} \left(\bigcup_{s \in \delta(r,\sigma)} Cl_\lambda(s) \right)$$

belongs to $\widehat{Q} \cup \widehat{R}$.

The "Subset Construction:" Why This Works

This implies that if $\widehat{R} = \emptyset$ then

5. $\delta^\star(q_0, \lambda) = Cl_\lambda(q_0) \in \widehat{Q}$ (this will correspond to the ***start state*** of the DFA being designed);
6. For every subset V of Q such that $V \in \widehat{Q}$, there is a string $\mu \in \Sigma^\star$ such that $V = \delta^\star(q_0, \mu)$ (...that is, we are not including states in our DFA that can never even be reached).
7. For every subset V of Q such that $V \in \widehat{Q}$ and for every symbol $\sigma \in \Sigma$, the set

$$W = \bigcup_{r \in V} \left(\bigcup_{s \in \delta(r,\sigma)} Cl_\lambda(s) \right)$$

also belongs to $\widehat{Q}$ (...so the "transition function" can actually be defined).

The "Subset Construction:" Why This Works

- Notice that $\widehat{Q}$ is initially empty and its size is increased by one during each execution of the body of the `while` loop.
- ***Conclusion:*** The execution of the loop must end (with $\widehat{R} = \emptyset$) after at most 2^k executions of the loop body if $k = |Q|$ — because $\widehat{Q}$ would have to include every set in $\mathcal{P}(Q)$ after that and, since $\widehat{R} \subseteq \mathcal{P}(Q)$ and $\widehat{Q} \cap \widehat{R} = \emptyset$, it would follow that $\widehat{R}$ would *have* to be empty, as claimed.
- The above properties 5–7 imply that the the transition function has been completely defined (and is a well-defined total function $\widehat{\delta} : \widehat{Q} \times \Sigma \to \widehat{Q}$ at that point, as required).

The "Subset Construction:" Choosing Final Sets

- Finally, for each state $\widehat{q} \in \widehat{Q} \subseteq \mathcal{P}(Q)$, include $\widehat{q}$ in the set $\widehat{F}$ ***if and only if*** $\widehat{q} \cap F \neq \emptyset$.
- *Application to our Example:* $\widehat{Q} = \{\widehat{q}_0, \widehat{q}_1, \widehat{q}_2, \widehat{q}_3, \widehat{q}_4, \widehat{q}_5\}$
  - $\widehat{q}_0 \cap F = \{q_0\} \cap \{q_3\} = \emptyset$, so $\widehat{q}_0 \notin \widehat{F}$.
  - $\widehat{q}_1 \cap F = \{q_0, q_1, q_2\} \cap \{q_3\} = \emptyset$, so $\widehat{q}_1 \notin \widehat{F}$.
  - $\widehat{q}_2 \cap F = \{q_0, q_2\} \cap \{q_3\} = \emptyset$, so $\widehat{q}_2 \notin \widehat{F}$.
  - $\widehat{q}_3 \cap F = \{q_0, q_1, q_2, q_3\} \cap \{q_3\} = \{q_3\}$, so $\widehat{q}_3 \in \widehat{F}$.
  - $\widehat{q}_4 \cap F = \{q_0, q_2, q_3\} \cap \{q_3\} = \{q_3\}$, so $\widehat{q}_4 \in \widehat{F}$.
  - $\widehat{q}_5 \cap F = \{q_0, q_3\} \cap \{q_3\} = \{q_3\}$, so $\widehat{q}_5 \in \widehat{F}$.

  Thus $\widehat{F} = \{\widehat{q}_3, \widehat{q}_4, \widehat{q}_5\}$.

The "Subset Construction:" Why This Works

- Recall that if $\omega \in \Sigma^\star$ then both
  - $\delta^\star(q_0, \omega)$ — the value of the extended transition function for the NFA M for the start state, q_0, and the string ω, and
  - $\widehat{\delta}^\star(\widehat{q}_0, \omega)$ — the value of the extended tradition function for the DFA $\widehat{M}$ for the start state, $\widehat{q}_0 = Cl_\lambda(q_0)$ and string ω

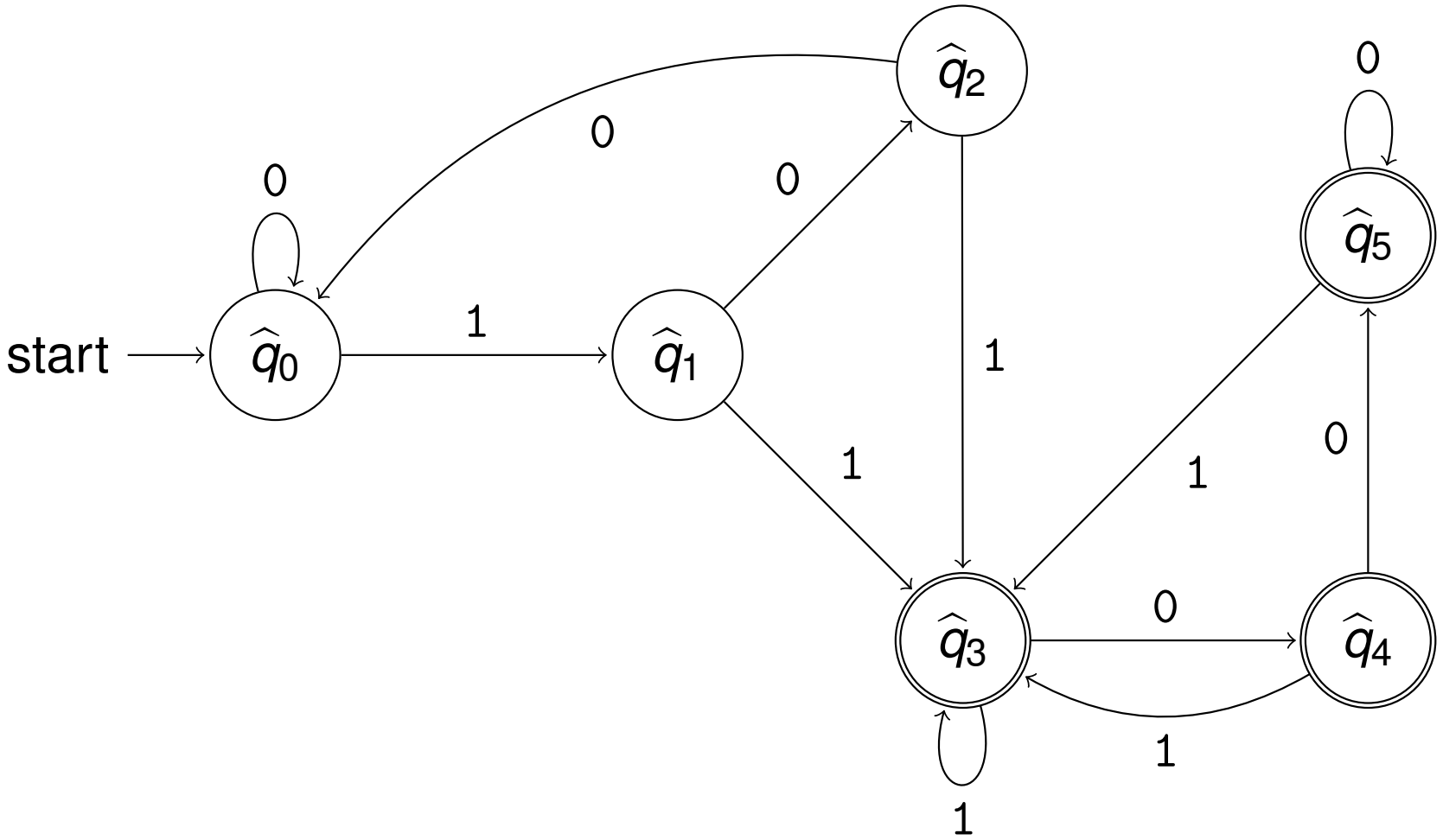
  are *sets* of states in Q — that is, elements of $\mathcal{P}(Q)$.
- It is possible to prove the following result (by induction on the length of the string ω):

$$\delta^\star(q_0, \omega) = \widehat{\delta}(\widehat{q}_0, \omega)$$

  for ***every*** string $\omega \in \Sigma^\star$
- If $\widehat{F}$ is as defined above then it follows that M accepts ω if and only if $\widehat{M}$ accepts ω, for all $\omega \in \Sigma^\star$. Thus $L(\widehat{M}) = L(M)$, as required.

Application to the Example

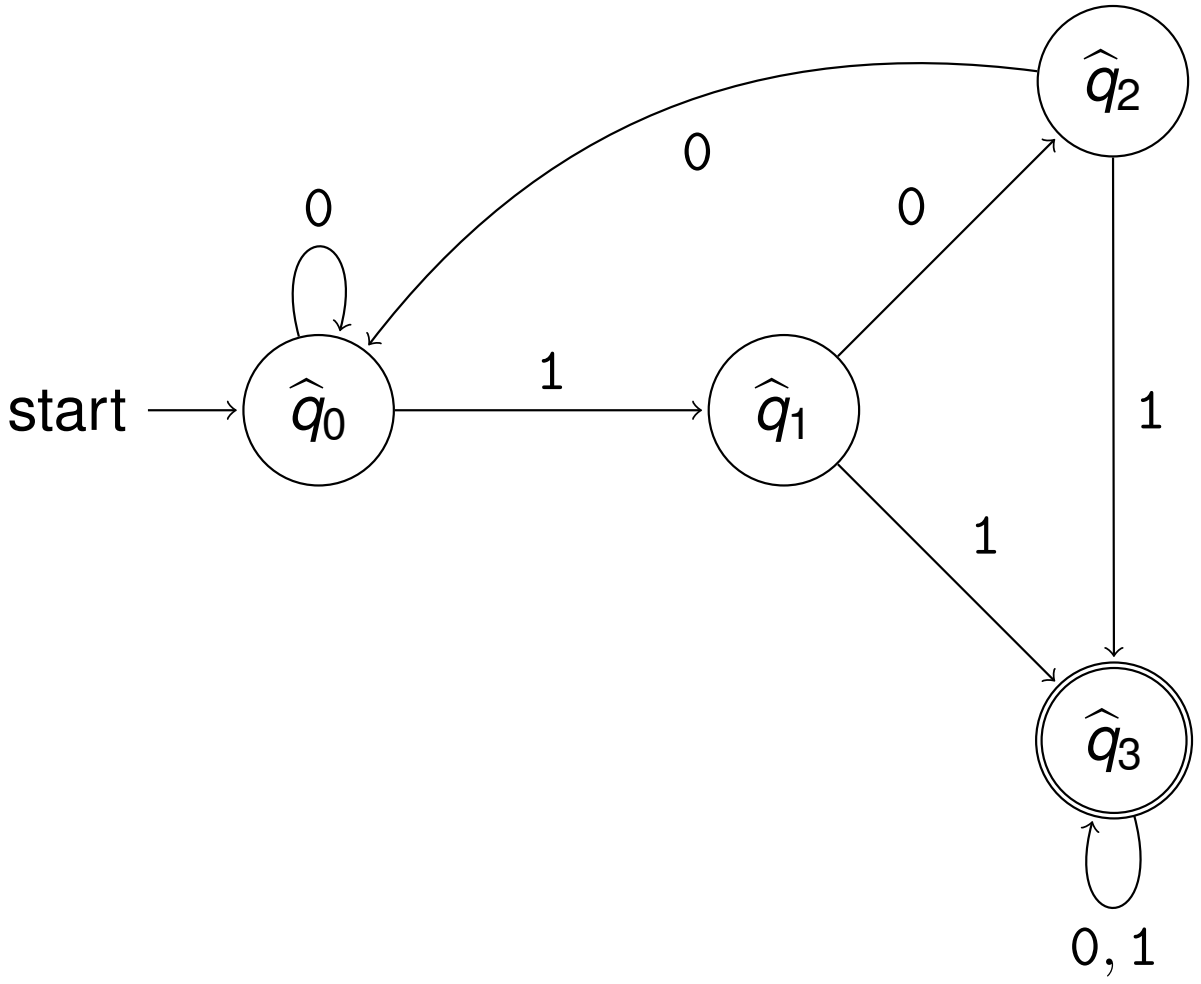
The resulting DFA is as follows.



Simplifying this Deterministic Finite Automaton

- Notice that, once this DFA enters an accepting state, it simply moves from one accepting state to another accepting state — it never returns to a state that is *not* in $\widehat{F}$.
- Thus all the accepting states can be replaced by a single one without changing the language of the DFA. The resulting, simpler DFA (for the same language) is as follows.

A Simpler DFA



Expectations for Students

- Students *will not* be asked to explain why the "subset construction" is correct (even though a proof of this is sketched in these online notes).
- Students *might* be given an NFA M for a language in $\Sigma^\star$ and asked to generate a DFA for the same language and explain why the languages of the two automata are the same. Explaining why the NFA and DFA have the same language is easy to do (because the explanation is "the subset construction is correct") if the subset construction is used to solve this problem.
- ***Note:*** The rest of this set of lecture notes is "for interest only" — and will probably not be covered (since there will probably not be time for it in class).

A Bad Case for the "Subset Construction"

- It follows from the above that if $L \subseteq \Sigma^\star$ and there is an NFA

$$M = (Q, \Sigma, \delta, q_0, F)$$

with language L such that M has k states (i.e., $|Q| = k$), then there is also a DFA

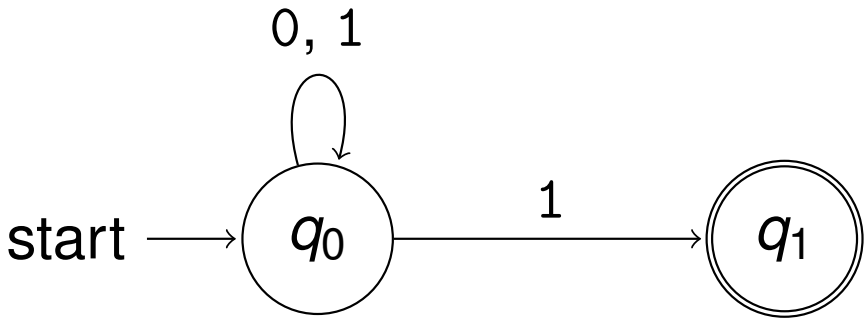
$$\widehat{M} = (\widehat{Q}, \Sigma, \widehat{\delta}, \widehat{F})$$

which also has language L, and which has *at most* 2^k states.

- Much of the time, the DFA generated using the "subset construction" will have a *much* smaller set of states than this suggest.
- ***Question:*** Is an exponential increase in the number of states ever really necessary?
- ***Answer:*** Yes!

A Bad Case for the "Subset Construction"

- Let $\Sigma = \{0, 1\}$.
- Let $L_1 = \{\omega \in \Sigma^\star \mid \omega \text{ ends with a } 1\}$. Then the following NFA has language L_1:



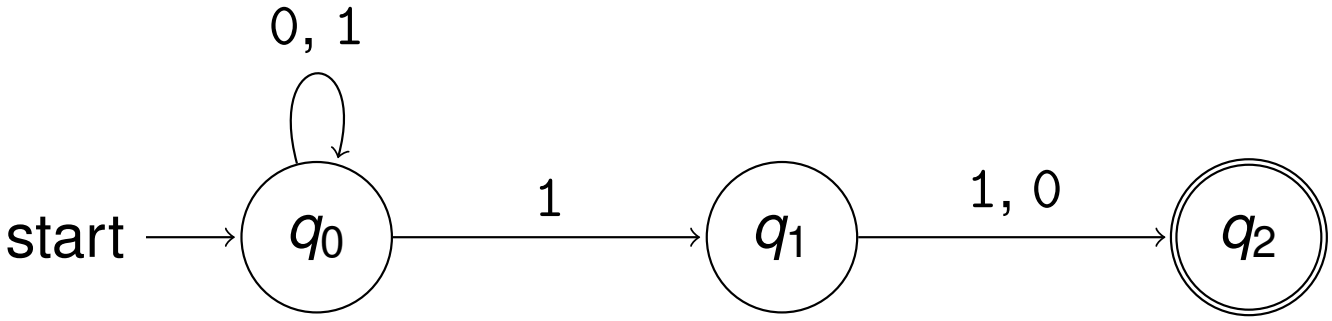
- Languages $L_2, L_3, L_4, \cdots \subseteq \Sigma^\star$ can be "inductively defined" by setting

$$L_{k+1} = \{\omega \cdot \sigma \mid \omega \in L_k \text{ and } \sigma \in \Sigma\}.$$

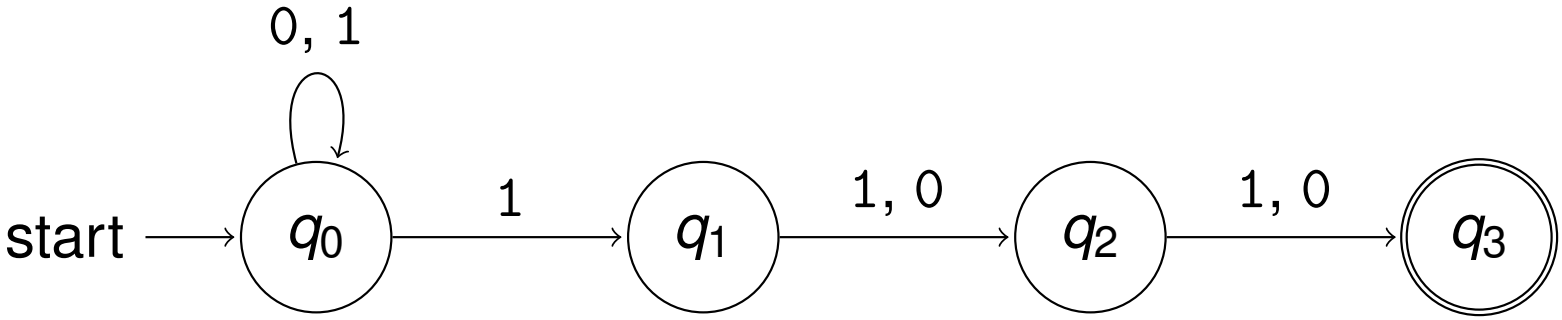
Then L_2 includes all strings in $\Sigma^\star$ whose *second-to-last* symbol is 1... and so on.

A Bad Case for the "Subset Construction"

- Now, the following NFA has language L_2...



- ... and the following NFA has language L_3:



A Bad Case for the "Subset Construction"

- For $k \geq 1$ consider an NFA $M_k = (Q_k, \Sigma, \delta_k, F_k)$ where
  - $Q_k = \{q_0, q_1, q_2, \ldots, q_k\}$, so that M_k has $k+1$ states.
  - $\delta_k(q_0, 0) = \{q_0\}$, $\delta_k(q_0, 1) = \{q_0, q_1\}$, and $\delta_k(q_0, \lambda) = \emptyset$;
  - For every integer j such that $1 \leq j \leq k-1$, $\delta_k(q_j, 0) = \delta_k(q_j, 1) = \{q_{j+1}\}$ and $\delta_k(q_j, \lambda) = \emptyset$;
  - $\delta_k(q_k, 0) = \delta_k(q_k, 1) = \delta_k(q_k, \lambda) = \emptyset$.
  - $F_k = \{q_k\}$

  Note that the NFA's on the previous slides are the NFA's M_2 and M_3, respectively.
- It is possible to prove the following (about M_k) by induction on i: For every integer i such that $1 \leq i \leq k$, and for every string $\omega \in \Sigma^\star$,

$$q_i \in \delta^\star(q_0, \omega) \text{ if and only if } \omega \in L_i.$$

- Thus $L(M_k) = L_k$ — so that L_k has an NFA with only $k+1$ states.

A Bad Case for the "Subset Construction"

Claim: Let $\widehat{M} = (\widehat{Q}, \Sigma, \widehat{\delta}, \widehat{q}_0, \widehat{F})$ be a DFA such that $L(\widehat{M}) = L_k$. Then $|\widehat{Q}| \geq 2^k$, that is, $\widehat{M}$ has *at least* 2^k states.

Proof: By contradiction.

- Suppose that there is a DFA $\widehat{M}$ whose language is L_k such that $\widehat{M}$ has *strictly fewer than* 2^k states.
- $\Sigma^\star$ has *exactly* 2^k strings with length k so it follows by the "Pigeonhole Principle" that there exist strings

$$\omega_1 = \sigma_1\sigma_2\ldots\sigma_k \text{ and } \omega_2 = \tau_1\tau_2\ldots\tau_k$$

in $\Sigma^\star$, both with length k, such that $\omega_1 \neq \omega_2$ but $\widehat{\delta}^\star(\widehat{q}_0, \omega_1) = \widehat{\delta}^\star(\widehat{q}_0, \omega_2)$.

A Bad Case for the "Subset Construction"

- Since $\omega_1 \neq \omega_2$ there is an integer i such that $1 \leq i \leq k$ and $\sigma_i \neq \tau_i$.
- Without loss of generality we may assume that $\sigma_i = 1$ and $\tau_i = 0$ (we can just switch ω_1 and ω_2 otherwise).
- Then $\omega_1 \in L_{k-i+1}$ and $\omega_2 \notin L_{k-i+1}$.
- For $\ell \geq 0$ let 1^ℓ denote a string of ℓ 1's — so that $1^0 = \lambda$, $1 = 1$, $1^2 = 11$, and so on.

A Bad Case for the "Subset Construction"

- Each of the following things can now be proved by induction on ℓ: For every integer $\ell \geq 0$,
  a) $\omega_1 \cdot 1^\ell \in L_{k+\ell-i+1}$ and $\omega_2 \cdot 1^\ell \notin L_{k+\ell-i+1}$ — so that (in particular, with $\ell = i-1$) $\omega_1 \cdot 1^{i-1} \in L_k$ and $\omega_2 \cdot 1^{i-1} \notin L_k$.
  b) $\widehat{\delta}(\widehat{q}_0, \omega_1 \cdot 1^\ell) = \widehat{\delta}(\widehat{q}_0, \omega_2 \cdot 1^\ell)$ — so that (in particular, with $\ell = i-1$) $\widehat{\delta}(\widehat{q}_0, \omega_1 \cdot 1^{i-1})$ and $\widehat{\delta}(\widehat{q}_0, \omega_2 \cdot 1^{i-1})$ are both equal to the same state $\widehat{q} \in \widehat{Q}$.
- Now, since $\omega_1 \cdot 1^{i-1} \in L_k$, $\widehat{\delta}(\widehat{q}_0, \omega_1 \cdot 1^{i-1}) = \widehat{q}$, and $L(\widehat{M}) = L_k$, it must be true that $\widehat{q} \in \widehat{F}$.
- Since $\widehat{\delta}(\widehat{q}_0, \omega_2 \cdot 1^{i-1}) = \widehat{q}$ it now follows that $\omega_2 \cdot 1^{i-1} \in L(\widehat{M}) = L_k$ as well.

A Bad Case for the "Subset Construction"

- We have a ***contradiction*** — because we already know that $\omega_2 \cdot 1^{i-1} \notin L_k$.
- So, an assumption that we made, along the way, must be incorrect. We only made one assumption, so *that* one must be false: "The DFA for L_k being considered has fewer than 2^k states."
- Since this was an arbitrarily chosen DFA whose language is L_k, it now follows that ***every*** DFA whose language is L_k must have at least 2^k states, as claimed.

Reference

The information about the "bad case for the subset construction" is taken from

- J. Hopcroft, J. Ullman, and R. Motwani
  Introduction to Automata Theory, Languages, and Computation
  Third Edition, Prentice Hall, 2006

What's Next?

Coming Up...

- A set of operations on languages, called ***regular operations***, will be introduced.
- ***Closure properties*** will also be introduced. Several "closure properties" will be stated and proved, for the set of regular languages.
- This will provide additional ways to show that a given language is regular.