1 Inductive Proofs for DFAs

1.1 Properties about DFAs

Deterministic Behavior

Proposition 1. *For a DFA* $M = (Q, \Sigma, \delta, q_0, F)$*, and any* $q \in Q$*, and* $w \in \Sigma^*$*,* $|\hat{\delta}_M(q, w)| = 1$*.*

Proof. Proof is by induction on $|w|$. Thus, S_i is taken to be

For every $q \in Q$, and $w \in \Sigma^i$, $|\hat{\delta}_M(q, w)| = 1$.

Base Case: We need to prove the case when $w \in \Sigma^0$. Thus, $w = \epsilon$. By definition $\xrightarrow{w}_M$, $q \xrightarrow{w}_M q'$ if and only $q' = q$. Thus, $|\hat{\delta}_M(q, w)| = |\{q\}| = 1$.

Ind. Hyp.: Suppose for every $q \in Q$, and $w \in \Sigma^*$ such that $|w| < i$, $|\hat{\delta}_M(q, w)| = 1$.

Ind. Step: Consider (without loss of generality) $w = a_1 a_2 \cdots a_i$, such that $a_i \in \Sigma$. Take $u = a_1 \cdots a_{i-1}$

$$\begin{aligned} q \xrightarrow{w}_M q' \quad & \text{iff there are } r_0, r_1, \ldots, r_i \text{ such that } r_0 = q,\ r_i = q',\ \text{and } \delta(r_j, a_{j+1}) = r_{j+1} \\ & \text{iff there is } r_{i-1} \text{ such that } q \xrightarrow{u}_M r_{i-1} \text{ and } \delta(r_{i-1}, a_i) = q' \end{aligned}$$

Now, by induction hypothesis, since $|\hat{\delta}_M(q, u)| = 1$, there is a unique r_{i-1} such that $q \xrightarrow{u}_M r_{i-1}$. Also, since from any state r_{i-1} on symbol a_i the next state is uniquely determined, $|\hat{\delta}_M(q, w)| = 1$.

□

DFA Computation

Proposition 2. *Let* $M = (Q, \Sigma, \delta, q_0, F)$ *be a DFA. For any* $q_1, q_2 \in Q$*,* $u, v \in \Sigma^*$*,* $q_1 \xrightarrow{uv}_M q_2$ *iff there is* $q \in Q$ *such that* $q_1 \xrightarrow{u}_M q$ *and* $q \xrightarrow{v}_M q_2$*.*

Proof. Let $u = a_1 a_2 \ldots a_i$ and $v = a_{i+1} \cdots a_{i+k}$. Observe that,

$$\begin{aligned} q_1 \xrightarrow{uv}_M q_2 \quad & \text{iff there are } r_0, r_1, \ldots, r_{i+k} \text{ such that } r_0 = q_1,\ r_{i+k} = q_2,\ \text{and } \delta(r_j, a_{j+1}) = r_{j+1} \\ & \text{iff there is } r_i\ (= q \text{ of the proposition}) \text{ such that } q_1 \xrightarrow{u}_M r_i \text{ and } r_i \xrightarrow{v}_M q_2 \end{aligned}$$

□

Conventions in Inductive Proofs

"We will prove by induction on $|v|$" is a short-hand for "We will prove the proposition by induction. Take S_i to be statement of the proposition restricted to strings v where $|v| = i$."

1.2 Proving Correctness of DFA Constructions

Proving Correctness of DFAs

Problem
Show that DFA M recognizes language L.

That is, we need to show that for all w, $w \in \mathbf{L}(M)$ iff $w \in L$. This is often carried out by induction on $|w|$.

Example I

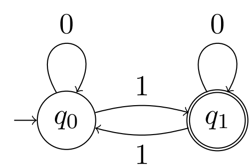

Figure 1: Transition Diagram of M_1

Proposition 3. $\mathbf{L}(M_1) = \{w \in \{0,1\}^* \mid w \textit{ has an odd number of 1s}\}$

Proof. We will prove this by induction on $|w|$. That is, let S_i be

For all $w \in \{0,1\}^i$. M_1 accepts w iff w has an odd number of 1s

Observe that M_1 accepts w iff $q_0 \xrightarrow{w}_{M_1} q_1$. So we could rewrite S_i as

For all $w \in \{0,1\}^i$. $q_0 \xrightarrow{w}_{M_1} q_1$ iff w has an odd number of 1s

Base Case: When $w = \epsilon$, w has an even number of 1s. Further, $q_0 \xrightarrow{\epsilon}_{M_1} q_0$, and so M_1 does not accept w.

Ind. Hyp.: Assume that for all w of length $< n$, $q_0 \xrightarrow{w}_{M_1} q_1$ iff w has an odd number of 1s.

Ind. Step: Consider w of length n; without loss of generality, w is either $0u$ or $1u$ for some string u of length $i-1$.

If $w = 0u$ then, w has an odd number of 1s iff u has an odd number of 1s, iff (by ind. hyp.) $q_0 \xrightarrow{u}_{M_1} q_1$ iff $q_0 \xrightarrow{w=0u}_{M_1} q_1$ (since $\delta(q_0, 0) = q_0$).

On the other hand, if $w = 1u$ then, w has an odd number of 1s iff u has an even number of 1s. Now $q_0 \xrightarrow{w=1u}_{M_1} q_1$ iff $q_1 \xrightarrow{u}_{M_1} q_1$. Does M_1 accept u that has an even number of 0s from state q_1? Unfortunately, we cannot use the induction hypothesis in this case, as the hypothesis does not say anything about what strings u are accepted when the automaton is started from state q_1; it only gives the behavior on strings when M_1 is started in the initial state q_0. We need to strengthen the hypothesis to make the proof work!! The strengthening will explicitly tell us the behavior of the machine on strings when starting from states other than the initial state.

New (correct) induction proof: Let S_i be

$$\forall w \in \{0,1\}^i. \quad q_0 \xrightarrow{w}_{M_1} q_1 \text{ iff } w \text{ has an odd number of 1s} \\ \text{and } q_1 \xrightarrow{w}_{M_1} q_1 \text{ iff } w \text{ has an even number of 1s}$$

We will prove this sequence of statements by induction.

Base Case: When $w = \epsilon$, w has an even number of 1s. Further, $q_0 \xrightarrow{\epsilon}_{M_1} q_0$ and $q_1 \xrightarrow{w}_{M_1} q_1$, and so M_1 does not accept w from state q_0, but accepts w from state q_1. This establishes the base case.

Ind. Hyp.: Assume that for all w of length $< n$, $q_0 \xrightarrow{w}_{M_1} q_1$ iff w has an odd number of 1s and $q_1 \xrightarrow{w}_{M_1} q_1$ iff w has an even number of 1s.

Ind. Step: Consider w of length n; without loss of generality, w is either $0u$ or $1u$ for some string u of length $i - 1$.

If $w = 0u$ then $q_0 \xrightarrow{0u}_{M_1} q_1$ iff $q_0 \xrightarrow{u}_{M_1} q_1$ (because $\delta(q_0, 0) = q_0$) iff u has an odd number of 1s (by ind. hyp.) iff $w = 0u$ has an odd number of 1s. Similarly, $q_1 \xrightarrow{0u}_{M_1} q_1$ iff $q_1 \xrightarrow{u}_{M_1} q_1$ (because $\delta(q_1, 0) = q_1$) iff u has an even number of 1s iff $w = 0u$ has an even number of 1s.

On the other hand, if $w = 1u$ then $q_0 \xrightarrow{w=1u}_{M_1} q_1$ iff $q_1 \xrightarrow{u}_{M_1} q_1$ (since $\delta(q_0, 1) = q_1$) iff (by ind. hyp.) u has an even number of 1s iff $w = 1u$ has an odd number of 1s. Similarly, $q_1 \xrightarrow{w=1u}_{M_1} q_1$ iff $q_0 \xrightarrow{u}_{M_1} q_1$ (since $\delta(q_1, 1) = q_0$) iff (by ind. hyp.) u has an odd number of 1s iff w has an even number of 1s.

□

Remark

The above induction proof can be made to work *without* strengthening if in the first induction proof step, we considered $w = ua$, for $a \in \{0,1\}$, instead of $w = au$ as we did. However, the fact that the induction proof works without strengthening here is a very special case, and does not hold in general for DFAs.

Example II

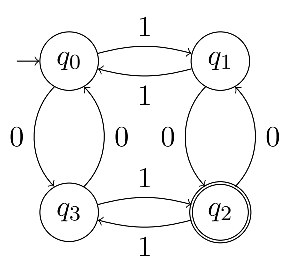

Figure 2: Transition Diagram of M_2

Proposition 4. $\mathbf{L}(M_2) = \{w \in \{0,1\}^* \mid$ *w has an odd number of 1s and odd number of 0s*$\}$

Proof. We will once again prove the proposition by induction on $|w|$. The straightforward proof would suggest that we take S_i to be

$$\text{For any } w \in \{0,1\}^i.\ M_2 \text{ accepts } w \text{ iff } w \text{ has an odd number of 1s and 0s}$$

Since M_2 accepts w iff $q_0 \xrightarrow{w}_{M_2} q_2$, we could rewrite the condition as "$q_0 \xrightarrow{w}_{M_2} q_2$ iff w has an odd number of 1s and 0s". The induction proof will unfortunately not go through! To see this, consider the induction step, when $w = 0u$. Now, $q_0 \xrightarrow{w}_{M_2} q$ iff $q_3 \xrightarrow{u}_{M_2} q$, because M_2 goes to state q_3 (from q_0) on reading 0. Since w and u have the same parity for the number of 1s, but opposite parity for the number of 0s, w must be accepted (i.e., reach state q_2) iff u is accepted from q_3 when u has an odd number of 1s and even number of 0s. But is that the case? The induction hypothesis says nothing about strings accepted from state q_3, and so the induction step cannot be established.

This is typical of many induction proofs. Again, we must *strengthen* the proposition in order to construct a proof. The proposition must not only characterize the strings that are accepted from the initial state q_0, but also those that are accepted from states q_1, q_2, and q_3.

We will show by induction on w that

(a) $q_0 \xrightarrow{w}_{M_2} q_2$ iff w has an odd number of 0s and odd number of 1s,

(b) $q_1 \xrightarrow{w}_{M_2} q_2$ iff w has odd number of 0s and even number of 1s,

(c) $q_2 \xrightarrow{w}_{M_2} q_2$ iff w has an even number of 0s and even number of 1s, and

(d) $q_3 \xrightarrow{w}_{M_2} q_2$ iff w has even number of 0s and odd number of 1s.

Thus in the our new induction proof, statement S_i says that conditions (a),(b),(c), and (d) hold for all strings of length i.

Base Case: When $|w| = 0$, $w = \epsilon$. Observe that w has an even number of 0s and 1s, and $q \xrightarrow{\epsilon}_{M_2} q$ for any state q. String ϵ is only accepted from state q_2, and thus statements (a),(b),(c), and (d) hold in the base case.

Ind. Hyp.: Suppose (a),(b),(c),(d) all hold for any string w of length $< n$.

Ind. Step: Consider w of length n. Without loss of generality, w is of the form au, where $a \in \{0,1\}$ and $u \in \{0,1\}^{n-1}$.

- *Case $q = q_0$, $a = 0$:* $q_0 \xrightarrow{0u}_{M_2} q_2$ iff $q_3 \xrightarrow{u}_{M_2} q_2$ iff u has even number of 0s and odd number of 1s (by ind. hyp. (d)) iff w has odd number of 0s and odd number of 1s.
- *Case $q = q_0$, $a = 1$:* $q_0 \xrightarrow{1u}_{M_2} q_2$ iff $q_1 \xrightarrow{u}_{M_2} q_2$ iff u has odd number of 0s and even number of 1s (by ind. hyp. (b)) iff w has odd number of 0s and odd number of 1s
- *Case $q = q_1$, $a = 0$:* $q_1 \xrightarrow{0u}_{M_2} q_2$ iff $q_2 \xrightarrow{u}_{M_2} q_2$ iff u has even number of 0s and even number of 1s (by ind. hyp. (c)) iff w has odd number of 0s and even number of 1s
- …And so on for the other cases of $q = q_1$ and $a = 1$, $q = q_2$ and $a = 0$, $q = q_2$ and $a = 1$, $q = q_3$ and $a = 0$, and finally $q = q_3$ and $a = 1$. □

Proving Correctness of a DFA

Proof Template
Given a DFA M having n states $\{q_0, q_1, \ldots q_{n-1}\}$ with initial state q_0, and final states F, to prove that $L(M) = L$, we do the following.

1. Come up with languages $L_0, L_1, \ldots L_{n-1}$ such that $L_0 = L$
2. Prove by induction on $|w|$, $\hat{\delta}_M(q_i, w) \cap F \neq \emptyset$ if and only if $w \in L_i$

---

2 Proving DFA Lower Bounds

Even length strings with at least 2 as

Problem
Design an automaton for the language $L^{\geq 2a}_{\text{even}} = \{w \in \{a,b\}^* \mid w$ has even length and contains at least 2 as$\}$.

Solution
What do you need to remember? We need to remember the numbers of as we have seen (either 0, 1, or ≥ 2), *and* the parity (odd or even) of the number of symbols we have seen. So the states will be $\langle 0, e\rangle$ (no as seen, and even number of total symbols), $\langle 0, o\rangle$ (no as seen and total symbols seen is odd), $\langle 1, e\rangle$ (one a seen and even number of symbols), $\langle 1, o\rangle$ (one a seen and odd number of symbols), $\langle 2, e\rangle$ (at least 2 as seen and even number of total symbols), and $\langle 2, o\rangle$ (at least 2 as seen and an odd number of symbols). The DFA is as follows.

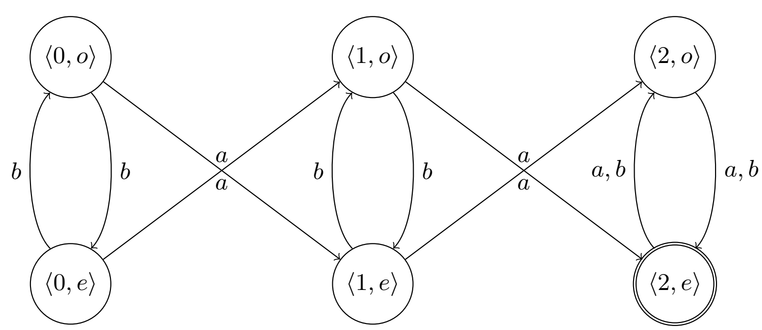

Figure 3: DFA recognizing $L^{\geq 2a}_{\text{even}}$

Tightness of DFA construction

Proposition 5. *Any DFA recognizing $L^{\geq 2a}_{\text{even}}$ has at least 6 states.*

Proof Idea
We will identify 6 strings $w_1, w_2, w_3, w_4, w_5, w_6$ that must take any DFA recognizing $L^{\geq 2a}_{\text{even}}$ to different states.

- How do we find such strings? Based on our intuition of what the DFA must remember in order to solve the problem.
- How do we prove that they must go to different states? We will argue that if some pair of strings (say w_i and w_j) go to the same state in some DFA M, then M cannot recognize the language $L^{\geq 2a}_{\text{even}}$. The crux of the argument is as follows. Suppose M ends up in the same state (say q) after reading w_i and w_j. Since the DFA behavior only depends on the current state (and not on how the current state was reached), M will give the same answer (either accept or reject) on strings $w_i u$ and $w_j u$, no matter what u is. Now, if we find a string u such that $w_i u \in L^{\geq 2a}_{\text{even}}$ and $w_j u \notin L^{\geq 2a}_{\text{even}}$ then M gives an answer on $w_i u$ and $w_j u$ (either both accept or both reject) that is inconsistent with what the problem $L^{\geq 2a}_{\text{even}}$ requires it to do.

We are now ready to formally prove the proposition based on the above idea.

Proof. Consider the strings $w_1 = \epsilon$, $w_2 = b$, $w_3 = a$, $w_4 = ab$, $w_5 = aa$, $w_6 = aab$. Our intuition tells us that a DFA recognizing $L^{\geq 2a}_{\text{even}}$ must remember different things for each of these strings: for w_1 the fact that we have seen no as and an even number of symbols; for w_2 the fact that we have seen no as but an odd number of symbols; for w_3 the fact that we have seen an a and an odd number of symbols; for w_4 the fact that we have seen an a and an even number of symbols; for w_5 the fact that we have seen at least 2 as and an even number of symbols; for w_6 the fact that we have seen at least 2 as and an odd number of symbols.

Suppose (for contradiction) the proposition does not hold. That is, there is a DFA M with < 6 states that recognizes $L^{\geq 2a}_{\text{even}}$. Let the initial state of M be (say) q_0. Now, by pigeon hole principle, there must be two strings $w_i, w_j \in \{w_1, w_2, w_3, w_4, w_5, w_6\}$ such that M, starting in state q_0, goes to the same state on both w_i and w_j, i.e., for some q, $\hat{\delta}_M(q_0, w_i) = \{q\} = \hat{\delta}_M(q_0, w_j)$. We will show that then M cannot recognize $L^{\geq 2a}_{\text{even}}$, contradicting our assumption that M does recognize $L^{\geq 2a}_{\text{even}}$ and thus proving our proposition. How does one prove this? Observe that if $\hat{\delta}_M(q_0, w_i) = \hat{\delta}_M(q_0, w_j)$ then no matter what u is, $\hat{\delta}_M(q_0, w_i u) = \hat{\delta}_M(q_0, w_j u)$, which means that either M accepts both $w_i u$ and $w_j u$, or rejects both of them. Now if we find a u such that $w_i u \in L^{\geq 2a}_{\text{even}}$ and $w_j u \notin L^{\geq 2a}_{\text{even}}$ then it means that M cannot recognize $L^{\geq 2a}_{\text{even}}$. We will identify u based on what w_i and w_j are.

For every pair of strings (among $w_1, w_2, w_3, w_4, w_5, w_6$) we need to find a "witness" string u with the above properties. That gives us 15 cases to consider, but we will combine many of the cases together.

- Case $w_i = w_5$ and $w_j \in \{w_1, w_2, w_3, w_4, w_6\}$. Suppose $\hat{\delta}_M(q_0, w_5) = \hat{\delta}_M(q_0, w_j)$. Then for all u, either $w_5 u$ and $w_j u$ are both accepted by M or neither one is. Take $u = \epsilon$. Now $w_5 u = w_5 \in L^{\geq 2a}_{\text{even}}$, whereas $w_j u \notin L^{\geq 2a}_{\text{even}}$ (when $w_j \in \{w_1, w_2, w_3, w_4, w_6\}$). Thus, M cannot recognize $L^{\geq 2a}_{\text{even}}$ giving us the desired contradiction.
- Case $w_i = w_1$, and $w_j \in \{w_2, w_4, w_6\}$. Once again, if $\hat{\delta}_M(q_0, w_1) = \hat{\delta}_M(q_0, w_j)$ then for every u, M either accepts both $w_1 u$ and $w_j u$ or neither one. If we take $u = aa$, then $w_1 u = aa \in L^{\geq 2a}_{\text{even}}$, but $w_j u \notin L^{\geq 2a}_{\text{even}}$ when $w_j \in \{w_2, w_4, w_6\}$, giving us the desired contradiction.

- Case $w_i = w_3$ and $w_j \in \{w_2, w_4, w_6\}$. Similar to the previous case we can take $u = aa$.
- Case $w_i = w_1$ and $w_j = w_3$. Take $u = ab$.
- Case $w_i = w_2$ and $w_j = w_4$. Take $u = a$.
- Case $w_i = w_2$ and $w_j = w_6$. Take $u = b$.
- Case $w_i = w_4$ and $w_j = w_6$. Take $u = b$.

$\square$

A One k-positions from end

Problem
Design an automaton for the language $L_k = \{w \mid k\text{th character from end of } w \text{ is } 1\}$

Solution
What do you need to remember? The last k characters seen so far!

Formally, $M_k = (Q, \{0,1\}, \delta, q_0, F)$

- States $= Q = \{\langle w \rangle \mid w \in \{0,1\}^k\}$
- $\delta(\langle w \rangle, b) = \langle w_2 w_3 \ldots w_k b \rangle$ where $w = w_1 w_2 \ldots w_k$
- $q_0 = \langle 0^k \rangle$
- $F = \{\langle 1 w_2 w_3 \ldots w_k \rangle \mid w_i \in \{0,1\}\}$

Lower Bound on DFA size

Proposition 6. *Any DFA recognizing L_k has at least 2^k states.*

Proof. Let M, with initial state q_0, recognize L_k and assume (for contradiction) that M has $< 2^k$ states.

- Number of strings of length $k = 2^k$
- There must be two distinct string w_0 and w_1 of length k such that for some state q, $q_0 \xrightarrow{w_0}_M q$ and $q_0 \xrightarrow{w_1}_M q$.

Let i be the first position where w_0 and w_1 differ. Without loss of generality assume that w_0 has 0 in the ith position and w_1 has 1.

$$\begin{aligned} w_0 0^{i-1} &= \ \ldots\ \overbrace{0\ \ldots\ 0^{i-1}}^{k} \\ w_1 0^{i-1} &= \underbrace{\ldots}_{i-1}\ 1 \underbrace{\ldots}_{k-i} 0^{i-1} \end{aligned}$$

$w_0 0^{i-1} \notin L_k$ and $w_1 0^{i-1} \in L_k$. Thus, M cannot accept both $w_0 0^{i-1}$ and $w_1 0^{i-1}$.

So far, $w_0 0^{i-1} \notin L_n$, $w_1 0^{i-1} \in L_n$, $q_0 \xrightarrow{w_0}_M q$, and $q_0 \xrightarrow{w_1}_M q$.

$$\begin{aligned} q_0 \xrightarrow{w_0 0^{i-1}}_M q_1 \quad &\text{iff} \quad q \xrightarrow{0^{i-1}}_M q_1 \\ &\text{iff} \quad q_0 \xrightarrow{w_1 0^{i-1}}_M q_1 \end{aligned}$$

Thus, M accepts or rejects both $w_0 0^{i-1}$ and $w_1 0^{i-1}$. Contradiction! □
