Recall: Wanted a canonical DFA recognizing any regular L.

Defined an equivalence $\sim_L$ on strings from Σ^*.

$x \sim_L y$ iff for all $z \in \Sigma^*$, $xz \in L$ iff $yz \in L$.

Myhill-Nerode theorem shows that

L is regular iff $\sim_L$ induces finitely many eq. classes, and

Any DFA recognizing L must have at least as many states as the index of $\sim_L$.

Today: An algorithm to minimize a given DFA.

For any DFA $M = (Q, \Sigma, \delta, q_0, F)$ s.t. $\mathcal{L}(M) = \mathcal{L}$,
we set up a surjective function from Q to Q_L.
* This shows that $|Q| \geq |Q_L|$, and since $|Q_L|$ is the index of $\sim_L$,
M_L is a DFA with the fewest states recognizing $\mathcal{L}$.

Assume wlog that $M = (Q, \Sigma, \delta, q_0, F)$ is a DFA
where every state is reachable from q_0, and $\mathcal{L}(M) = \mathcal{L}$.
Define $f : Q \to Q_L$ as follows: $f(q) = [x]$ iff $\hat{\delta}(q_0, x) = q$, $x \in \Sigma^*$
Show that f is well-defined, and surjective
if $\hat{\delta}(q_0, x) = \hat{\delta}(q_0, y)$, $x \sim_L y$
every $[x]$ is the image of some y under f

This also gives us a criterion for the minimality of a DFA.

If $f: Q \to Q_L$ is the above kind of map from the states of an arbitrary DFA to those of the Myhill-Nerode DFA, if we show that f is actually a bijection, the chosen DFA is minimal.

We would like a minimization procedure as well.
Given an arbitrary M, how does one get M_L from it?
Cannot try all strings on M to determine $[x]$ for all $x \in \Sigma^*$.
Instead, we want to capture the effect of $\sim_L$ via some relation on the states of M.

Let $M = (Q, \Sigma, \delta, q_0, F)$ s.t. $\mathcal{L}(M) = \mathcal{L}$.

Let $x, y \in \Sigma^*$ and $\hat{\delta}(q_0, x) = p \in Q$, and $\hat{\delta}(q_0, y) = q \in Q$.

Then, we want a relation $\sim$ s.t. if $x \sim_L y$, then $p \sim q$.

What might such a $\sim$ look like?

What does $x \sim_L y$ mean?

So, $p \sim q$ iff $\mathcal{L}(p) = \mathcal{L}(q)$, where

$$\mathcal{L}(s) = \{ z \in \Sigma^* \mid \hat{\delta}(s, z) \in F \}, \text{ for } s \in Q.$$

Lemma: Suppose $\mathcal{L}$ is regular and $M = (Q, \Sigma, \delta, q_0, F)$ is a DFA accepting $\mathcal{L}$. Let $x, y \in \Sigma^*$, $p = \hat{\delta}(q_0, x)$, $q = \hat{\delta}(q_0, y)$. Then, $x \sim_L y \Rightarrow \mathcal{L}(p) = \mathcal{L}(q)$. Prove this!

Now, recall our surjective function $f : Q \to Q_L$, defined as
$$f(q) = [x] \text{ iff } \hat{\delta}(q_0, x) = q$$

Suppose M is not a minimal DFA for $\mathcal{L}$. Then, f is not bijective. What does this imply about $\sim$ and $\mathcal{L}(\cdot)$?

If f is not bijective, it is not injective. So, there exist $p, q \in Q$ s.t. $f(p) = f(q)$, but $p \neq q$.

Let $x, y \in \Sigma^*$ be such that $\hat{\delta}(q_0, x) = p$ and $\hat{\delta}(q_0, y) = q$.

Then, if $f(p) = f(q)$, $[x] = [y]$ (by the definition of f)

But then, by the earlier lemma, $\alpha(p) = \alpha(q)$.

So, if M is not minimal, there are two distinct states $p, q \in Q$ s.t. $\alpha(p) = \alpha(q)$.

This gives us a nice criterion on states to show minimality.

If $\alpha(p) \neq \alpha(q)$ for any distinct $p, q \in Q$, then M is minimal.

Given a DFA $M = (Q, \Sigma, \delta, q_0, F)$, how do we compute $\sim$?

Define a sequence $\sim_0, \sim_1, \sim_2, \ldots$ on Q as follows:

- $p \sim_0 q$ iff ($p \in F$ iff $q \in F$)
- For $i \geq 0$, $p \sim_{i+1} q$ iff $p \sim_i q$ and

  $\forall a \in \Sigma : \delta(p, a) \sim_i \delta(q, a)$.

Note that $\sim_{i+1} \subseteq \sim_i$.

What is the maximum index of an equivalence relation on Q?

So the algorithm sets $\sim$ to be the first $\sim_k$ in the sequence s.t. for all $l \geq k$, $\sim_k = \sim_l$.

The following lemmas will help us prove the correctness of this construction.

Lemma: For $i \geq 0$, $p \sim_i q$ iff $\alpha_i(p) = \alpha_i(q)$, where, for $s \in Q$,
$$\alpha_i(s) = \{x \in \Sigma^* \mid |x| \leq i, \text{ and } \hat{\delta}(s, x) \in F\}$$

Lemma: Suppose $\sim_k = \sim_{k+1}$ for some k. Then,
$$p \sim_k q \text{ iff } \alpha(p) = \alpha(q).$$

Try to prove these statements!

https://www.youtube.com/watch?v=0XaGAkY09Wc

Brzozowski's algorithm

- Start with a DFA $M = (Q, \Sigma, \delta, q_0, F)$
- Construct the NFA $M' = (Q, \Sigma, \Delta, Q_0, F')$ as follows

  $Q_0 = \{q \mid q \in F\}$ $\qquad F' = \{q_0\}$

  $(q, a, q') \in \Delta$ if $\delta(q', a) = q$

  What is $\mathcal{L}(M')$?
- Determinize M' to get a DFA $M_2 = (Q_2, \Sigma, \delta_2, q_0^2, F_2)$
- Construct the NFA $M_2' = (Q_2, \Sigma, \Delta_2, Q_0^2, F_2')$ as follows

  $Q_0^2 = \{q \mid q \in F_2\}$ $\qquad F_2' = \{q_0^2\}$

  $(q, a, q') \in \Delta_2$ if $\delta_2(q', a) = q$.
- Determinize M_2' to get M_{fin}, the minimal DFA recognizing $\mathcal{L}(M)$!