Proof of Correctness for NFA to DFA Construction

This proof is taken from *Elements of the Theory of Computation* by Lewis and Papadimitriou, second edition. Any errors are probably due to my faulty transcription.

We are given an NFA $N = (Q, \Sigma, \delta, q_0, F)$. The construction discussed in class produces a DFA $D = (Q', \Sigma, \delta', q_0', F')$ where $Q' = 2^Q$, $q_0' = E(q_0)$ and $F = \{q' \in Q' | q' \cap F \neq \emptyset\}$ (remember $E(q)$ is the ϵ-closure of q). We need to show that $\mathcal{L}(D) = \mathcal{L}(N)$; we will do this by proving that for any string $w \in \Sigma^*$, $w \in \mathcal{L}(D)$ if and only if $w \in \mathcal{L}(N)$. In other words, if we begin at q_0' and trace a path through D on the string w, we will be in a final state if and only if there is a path in N starting at q_0 that ends in a final state of N. We will in fact prove something stronger:

Lemma 1 *If we start in* any *state q of N and trace a path on w, we will end at state p if and only if there is a path in D on w from a state containing $E(q)$ to state P, where P contains p.*

Proof: By induction on $|w|$.

Base Case. Take $|w| = 0$, that is, $w = \epsilon$. We need to show that tracing through N on ϵ, starting from q, leads to p if and only if tracing through D on ϵ, starting from a state containing $E(q)$, leads to P with $p \in P$. As D is deterministic, it can't go anywhere on ϵ, that is, $P = E(q)$. What could p be? Any state reachable in N from q on ϵ, that is, $p \in E(q)$, which equals P. This completes the proof of the base case.

Inductive Hypothesis. Assume the Lemma is true for all strings of length k or less for some $k \geq 0$.

Inductive Step. Consider a string w with $|w| = k + 1$. Let $w = va$, where $v \in \Sigma^*$ and $a \in \Sigma$. Note that $|v| = k$, so the Inductive Hypothesis applies to it.

For the *only if* direction of the proof, we have a path from q to p on w in N. What could this path look like? First, we follow a path on v, and possibly do some ϵ-transitions, to end at state r_1. Then we read a and get to state r_2, then we take more ϵ-transitions and get to p. See Figure 1. Now consider a path in D. By the induction hypothesis, after reading v we are in a state R that contains r_1. By our construction, after reading a we are in a state containing $E(r_2)$. But by the definition of ϵ-closure, $p \in E(a)$.

$$N: \qquad q \rightsquigarrow_v r_1 \rightarrow_a r_2 \rightarrow_\epsilon p$$

$$\Updownarrow$$

$$D: \quad E(q) \rightsquigarrow_v \underbrace{\{\ldots, r_1, \ldots\}}_{R} \rightarrow_a \underbrace{\{\ldots, E(r_2), \ldots\}}_{P} \supset \{p\}$$

Figure 1: Paths through N and D on $w = va$.

For the *if* direction of the proof, we have a path in D from a state containing $E(q)$ to P. As above, this can be broken down into a path from a state containing $E(q)$ to R on v, then an arc from R to P on a. By the induction hypothesis, there is a path in N from q to some $r_1 \in R$ on v. When we take the transition from r_1 to some other state r_2 on a, and then any ϵ-transitions, by our construction we must be in a state in P.

$\square$