Example 2.13: Consider the NFA N of Fig 2 15 $L(N)$ is the set of all strings of 0 s and 1 s such that the nth symbol from the end is 1 IntuitivelyΣa DFA D that accepts this language must remember the last n symbols it has read Since any of 2 subsets of the last n symbols could have been 1Σif D has fewer

than 2 statesΣthen there would be some state q such that D can be in state q after reading two different sequences of n bitsΣsay $a_8 a_($ a and $b_8 b_($ b

Since the sequences are differentΣthey must differ in some positionΣsay $a = b$ Suppose (by symmetry) that $a = 1$ and $b = 0$ If $i = 1$Σthen q must be both an accepting state and a nonaccepting stateΣsince $a_8 a_($ a is accepted (the nth symbol from the end is 1) and $b_8 b_($ b is not If i 1Σ then consider the state p that D enters after reading i 1 0 s Then p must be both accepting and nonacceptingΣsince $a\ a_{!\ 8}$ a 00 0 is accepted and $b\ b_{!\ 8}$ b 00 0 is not

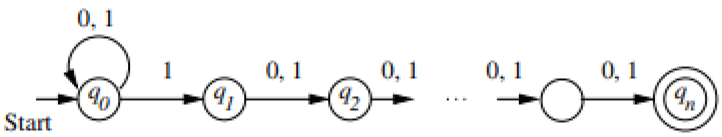

Figure 2 15: This NFA has no equivalent DFA with fewer than 2 states

NowΣlet us see how the NFA N of Fig 2 15 works There is a state q_0 that the NFA is always inΣregardless of what inputs have been read If the next input is 1ΣN may also ˆguess" that this 1 will be the nth symbol from the endΣ so it goes to state q_8 as well as q_0 From state q_8Σany input takes N to $q_($Σ the next input takes it to $q_)$Σand so onΣuntil n 1 inputs laterΣit is in the accepting state q The formal statement of what the states of N do is:

1 N is in state q_0 after reading any sequence of inputs w

2 N is in state q Σfor $i = 1$ c2cδδδcnΣafter reading input sequence w if and only if the ith symbol from the end of w is 1; that isΣw is of the form $x1a_8a_($ $a_{_8}$Σwhere the a s are each input symbols

We shall not prove these statements formally; the proof is an easy induction on w Σmimicking Example 2 9 To complete the proof that the automaton accepts exactly those strings with a 1 in the nth position from the endΣwe consider statement (2) with $i = n$ That says N is in state q if and only if the nth symbol from the end is 1 But q is the only accepting stateΣso that condition also characterizes exactly the set of strings accepted by N □

Pigeonhole principle

The Pigeonhole Principle

In Example 2 13 we used an important reasoning technique called the *pigeonhole principle* ColloquiallyΣif you have more pigeons than pigeon-holesΣand each pigeon flies into some pigeonholeΣthen there must be at least one hole that has more than one pigeon In our exampleΣthe ˆpi-geons" are the sequences of n bitsΣand the ˆpigeonholes" are the states Since there are fewer states than sequencesΣone state must be assigned two sequences

The pigeonhole principle may appear obviousΣbut it actually depends on the number of pigeonholes being finite ThusΣit works for finite-state automataΣwith the states as pigeonholesΣbut does not apply to other kinds of automata that have an infinite number of states

To see why the finiteness of the number of pigeonholes is essentialΣ consider the infinite situation where the pigeonholes correspond to integers 1α2αδδδ Number the pigeons 0α1α2αδδδΣso there is one more pigeon than there are pigeonholes HoweverΣwe can send pigeon i to hole $i+1$ for all i 0 Then each of the infinite number of pigeons gets a pigeonholeΣand no two pigeons have to share a pigeonhole

Example 2.10: Let N be the automaton of Fig 2 9 that accepts all strings that end in 01 Since N s set of states is $q_0 α q_8 α q_($ Σthe subset construction produces a DFA with $2^{)} = 8$ statesΣcorresponding to all the subsets of these three states Figure 2 12 shows the transition table for these eight states; we shall show shortly the details of how some of these entries are computed

Notice that this transition table belongs to a deterministic finite automaton Even though the entries in the table are setsΣthe states of the constructed DFA *are* sets To make the point clearerΣwe can invent new names for these statesΣ e g ΣA for ΣB for q_0 Σand so on The DFA transition table of Fig 2 13 defines exactly the same automaton as Fig 2 12Σbut makes clear the point that the entries in the table are single states of the DFA

Of the eight states in Fig 2 13Σstarting in the start state BΣwe can only reach states BΣEΣand F The other five states are inaccessible from the start state and may as well not be there We often can avoid the exponential-time step of constructing transition-table entries for every subset of states if we perform ˆlazy evaluation" on the subsetsΣas follows

BASIS: We know for certain that the singleton set consisting only of N s start state is accessible

	0	1
A	A	A
B	E	B
C	A	D
D	A	A
E	E	F
F	E	B
G	A	D
H	E	F

Figure 2 13: Renaming the states of Fig 2 12

INDUCTION: Suppose we have determined that set S of states is accessible Then for each input symbol aΣcompute the set of states $(S\alpha a)$; we know that these sets of states will also be accessible

For the example at handΣwe know that q_0 is a state of the DFA D We find that $(q_0 \alpha 0) = q_0 \alpha q_8$ and $(q_0 \alpha 1) = q_0$ Both these facts are established by looking at the transition diagram of Fig 2 9 and observing that on 0 there are arcs out of q_0 to both q_0 and q_8Σwhile on 1 there is an arc only to q_0 We thus have one row of the transition table for the DFA: the second row in Fig 2 12

One of the two sets we computed is "old"; q_0 has already been considered HoweverΣthe other — $q_0 \alpha q_8$ — is new and its transitions must be computed We find $(q_0 \alpha q_8 \alpha 0) = q_0 \alpha q_8$ and $(q_0 \alpha q_8 \alpha 1) = q_0 \alpha q_($ For instanceΣ to see the latter calculationΣwe know that

$$(q_0 \alpha q_8 \alpha 1) = {}_N(q_0 \alpha 1) \quad {}_N(q_8 \alpha 1) = q_0 \quad q_(= q_0 \alpha q_($$

We now have the fifth row of Fig 2 12Σand we have discovered one new state of DΣwhich is $q_0 \alpha q_($ A similar calculation tells us

$$(q_0 \alpha q_(\alpha 0) = {}_N(q_0 \alpha 0) \quad {}_N(q_(\alpha 0) = q_0 \alpha q_8 \quad = q_0 \alpha q_8$$
$$(q_0 \alpha q_(\alpha 1) = {}_N(q_0 \alpha 1) \quad {}_N(q_(\alpha 1) = q_0 \quad = q_0$$

These calculations give us the sixth row of Fig 2 12Σbut it gives us only sets of states that we have already seen

ThusΣthe subset construction has converged; we know all the accessible states and their transitions The entire DFA is shown in Fig 2 14 Notice that it has only three statesΣwhich isΣby coincidenceΣexactly the same number of states as the NFA of Fig 2 9Σfrom which it was constructed HoweverΣthe DFA of Fig 2 14 has six transitionsΣcompared with the four transitions in Fig 2 9 □

Theorem 2.11: If $D = (Q \ \alpha \ \alpha \ \ \alpha \, q_0 \ \alpha F \ )$ is the DFA constructed from NFA $N = (Q_N \alpha \ \ \alpha \ _N \alpha q_0 \alpha F_N)$ by the subset constructionΣthen $L(D) = L(N)$

PROOF: What we actually prove firstΣby induction on w Σis that

$$(\ q_0 \ \ \alpha w) = \ _N(q_0 \alpha w)$$

Notice that each of the functions returns a set of states from Q_N Σbut interprets this set as one of the states of Q (which is the power set of Q_N)Σ while $_N$ interprets this set as a subset of Q_N

BASIS: Let $w = 0$; that isΣ$w =$ By the basis definitions of for DFA s and NFA sΣboth $(\ q_0 \ \ \alpha \ )$ and $_N(q_0 \alpha \ )$ are q_0

INDUCTION: Let w be of length $n + 1$Σand assume the statement for length n Break w up as $w = xa$Σwhere a is the final symbol of w By the inductive hypothesisΣ $(\ q_0 \ \ \alpha x) = \ _N(q_0 \alpha x)$ Let both these sets of N s states be $p_8 \alpha p_(\alpha \delta \delta \delta \alpha p$

The inductive part of the definition of for NFA s tells us

$$_N(q_0 \alpha w) = \sum_{9 \ 8} \ _N(p \ \alpha a) \tag{2 2}$$

The subset constructionΣon the other handΣtells us that

$$(\ p_8 \alpha p_(\alpha \delta \delta \delta \alpha p \ \ \ \alpha a) = \sum_{9 \ 8} \ _N(p \ \alpha a) \tag{2 3}$$

NowΣlet us use (2 3) and the fact that $(\ q_0 \ \ \alpha x) = p_8 \alpha p_(\alpha \delta \delta \delta \alpha p$ in the inductive part of the definition of for DFA s:

$$(\ q_0 \ \ \alpha w) = \ \ (\ q_0 \ \ \alpha x) \alpha a \ = \ (\ p_8 \alpha p_(\alpha \delta \delta \delta \alpha p \ \ \ \alpha a) = \sum_{9 \ 8} \ _N(p \ \alpha a) \tag{2 4}$$

ThusΣEquations (2 2) and (2 4) demonstrate that $(\ q_0 \ \ \alpha w) = \ _N(q_0 \alpha w)$ When we observe that D and N both accept w if and only if $(\ q_0 \ \ \alpha w)$ or $_N(q_0 \alpha w)$ΣrespectivelyΣcontain a state in F_N Σwe have a complete proof that $L(D) = L(N)$ □

Exercise:

* **Exercise 2.3.1**: Convert to a DFA the following NFA:

	0	1
p	$p \alpha q$	p
q	r	r
r	s	
s	s	s

Exercise 2.3.2: Convert to a DFA the following NFA:

	0	1
p	$q \alpha s$	q
q	r	$q \alpha r$
r	s	p
s		p

Example 2.20: Let us compute $(q_0 \alpha\ 5\ )$ for the -NFA of Fig 2 18 A summary of the steps needed are as follows:

- $(q_0 \alpha\ ) = \text{ECLOSE}(q_0) = q_0 \alpha q_8$
- Compute $(q_0 \alpha\ )$ as follows:
  1. First compute the transitions on input 5 from the states q_0 and q_8 that we obtained in the calculation of $(q_0 \alpha\ )$Σabove That isΣwe compute $(q_0 \alpha\ )$ $(q_8 \alpha\ ) = q_8 \alpha q_6$
  2. NextΣ -close the members of the set computed in step (1) We get $\text{ECLOSE}(q_8)$ $\text{ECLOSE}(q_6) = q_8$ $q_6 = q_8 \alpha q_6$ That set is $(q_0 \alpha\ )$ This two-step pattern repeats for the next two symbols
- Compute $(q_0 \alpha\ 5)$ as follows:
  1. First compute $(q_8 \alpha\ 5)$ $(q_6 \alpha\ 5) = q_($ $q_)$ $= q_(\alpha q_)$
  2. Then compute

     $(q_0 \alpha\ 5) = \text{ECLOSE}(q_()$ $\text{ECLOSE}(q_)) = q_($ $q_) \alpha q_7 = q_(\alpha q_) \alpha q_7$
- Compute $(q_0 \alpha\ 5\ )$ as follows:
  1. First compute $(q_(\alpha\ )$ $(q_) \alpha\ )$ $(q_7 \alpha\ ) = q_)$ $q_) = q_)$
  2. Then compute $(q_0 \alpha\ 5\ ) = \text{ECLOSE}(q_)) = q_) \alpha q_7$

□

Example 2.21: Let us eliminate -transitions from the -NFA of Fig 2 18Σ which we shall call E in what follows From EΣwe construct an DFA DΣwhich is shown in Fig 2 22 HoweverΣto avoid clutterΣwe omitted from Fig 2 22 the dead state and all transitions to the dead state You should imagine that for each state shown in Fig 2 22 there are additional transitions from any state to on any input symbols for which a transition is not indicated AlsoΣthe state has transitions to itself on all input symbols

Since the start state of E is q_0Σthe start state of D is $\text{ECLOSE}(q_0)$Σwhich is $q_0 \alpha q_8$ Our first job is to find the successors of q_0 and q_8 on the various symbols in ; note that these symbols are the plus and minus signsΣthe dotΣ and the digits 0 through 9 On + and Σq_8 goes nowhere in Fig 2 18Σwhile q_0 goes to q_8 ThusΣto compute $(q_0 \alpha q_8\ \alpha +)$ we start with q_8 and -close it Since there are no -transitions out of q_8Σwe have $(q_0 \alpha q_8\ \alpha +) = q_8$ SimilarlyΣ $(q_0 \alpha q_8\ \alpha\ ) = q_8$ These two transitions are shown by one arc in Fig 2 22

Theorem 2.22: A language L is accepted by some -NFA if and only if L is accepted by some DFA

PROOF: (If) This direction is easy Suppose $L = L(D)$ for some DFA Turn D into an -NFA E by adding transitions $(q\alpha\ ) =$ for all states q of D TechnicallyΣwe must also convert the transitions of D on input symbolsΣe g Σ $(q\alpha a) = p$ into an NFA-transition to the set containing only pΣ that is $(q\alpha a) = p$ ThusΣthe transitions of E and D are the sameΣbut E explicitly states that there are no transitions out of any state on

(Only-if) Let $E = (Q\ \alpha\ \alpha\ \ \alpha q_0\alpha F\ )$ be an -NFA Apply the modified subset construction described above to produce the DFA

$$D = (Q\ \ \alpha\ \ \alpha\ \ \ \alpha q\ \ \alpha F\ \ )$$

We need to show that $L(D) = L(E)$Σand we do so by showing that the extended transition functions of E and D are the same FormallyΣwe show $(q_0\alpha w) = (q\ \ \alpha w)$ by induction on the length of w

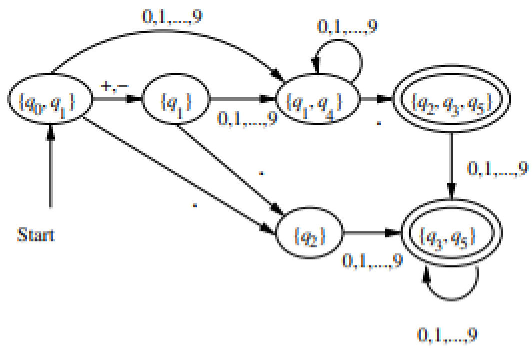

Figure 2 22: The DFA D that eliminates -transitions from Fig 2 18

NextΣwe need to compute $(\ q_0\alpha q_8\ \ \alpha\ 5)$ Since q_0 goes nowhere on the dotΣand q_8 goes to $q_($ in Fig 2 18Σwe must -close $q_($ As there are no -transitions out of $q_($Σthis state is its own closureΣso $(\ q_0\alpha q_8\ \ \alpha\ 5) = \ q_($

FinallyΣwe must compute $(\ q_0\alpha q_8\ \ \alpha 0)$Σas an example of the transitions from $q_0\alpha q_8$ on all the digits We find that q_0 goes nowhere on the digitsΣbut q_8 goes to both q_8 and q_6 Since neither of those states have -transitions outΣ we conclude $(\ q_0\alpha q_8\ \ \alpha 0) = \ q_8\alpha q_6$ Σand likewise for the other digits

We have now explained the arcs out of $q_0\alpha q_8$ in Fig 2 22 The other transitions are computed similarlyΣand we leave them for you to check Since q_7 is the only accepting state of EΣthe accepting states of D are those accessible states that contain q_7 We see these two sets $q_)\alpha q_7$ and $q_(\alpha q_)\alpha q_7$ indicated by double circles in Fig 2 22 □

BASIS: If $|w| = 0$Σthen $w =$ We know $(q_0 α) =$ ECLOSE(q_0) We also know that $q =$ ECLOSE(q_0)Σbecause that is how the start state of D is defined Finally Σfor a DFA Σwe know that $(pα) = p$ for any state pΣso in particular Σ $(q\ α) =$ ECLOSE(q_0) We have thus proved that $(q_0 α) = (q\ α)$

INDUCTION: Suppose $w = xa$Σwhere a is the final symbol of wΣand assume that the statement holds for x That isΣ $(q_0 αx) = (q\ αx)$ Let both these sets of states be $p_8 αp_(αδδδαp$

By the definition of for -NFA sΣwe compute $(q_0 αw)$ by:

1 Let $r_8 αr_(αδδδαr$ be $\bigcup_{9\ 8}$ $(p\ αa)$

2 Then $(q_0 αw) =$ ECLOSE$(\ r_8 αr_(αδδδαr\ )$

If we examine the construction of DFA D in the modified subset construction above Σwe see that $(\ p_8 αp_(αδδδαp\ \ αa)$ is constructed by the same two steps (1) and (2) above ThusΣ $(q\ \ αw)$Σwhich is $(\ p_8 αp_(αδδδαp\ \ αa)$ is the same set as $(q_0 αw)$ We have now proved that $(q_0 αw) = (q\ \ αw)$ and completed the inductive part □