The algorithm of Fig. 3.8 is more efficient than the obvious marking algorithm, although it is not the most efficient possible. Let Σ have k symbols and Q have n states. Line 1 takes $0(n^2)$ steps.† The loop of lines 2 through 7 is executed $0(n^2)$ times, at most once for each pair of states. The total time spent on lines 2 through 4, 6, and 7 is $0(kn^2)$. The time spent on line 5 is the sum of the length of all lists. But each pair (r, s) is put on at most k lists, at line 7. Thus the time spent on line 5 is $0(kn^2)$, so the total time is also $0(kn^2)$.

Theorem 3.11 The DFA constructed by the algorithm of Fig. 3.8, with inaccessible states removed, is the minimum state DFA for its language.

Proof Let $M = (Q, \Sigma, \delta, q_0, F)$ be the DFA to which the algorithm is applied and $M' = (Q', \Sigma, \delta', [q_0], F')$ be the DFA constructed. That is,

$$Q' = \{[q] \mid q \text{ is accessible from } q_0\},$$

$$F' = \{[q] \mid q \text{ is in } F\}$$

and

$$\delta'([q], a) = [\delta(q, a)].$$

It is easy to show that δ' is consistently defined, since if $q \equiv p$, then $\delta(q, a) \equiv \delta(p, a)$. That is, if $\delta(q, a)$ is distinguished from $\delta(p, a)$ by x, then ax distinguishes q from p. It is also easy to show that $\delta'([q_0], w) = [\delta(q_0, w)]$ by induction on $|w|$. Thus $L(M') = L(M)$.

Now we must show that M' has no more states than R_L has equivalence classes, where $L = L(M)$. Suppose it did; then there are two accessible states q and p in Q such that $[q] \neq [p]$, yet there are x and y such that $\delta(q_0, x) = q$, $\delta(q_0, y) = p$, and $xR_L y$. We claim that $p \equiv q$, for if not, then some w in Σ^* distinguishes p from q. But then $xwR_L yw$ is false, for we may let $z = \epsilon$ and observe that exactly one of xwz and ywz is in L. But since R_L is right invariant, $xwR_L yw$ is true. Hence q and p do not exist, and M' has no more states than the index of R_L. Thus M' is the minimum state DFA for L. □

EXERCISES

3.1 Which of the following languages are regular sets? Prove your answer.

a) $\{0^{2n} \mid n \geq 1\}$
b) $\{0^m 1^n 0^{m+n} \mid m \geq 1 \text{ and } n \geq 1\}$
c) $\{0^n \mid n \text{ is a prime}\}$
d) the set of all strings that do not have three consecutive 0's.
e) the set of all strings with an equal number of 0's and 1's.
f) $\{x \mid x \text{ in } (\mathbf{0} + \mathbf{1})^*, \text{ and } x = x^R\}$ x^R is x written backward; for example, $(011)^R = 110$.
g) $\{xwx^R \mid x, w \text{ in } (\mathbf{0} + \mathbf{1})^+\}$
*h) $\{xx^Rw \mid x, w \text{ in } (\mathbf{0} + \mathbf{1})^+\}$

† We say that $g(n)$ is $0(f(n))$ if there exist constants c and n_0 such that $g(n) \leq cf(n)$ for all $n \geq n_0$.

3.2 Prove the following extension of the pumping lemma for regular sets. Let L be a regular set. Then there exists a constant n such that for each z_1, z_2, z_3, with $z_1 z_2 z_3$ in L and $|z_2| = n$, z_2 can be written $z_2 = uvw$ such that $|v| \geq 1$ and for each $i \geq 0$, $z_1 uv^i w z_3$ is in L.

3.3 Use Exercise 3.2 to prove that $\{0^i 1^m 2^m \mid i \geq 1,\ m \geq 1\}$ is nonregular.

* **3.4** Let L be a regular set. Which of the following sets are regular? Justify your answers.

a) $\{a_1 a_3 a_5 \cdots a_{2n-1} \mid a_1 a_2 a_3 a_4 \cdots a_{2n} \text{ is in } L\}$
S b) $\{a_2 a_1 a_4 a_3 \cdots a_{2n} a_{2n-1} \mid a_1 a_2 \cdots a_{2n} \text{ is in } L\}$
c) $\mathrm{CYCLE}(L) = \{x_1 x_2 \mid x_2 x_1 \text{ is in } L \text{ for strings } x_1 \text{ and } x_2\}$
d) $\mathrm{MAX}(L) = \{x \text{ in } L \mid \text{for no } y \text{ other than } \epsilon \text{ is } xy \text{ in } L\}$
e) $\mathrm{MIN}(L) = \{x \text{ in } L \mid \text{no proper prefix of } x \text{ is in } L\}$
f) $\mathrm{INIT}(L) = \{x \mid \text{for some } y,\ xy \text{ is in } L\}$
g) $L^R = \{x \mid x^R \text{ is in } L\}$
h) $\{x \mid xx^R \text{ is in } L\}$

* **3.5** Let value(x) be the result when the symbols of x are multiplied from left to right according to the table of Fig. 2.31.

a) Is $L = \{xy \mid |x| = |y| \text{ and value}(x) = \text{value}(y)\}$ regular?
b) Is $L = \{xy \mid \text{value}(x) = \text{value}(y)\}$ regular?

Justify your answers.

* **3.6** Show that $\{0^i 1^j \mid gcd(i, j) = 1\}$ is not regular.

** **3.7** Let L be any subset of 0^*. Prove that L^* is regular.

3.8 A set of integers is *linear* if it is of the form $\{c + pi \mid i = 0, 1, 2, \ldots\}$. A set is *semilinear* if it is the finite union of linear sets. Let $R \subseteq 0^*$ be regular. Prove that $\{i \mid 0^i \text{ is in } R\}$ is semilinear.

3.9 Is the class of regular sets closed under infinite union?

3.10 What is the relationship between the class of regular sets and the least class of languages closed under union, intersection, and complement containing all finite sets?

* **3.11** Give a finite automaton construction to prove that the class of regular sets is closed under substitution.

** **3.12** Is the class of regular sets closed under inverse substitution?

3.13 Let h be the homomorphism $h(a) = 01$, $h(b) = 0$.

a) Find $h^{-1}(L_1)$, where $L_1 = (\mathbf{10} + \mathbf{1})^*$
b) Find $h(L_2)$, where $L_2 = (\mathbf{a} + \mathbf{b})^*$
c) Find $h^{-1}(L_3)$, where L_3 is the set of all strings of 0's and 1's with an equal number of 0's and 1's.

3.14 Show that 2DFA with endmarkers (see Exercise 2.20) accept only regular sets by making use of closure properties developed in this chapter.

** **3.15** The use of $\cap$ with regular expressions does not allow representation of new sets. However it does allow more compact expression. Show that $\cap$ can shorten a regular expression by an exponential amount. [*Hint:* What is the regular expression of shortest length describing the set consisting of the one sentence $(\ldots((a_0^2 a_1)^2 a_2)^2 \cdots)^2$?]

** **3.16** Let L be a language. Define $\frac{1}{2}(L)$ to be

$$\{x \mid \text{for some } y \text{ such that } |x| = |y|,\ xy \text{ is in } L\}.$$

That is, $\frac{1}{2}(L)$ is the first halves of strings in L. Prove for each regular L that $\frac{1}{2}(L)$ is regular.

** **3.17** If L is regular, is the set of first thirds of strings in L regular? What about the last third? Middle third? Is the set

$$\{xz \mid \text{for some } y \text{ with } |x| = |y| = |z|,\ xyz \text{ is in } L\}$$

regular?

** **3.18** Show that if L is regular, so are

a) $\mathrm{SQRT}(L) = \{x \mid \text{for some } y \text{ with } |y| = |x|^2,\ xy \text{ is in } L\}$
b) $\mathrm{LOG}(L) = \{x \mid \text{for some } y \text{ with } |y| = 2^{|x|},\ xy \text{ is in } L\}$

* **3.19** A *one-pebble* 2DFA is a 2DFA with the added capability of marking a tape square by placing a pebble on it. The next state function depends on the present state, the tape symbol scanned, and the presence or absence of a pebble on the tape square scanned. A move consists of a change of state, a direction of head motion, and possibly placing or removing the pebble from the scanned tape cell. The automaton "jams" if it attempts to place a second pebble on the input. Prove that one-pebble 2DFA's accept only regular sets. [*Hint:* Add two additional tracks to the input that contain tables indicating for each state p, the state q in which the 2DFA will return if it moves left or right from the tape cell in state p, under the assumption that the pebble is not encountered. Observe that the one-pebble 2DFA operating on the augmented tape need never leave its pebble. Then make use of a homomorphic mapping to remove the additional tracks.]

* **3.20** In converting an NFA to a DFA the number of states may increase substantially. Give upper and lower bounds on the maximum increase in number of states for an n-state NFA. [*Hint:* Consider Exercises 2.5(e) and 2.8(c).]

3.21 Give a decision procedure to determine if the set accepted by a DFA is

a) the set of all strings over a given alphabet,
b) *cofinite* (a set whose complement is finite).

** **3.22** Consider a DFA M. Suppose you are told that M has at most n states and you wish to determine the transition diagram of M. Suppose further that the only way you can obtain information concerning M is by supplying an input sequence x and observing the prefixes of x which are accepted.

a) What assumptions must you make concerning the transition diagram of M in order to be able to determine the transition diagram?
b) Give an algorithm for determining the transition diagram of M (except for the start state) including the construction of x under your assumptions in part (a).

S **3.23 Give an efficient decision procedure to determine if x is in the language denoted by an *extended regular expression* (a regular expression with operators $\cup$, $\cdot$ (concatenation), *, $\cap$, and $\neg$, that is complement).

3.24 Give an efficient decision procedure for determining if a *semi-extended regular expression* r (a regular expression with $\cup$, $\cdot$, *, $\cap$) denotes a nonempty set. [*Hint:* Space $0(|r|)$ and time $0(2^{|r|})$ are sufficient.]

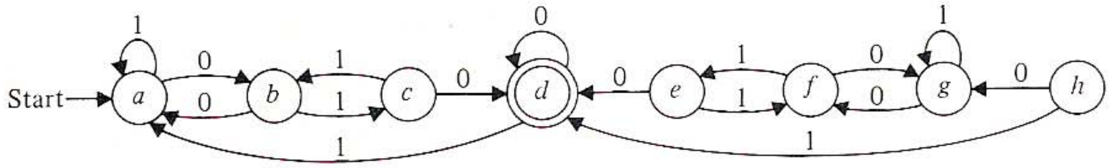

Fig. 3.9 A finite automaton.

3.25 Find the minimum-state finite automaton equivalent to the transition diagram of Fig. 3.9.

3.26

a) What are the equivalence classes of R_L in the Myhill-Nerode theorem (Theorem 3.9) for $L = \{0^n1^n \mid n \geq 1\}$?

b) Use your answer in (a) to show $\{0^n1^n \mid n \geq 1\}$ not regular.

c) Repeat (a) for $\{x \mid x$ has an equal number of 0's and 1's$\}$.

* **3.27** R is a *congruence* relation if xRy implies $wxzRwyz$ for all w and z. Prove that a set is regular if and only if it is the union of some of the congruence classes of a congruence relation of finite index.

* **3.28** Let M be a finite automaton with n states. Let p and q be distinguishable states of M and let x be a shortest string distinguishing p and q. How long can the string x be as a function of n?

** **3.29** In a two-tape FA each state is designated as reading tape 1 or tape 2. A pair of strings (x, y) is accepted if the FA, when presented with strings x and y on its respective tapes, reaches a final state with the tape heads immediately to the right of x and y. Let L be the set of pairs accepted by a two-tape FA M. Give algorithms to answer the following questions.

a) Is L empty? b) Is L finite?

c) Do there exist L_1 and L_2 such that $L = L_1 \times L_2$?

3.30

a) Prove that there exists a constant $c > 0$ such that the algorithm of Fig. 3.8 requires time greater than cn^2 for infinitely many DFA where n is the number of states and the input alphabet has two symbols.

** b) Give an algorithm for minimizing states in a DFA whose execution time is $0(|\Sigma| n \log n)$. Here Σ is the input alphabet. [*Hint:* Instead of asking for each pair of states (p, q) and each input a if $\delta(p, a)$ and $\delta(q, a)$ are distinguishable, partition the states into final and nonfinal states. Then refine the partition by considering all states whose next state under some input symbol is in one particular block of the partition. Each time a block is partitioned, refine the partition further by using the smaller sub-block. Use list processing to make the algorithm as efficient as possible.]

Solutions to Selected Exercises

3.4(b) $L' = \{a_2 a_1 a_4 a_3 \cdots a_{2n} a_{2n-1} \mid a_1 a_2 \cdots a_{2n}$ is in $L\}$ is regular. Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA accepting L. We construct a DFA M' that accepts L'. M' will process tape symbols in pairs. On seeing the first symbol a in a pair, M' stores a in its finite control. Then