计算理论导论

习题三: 正则语言的特性

中国人民大学　信息学院　崔冠宇　2018202147

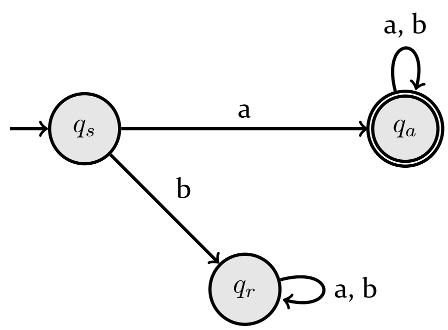
1. **1.4** (e)(g).

(e) $\{w|\ w$ starts with an a and has at most one b$\}$,

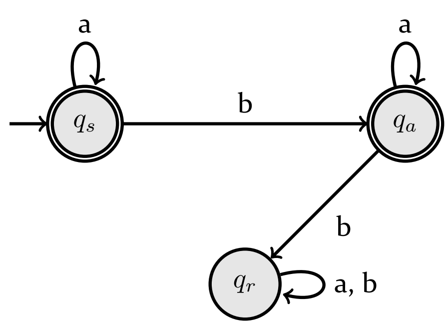
(g) $\{w|\ w$ has even length and an odd number of a's$\}$.

解: 思路: 先设计识别两个语言的 DFA，再利用笛卡尔积构造识别两个语言的交的 DFA。

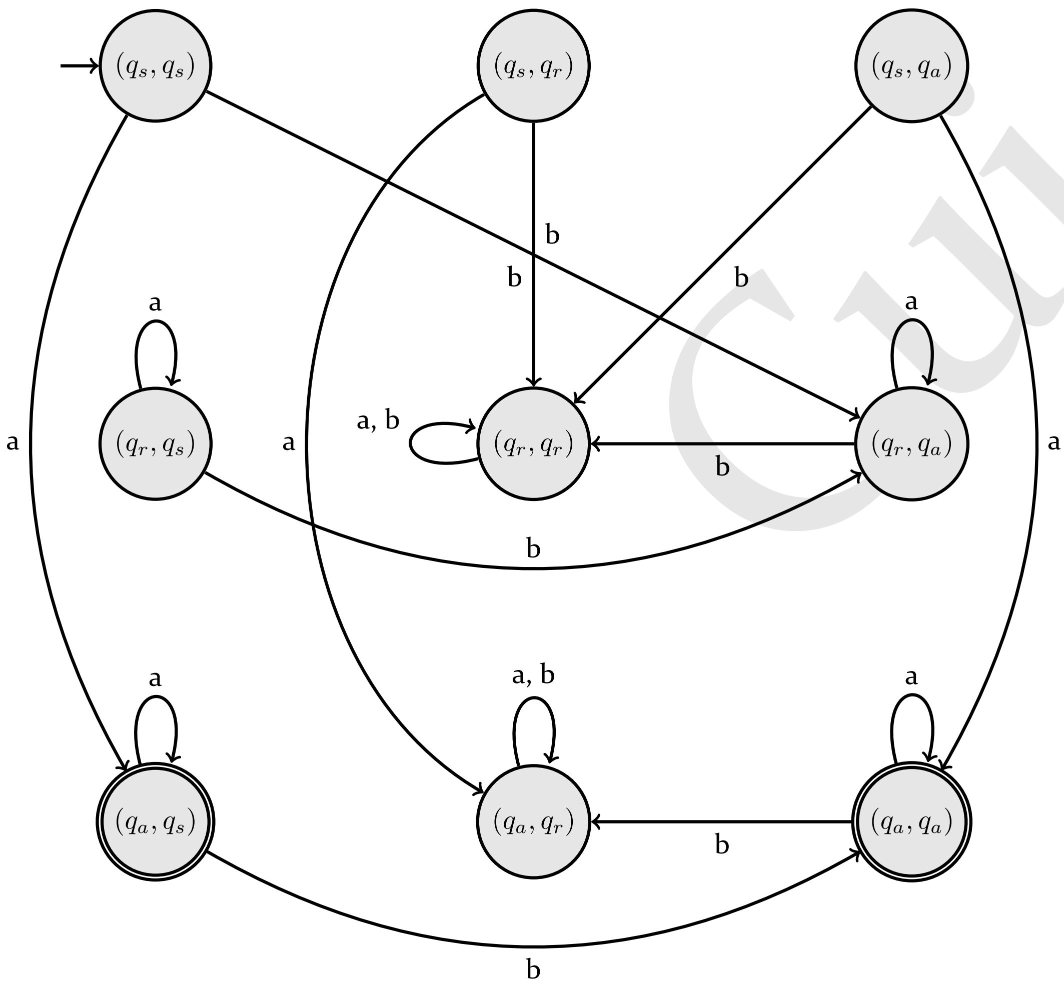
(e) 识别“以 a 开头的字符串”的 DFA:



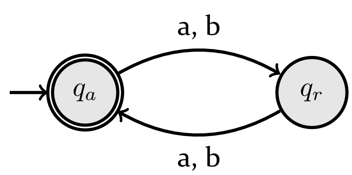
识别“至多含有一个 b”的字符串的 DFA:



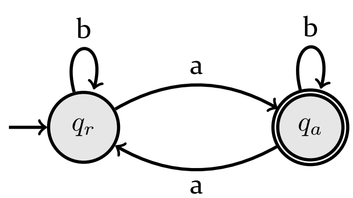
二者的笛卡尔积（其中 (q_r, q_s) 从起始状态不可达，可以去掉）：



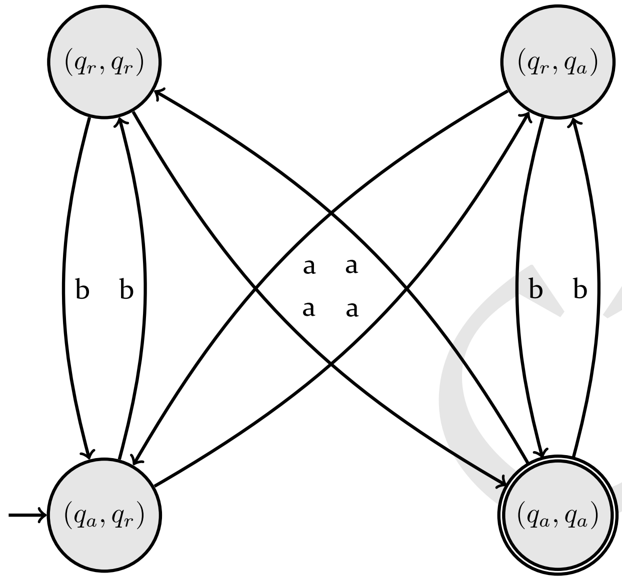
(g) 识别“偶数长度的字符串”的 DFA:



识别“含有奇数个 a 的字符串”的 DFA



二者的笛卡尔积:

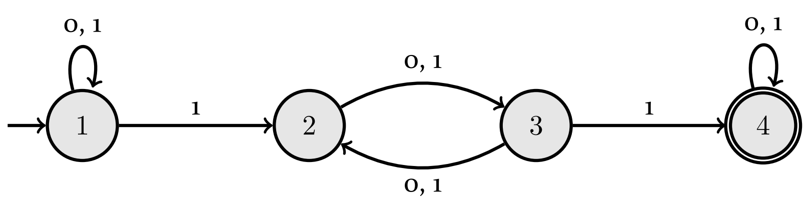


2. **1.13** Let F be the language of all strings over $\{0, 1\}$ that do not contain a pair of 1s that are separated by an odd number of symbols. Give the state diagram of a DFA with five states that recognizes F. (You may find it helpful first to find a 4-state NFA for the complement of F.)

解: 思路: 先设计接受 $\bar{F}$ 的 NFA，然后将其改造为 DFA，再通过交换接受状态与非接受状态得到结果。

(1) $\bar{F} = \{w|\ w$ 含有被奇数个符号分隔的一对1$\}$，写成正则表达式为 (0+1)*1(0+1)((0+1)(0+1))*1(0+1)*。

设计识别 $\bar{F}$ 语言的四状态 NFA $= (Q, \{0, 1\}, \delta, 1, \{4\})$:

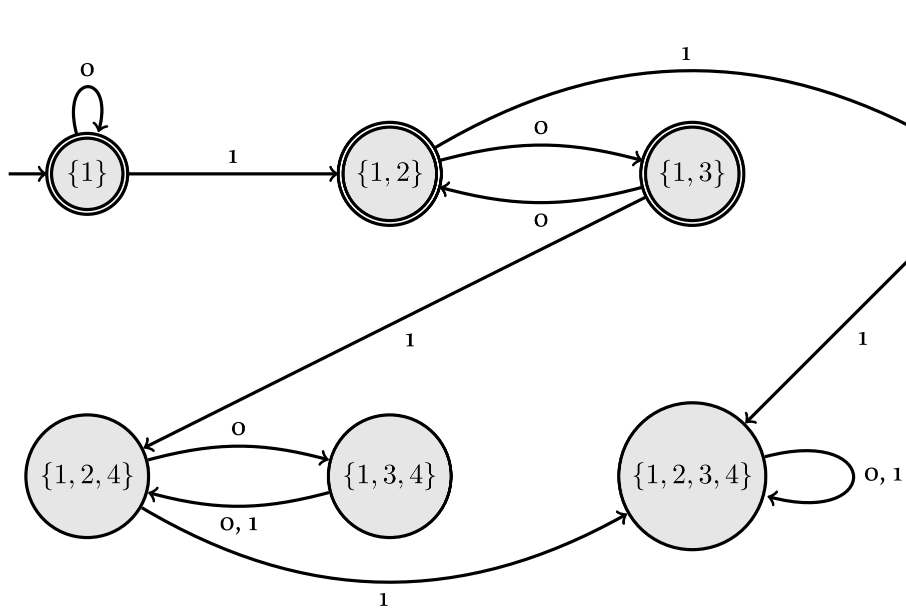


使用 **Theorem 1.39** 将其改造为等价的 DFA $= (Q' = \mathcal{P}(Q), \{0, 1\}, \delta', \{1\}, F')$，其中，

$F' = \{\{4\}, \{1, 4\}, \{2, 4\}, \{3, 4\}, \{1, 2, 4\}, \{1, 3, 4\}, \{2, 3, 4\}, \{1, 2, 3, 4\}\}$，$\delta'$ 由下表给出:

	0	1
$\emptyset$	$\emptyset$	$\emptyset$
$\{1\}$	$\{1\}$	$\{1,2\}$
$\{2\}$	$\{3\}$	$\{3\}$
$\{3\}$	$\{2\}$	$\{2,4\}$
$\{4\}$	$\{4\}$	$\{4\}$
$\{1,2\}$	$\{1,3\}$	$\{1,2,3\}$
$\{1,3\}$	$\{1,2\}$	$\{1,2,4\}$
$\{1,4\}$	$\{1,4\}$	$\{1,2,4\}$
$\{2,3\}$	$\{2,3\}$	$\{2,3,4\}$
$\{2,4\}$	$\{3,4\}$	$\{3,4\}$
$\{3,4\}$	$\{2,4\}$	$\{2,4\}$
$\{1,2,3\}$	$\{1,2,3\}$	$\{1,2,3,4\}$
$\{1,2,4\}$	$\{1,3,4\}$	$\{1,2,3,4\}$
$\{1,3,4\}$	$\{1,2,4\}$	$\{1,2,4\}$
$\{2,3,4\}$	$\{2,3,4\}$	$\{2,3,4\}$
$\{1,2,3,4\}$	$\{1,2,3,4\}$	$\{1,2,3,4\}$

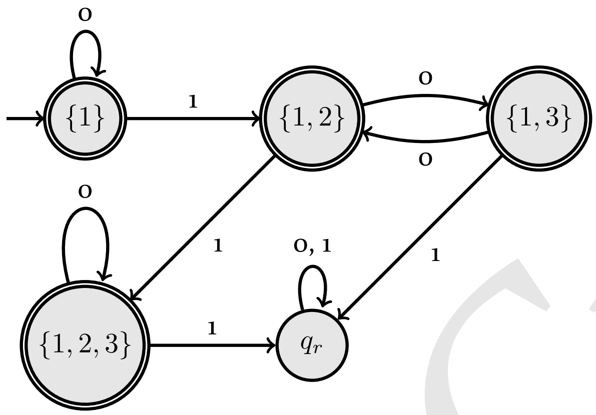
从起始状态开始遍历，将可达的状态标红。可见仅有 $\{1\},\{1,2\},\{1,3\},\{1,2,3\},\{1,2,4\},\{1,3,4\}$, $\{1,2,3,4\}$ 七个状态可达。交换 DFA 的接受状态与非接受状态，可得接受 F 的 DFA $= (Q', \{0,1\}, \delta', \{1\}, Q'-F)$ 如下图所示:



这个 DFA 还不是五状态，考虑使用最小化的方法。

先求出状态之间的等价关系：$\{1,2,4\} \equiv \{1,3,4\} \equiv \{1,2,3,4\}$。

将这些状态合并为 q_r，得到五状态 DFA:



3. 1.14

a. Show that if M is a DFA that recognizes language B, swapping the accept and nonaccept states in M yields a new DFA recognizing the complement of B. Conclude that the class of regular languages is closed under complement.

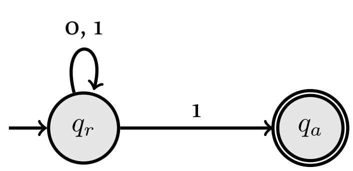
b. Show by giving an example that if M is an NFA that recognizes language C, swapping the accept and nonaccept states in M doesn't necessarily yield a new NFA that recognizes the complement of C. Is the class of languages recognized by NFAs closed under complement? Explain your answer.

解:

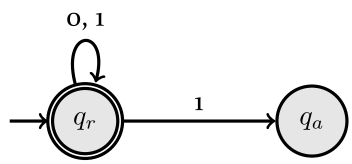
(a) 设 $M = (Q, \Sigma, \delta, q_0, F)$ 是识别 B 的 NFA，交换 M 的接受状态与非接受状态，得到 $\bar{M} = (Q, \Sigma, \delta, q_0, F' = Q - F)$，下面证明 $\bar{M}$ 接受 $\bar{B} = \{w \in \Sigma^* | w \notin B\}$。

因为 M 接受 B，即对任意 w，$w \in B \leftrightarrow \delta^*(q_0, w) \in F$，即 $w \in \bar{B} \leftrightarrow \delta^*(q_0, w) \in Q - F$，因此 $\bar{M}$ 识别语言 $\bar{B}$。因此，对于任意正则语言 L，都能构造一台识别它的补语言 $\bar{L}$ 的 DFA，即 $\bar{L}$ 也是正则语言。或者说正则语言在补运算下封闭。

(b) 考虑如下识别“以 1 结尾的字符串”的 NFA A:



如果仅仅交换接受状态与非接受状态，得到下列 NFA A':



注意到 1, 11 等字符串也能被接受（自环）。由此说明，对于 NFA 不能通过简单地“交换接受状态与非接受状态”获得识别补语言的 NFA。

虽然有上面的“反例”存在，但是由于每一台 NFA 都存在与之等价的 DFA，因此 NFA 和 DFA 识别相同的语言类（正则语言），所以 NFA 识别的语言类在补操作下仍然封闭。

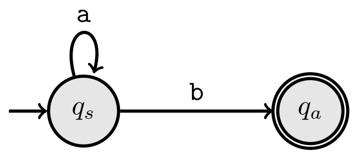
4. **1.15** Give a counterexample to show that the following construction fails to prove **Theorem 1.49**, the closure of the class of regular languages under the star operation. [1]

> Let $N_1 = (Q_1, \Sigma, \delta_1, q_1, F_1)$ recognize A_1. Construct $N = (Q_1, \Sigma, \delta, q_1, F)$ as follows. N is supposed to recognize A_1^*.
>
> **a.** The states of N are the states of N_1.
>
> **b.** The start state of N is the same as the start state of N_1.
>
> **c.** $F = \{q_1\} \cup F_1$. The accept states F are the old accept states plus its start state.
>
> **d.** Define δ so that for any $q \in Q_1$ and any $a \in \Sigma_\varepsilon$,
>
> $$\delta(q, a) = \begin{cases} \delta_1(q, a) & q \notin F_1 \text{ or } a \neq \varepsilon \\ \delta_1(q, a) \cup \{q_1\} & q \in F_1 \text{ and } a = \varepsilon. \end{cases}$$

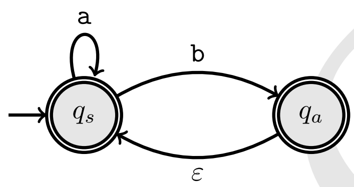
(Suggestion: Show this construction graphically, as in **Figure 1.50**.)

解: 取 $\Sigma = \{\mathtt{a}, \mathtt{b}\}$，$A_1 = \{\mathtt{a}^n\mathtt{b} | n \geq 0\}$，$N_1$ 如下:

[1]In other words, you must present a finite automaton, N_1, for which the constructed automaton N does not recognize the star of N_1's language.



按照上面的方法构造 N:



可见 N 接受了串 $\mathtt{a} \notin A_1^*$，与预期不符。

5. Find a minimal equivalent DFA for the following DFA, where A is the start state, C and F are accepting states:

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

解: 思路: 利用课堂讲解的算法。

第 0 次迭代:

$\equiv_Q = Q \times Q$（仅写出“上三角”部分，即 (a, b)，(b, a) 只写出其一，下同），即

$\equiv_Q = \{(A, A), (A, B), (A, C), (A, D), (A, E), (A, F), (B, B), (B, C), (B, D), (B, E), (B, F), (C, C), (C, D), (C, E), (C, F), (D, D), (D, E), (D, F), (E, E), (E, F), (F, F)\}$。

对每一个接受态 q_a，删去所有 $(q, q_a)(q \notin F)$ 和 (q_a, q)，可得

$\equiv_Q = \{(A, A), (A, B), (A, D), (A, E), (B, B), (B, D), (B, E), (C, C), (C, F), (D, D), (D, E), (E, E), (F, F)\}$。

之后的每次迭代，对 $\equiv_Q$ 中的每一对状态 (p, q)，考察二者是否等价，即 $\forall a \in \Sigma$，

都有 $(\delta(p, a), \delta(q, a)) \in \equiv_Q$。等价则保留，不等价则删除。

第 1 次迭代:

$\equiv_Q = \{(A,A),(A,D),(B,B),(B,E),(C,C),(C,F),(D,D),(E,E),(F,F)\}$。

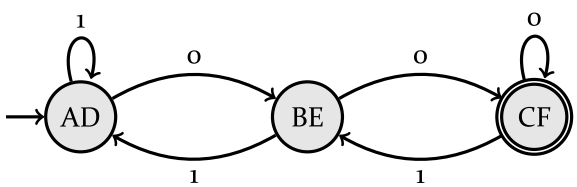
第 2 次迭代:

$\equiv_Q = \{(A,A),(A,D),(B,B),(B,E),(C,C),(C,F),(D,D),(E,E),(F,F)\}$，与上次相比没有变化，迭代结束。因此状态间非平凡的等价关系有: $A \equiv D, B \equiv E, C \equiv F$。

将等价的状态合并后的 δ 为:

	0	1
AD	BE	AD
BE	CF	AD
CF	CF	BE

画出此时的 DFA 的状态图:



6. 1.32 Let

$$\Sigma_3 = \left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \cdots, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$$

Σ_3 contains all size 3 columns of 0s and 1s. A string of symbols in Σ_3 gives three rows of 0s and 1s. Consider each row to be a binary number and let

$$B = \{w \in \Sigma_3 | \text{ the bottom row of } w \text{ is the sum of the top two rows}\}.$$

For example,

$$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \in B, \qquad \text{but} \qquad \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \notin B$$

Show that B is regular. (Hint: Working with $B^{\mathcal{R}}$ is easier. You may assume the result claimed in **Problem 1.31**.)

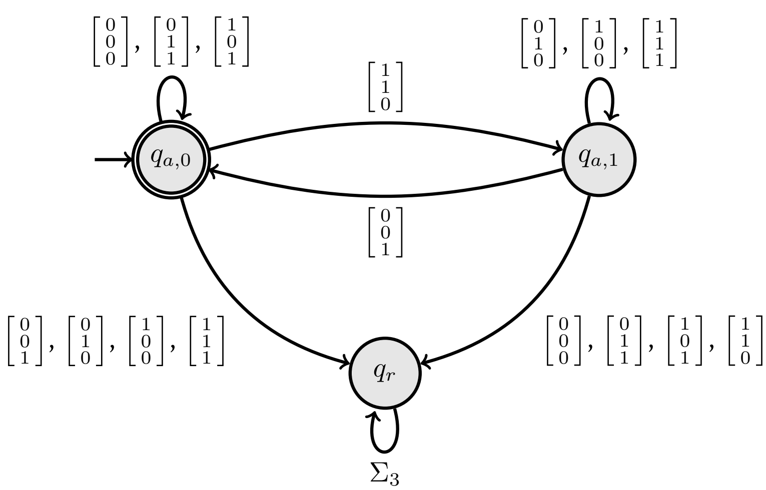
证明: 思路: 本题是证明“符合二进制加法的字符串”是正则语言。可以模仿计算机做加法的行为，设置一个保存进位的状态，再从 $w^{\mathcal{R}}$ 入手，从“低位”向“高位”验证正确性。

由于正则语言对字符串反转封闭，所以往证 $B^{\mathcal{R}}$ 是正则语言，尝试构造一台 DFA $A = (Q, \Sigma_3, \delta, q_{a,0}, F)$ 识别 $B^{\mathcal{R}}$:

- $Q = \{q_{a,0}, q_{a,1}, q_r\}$，其中 $q_{a,0}$ 表示和正确，进位位为 0；$q_{a,1}$ 表示和正确，进位位为 1；q_r 表示和不正确。
- $F = \{q_{a,0}\}$
- $\delta : Q \times \Sigma_3 \to Q$ 参照二进制加法法则，如下表所示:

	$\begin{bmatrix}0\\0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\\0\end{bmatrix}$	$\begin{bmatrix}0\\1\\1\end{bmatrix}$	$\begin{bmatrix}1\\0\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\\1\end{bmatrix}$	$\begin{bmatrix}1\\1\\0\end{bmatrix}$	$\begin{bmatrix}1\\1\\1\end{bmatrix}$
$q_{a,0}$	$q_{a,0}$	q_r	q_r	$q_{a,0}$	q_r	$q_{a,0}$	$q_{a,1}$	q_r
$q_{a,1}$	q_r	$q_{a,0}$	$q_{a,1}$	q_r	$q_{a,1}$	q_r	q_r	$q_{a,1}$
q_r	q_r							

画出 DFA 的图表示:



因此，$B^{\mathcal{R}}$ 可被 A 接受，于是 $B^{\mathcal{R}}$ 是正则语言。根据正则语言对于反转操作的封闭性，$B = (B^{\mathcal{R}})^{\mathcal{R}}$ 也是正则语言。□

7. **1.35** Let Σ_2 be the same as in **Problem 1.33**. Consider the top and bottom rows to be strings of 0s and 1s, and let

$$E = \{w \in \Sigma_2 | \text{ the bottom row of } w \text{ is the reverse of the top row of } w\}.$$

Show that E is not regular.

证明: 思路: 用泵引理。

反证法，若 E 是正则语言，设 p 是由泵引理给出的泵长度。

取 $w = \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]^p \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]^p$，则根据泵引理的要求，$w$ 可被写成 xyz 三部分，其中 $|y| > 0$ 且 $|xy| \le p$。于是 y 只能具有 $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]^k$ $(k > 0)$ 的形式，但是 $xyyz = \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]^{p+k} \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]^p \notin E$，矛盾。于是 E 不是正则语言。□

8. 【选做】**1.38** An ***all***-NFA M is a 5-tuple $(Q, \Sigma, \delta, q_0, F)$ that accepts $x \in \Sigma^*$ if every possible state that M could be in after reading input x is a state from F. Note, in contrast, that an ordinary NFA accepts a string if some state among these possible states is an accept state. Prove that all-NFAs recognize the class of regular languages.

解: 根据定义，每一台 DFA 都是特殊的 all-NFA，因此只需要证明任意 all-NFA 都能转化为等价的 DFA。这个转化过程类似于 **Theorem 1.39**，对于某 all-NFA $M = (Q, \Sigma, \delta, q_0, F)$，它的等价的 DFA 为 $A = (Q', \Sigma, \delta', \{q_0\}, F')$，其中:

- $Q' = \mathcal{P}(Q)$;
- $\delta'(R, a) = \bigcup\limits_{r \in R} E(\delta(r, a))$;
- $F' = \{q \in Q' | q \subseteq F\}$。

最后的 F' 的修改是为了保证读入串 x 后的所有状态都包含于 F。

9. 【选做】**1.43** Let A be any language. Define *DROP-OUT*(A) to be the language containing all strings that can be obtained by removing one symbol from a string in A. Thus, *DROP-OUT*$(A) = \{xz | xyz \in$

A where $x, z \in \Sigma^*, y \in \Sigma\}$. Show that the class of regular languages is closed under the *DROP-OUT* operation. Give both a proof by picture and a more formal proof by construction as in **Theorem 1.47**.

解: 思路: 直接在原 DFA 的边上修改符号是错误的。考虑拷贝一台 DFA，在二者之间增加一些边来实现"跳过字符"。这里只给出形式化的证明。

任意给定正则语言 A，由定义，存在识别它的有限状态自动机 $M = (Q, \Sigma, \delta, q_0, F)$，构造一台 NFA $N = (Q \cup Q', \Sigma, \delta', q_0, F')$，其中 $Q' = \{q' | q \in Q\}$ 是原机器的状态的拷贝；δ' 如下所示:

$$\delta'(p, \alpha) = \begin{cases} \bigcup\limits_{\beta \in \Sigma} \{(\delta(\delta(q, \beta), \alpha))'\} \cup \{\delta(q, \alpha)\}, & t = q \in Q \\ \\ (\delta(q, \alpha))', & t = q' \in Q' \end{cases}$$

即每一个原状态 q 识别字符 α 时，给它增加一些指向它的所有后继状态 $\bigcup\limits_{\beta \in \Sigma} \{\delta(q, \beta)\}$ 接受字符 α 的后继状态 $\bigcup\limits_{\beta \in \Sigma} \{\delta(\delta(q, \beta), \alpha)\}$ 在复制状态集 Q' 中对应的状态 $\bigcup\limits_{\beta \in \Sigma} \{(\delta(\delta(q, \beta), \alpha))'\}$。

正确性的证明:

设 $xyz \in A$，其中 $x, z \in \Sigma^*, y \in \Sigma$，于是在原自动机 M 上存在状态序列 q_0, q_1, q_2, q_a，满足 $q_1 = \delta^*(q_0, x), q_2 = \delta(q_1, y), q_3 = \delta^*(q_2, z) \in F$。$N$ 读取字符串 xz 时，有一条通路经过 q_0, q_1, q_3'，因而 N 识别该串。反之亦然。