形式语言与自动机复习

一. 概论

数学概念

① $\Sigma=\{0,1\}$　　Σ: 符号的有穷非空序列

$\Sigma\times\Sigma=\{(0,0),(0,1),(1,0),(1,1)\}$

② $A=\{0,1\}$　　一切子集的集合

$P(A)=\{\{\},\{0\},\{1\},\{1,0\}\}$

$|P(A)|=2^{|A|}$　　个数

$|A\times A|=|A|^2$

③ Σ^k: 长度为k的串的集合.

$\Sigma^3=\{000,001,010,011,\cdots\cdots,111\}$

④克林闭包

$$\Sigma^*=\bigcup_{i=1}^{\infty}\Sigma^i=\Sigma^0\cup\Sigma^1\cup\Sigma^2\cdots\quad(L^0=\{\varepsilon\})$$

⑤语言: 若Σ是字母表, $L\in\Sigma^*$ 则L是Σ上的语言.

思考

1. $\phi^0=\{\varepsilon\}$　$\{\varepsilon\}^0=\{\varepsilon\}$　$\phi^*=\{\varepsilon\}$　$\{\varepsilon\}^*=\{\varepsilon\}$

2. Σ^*一定不等于Σ^+

$\because\ \varepsilon\in\Sigma^*$ 但 $\varepsilon\notin\Sigma^+$

☆3. $\varepsilon A=A\varepsilon=A$，$\phi A=A\cdot\phi=\phi$　　$\phi A=\{xy\mid x\in\phi,\ y\in A\}$

5. $\Sigma^* 1 \Sigma^* = \{w \mid w中至少出现1次1\}$

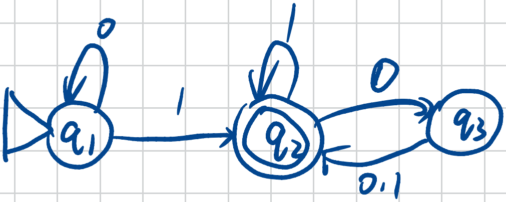
6. $(\Sigma\Sigma)^* = \{w \in \Sigma^* \mid |w|为偶数\}$

二、有穷状态自动机

1. DFA

def1. DFA由5元组 $M=(Q, \Sigma, \delta, q_0, F)$ 构成

Q：状态集　Σ：字母表　δ：转移函数　q_0：起始状态　F：接受状态集合



$Q = \{q_1, q_2, q_3\}$

$\Sigma = \{0, 1\}$

$q_0 = q_1$

$F = q_2$

δ :

	0	1
q_1	q_1	q_2
q_2	q_3	q_2
q_3	q_2	q_2

def2：正则语言 $\longleftrightarrow$ 被有穷自动机FA接受的语言

def3：M_1, M_2 为FA，若 $L(M_1) = L(M_2)$ 则 M_1 与 M_2 等价

2. NFA.

·与DFA区别 ① 一入多出

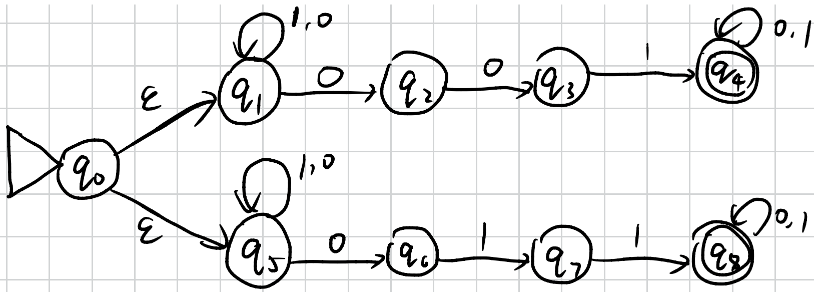
② 有ε自动迁移

③ 可作并行计算或树

EXP2-8:构造一个接受“含有子串011或001”的NFA，字母表为{0, 1}。要求：

① 采用文字形式，简要说明设计思路。

② 给出所设计的非确定性自动机的状态转移图或形式化定义。



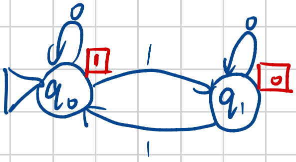
3. Moore 机与 Mealy 机

$M=(Q,\Sigma,\Delta,\delta,\lambda,q_0)$

Moore：输出由状态决定，直接将输出标签写在状态上.

Mealy：输出由状态、输入共同决定.

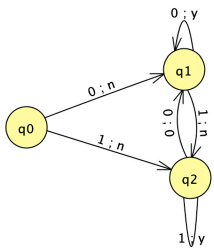
Moore



(偶数1 → 输出1；奇数1 → 输出0)

Mealy

EXP2.13　给出一个0，1串的集合S，该集合中的串都以00或11结尾。要求设计一个只有两个输出符号（Δ={y,n}）的Mealy机，当它读属于集合S的串时，输出y，表示接受；当它读不属于集合S的串时，输出n，表示不接受。



与该Mealy机等价的DFA?

三、文法

def: $G=(V, T, P, S)$

变元集　终结符集　产生式　开始符号

分类（区别在于：产生式的形式）

① 0型 — 无任何限制　　短语结构文法

② 1型 — 0型+不能缩短　　上下文有关文法

③ 2型 — 1型+左边仅1个非终结符　　上下文无关文法

④ 3型 — 2型+右边线性（$A\to aB$ 或 $A\to a$）　　正则文法

限制增加

3型 ⊂ 2型 ⊂ 1型 ⊂ 0型

例　$L=\{a^nb^n, n\geq 1\}$

例　$L=\{a^nb^nc^n, n\geq 1\}$

四、正则表达式

正则表达式 = 基础字符($\Sigma, \varepsilon, \phi$) + 三大运算($\cup, \cdot, *$)

① $\cup$: $A\cup B = \{x \mid x\in A \text{ or } x\in B\}$

② $\cdot$: $A\cdot B = \{xy \mid x\in A \text{ and } y\in B\}$ $xy \neq yx$

③ $*$: $A^* = \{x_1x_2\cdots x_k \mid k\geq 0 \text{ and each } x_i\in A\} = A^0\cup A^1\cup A^2\cup\cdots$

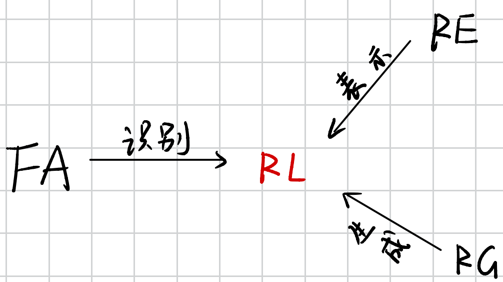
注: $A=\{\varepsilon\}$ $A^*=\{\varepsilon\}$ (ε 拼多少次都是空串)

$A=\phi$ $A^*=\{\varepsilon\}$ ($A^0=\{\varepsilon\}$; $k\geq1$时 $A^k=\phi$)

$R\cdot\phi=\phi$, $R\cdot\varepsilon=R$ $R+\phi=R$

FA & RL & RE & RG

自动机 语言 表达式 文法

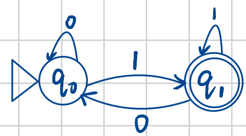


例: $\Sigma=\{0,1\}$

$L=\{w\in\Sigma^* \mid w\text{以}1\text{结尾}\}$

① RL: L本身,如 $\{1, 01, 011\}$

② FA:



③ RG: $S\rightarrow 0S \mid 1S \mid 1$

④ RE: $(0|1)^*1$

ex: 写出RE for L

L: $\{w \mid w\in\{0,1\}^* \text{ and } w$ 没有连续的0出现 $\}$

$r = (1+01)^*(0+\varepsilon)$

L: $\{w \mid w\in\{0,1\}^* \text{ and } w$ 有连续的0出现 $\}$

$r = (1+0)^*00(1+0)^*$

相关定律

$$L+L=L$$

$$(L^*)^* = L^*$$

$$L^+ = LL^*$$

$$L^* = L^+ + \varepsilon$$

五. 正则语言性质

★泵引理 (正则语言一定符合泵引理，不符合则不是正则语言)

若 L 是正则的，则存在正整数 N，对 $\forall w \in L$，只要 $|w| \geq N$，可将 w 分成 $w=xyz$ 满足:

① $y \neq \varepsilon$ $(|y|>0)$

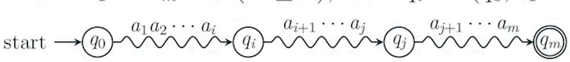
② $|xy| \leq N$

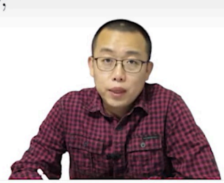
③ $\forall k \geq 0$, $xy^kz \in L$

证明:

❶ 如果 L 正则, 那么存在有 n 个状态 DFA A 使 $\mathbf{L}(A) = L$;

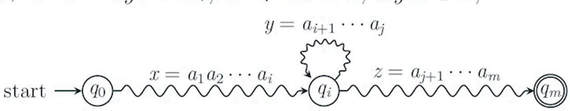
❷ 取 $w = a_1 \dots a_m \in L$ $(m \geq n)$, 定义 $q_i = \hat{\delta}(q_0, a_1 \dots a_i)$;





❸ 由鸽巢原理, 必有两状态相同 $q_i = q_j$ $(0 \leq i < j \leq n)$;

❹ 那么 $w = xyz$ 如图, 且有 $\forall k > 0$, $xy^kz \in L$;



❺ 而因为 $i < j$ 所以 $y \neq \varepsilon$ (即 $|y| > 0$), 因为 $j \leq n$ 所以 $|xy| \leq n$. □

例 (往少3泵)

例 5. 证明 $L = \{0^i 1^j \mid i > j\}$ 不是正则的.

证明:

❶ 假设 L 是正则的.

❷ 那么, 存在 $N \in \mathbb{Z}^+$, 对 $\forall w \in L\,(|w| \geq N)$ 满足泵引理.

❸ 从 L 中取 $w = 0^{N+1}1^N$, 则 $w \in L$ 且 $|w| = 2N+1 \geq N$.

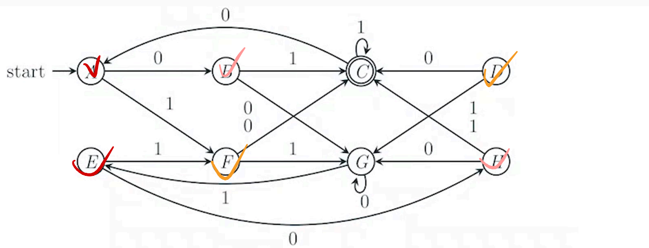
❹ 由泵引理, w 可被分为 $w = xyz$, 且 $|xy| \leq N$ 和 $y \neq \varepsilon$.

❺ 那么, y 只能是 0^m 且 $m \geq 1$.

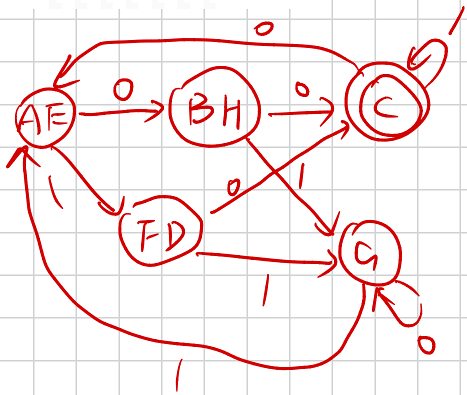
❻ 那么, $xz = xy^0z = 0^{N+1-m}1^N \notin L$, 因为 $N+1-m \leq N$, 而由泵引理 $xy^0z \in L$, 矛盾.

最小DFA

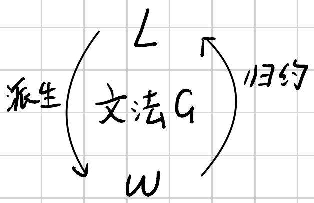
例 16. 用填表算法找到如图 DFA 中全部可区分状态对.



	A	B	C	D	E	F	G
B	x						
C	x	x					
D	x	x	x				
E		x	x	x			
F	x	x	x		x		
G	x	x	x	x	x	x	
H	x		x	x	x	x	x



六、上下文无关文法



[例] 文法：$S \to aSb \mid ab$ 生成语言 $\{a^n b^n \mid n \geq 1\}$

对串aabb

派生：$S \Rightarrow aSb \Rightarrow aabb$

归约：$aabb \Rightarrow aSb \Rightarrow S$

CFG：上下文无关文法、$G=(V,T,P,S)$ $L(G)=\{w \in T^* \mid S \overset{*}{\Rightarrow} w\}$

★ 语言转文法

① $L=\{0^n 1^n \mid n \in N\}$

$G:(S,\{0,1\},P,S)$

$P: S \to 0S1,\ S \to \varepsilon$

② $L=\{0^n 1^{2n+1} \mid n \in N\}$

$G=(S,\{0,1\},P,S)$

$P: S \to 1,\ S \to 0S11$

③ $L=\{w \in \{0,1\}^* \mid w$ 至少有 3个1$\}$

$RE=(1+0)^* 1 (1+0)^* 1 (1+0)^* 1 (1+0)^*$

$G:(\{A,S\},\{0,1\},P,S)$

$P:\ A \to 0A \mid A1 \mid \varepsilon$

$S \to A1A1A1A$

④ $L=\{w \in \{0,1\}^* \mid w$ 中0、1个数相等$\}$

$G:(\{A,S\},\{0,1\},P,S)$

$P: S \to 1S0 \mid 0S1 \mid SS$

证明：$RL \subset CFL$

思路：即任意一个 $L \subseteq RL$，L的表达式为E，都找到一个CFL表达示S

使 $L=L(E)=L(S)$

证：(归纳法) ① $|E| \leq 1$ 时易证得

② 设比 $|E|$ 小的正则表达式定理都成立。

最后一步，假设 $L(E_1)=L(S_1)$，$L(E_2)=L(S_2)$

① $E = E_1 + E_2$: $S \to S_1 \mid S_2$

② $E = E_1 \cdot E_2$: $S \to S_1 S_2$

③ $E = E_1^*$: $S \to S_1 S \mid \varepsilon$

$\therefore L(S) = L(E)$

(S)((S)(S))

语法分析树

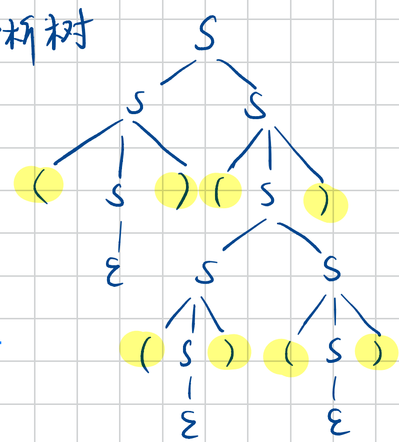
例: $S \to \varepsilon \mid (S) \mid SS$ 推出 ()(()())

$$S \underset{lm}{\Rightarrow} SS \underset{lm}{\Rightarrow} (S)S \underset{lm}{\Rightarrow} ()S \underset{lm}{\Rightarrow} ()(S) \underset{lm}{\Rightarrow} ()(SS) \underset{lm}{\Rightarrow} ()((S)S)$$

$$\Downarrow_{lm}$$

$$()(()()) \Leftarrow ()(()(S)) \Leftarrow ()(()S)$$

语法分析树



二义性/歧义性：$w \in L(G)$，w 至少有2棵不同分析树

• 乔姆斯基范式:

$A \to BC$ 一分为二

$A \to x$ 或终极化

$S \to \varepsilon$

不存在二义性

七、下推自动机PDA

PDA = NFA + stack

def：六元组 $(Q, \Sigma, \Gamma, \delta, q_0, F)$

Γ：可压栈字符表

$\delta: Q \times \Sigma_\varepsilon \times \Gamma_\varepsilon \longrightarrow P(Q \times \Gamma_\varepsilon)$ 是一个超集.

$\Sigma_\varepsilon = \Sigma \cup \{\varepsilon\}$ 读入/不读

$\Gamma_\varepsilon = \Gamma \cup \{\varepsilon\}$ 弹入/不弹

• PDA 与 CFG 等价

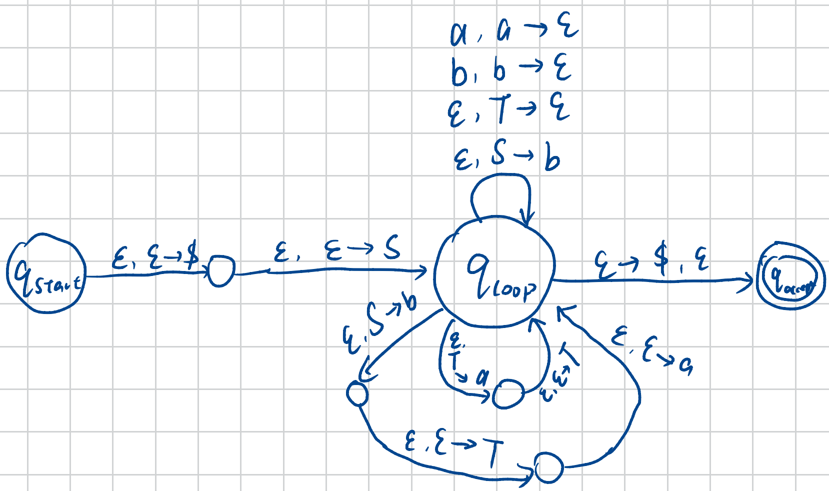
[构造PDA]

EXP：把CFG G转换成PDA P，其中

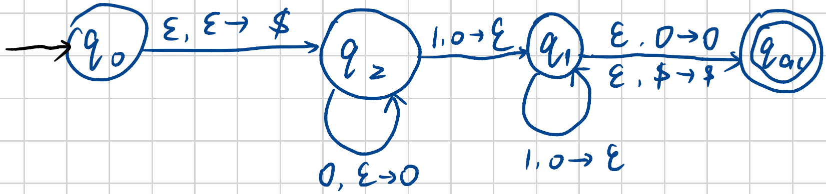
G：

S → aTb | b

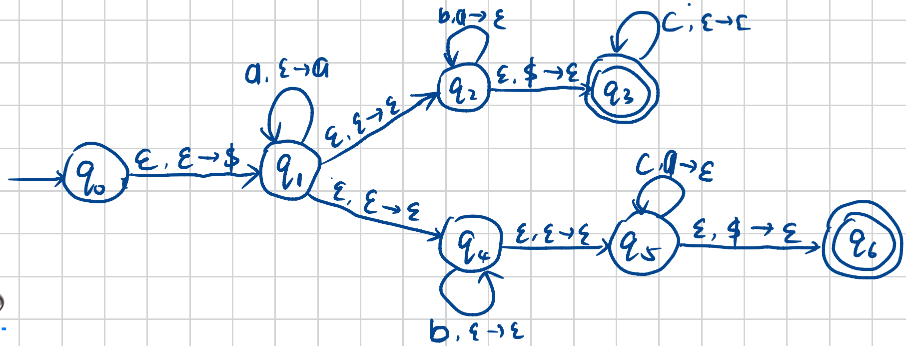
T → Ta |ε



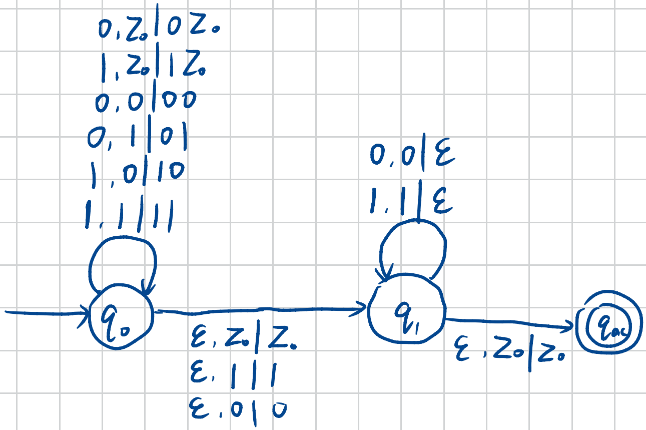
Ex：设计 PAD P，识别 $L=\{0^i1^j \mid i \geq j \geq 1\}$



Ex：$\{a^ib^jc^k \mid i,j,k \geq 0 \text{ and } i=j \text{ or } i=k\}$，设计 PDA

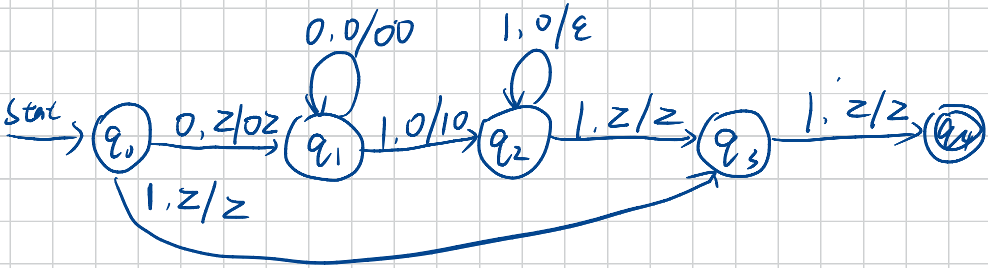


Ex：$L=\{ww^R \mid w\in\{0,1\}^*\}$，设计 PDA



[构造DPDA]

构造 DPDA M，使接受语言 $\{0^n 1^{n+2} \mid n \geq 0\}$



八、图灵机

• 电子计算机三大基础

- 冯·诺依曼结构　　系统基础
- 布尔电路　　物理基础
- 图灵机　　数学基础

• 形式化定义 7元组 $(Q, \Sigma, \Gamma, \delta, q_0, q_{accept}, q_{reject})$

$$\delta: Q \times \Gamma \rightarrow Q \times \Gamma \times \{L, R\}$$

一次计算三种可能

- accept
- reject
- loop

九、可判定性

可判定：对任意输入都停机，$w\in L$ 接受，$w\notin L$ 拒绝

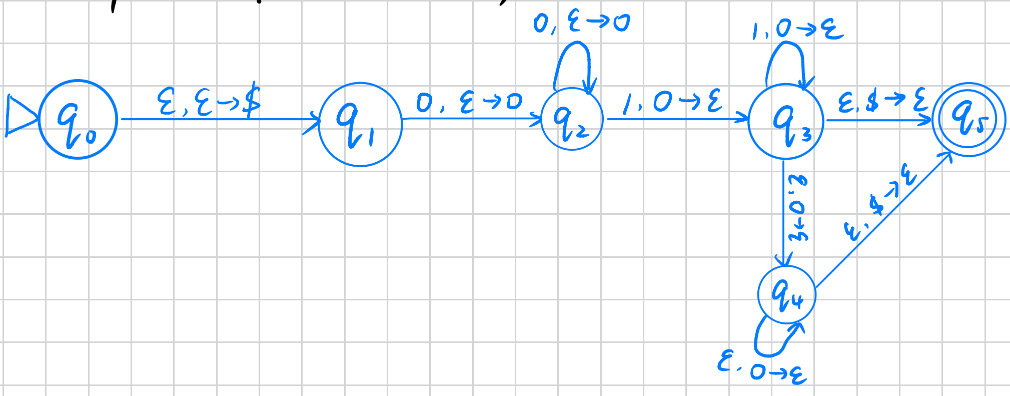
可识别：对任意输入都可能出现 accept、reject、loop 三种行为

A_{TM}、$HALT_{AM}$：可识别 不可判定

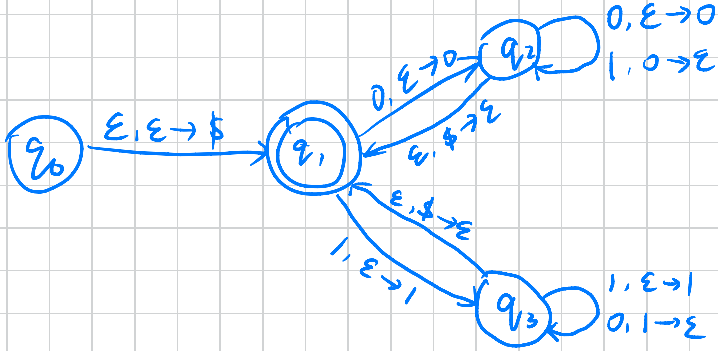
重点题型

[语言 $L \longrightarrow$ PDA]

① $\{0^n 1^m \mid n \geq m \geq 1\}$

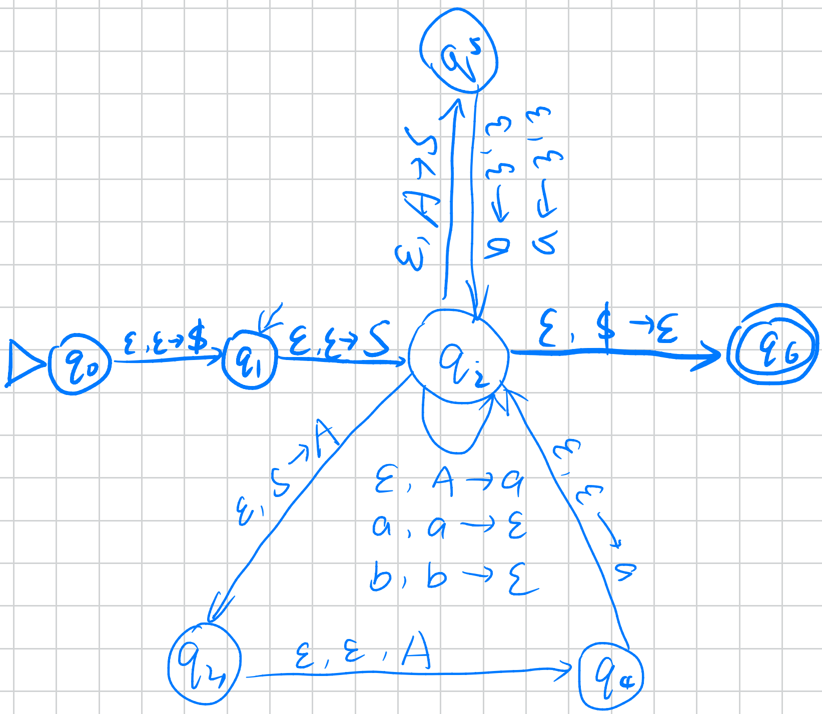


② 含有0的个数和1的个数相同的所有0,1串.



[文法G ⟶ PDA]

① $S \to aAA$, $A \to aS \mid bS \mid a$



[M-N定理]

1. 利用 Myhill-Nerode 定理证明下列语言是否正则语言，如果是正则语言，请构造其 FA、RE 及 RG。

   1) $\{x \mid x=x^R,\ x\in\{0,1\}^+\}$

   2) $\{x \mid x$ 中 0 的个数不少于 1 的个数，$x\in\{0,1\}\}$

   3) $\{xx^Rw \mid x,w\in\{0,1\}^+\}$

1) 证：设 $S=\{0^k \mid k\geq 1\}$

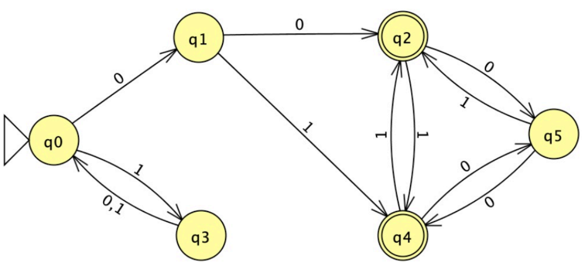
可取 $x_1, x_2 \in S$，$x_1=0^n$，$x_2=0^m$ $(m\neq n$ 且 $m,n\geq 1)$

取后缀 $z=10^n$ 则 $x_1z=0^n10^n\in L$

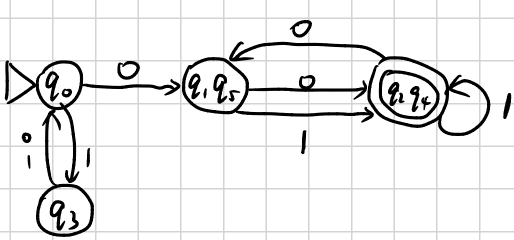
$x_2z=0^m10^n\notin L$

故 0^n 与 0^m 所在等价类是两个不同等价类，这样的等价类有无穷个，因此 L 不是正则语言。

[极小状态DFA]

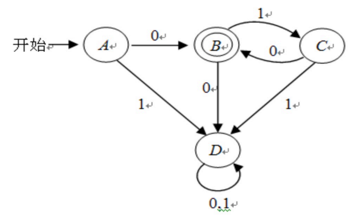


q_1	X				
q_2	X	X			
q_3	X	X	X		
q_4	X	X		X	
q_5	X		X	X	X
	q_0	q_1	q_2	q_3	q_4



[DFA → 正则文法RG]

EXP5.1 将下列DFA转化为等价的正则文法。



A → 0B | 0
A → 1D
B → 0D
B → 1C
C → 1D
C → 0B | 0
D → 0D
D → 1D

⇒

A → 0B | 1D | 0
B → 0D | 1C
C → 1D | 0B | 0
D → 0D | 1D

[正则文法RG → DFA]

1. 给出下列的正则文法 G，求出对应的 DFA M，使得 L(M) = L(G)。

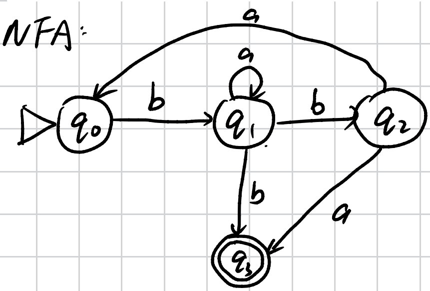
(1) G_1 =（V， T， P_1， S）

P_1: S→bB, B→aB | bA | b, A→a | aS

(2) G2= V， T， P_2， S）

P_2： S→aS | bB | a, B→bA | aB | aS

(1) NFA:



		a	b
A	0		1 = B
B	1	1 : B	23 = C
C	23	03 = D	
D	03		1 = B