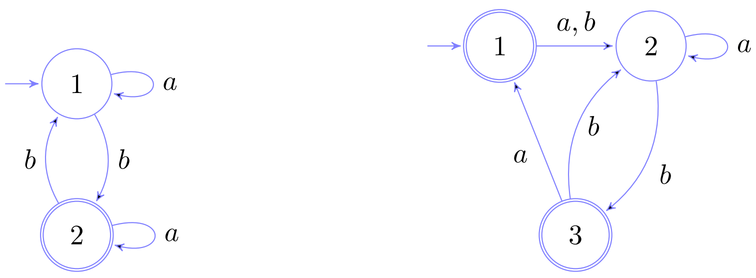
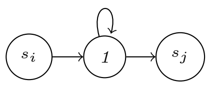
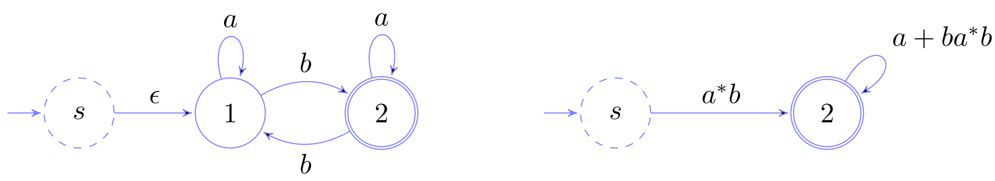
1. 把下述两个 NFA（如图 1 所示）分别转换成等价的正则表达式。



(a) (b)

图 1

解. (a) 对于仅剩初始状态 *1* 和接受状态 *2* 的情况，首先添加辅助初始状态 s（如图 *2a* 所示），然后消去状态 *1*。消去状态 *1* 的方法如下，考虑每一条形如  的路径



(a) 添加辅助初始状态 (b) 消去状态 1

图 2: NFA(a) 转化过程

（后面简称 $s_i \to 1 \to s_j$），得到状态 $s_i \rightsquigarrow s_j$ 的转换：例如路径 $s \to 1 \to 2$ 化简得到正则表达式 a^*b；路径 $2 \to 1 \to 2$ 化简得到正则表达式 ba^*b，和状态 *2* 本身的循环边合并为 $a + ba^*b$。类似的消去状态方法后面不在赘述，消去状态 *1* 后如图 *2b* 所示。

最后显然得到图 *1a* 等价的正则表达式$a^*b(a + ba^*b)^*$。

(b) 对于有多个接收状态的情况，首先添加辅助接受状态 *4* 如图 *3* 所示，使 NFA 仅含一个接收状态。按照状态 *1*，*2* 顺序消去，依次得代中间 NFA 如图图 *4a* 和图 *4b* 所示。

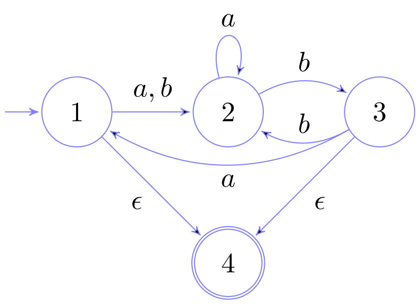
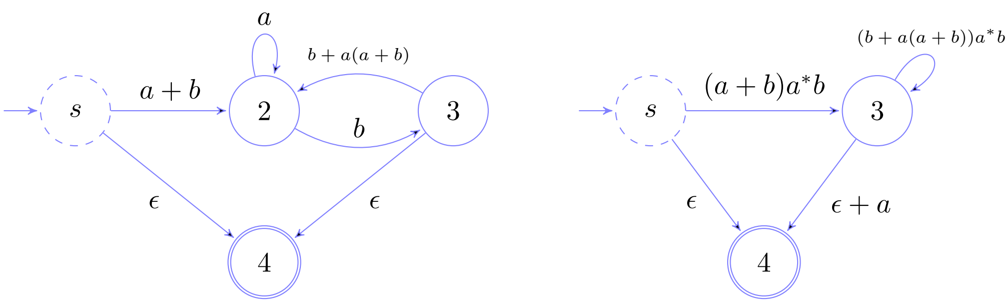


图 3: 接收状态唯一化

最后消去状态 *3*，得到图 *1b* 等价的正则表达式$\epsilon + ((a + b)a^*b((b + a(a + b))a^*b)^*(\epsilon + a))$。

注 1. 本题答案形式不唯一。具体来说，最终答案形式与消去状态的顺序有关，例如，如果按照状态 *2*，*3*，*1* 的顺序消去，最终答案是$((a + b)a^*b(ba^*b)^*a)^*((a + b)a^*b(ba^*b)^* + \epsilon)$。



(a) 中间过程 1 (b) 中间过程 2

图 4: NFA(b) 转化过程

2. 对语言 $L = \{(aab^*ab)^*\}$ 构造等价的右线性文法。

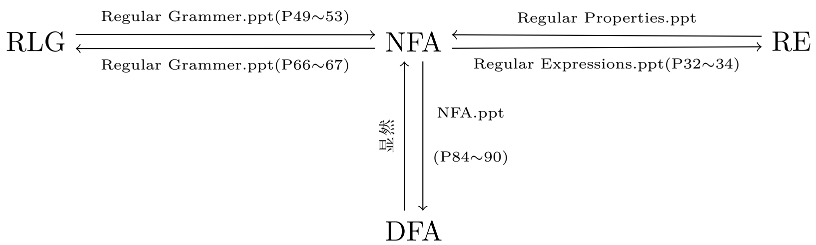
解. （法一，观察法）观察正则表达式 $(aab^*ab)^*$ 的形式，通过找规律方法容易写出与 L 等价的非线性文法 G'：

$$
\begin{aligned}
S &\rightarrow \lambda \mid aaBabS \\
B &\rightarrow \lambda \mid bB
\end{aligned}
$$

其中 λ 表示空串（下同），然后通过修正将 G' 转换为右线性文法 G：

$$
\begin{aligned}
S &\rightarrow \lambda \mid aaA \\
A &\rightarrow bA \mid abS
\end{aligned}
$$

注 2. 首先，本题答案并不唯一。就解法一而言，整个过程看起来简单，但是并不具备一般性，因为有时候将非线性文法转换为线性文法并不显然。但是通过 DFA、NFA、正则表达式和右（左）线性文法的学习，我们知道它们能够接收（表示）的语言集合相同，也学习了它们之间的一般性转换构造方法，可以总结为：



其中 RLG 表示右线性文法，RE 表示正则表达式。于是本题可以拆分为两个步骤：先将正则表达式转换为等价的 NFA（RE→NFA），再由 NFA 直接转换为右线性文法（NFA→RLG），详见解法二。

解. （法二，转化为等价 NFA）由正则表达式 $(aab^*ab)^*$ 容易构造等价的 NFA（图 *5* ）初始状态用起始变量 S 表示，状态 i 分别用变量 V_i 表示（$i = 1, 2, 3, 4$），由图 *5* 容易构造与之等价的右线性文法 G：

$$
\begin{aligned}
S &\rightarrow \lambda \mid aV_1 \\
V_1 &\rightarrow aV_2 \\
V_2 &\rightarrow bV_2 \mid aV_3
\end{aligned}
$$

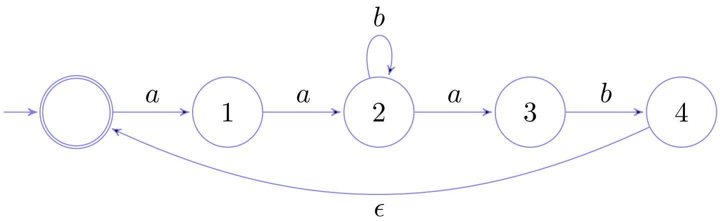


图 5: 问题二正则表达式等价 NFA

$$\begin{aligned} V_3 &\to bV_4 \\ V_4 &\to S \end{aligned}$$

简单化简 G 可得：

$$\begin{aligned} S &\to \lambda \ \mid \ aaV_2 \\ V_2 &\to bV_2 \ \mid \ abS \end{aligned}$$

3. 利用正则语言的泵引理证明 $L = \{www|w \in \{0,1\}^*\}$ 不是正则语言。

证明. 采用反证法。假设 L 是正则语言，设 p 是泵引理中的关键路径。考虑 L 中的串 $s = 0^p10^p10^p1$。因为 $|s| \geq p$，由泵引理可知，$s = xyz$，其中 $|xy| \leq p, |y| \geq 1$。展开观察，

$$s = \underbrace{\overbrace{0\cdots\cdots0}^{p}}_{xy}\underbrace{10^p10^p1}_{z}$$

显然，$y = 0^k$，其中 $1 \leq k \leq p$。于是考察 $xyyz$，发现 $xyyz = 0^{p+k}10^p10^p1 \notin L$，而由泵引理知 $xyyz$ 必属于 L，产生矛盾！故 L 不是正则语言。 □