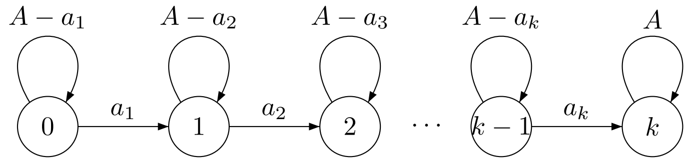
Solutions to problems of Chapter 1

Section 1.1

1.1.1 The minimal deterministic automaton recognizing the set X of words on an alphabet A with q symbols having $w = a_1a_2\cdots a_k$ as a subword has the form indicated below. This shows that



$$X = (A - a_1)^*a_1(A - a_2)^* \cdots (A - a_k)^*a_kA^*$$

is an unambiguous regular expression for X. Thus, the generating sequence of the number of words of X has the form

$$\begin{aligned} f_X(z) &= \frac{z^k}{(1-(q-1)z)^k(1-qz)} \\ &= \sum_{i=0}^{n-k} \binom{n-i-1}{k-1}(q-1)^{n-i-k}q^i \end{aligned}$$

since

$$\frac{1}{(1-z)^k} = \sum \binom{n+k-1}{k-1} z^n.$$

1.1.2 The axioms of a distance are easy to verify.

Section 1.2

1.2.1 The call $b \leftarrow Border(x)$ can be replaced by $b \leftarrow BorderSharp(x)$

Section 1.3

1.3.1 It is enough to keep track for each state p and for $0 \le i \le n$ of a word $x(p,i) \in X$ which realizes the distance $d(p,i)$ with $a_0 \cdots a_{i-1}$. Th result is the word $x(s,n)$ where s is the terminal state such that $d(s,n)$ is minimal.

1.3.2 If $u \in p^{-1}\mathcal{S}(w) \cap q^{-1}\mathcal{S}(w)$, then $pu = qu \in \mathcal{S}(w)$. Thus one the words, say pu, is a suffix of the other. This implies that p is a suffix of q and thus that $q^{-1}\mathcal{S}(w) \subset p^{-1}\mathcal{S}(w)$. Thus we can arrange the states of the automaton, which are the sets $p^{-1}\mathcal{S}(w)$, as the nodes of a tree reflecting the ordering of the set of states by inclusion. This tree has at most $n+1$ leaves and thus a total number of nodes at most $2n$.

Section 1.5

1.5.1 The states of $\mathfrak{B}$ are the sets

$$P = \{(w,q) \mid \text{ there exists a path } i \xrightarrow{u|vw} q \text{ with } i \in I\}.$$

Let (w,q) be a pair appearing in P, with a path $i \xrightarrow{u|vw} q$. If $|u| < n^2$, then $|w| < n^2M$. Otherwise, by definition of the transitions of $\mathfrak{B}$, there is another pair (w',q') with a path $i' \xrightarrow{u|vw'} q'$ such that w and w' have no common prefix. Since $|u| \geq n^2$, there are two decompositions

$$i \xrightarrow{u_1|v_1} p \xrightarrow{u_2|v_2} p \xrightarrow{u_3|v_3} q$$

and

$$i' \xrightarrow{u_1|v_1'} p' \xrightarrow{u_2|v_2'} p' \xrightarrow{u_3|v_3'} q'$$

with $vw = v_1v_2v_3, vw' = v_1'v_2'v_3'$ in which we may assume $|u_2u_3| \leq n^2$. The twinning property implies that v_1v_2 and $v_1'v_2'$ are prefix one of the other. This implies that w is a suffix of v_2v_3 and thus $|w| \leq n^2M$.

Section 1.7

1.7.1 The form of ${(I-Mz)^{-1}}_{1,1}$ results easily from the formula for the star of a matrix. If we multiply both sides of the formula $I+(I-Mz)^{-1}Mz = (I-Mz)^{-1}$ by $(\sigma - z)$ and take the value of the any row for $z = \sigma$, we obtain $vM\sigma = v$.

Section 1.8

1.8.1 Let $f_\ell : (\mathcal{A}^\ell)^* \to (\mathcal{A}^l)^*$ be the morphism defined as follows. For $x = a_1 \cdots a_\ell \in \mathcal{A}^\ell$, let $f(x) = b_1b_2 \cdots b_n$ and let $m = |f(a_1)|$. Then $f_l(x) = y_1y_2 \cdots y_m$ with $y_j = b_jb_{j+1} \cdots b_{j+l}$. The matrix $M^{(\ell)}$ is defined by $M^{(\ell)}_{xy} = |f_\ell(x)|_y$.

The entry (ab, y) of both sides of the equality $UM^{(\ell)} = M^{(2)}U$ is the number of occurrences of y in $f^{p+1}(ab)$ that begin in the prefix $f^p(a)$. The other assertion follows from the fact that if $v_2M = \rho M$, then

$$v_\ell M^{(\ell)} = v_2UM^{(\ell)} = v_2M^{(2)}U = \rho v_2 = v_\ell.$$

1.8.2 This follows from the previous problem since the vector $v_5 = v_2U$ has all entries equal to 1. Note that we can also use $p = 2$ to define U. Indeed, $|f^2(a)| = 4 > \ell - 2 = 3$. The assertion on the frequencies of the factors of length 5 follows. It can also be obtained using the function π of example 1.8.4.

1.8.3 Let $T(w) = b_1b_2 \cdots b_n$. We may assume (up to conjugacy) that w is the first row of the array. Let $z = c_1c_2 \cdots c_n$ be nondecreasing rearrangment of w (which is also the first column of the array). The first symbol of w is clearly c_1. Let j be the smallest index such that $c_1 = b_j$. then we have $a_2 = c_j$. More generally, we have $a_i = c_{\pi^{i-1}}(1)$ where the permutation π is defined as follows. For each index i, we define $\pi(i)$ as the least integer $i \geq 1$ such that $c_i = b_j$ and such that the numbers of symbols equal to c_i in $c_1 \cdots c_i$ and $b_1 \cdots b_j$ are equal.

1.8.4 This results simply from the fact that

$$S(z) = \sum_{w \in \mathcal{S}} (z/q)^{|w|}$$

1.8.5 This results from the previous problem. Indeed, we have $u_{\mathcal{F}}(z) = F(2z) = (1 - C(2z))/(1 - 2z)$ whence the desired formula.