Option Informatique en Spé MP et MP*
Palindromes : le corrigé

Première approche

Question 1 • Si $n = 2p$, alors $p(n) = |\Sigma|^p$; si $n = 2p+1$, alors $p(n) = |\Sigma|^{p+1}$. Nous pouvons résumer par la formule unique $p(n) = |\Sigma|^{\lceil n/2 \rceil}$.

Question 2 • Raisonnons par l'absurde, en utilisant le lemme de l'étoile. Supposons Pal rationnel ; il existe alors une constante N telle que tout mot $u \in \mathsf{Pal}$ de longueur au moins N se décompose en $u = xyz$, avec $|xy| \leqslant N$, $y \neq \varepsilon$ et $xy^kz \in \mathsf{Pal}$ quel que soit k. Appliquons ceci au mot a^Nba^N, qui est manifestement palindrome. Comme $|xy| \leqslant |a^N|$, nous avons certainement $x = a^p$, $y = a^q$ et $z = a^{N-p-q}ba^N$ avec $q \neq 0$. Considérons alors le mot $xz = a^{N-p}ba^N$; il devrait appartenir à Pal, or il est clair qu'il n'est pas palindrome.

Question 3 • Notons $n = |s|$. Il suffit de vérifier que $s_i = s_{n-1-i}$ pour chaque i vérifiant $0 \leqslant i < n-1-i$, soit $i \in [\![0, \lfloor (n-2)/2 \rfloor]\!]$.

```
let est_palindrome s =
 let n = string_length s in
 try
  for i = 0 to (n - 2)/2 do
   if s.[i] <> s.[n-1-i] then failwith ""
  done; true
 with _ -> false ;;
```

Question 4 • Le lemme de l'étoile énonce une condition *nécessaire* pour qu'un langage soit rationnel. Il ne peut donc pas être utilisé pour montrer qu'un langage donné est rationnel.

Question 5 • La réponse est négative. Soit $N = 4$. Considérons un mot vv^Rw de $\mathcal{Z}$, de longueur au moins égale à N. Comme $v \neq \varepsilon$, nous pouvons écrire $v = au$ où a est une lettre ; deux cas de figure se présentent, selon que u est ou non égal au mot vide.

• Si $u \neq \varepsilon$, prenons $x = \varepsilon$, $y = a$ et $z = uu^Raw$; nous avons bien $|xy| = 1 \leqslant N$, $y \neq \varepsilon$ et $xy^nz \in \mathcal{Z}$ pour tout $n \in \mathbb{N}$ puisque $xy^0z = uu^Raw$ (avec $u \neq \varepsilon$ et $aw \neq \varepsilon$) et $xy^nz = a^nuu^Raw = aa^Ra^{n-2}uu^Raw$ si $n \geqslant 2$ (avec $a \neq \varepsilon$ et $a^{n-2}uu^Raw \neq \varepsilon$).

• Si $u = \varepsilon$, prenons $x = a^2$, $y = b$ et $z = w'$ où b est la première lettre de w et w' est tel que $w = bw'$; Remarquons que $|z| \geqslant 1$, donc $z \neq \varepsilon$. Nous avons bien $|xy| = 3 \leqslant N$, $y \neq \varepsilon$, et $xy^nz \in \mathcal{Z}$ pour tout $n \in \mathbb{N}$ puisque $xy^nz = aa^Rb^nz$ avec $a \neq \varepsilon$ et $y^nz \neq \varepsilon$.

• Dans les deux cas, $\mathcal{Z}$ vérifie la *conclusion* du lemme de l'étoile ; nous ne pouvons donc pas espérer utiliser celui-ci pour montrer que ce langage n'est pas rationnel.

Question 6 • Nous pouvons supposer que $\mathcal{A} = (Q, \delta, i, F)$ est déterministe complet. Choisissons un palindrome u reconnu par $\mathcal{A}$, tel que sa longueur p soit minimale. Nous allons montrer que l'hypothèse $p \geqslant 2n^2$ mène à une contradiction.

• Pour $j \in [\![0, p]\!]$, notons $q_j = \delta^*\big(i, u[1..j]\big)$. Remarquons que $j \leqslant n^2$ implique $p - j \geqslant 2n^2 - j \geqslant n^2 \geqslant j$. Considérons alors l'ensemble des couples (q_j, q_{p-j}), où j décrit l'intervalle $[\![0, n^2]\!]$: d'après le principe des tiroirs, au moins deux de ces couples sont égaux ; notons donc j et k deux indices tels que $0 \leqslant j < k \leqslant n^2$, $q_j = q_k$ et $q_{p-j} = q_{p-k}$. Nous aurons $n^2 \leqslant p - k < p - j \leqslant 2n^2$. Notons $x = u[1..j]$, $y = u[j+1..k]$ et $z = u[k+1..p-k]$; remarquons que z est palindrome. Comme u est palindrome, $x^R = u[p-j+1..p]$ et $y^R = u[p-k+1..p-j]$; ainsi, u est l'étiquette du calcul réussi suivant de $\mathcal{A}$:

$$i \xrightarrow[*]{x} q_j \xrightarrow[*]{y} q_k \xrightarrow[*]{z} q_{p-k} \xrightarrow[*]{y^R} q_{p-j} \xrightarrow[*]{x^R} q \in F$$

Observons le mot $v = xzx^R$: il est plus court que u, car $|y| = k - j > 0$; il est palindrome, puisque z l'est ; enfin, v est l'étiquette du calcul réussi de $\mathcal{A}$ suivant :

$$i \xrightarrow[*]{x} q_j = q_k \xrightarrow[*]{z} q_{p-k} = q_{p-j} \xrightarrow[*]{x^R} q \in F$$

D'où la contradiction. Fin de la preuve.

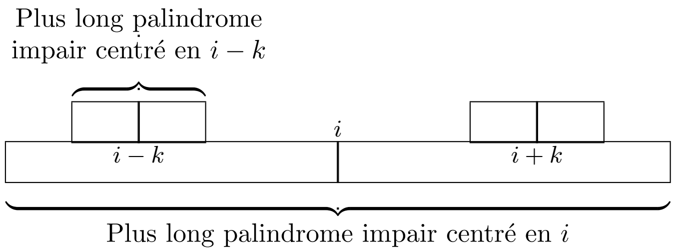
L'algorithme de Manacher

Question 7 • Notons $n = |u|$. Il suffit, pour chaque $i \in [\![1, |u|]\!]$, de tester si $u[1..i]$ est palindrome ; ceci peut se faire pour un coût $\sum_{1 \leqslant i \leqslant n} \lfloor i/2 \rfloor$, et vous vérifierez que cette somme vaut $\lfloor n^2/4 \rfloor$.

Question 8 • Facile à écrire, avec les indispensables `intervalle` et `filtre`.

```
let préfixes_palindromes s =
 let l = intervalle 0 (string_length s - 1) in
 let f k = est_palindrome (sub_string s 0 k)
 in filtre f l ;;
```

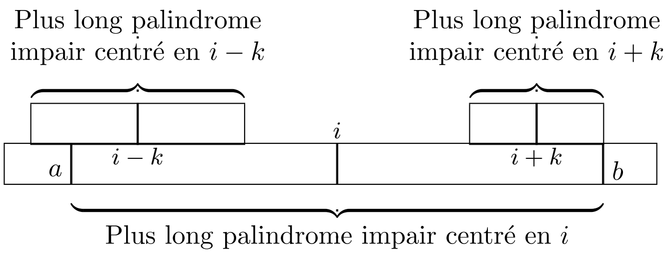
Question 9 • Nous avons $\mathrm{Ri}(i-k) < \mathrm{Ri}(i) - k$. Le plus long palindrome impair centré en $i-k$ est strictement recouvert par le plus long palindrome impair centré en i ; par raison de symétrie, le plus long palindrome impair centré en $i+k$ a même rayon d'où $\mathrm{Ri}(i+k) = \mathrm{Ri}(i-k)$.



Question 10 • Nous avons maintenant $\mathrm{Ri}(i-k) > \mathrm{Ri}(i) - k$. Par raison de symétrie, il y a un palindrome impair de rayon $\mathrm{Ri}(i) - k$ centré en $i+k$; et ce rayon est maximal, car :

- ou bien ce palindrome est un suffixe du mot considéré ;
- ou bien la lettre b qui le suit est distincte de la lettre a qui précède le palindrome impair de rayon $\mathrm{Ri}(i)$ centré en i.

D'où $\mathrm{Ri}(i+k) = \mathrm{Ri}(i) - k$.



Question 11 • Il suffit de montrer que $u_{i-k} = u_{i-k+2}$ pour $2 \leqslant k \leqslant r$. Or $u[i-r..i+r-2]$ est un palindrome centré en $i-1$; donc $u_{i-k} = u_{(i-1)-(k-1)} = u_{(i-1)+(k-1)} = u_{i+k-2}$. Mais $u[i-r..i+r]$ est un palindrome centré en i ; donc $u_{i+k-2} = u_{i-(k-2)} = u_{i-k+2}$. Par transitivité, $u_{i-k} = u_{i-k+2}$. Nous avons ici une version «discrète» d'un résultat élémentaire de géométrie : la composée de deux symétries centrales est une translation.

Question 12 • Notons $x = u_{i+r+1}$. Nous avons $u[i-r..i+r+1] = a^{2r+1}x$. Pour $1 \leqslant k \leqslant r$, nous aurons $u[i+2k-r..i+r+1] = a^{2r-2k}x$ avec $x \neq a$, et par suite $\mathrm{Ri}(i+k) = r-k$.

Question 13 • Notons q le plus grand indice tel que $q \leqslant |u|$ et $u_q = a$. Remarquons que $u_{i-r-1} \neq a$, de par la définition de r. Alors $u[i-r..q] = a^{q+i-r}$. Soit $j \in [\![i+1, q]\!]$: à gauche de j, nous trouvons $j-i+r$ occurrences de a ; à droite, nous en trouvons $q-j$. Donc $\mathrm{Ri}(k) = \max(j-i+r, q-j)$. Remarquons que les cases d'indice $j \in [\![i+1, i+r]\!]$ sont remplies sans frais ; chaque case d'indice $j \in [\![i+r+1, q]\!]$ coûte une comparaison de caractères.

Question 14 • Par observation directe, nous déterminons le plus grand indice $q \in [\![i+1, |u|]\!]$ tel que $u[i-r..q]$ soit préfixe de $(ab)^{|u|}$. Le coût est le même, à savoir $q+1-(i+r)$ comparaisons. Nous pouvons ensuite remplir les cases d'indice $j \in [\![i+1, q]\!]$.

Question 15 • Nous avons déjà $\mathrm{Ri}(1) = 0$.

• Par observation directe, nous obtenons $\mathrm{Ri}(2) = 1$; fixons $i = 2$. Pour $k = 1$, nous constatons que $\mathrm{Ri}(i - k) = \mathrm{Ri}(i) - k = 0$: nous ne pouvons donc rien déduire des informations obtenues jusqu'ici.

• Par observation directe, nous obtenons $\mathrm{Ri}(3) = 0$ puis $\mathrm{Ri}(4) = 1$. Fixons alors $i = 4$; en prenant $k = 1$, nous avons $\mathrm{Ri}(i - k) = \mathrm{Ri}(3) = 0$ et $\mathrm{Ri}(i) - k = 0$: il n'y a rien à exploiter.

• Par observation directe, nous obtenons $\mathrm{Ri}(5) = 4$: ici, il y a espoir de faire de bonnes affaires. Fixons $i = 5$; pour $k = 1$, il vient $\mathrm{Ri}(i - k) = \mathrm{Ri}(4) = 1$ et $\mathrm{Ri}(i) - k = 3$, donc $\mathrm{Ri}(6) = \mathrm{Ri}(i + k) = 1$; pour $k = 2$, il vient $\mathrm{Ri}(i - k) = \mathrm{Ri}(3) = 0$ et $\mathrm{Ri}(i) - k = 2$, donc $\mathrm{Ri}(7) = \mathrm{Ri}(i + k) = 0$; enfin, pour $k = 3$, il vient $\mathrm{Ri}(i - k) = \mathrm{Ri}(2) = 1$ et $\mathrm{Ri}(i) - k = 1$, donc $\mathrm{Ri}(8) = \mathrm{Ri}(i + k) \geqslant 1$: l'observation directe donne $\mathrm{Ri}(8) = 2$.

• Par observation directe, nous obtenons $\mathrm{Ri}(9) = 0$ et $\mathrm{Ri}(10) = 0$.

Question 16 • Le calcul des deux tables Ri et Rp se fait en temps linéaire. Les préfixes palindromes impairs de u sont *centrés* sur les positions i telles que $\mathrm{Ri}[i] = i - 1$; les préfixes palindromes pairs de u sont *calés* sur les positions i telles que $\mathrm{Rp}[i] = i - 1$. Ceci peut se tester par un simple parcours de ces tables, donc en temps linéaire.

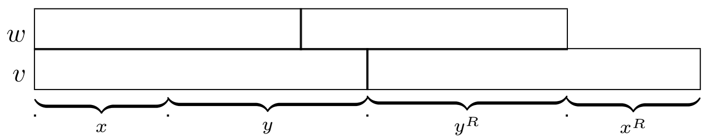
Question 17 • Nous ne traitons que le cas où u est de longueur paire : $|u| = 2p$, avec $p \geqslant 2$. Commençons par construire la table Ri, et voyons si u est le produit de deux palindromes impairs : ceci revient à dire qu'il existe $i \in [\![2, p - 1]\!]$ tel que $\mathrm{Ri}[i] = i - 1$ et $\mathrm{Ri}[i + p] = p - i$. Si ce test échoue, nous regardons si u est le produit de deux palindromes pairs : ceci revient à dire qu'il existe $i \in [\![2, p - 1]\!]$ tel que $\mathrm{Rp}[i] = i$ et $\mathrm{Rp}[i + p] = p - i$. Chacune de ces étapes peut se faire en temps linéaire.

Factorisation en produit de palindromes pairs

Question 18 • Le cas $n = 2p$ est immédiat : alors $v = ww^R = ww$ est le produit de deux palindromes pairs, ce qui contredit le caractère minimal de v.

• Examinons maintenant le cas où $n > 2p$. Notons $y = v[2p + 1..n]$; comme v est palindrome, nous aurons $v[n + 1..2n - 2p] = y^R$ et $v[n + 2p + 1..2n] = w^R = w$. Finalement, $v = w(yy^R)w$ est le produit de trois palindromes pairs : ceci contredit le caractère minimal de v.

Question 19 • Si $p < n < 2p$, alors le mot w «déborde» au-delà de la moitié de v. Notons $x = v[1..2n - 2p]$ et $y = v[2n - 2p + 1..n]$; comme v est palindrome, nous aurons $v[n + 1..2p] = y^R$. Le «débordement» est $v[2p + 1..2n] = x^R$. Alors $w = w[1..2n - 2p]w[2n - 2p + 1..n]w[n + 1..2p] = xyy^R$; mais w est palindrome, donc $w = yy^Rx^R$: yy^R est un préfixe palindrome pair de v, plus court que w : ceci contredit le caractère minimal de ce dernier.



Question 20 • Concluons : $n = p$, donc $w = v$. Par conséquent, le plus court facteur palindrome pair d'un palstar u est le premier facteur d'au moins une décomposition de u en produit de palindromes pairs.

Question 21 • L'algorithme suit la démarche «gloutonne» :

- si $u = \varepsilon$, c'est fini (fin de la décomposition) ;
- sinon : si u ne possède aucun préfixe palindrome pair v, c'est fini (échec) ;
- sinon : notons v le plus court préfixe palindrome pair de u ; u est un palstar pair ssi le mot w défini par $u = vw$ est lui aussi un palstar pair : il suffit d'appliquer l'algorithme à w. L'inégalité $|w| < |u|$ assure la terminaison de cet algorithme.

Factorisation en produit de palindromes quelconques

Question 22 • L'algorithme trouve *aba* comme plus court préfixe palindrome, puis échoue face au mot *babb* ; or $ababb = (abab a)(bb)$ est un palstar.

Question 23 • L'algorithme trouve *aa* comme plus court préfixe palindrome, puis échoue face au mot *baaaa* ; or $aabaaaa = (aabaa)(aa)$ est un palstar.

Question 24 • Le cas $n = 2p$ a déjà vu à la question **??**.

Question 25 • Cas $n > 2p+1$. Come v et w sont palindromes, nous avons $v[n-p+1..n] = v^R = w$. Notons $z = v[p+1..n-p]$, ainsi $v = wzw$; il est clair que z est palindrome, non banal car $|z| \geqslant 2$. Ainsi v est le produit de trois palindromes, ce qui contredit son caractère minimal.

Question 26 • Cas $p < n < 2p-1$. Supposons n pair : $n = 2q$; alors $2q < 2p-1$, donc $q < p$. Notons $x = v[1..n-p]$ et $y = v[n-p+1..q]$. Comme v est un palindrome pair calé sur q, nous aurons $v[q+1..p] = y^R$. Alors $w = v[1..p] = xyy^R$ puis comme w est palindrome : $w = w^R = yy^Rx^R$. Le mot yy^R est un préfixe palindrome de u, et il est plus court que w puisque $|x| = n-p > 0$. Ceci contredit le caractère minimal de w.
• Supposons maintenant n impair : $n = 2q+1$; alors $2q+1 < 2p-1$, donc $q < p-1$. Notons $x = v[1..n-p]$ et $y = v[n-p+1..q]$. Comme v est un palindrome impair centré en $q+1$, nous aurons $v[q+2..p] = y^R$, d'où en notant $a = w_{q+1}$: $w = v[1..p] = xyay^R$ puis comme w est palindrome : $w = w^R = yay^Rx^R$. La fin du raisonnement est la même que pour le cas n pair.

Question 27 • $\pi(i)$ est non nul lorsque l'une au moins des conditions suivantes est satisfaite :

- il existe un palindrome pair calé en $j \geqslant i$ et de rayon $\mathrm{Rp}(j) \geqslant j-i+1$; alors $u[i..2j-i-1]$ est palindrome ;
- il existe un palindrome impair centré en $j > i$ et de rayon $\mathrm{Ri}(j) \geqslant j-i$; alors $u[i..2j-i]$ est palindrome.

L'algorithme gérera une pile d'indices en attente ; initialement, cette pile est vide. Le tableau π est initialisé à 0. Le mot u est lu de gauche à droite. Pour chaque position j, de 1 à ℓ :

- si $\mathrm{Rp}(j) = \mathrm{Ri}(j) = 0$, l'algorithme ne fait rien ;
- sinon, il extrait de la pile chaque i vérifiant $\mathrm{Rp}(j) \geqslant j-i+1$ ou $\mathrm{Ri}(j) \geqslant j-i$ et affecte à $\pi(i)$ la «bonne» valeur ; plus précisément : si un seul des tests est positif, c'est la valeur correspondante qui est attribuée à $\pi(i)$; si les deux tests sont positifs, c'est la plus petite des deux valeurs ;
- ensuite, j est mis sur la pile.

Le coût est clairement un $\mathcal{O}(\ell)$.

Question 28 • Nous allons remplir le tableau pal de droite à gauche. Initialisation : $\mathrm{pal}(\ell) \leftarrow$ FAUX et $\mathrm{pal}(\ell+1) \leftarrow$ VRAI. Itération : supposons $\mathrm{pal}(j)$ calculé pour $i < j \leqslant \ell$. Notons $k = \pi(i)$; si $k = 0$, alors $\mathrm{pal}(i) \leftarrow$ FAUX. Sinon, $\mathrm{pal}(i) \leftarrow$ VRAI ssi nous sommes dans un des trois cas suivants :

- $\mathrm{pal}(i+k) =$ VRAI ;
- $u[i..i+2k-2]$ est palindrome et $\mathrm{pal}(i+2k-1) =$ VRAI ;
- $u[i..i+2k]$ est palindrome et $\mathrm{pal}(i+2k+1) =$ VRAI.

Complément de programme

Question 29 • Preuve : seul le sens direct demande une preuve. Supposons u non équilibré, et considérons deux facteurs de u, de même longueur, en déséquilibre ; nous dirons qu'ils forment une *paire critique*. S'ils sont de longueur 2, ce sont 00 et 11 si bien que $x = \varepsilon$ convient. Sinon, supposons-les de longueur minimale. Nous pouvons écrire ces deux facteurs $0x0$ et $1y1$ respectivement, avec $|x| = |y| \geqslant 1$. Supposons x et y distincts, et notons z le plus long préfixe commun à x et y. Deux cas se présentent : $x = z0x'$ et $y = z1y'$, alors $0x'0$ et $1y'1$ forment une paire critique ; si $x = z1x'$ et $y = z0y'$, ce sont $x'0$ et $y'1$ qui forment une paire critique ; dans chaque cas, nous mettons en évidence une contradiction, et donc $x = y$. Supposons x non palindrome, nous pouvons l'écrire $x = ztz^R$, avec z de longueur maximale pour cette propriété : donc $|t| \geqslant 1$, et $t_1 \neq t_{|t|}$. Distinguons à nouveau deux cas : si $t = 0t'1$, ce sont les facteurs $0z0$ et $1z^R1$ des mots $0z0t'1z^R0$ et $1z0t'1z^R1$ qui forment une paire critique ; si $t = 1t'0$, ce sont les facteurs $1z1$ et $0z^R0$ des mots $0z1t'0z^R0$ et $1z1t'0z^R1$ qui forment une paire critique ; dans chaque cas, nous mettons en évidence une contradiction, et donc x est palindrome.

Références bibliographiques

▶ Source principale d'inspiration : le livre *Jewels of Stringology*, de Crochemore et Rytter.
▶ L'algorithme de Manacher a été publié dans le JACM en 1975.
▶ L'étude des palstars a été publiée par Galil et Seiferas dans le JACM en 1978.
▶ Le résultat de la question 29 est classique.

FIN