## 2.10 Problème de correspondance de Post

Le problème de correspondance de Post (PCP) :

**Donnée :** Deux suites de mots finies $(u_1, \ldots, u_n)$ et $(v_1, \ldots, v_n)$ de même longueur

**Question :** existe-t-il un entier $k$ et une suite d'indices $i_1, \ldots, i_k$ tels que

$$u_{i_1} \cdots u_{i_k} = v_{i_1} \cdots v_{i_k}$$

Par exemple : soient

| $i$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $u_i$ | $a$ | $b$ | $ca$ | $abc$ |
| $v_i$ | $ab$ | $ca$ | $a$ | $c$ |

Cette instance de PCP a une solution (12314).

**Théorème 2.14** *PCP est indécidable.*

On commence par montrer que le problème de correspondance de Post *modifié*, dans lequel on fixe $i_1 = 1$ est indécidable.

On réduit le problème de l'arrêt.

Soient $M$ une machine de Turing et $w \in \Sigma^*$. On note $q_f$ le (seul état d'arrêt de $M$)

Les mots de PCP modifié :

| | mots $v_i$ (dans l'ordre) | mots $u_i$ correspondant | |
|---|---|---|---|
| 1. | $\lhd$ | $\lhd\, q_0\$w\star$ | |
| 2. | $a$ | $a$ | pour $a \in \Sigma$ |
| 3. | $qa$ | $a'q'$ | Si $\delta(q,a) = q', a', \rightarrow$ |
| 4. | $aqb$ | $q'ab'$ | si $\delta(q,b) = q', b', \leftarrow$ |
| 5. | $qa$ | $q'a'$ | si $\delta(q,a) = q', a', \downarrow$ |
| 6. | $q\star$ | $aq'\star$ | si $\delta(q,B) = q', a', \rightarrow$ |
| 7. | $bq\star$ | $q'ba'\star$ | si $\delta(q,B) = q', a', \leftarrow$ |
| 8. | $q\star$ | $q'a\star$ | si $\delta(q,B) = q', a', \downarrow$ |
| 9. | $\$q_e\star \rhd$ | $\rhd$ | |
| 10. | $\star$ | $\star$ | |
| 11. | $aq_e\star$ | $q_e\star$ | |
| 12. | $q_f a$ | $aq_f$ | |
| 13. | $q_f\star$ | $q_e\star$ | |

Montrons que PCP a une solution ssi $M$ s'arrête sur $w$ :

**Si $M$ s'arrête sur $w$** Soit

$$s_1 = q_0\$w \vdash \cdots \vdash w_k q_f w'_k = s_k$$

le calcul de $M$ sur $w$ (les blancs en fin de ruban n'apparaissent pas dans les configurations). Pour tout $i < k$, $s_i, s_{i+1}$ sont de l'une des formes suivantes :

- $s_i = t_i q a w_i$ et $s_{i+1} = t_i a' q' w_i$ avec $\delta(q,a) = q', a', \rightarrow$
- $s_i = t_i a q b w_i$ et $s_{i+1} = t_i q' a b' w_i$ avec $\delta(q,b) = q', b', \leftarrow$
- $s_i = t_i q a w_i$ et $s_{i+1} = t_i q' a' w_i$ avec $\delta(q,a) = q', a', \downarrow$
- L'un des cas ci-dessus en retirant la partie à droite de $q$ (tête de lecture à droite du ruban)...

Dans chaque cas, on remarque que $s_i\star$ peut s'écrire comme la concaténation $v_{m_1}\cdots v_{m_i}$ et $s_{i+1}\star$ s'écrit comme la concaténation $u_{m_1}\cdots u_{m_i}$. Par exemple, dans le premier cas : On utilise $|t_i|$ fois la correspondance 2, puis une fois la correspondance 3, puis $|w_i|$ fois la correspondance 2, puis la correspondance 10.

On obtient alors une solution de PCP modifié comme suit :

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2*3 ... | $\cdots$ | | | | | | | |
| $\lhd\ q_0\$w\star$ | $s_1\star$ | $\cdots$ | $s_k\star$ | $w_k a q_f w'_k\star$ | $\cdots$ | $w_n q_e\star$ | $w_{n-1}q_e\star$ | $\cdots$ | $\rhd$ |
| $\lhd$ | $s_0\star$ | $\cdots$ | $s_{k-1}\star$ | $w_k q_f a w'_k\star$ | $\cdots$ | $w_n q_f\star$ | $w_{n-1}aq_e\star$ | $\cdots$ | $\$q_e\star\rhd$ |

**Réciproquement, si PCP modifié a une solution** Alors montrons que $M$ s'arrête sur $w$.

Soit $u_{i_1}\cdots u_{i_m} = v_{i_1}\cdots v_{i_m}$. On montre, par récurrence sur $k$ qu'il existe un $p$, et des mots $t, t'$ tels que $u_{i_1}\cdots u_{i_k} = \lhd\ s_1\star\cdots s_p\star t$, $v_{i_1}\cdots v_{i_k} = \lhd\ s_1\star\cdots s_{p-1}\star t'$ et ou bien $s_p$ est une configuration finale de la machine, ou bien
- $t'$ est un préfixe de $s_p$ et $t$ est un préfixe de $s_{p+1}$
- Si $t'$ ne contient pas de symbole d'état et ne se termine pas par $\star$, alors $t' = t$, sinon $t' \vdash_M t$.

Pour $k = 1$, $u_{i_1} = u_1 = \lhd\ q_0\$w\star$ et $v_{i_1} = v_1 = \lhd$. On a bien, pour $p = 1$, $u_{i_1} = \lhd\ s_1\star$, $t = t' = \epsilon$.

Si la propriété est vraie pour $k$, considérons la paire suivante $u_{i_{k+1}}, v_{i_{k+1}}$. Si $s_p$ était déjà une configuration finale, il n'y a rien à faire. Sinon, $s_p = t'\cdot t'_p$ et $s_{p+1} = t\cdot t_p$. De plus, comme on a une solution de PCP modifié, $v_{i_{k+1}}$ est un préfixe de $t'_p\star t u_{i_{k+1}}$. Considérons successivement tous les $v_{i_{k+1}}$ possibles :
- 1. est impossible car $\lhd$ n'apparait pas dans $t'_p\star t u_{i_{k+1}}$
- 2. Dans ce cas, $t'_p = v_{i_{k+1}}t''_p$ et on a la propriété voulue
- 3, 4,5 : $t'_p = v_{i_{k+1}}t''_p$, $t'$ ne contient pas de symbole d'état, et donc $t = t'$ et $t'\cdot v_{i_{k+1}} \vdash_M t\cdot u_{i_{k+1}}$
- 6,7,8 : comme $t'_p$ ne contient pas $\star$, $t'_p\star = v_{i_{k+1}}$. Donc $s_p\star = t'\cdot v_{i_{k+1}}$. De plus, $t = t'$ et $t\cdot u_{i_{k+1}} = s_{p+1}\star$. Il suffit alors de choisir des mots vides pour les nouveaux $t, t'$
- 9 : $t'$ doit être vide puisque $v_{i_{k+1}}$ est un préfixe de $t'_p\star t u_{i_{k+1}}$.De plus, on doit avoir $t'_p = \$q_e$, ce qui n'est pas possible. Ce cas n'a pas lieu
- 10. $t'_p$ est vide et il suffit de choisir des mots vides pour les nouveaux $t, t'$
- 11. Impossible pour des raisons identiques à 9.
- 12,13 : l'état final a été atteint.

Maintenant, reste à montrer que PCP lui-même est indécidable. Pour cela on réduit PCP modifié à PCP comme suit, en supposant (sans perte de généralité) qu'il n'y a pas de paire $(\epsilon, \epsilon)$.

Si $(u_1, \ldots, u_n)$, $(v_1, \ldots, v_n)$ est une instance de PCP modifié, on considère un alphabet augmenté des lettres $\bullet, \rhd, \lhd$ et l'instance de PCP : $(\lhd\ \overline{u_1}, \overline{u_1}, \ldots, \overline{u_n}, \bullet\ \rhd), (\lhd\ \bullet\widetilde{v_1}, \widetilde{v_1}, \ldots, \widetilde{v_n}, \rhd$ $)$ où $\overline{\epsilon} = \widetilde{\epsilon} = \epsilon$ et $\overline{a\cdot w} = \bullet a\overline{w}$ et $\widetilde{a\cdot w} = a\bullet\widetilde{w}$.

Si PCP modifié a une solution $u_{i_1}\cdots u_{i_m} = v_{i_1}\cdots v_{i_m}$, alors $\lhd\ \overline{u_{i_1}}\cdots\overline{u_{i_m}}\bullet\ \rhd = \lhd\ \bullet\widetilde{v_{i_1}}\cdots\widetilde{v_{i_m}}\ \rhd$ est une solution de PCP.

Réciproquement, si PCP admet une solution, notons $(u'_0, \ldots, u'_n, u'_{n+1})$ et $(v'_0, \ldots, v'_{n+1})$ l'instance du problème : il existe une suite d'indices telle que $u'_{i_1}\cdots u'_{i_m} = v'_{i_1}\cdots v'_{i_m}$. Notons que, pour tout $i$, $u'_i = \epsilon$ ou bien $u'_i$ commence par $\bullet$, ou bien $i = 1$. Si $u'_{i_1} = \epsilon$, alors soit $k$ le plus petit indice tel que $u'_{i_k} \neq \epsilon$. Par hypothèse et par construction, $v'_{i_1} \neq \epsilon$ et sa première lettre est $a \notin \{\bullet, \lhd\}$. À l'inverse, la première lettre de $u'_{i_k}$ est dans $\{\bullet, \lhd\}$. Ce qui est absurde. Il en résulte que $u'_{i_1} \neq \epsilon$. Dans ce cas, la première lettre de $u'_{i_1}$ est dans $\{\bullet, \lhd\}$ et donc aussi la première lettre du premier $v_{i_k}$ non vide. Ce n'est possible que si $i_1 = 1$ et cette première lettre est $\lhd$.

Soit maintenant $\phi$ le morphisme défini sur $(\Sigma \cup \{\bullet, \lhd, \rhd\})^*$ par $\phi(\bullet) = \phi(\lhd) = \phi(\rhd) = \epsilon$ et $\phi(a) = a$ sinon. On montre, par récurrence sur $k$ que $\phi(u'_{i_1} \cdots u'_{i_k}) = u_1 \cdot u_{i_2} \cdots u_{i_k}$ et $\phi(v_{i_1} \cdots v_{i_k}) = v_1 \cdot v_{i_2} \cdots v_{i_k}$, si $1 \leq k < m$. Il en résulte que, si PCP a une solution, alors PCP modifié aussi, en prenant l'image par $\phi$.

## 2.11 Problèmes de pavages

Le problème de pavage de $\mathbb{N} \times \mathbb{N}$ (les définitions sont semblables pour d'autres sous-ensembles de $\mathbb{Z} \times \mathbb{Z}$) est défini par :

**Donnée** : un ensemble fini $T = \{t_0, \ldots, t_k\}$ de *tuiles* et deux sous-ensembles $H, V$ de $T \times T$. (Les relations de compatibilité horizontale et verticale).

**Question** : existe-t-il une fonction de pavage $f : \mathbb{N} \times \mathbb{N} \mapsto T$ telle que $f(1,1) = t_0$ et, pour tout $i, j \in \mathbb{N}$, $(f(i,j), f(i+1,j)) \in H$, $(f(i,j), f(i,j+1)) \in V$.

**Théorème 2.15** *Le problème de savoir, étant donnés $T, V, H$ si $\mathbb{N} \times \mathbb{N}$ est pavable, est indécidable.*

On réduit le problème de l'arrêt. Soit $M, w$ les données du problème de l'arrêt et $k = \max(1, |w|)$. On suppose sans perte de généralité que la machine ne rentre jamais dans l'état initial (autrement dit, seul le premier mouvement utilise l'état initial) et qu'elle n'écrit jamais de blanc.

On choisit pour $T = (\Sigma \cup Q)^{k+3} \cap \Sigma^*(Q + \epsilon)\Sigma^*$ (autrement dit, les mots de longueur $k+3$ contenant au plus un symbole de $Q$).

$$H = \{(atc, tcb) \mid atc, tcb \in T, a, c, b \in \Sigma \cup Q, c = B \Rightarrow b = B\}$$

$$\begin{array}{rlllll}
V = & \{ & (\alpha q a \beta, & \alpha a' q' \beta) & \mid \alpha\beta \in \Sigma^{k+1},\ a \in \Sigma,\ q \in Q,\ \delta(q,a) = (q', a', \rightarrow) & \} \\
\cup & \{ & (\alpha q, & \alpha a') & \mid \alpha \in \Sigma^{k+2},\ a' \in \Sigma,\ q \in Q,\ \delta(q,a) = (q', a', \rightarrow) & \} \\
\cup & \{ & (a\alpha, & q'\alpha) & \mid \alpha \in \Sigma^{k+2},\ q \in Q,\ \delta(q,a) = (q', a', \rightarrow) & \} \\
\cup & \{ & (\alpha b q a \beta, & \alpha q b a' \beta) & \mid \alpha\beta \in \Sigma^{k},\ a, b \in \Sigma,\ q \in Q,\ \delta(q,a) = (q', a', \leftarrow) & \} \\
\cup & \{ & (q a \beta, & b a' \beta) & \mid \beta \in \Sigma^{k+1},\ a, b \in \Sigma,\ q \in Q,\ \delta(q,a) = (q', a', \leftarrow) & \} \\
\cup & \{ & (\beta a, & \beta q') & \mid \beta \in \Sigma^{k+2},\ a \in \Sigma,\ \delta(q,a) = (q', a', \leftarrow) & \} \\
\cup & \{ & (\alpha q a \beta, & \alpha q' a' \beta) & \mid \alpha\beta \in \Sigma^{k+1},\ a \in \Sigma,\ q \in Q,\ \delta(q,a) = (q', a', \downarrow) & \} \\
\cup & \{ & (\alpha q, & \alpha q') & \mid \alpha \in \Sigma^{k+2},\ q \in Q,\ \delta(q,a) = (q', a', \downarrow) & \} \\
\cup & \{ & (\alpha, & \alpha) & \mid \alpha \in \Sigma^{k+3} & \}
\end{array}$$

Enfin, on demande que le pavage satisfasse $f(1,1) = q_0\$wB$.

Un pavage des $n$ premières lignes horizontales du quart de plan supérieur correspond alors à une succession de $n$ configurations de la machine de Turing, à partir de la configuration initiale. Plus précisément, si l'on ne retient que la première lettre de chaque tuile de la ligne $n$, on obtient la $n$ième configuration de $M$.

On montre ce résultat par récurrence sur $n$ :

– pour $n = 1$, la première ligne contient la configuration initiale ; $f(1, 1) = q_0\$wB$ et donc $f(1, 2) = \$wBB$. Par récurrence sur $m \geq 1$, $f(1, m + 1) = w_2B^{m+1}$ où $w_2$ est un suffixe de $\$wB$ de longueur $k + 2 - m$. En effet, étant donné $f(1, m) = w_2B^{m+1}$ avec $m \geq 1$, une seule tuile a pour membre gauche $f(1, m)$ : $(w_2B^{m+1}, w_3B^{m+2})$ où $w_2 = aw_3$ avec $a \in \Sigma$.
– Si la suite des premières lettres des tuiles de la ligne $n$ est une configuration $\gamma = w_1qaw_2$ de la machine de Turing, alors la suite des première lettres des tuiles de la ligne $n + 1$ est la configuration suivante de la machine : soit $m$ la position de $q$ dans $\gamma$. Supposons d'abord que $m > 1$. Pour $m - 1 \geq p \geq \max(1, m - k - 1)$, $f(n, p) = w_1'bqaw_2'$ où $w_1'b$ est le suffixe de longueur $m - p$ de $w_1$ et $w_2'$ est le préfixe de longueur $k + p + 1 - m$ de $w_2$, éventuellement complété de blancs.
  $f(n + 1, p)$ est alors donné par :
  – $w_1'qba'w_2'$ si $\delta(q, a) = (q', a', \leftarrow)$,
  – $w_1'bqa'w_2'$ si $\delta(q, a) = (q', a', \downarrow)$,
  – $w_1'ba'q'w_2'$ si $\delta(q, a) = (q', a', \rightarrow)$.
  On conclut, grâce aux compatibilités verticales. Par exemple, $p \geq m + 2$, alors $f(n, p) \in \Sigma^{k+3}$ et $f(n + 1, p) = f(n, p)$. En effet, les seules tuiles verticales compatibles sont $(\alpha, \alpha)$ et $(\alpha b, \alpha q')$ (avec $\delta(q, a) = (q', a', \leftarrow)$). Mais on montre que ce dernier cas n'est pas possible en considérant $f(n, p+1)$ : comme $p > m+1$, $f(n, p+1) \in \Sigma^{k+3}$ et donc, par compatibilité verticale, $f(n+1, p+1) \in \Sigma^{k+3} \cup Q \cdot \Sigma^{k+2}$. Si l'on avait $f(n+1, p) = \alpha q'$, on aurait une tuile horizontale de la forme $(\alpha q', \beta q'')$ ou $(\alpha q', \beta)$, ce qui n'est pas le cas. De même, pour $p < m - k - 2$, $f(n + 1, p) = f(n, p)$. Restent les cas $p = m - k - 2$, $p = m$, $p = m + 1$ pour lesquels on montre par compatibilités horizontale et verticale que l'on obtient les lettres voulues.
  Enfin, si $m = 1$, le seul mouvement possible est à droite et la seule tuile verticale possible impose $f(n + 1, 1) = \$q'w'$ si $\delta(q, \$) = (q', \$, \rightarrow)$. Puis, pour $p > 2$, $f(n + 1, p) = f(n, p)$ comme ci-dessus et $f(n + 1, 2) = q'w'b$ par compatibilités verticale et horizontale.