Uniwersytet Warszawski
Wydział Matematyki, Informatyki i Mechaniki

Franciszek Grzegorek
Nr albumu: 181045

Własności rozwiązań równań w półgrupie wolnej z jedną i dwoma niewiadomymi

Praca magisterska
na kierunku INFORMATYKA

Praca wykonana pod kierunkiem
dra hab. Wojciecha Plandowskiego
Instytut Informatyki

maj 2003

Pracę przedkładam do oceny

Data Podpis autora pracy:

Praca jest gotowa do oceny przez recenzenta

Data Podpis kierującego pracą:

Streszczenie

Praca składa się z trzech części. W pierwszej części zajmę się równaniami z dwoma niewiadomymi o szczególnej postaci $A(X)Y = YB(X)$. Oznacza to, że lewa strona równania zawiera wystąpienia jednej niewiadomej oddzielone stałymi (dopuszcza się słowa puste) i na końcu występuje druga niewiadoma. Prawą stronę równania rozpoczyna druga niewiadoma a dalej występuje dowolna kombinacja stałych i pierwszej niewiadomej. Ponadto liczba wystąpień niewiadomych jest taka sama po lewej i prawej stronie równania. Równania w półgrupie wolnej definiują języki. Dla każdej niewiadomej jeden język. Badam, które pary złożoności języków definiowanych przez niewiadome X i Y (podane w [1]) są możliwe do uzyskania. W drugiej części pracy pokażę uogólniony dowód na ścisłość ograniczenia na długość minimalnego rozwiązania równania z jedną niewiadomą, zarówno pod względem długości równania, jak i pod względem liczby wystąpień w tym równaniu niewiadomej. W trzeciej części pracy przedstawię sposób badania w czasie liniowym, czy zbiór rozwiązań równania z jedną niewiadomą jest skończony.

Słowa kluczowe

równania w półgrupie wolnej, języki formalne, relacje, funkcje złożoności, wyrażalność, algorytmy.

Klasyfikacja tematyczna

68Q70 Algebraiczna teoria języków i automatów,
68Q45 Języki formalne i automaty,
68R15 Kombinatoryka słów,
68W40 Analiza algorytmów.

Spis treści

1. Wprowadzenie 5

2. Podstawowe definicje 7

3. Równania postaci $A(X)Y = YB(X)$ 13

4. Długość minimalnego rozwiązania równania z jedną niewiadomą . . 23

5. Badanie skończoności zbioru rozwiązań równania z jedną niewiadomą 27

Rozdział 1

Wprowadzenie

Równania w półgrupie wolnej stanowią jeden z podstawowych tematów kombinatoryki słów. Mogą być używane do opisania pewnych języków, uogólniają problem wyszukiwania wzorca z niewiadomymi. Przykładowe równanie:

$$abX = Xba$$

ma rozwiązania, między innymi, $x = a$ (ponieważ $aba = aba$), oraz $x = aba$ (ponieważ $ababa = ababa$).

Podstawowym problemem dotyczącym równań, w szczególności równań w półgrupie wolnej, jest sprawdzenie czy dane równanie ma jakiekolwiek rozwiązanie. Problem ten jest rozstrzygalny, co pokazał Makanin [2]. W pracy [1] rozpatrywane są złożoności języków definiowanych przez poszczególne niewiadome w równaniu z dwoma niewiadomymi. Przez złożoność rozumie się liczbę słów tej samej długości. W rozpatrywanym przeze mnie szczególnym przypadku $A(X)Y = YB(X)$ złożoności są następujące: stała, typu $\mathcal{D}_1$, liniowa lub wykładnicza dla niewiadomej X, stała, typu $\mathcal{D}_1$, typu $\mathcal{D}_2$, liniowa lub wykładnicza dla niewiadomej Y. Typy $\mathcal{D}_1$ i $\mathcal{D}_2$ oznaczają funkcje większe od stałej, ale mniejesze od liniowej. Są one charakteryzowane przez liczby rozwiązań pewnych równań diofantycznych

W mojej pracy zajmę się badaniem wzajemnej zależności pomiędzy złożonościami języków definiowanych przez niewiadomą X i niewiadomą Y w równaniu postaci $A(X)Y = YB(X)$. Pokażę, które kombinacje są możliwe do uzyskania oraz udowodnię, że niektórych kombinacji nie da się uzyskać żadnym równaniem tej postaci.

Wykluczenie pewnych kombinacji złożoności wprowadza nową metodę sprawdzania, czy relacja jest wyrażalna przez równania postaci $A(X)Y = YB(X)$.

K. Baba, S. Tsuruta, A. Shinohara i M. Takeda pokazali w [7] zależność długości minimalnego (najkrótszego) rozwiązania równania z jedną niewiadomą od różnic w pozycjach wystąpień kolejnych wystąpień niewiadomej po lewej i po prawej stronie równania.

W cytowanej pracy dowód optymalności ograniczenia podanego w twierdzeniu polega na wskazaniu pewnego równania realizującego optymalność ograniczenia. Pokażę ciąg równań dowodzący optymalności w bardziej ogólnym sensie, gdyż zarówno długość obydwu stron równania, jak i liczba wystąpień niewiadomej, jak i długość minimalnego rozwiązania będą dowolnie duże.

Pokażę granicę górną na długość rozwiązania równania z jedną niewiadomą w przypadku, gdy zbiór rozwiązań jest skończony. Korzystając z tej granicy przedstawię sposób badania skończoności zbioru rozwiązań w czasie liniowym ze względu na rozmiar równania.

Rozdział 2

Podstawowe definicje

Ustalę najpierw terminologię.

Alfabet Σ jest to niepusty i skończony zbiór, którego elementy nazywane są *literami*. (np. $\Sigma = \{a, b, c\}$)

Słowo jest to ciąg liter z Σ, na przykład $(a, b, c, a, a) = abcaa$. Ciąg ten może być skończony lub nieskończony (w prawo). Słowa nieskończone nazywane są ω-*słowami*. W dalszej części pracy będę zajmował się tylko skończonymi słowami. 1 oznacza *słowo puste*, czyli ciąg liter długości zero.

Σ^*, Σ^+ to zbiory odpowiednio wszystkich skończonych i skończonych niepustych słów nad alfabetem Σ. Przez $\mathrm{card}(\Sigma)$ oznaczę liczność zbioru Σ.

Dla $w \in \Sigma^*$, $|w|$ oznacza długość słowa w, $|1| = 0$. Dla $a \in \Sigma$, $|w|_a$ oznacza liczbę wystąpień a w w.

Definiuje się dwuargumentową operację *konkatenacji*

$$a_1 \dots a_n \cdot b_1 \dots b_m = a_1 \dots a_n b_1 \cdots b_m,$$

przy czym symbol tej relacji $\cdot$ często się opuszcza pisząc, dla słów w_1, w_2, $w_1 w_2$ zamiast $w_1 \cdot w_2$. Oczywiście operacja konkatenacji jest łączna i słowo puste jest jej elementem neutralnym. Ponadto wynik operacji konkatenacji na dwóch skończonych słowach też jest skończonym słowem (pustym, gdy obydwa słowa puste). Zatem $(\Sigma^*, \cdot)$ jest monoidem oraz $(\Sigma^+, \cdot)$ jest półgrupą. Co więcej, są one wolne, gdyż zawierają podzbiór zwany wolnym zbiorem generatorów (tym podzbiorem jest zbiór słów jednoliterowych, czyli alfabet Σ), taki że każdy element z Σ^* (odpowiednio Σ^+), z wyjątkiem elementu neutralnego, może być zapisany w sposób jednoznaczny jako konkatenacja elementów Σ.

Przez w^n oznaczamy n-tą *potęgę* słowa w, czyli n-krotne złożenie konkatenacji

$$w^n = \underbrace{ww \dots w}_{n}.$$

Jeżeli istnieje słowo v, takie że $w = uv$, wtedy u nazywamy *prefiksem* słowa w, a jeżeli istnieje słowo v, takie że $w = vu$, wtedy u nazywamy *sufiksem* słowa w.

Prefiks długości k, dla $0 \leqslant k \leqslant |w|$, jest oznaczany przez $\mathrm{pref}_k(w)$. Jeżeli $w = uv$, wtedy definiuje się lewy iloraz $u^{-1}w = v$ oraz prawy iloraz $wv^{-1} = u$. Dla słowa $w = a_1a_2 \ldots a_n$, $a_i \in \Sigma$, $1 \leqslant i \leqslant n$ oznaczę *podsłowo* $a_la_{l+1} \ldots a_{m-1}a_m$, $1 \leqslant l \leqslant m \leqslant n$, przez $w[l..m]$.

Kwadratem jest słowo postaci uu, gdzie $u \in \Sigma^+$, a *sześcianem* jest słowo postaci uuu, gdzie $u \in \Sigma^+$.

Słowo jest *bezkwadratowe*, jeżeli nie zawiera jako podsłowa żadnego kwadratu, oraz *bezsześcienne*, jeśli nie zawiera sześcianu. Np. słowo *aab* jest bezsześcienne, ale nie jest bezkwadratowe, bo zawiera *aa*.

Liczba p jest *okresem* słowa $w = a_1a_2 \ldots a_n$, jeżeli

$$a_i = a_{i+p} \text{ dla } i = 1, \ldots, n-p.$$

Zachodzi następujący lemat

Lemat 1 (O okresowości) *Jeżeli słowo w ma dwa okresy p i q oraz*

$$|w| \geqslant p + q - nwd(p, q),$$

wtedy ma także okres $nwd(p, q)$.

Słowo $w \in \Sigma^*$ jest *pierwotne*, jeżeli

$$\forall_{z \in \Sigma^*}(w = z^n \Rightarrow n = 1).$$

Symbol $\rho(w)$ dla słowa $w \in \Sigma^+$ oznacza *pierwiastek pierwotny* słowa w, czyli jedyne słowo pierwotne, takie że $w = \rho(w)^n$ dla pewnego $n \geqslant 1$.

Dla dwóch alfabetów Σ i Γ określamy *morfizm* $h : \Sigma^* \to \Gamma^*$ (lub $\Sigma^+ \to \Gamma^+$) jako przekształcenie, które spełnia

$$h(ww') = h(w)h(w') \text{ dla wszystkich } w, w' \in \Sigma^* \text{ (odpowiednio } \Sigma^+).$$

Język L nad alfabetem Σ to dowolny podzbiór Σ^*.

Będę rozpatrywał dwa rozłączne, niepuste i skończone alfabety: Σ zawierający stałe i Ξ zawierający niewiadome.

Równaniem e ze zbiorem Ξ jako zbiorem niewiadomych nad alfabetem Σ jest para

$$(\varphi, \psi) \in (\Sigma \cup \Xi)^* \times (\Sigma \cup \Xi)^*$$

oznaczana $e : \varphi = \psi$. *Rozmiar* równania e, oznaczany $|e|$, jest sumą długości φ i ψ. Równanie jest *zredukowane*, jeżeli φ i ψ zaczynają się od różnych liter oraz kończą

się różnymi literami jako słowa nad alfabetem $\Sigma \cup \Xi$. W dalszej części pracy będę rozpatrywał tylko równania zredukowane.

Rozwiązaniem równania e jest morfizm $h : (\Sigma \cup \Xi)^* \to \Sigma^*$ spełniający $h(a) = a$ dla każdego $a \in \Sigma$ i $h(\varphi) = h(\psi)$. Zbiór rozwiązań e jest oznaczany przez $\mathrm{Sol}(e)$.

Rozwiązanie równania może być podane także jako uporządkowana krotka, odpowiednia składowa dla odpowiedniej niewiadomej. Dla niewiadomej $X \in \Xi$, X-składową jest zbiór (język)

$$L_X(e) = \{\, h(X) \mid h \text{ jest rozwiązaniem } e \,\}.$$

Zbiór L_X jest nazywany językiem *definiowanym* przez niewiadomą X w równaniu e. Język $L \subseteq \Sigma^*$ jest *wyrażalny*, jeżeli istnieje równanie e i niewiadoma X, takie że

$$L = L_X(e).$$

Przykładem języka wyrażalnego może być zbiór słów rozpoczynających się od litery a. Niewiadoma X w równaniu:

$$aY = X$$

definiuje ten język.

Przykładem języka *niewyrażalnego* przez równania w półgrupie wolnej, to znaczy takiego, którego nie można zapisać w postaci składowej rozwiązania równania, są słowa bezsześcienne nad przynajmniej dwuliterowym alfabetem, oraz słowa bezkwadratowe nad przynajmniej trzyliterowym alfabetem [4].

Relacja $\mathcal{R} \subseteq (\Sigma^*)^k$ jest *wyrażalna*, jeżeli istnieje równanie e i niewiadome X_i, $1 \leqslant i \leqslant k$, takie że

$$\mathcal{R} = \{\, (h(X_1), h(X_2), \ldots, h(X_k)) \mid h \text{ jest rozwiązaniem } e \,\}.$$

Przykładem relacji wyrażalnej jest relacja sprzężenia. Dwa słowa x i y są sprzężone wtedy i tylko wtedy, gdy istnieją słowa p i q, takie że $x = pq$ i $y = qp$. Następujące równanie definiuje (poprzez niewiadome X i Y) relację sprzężenia

$$XaYXbY = PQaQPPQbQP.$$

Powstaje ono z połączenia w jedno równanie (metoda opisana w [5]) układu równań [4]

$$\begin{cases} X = PQ \\ Y = QP \end{cases}.$$

Przykładem relacji niewyrażalnej (o liczbie argumentów większej od 1) jest relacja bycia podciągiem liter, tzn. relacja $\leqslant_s$ [4] zdefiniowana dla $u, v \in \Sigma^*$ przez

$$\begin{aligned} u \leqslant_s v \iff u &= a_1 a_2 \ldots a_n, n \geqslant 0, a_i \in \Sigma, 1 \leqslant i \leqslant n, \\ v &= v_1 a_1 v_2 a_2 \ldots v_n a_n v_{n+1}, v_i \in \Sigma^*, 1 \leqslant i \leqslant n+1. \end{aligned}$$

Funkcją złożoności języka $L \subseteq \Sigma^*$ jest funkcja

$$\#_L : \mathbf{N} \to \mathbf{N}$$

zdefiniowana wzorem

$$\#_L(n) = \mathrm{card}\{\, w \in L \mid |w| = n \,\}.$$

Przykładowe równanie $e : XX = Y$ pokazuje, że funkcja zdefiniowana w powyższy sposób może być bardzo nieregularna:

$$\#_{L_Y(e)}(n) = \begin{cases} \mathrm{card}(\Sigma)^{\frac{n}{2}}, & \text{jeżeli } n \text{ jest parzyste,} \\ 0, & \text{jeżeli } n \text{ jest nieparzyste.} \end{cases}$$

Z tego powodu rozpatruje się funkcję $\bar{\#}_L$, która jest określona wzorem

$$\bar{\#}_L(n) = \max_{1 \leqslant i \leqslant n} \#_L(i).$$

Funkcja f jest *stała* jeżeli $f(n) = \Theta(1)$, *liniowa* jeżeli $f(n) = \Theta(n)$ i *wykładnicza* jeżeli $f(n) = 2^{\Theta(n)}$.

Funkcja $f : \mathbf{N} \to \mathbf{N}$ jest *typu dzielnikowego*, jeżeli istnieją liczby nieujemne $c_i, 1 \leqslant i \leqslant 4$ i $f(k)$ jest liczbą rozwiązań równania diofantycznego o niewiadomych n i m

$$(c_1 n + c_2)m + c_3 n + c_4 = k.$$

Funkcja f jest *typu* $\mathcal{D}_1$, jeżeli istnieją funkcje $g_1, g_2, \ldots, g_d$ typu dzielnikowego, takie że $f = \Theta(\bar{g}_1 + \bar{g}_2 + \ldots + \bar{g}_d)$, gdzie $\bar{g}_i(n) = \max_{1 \leqslant j \leqslant n} g(j)$.

Funkcja $f : \mathbf{N} \to \mathbf{N}$ jest *drugiego typu dzielnikowego*, jeżeli istnieją liczby całkowite $c_i, 1 \leqslant i \leqslant 8$, takie że $f(k)$ jest liczbą rozwiązań równania diofantycznego o niewiadomych n, m i p

$$((c_1 n + c_2)m + c_3 n + c_4)p + (c_5 n + c_6)m + c_7 n + c_8 = k.$$

Jak widać z definicji, każda funkcja typu dzielnikowego jest także drugiego typu dzielnikowego. Funkcja f jest *typu* $\mathcal{D}_2$, jeżeli istnieją funkcje $g_1, g_2, \ldots, g_d$ drugiego typu dzielnikowego, takie że $f = \Theta(\bar{g}_1 + \bar{g}_2 + \ldots + \bar{g}_d)$, gdzie $\bar{g}_i(n) = \max_{1 \leqslant j \leqslant n} g(j)$.

Z definicji funkcji $\mathcal{D}_1$ i $\mathcal{D}_2$ wynika następujący lemat

Lemat 2 *Suma dwóch funkcji typu* $\mathcal{D}_1$ *jest także funkcją typu* $\mathcal{D}_1$*, suma dwóch funkcji typu* $\mathcal{D}_2$ *jest także funkcją typu* $\mathcal{D}_2$*. Suma funkcji typu* $\mathcal{D}_1$ *i funkcji typu* $\mathcal{D}_2$ *jest funkcją typu* $\mathcal{D}_2$*.*

Będę stosował notację:
$a, b, \ldots \in \Sigma$ są ustalonymi literami;
$A, B, \ldots \in \Sigma^*$ są ustalonymi słowami;
$X, Y, \ldots \in \Xi$ są niewiadomymi;
$x, y, \ldots \in \Sigma^*$ mogą oznaczać jakieś ustalone słowa, albo obrazy niewiadomych w morfizmie $h : (\Sigma \cup \Xi)^* \to \Sigma^*$, to jest $x = h(X)$, $y = h(Y)$, itd;
$\varphi, \psi \ldots \in (\Sigma \cup \Xi)^*$ są słowami mogącymi zawierać zarówno stałe jak i niewiadome.

W celu poprawienia czytelności zapisu będę stosował następującą notację (według [6]). Dla $\alpha_i \in (\Sigma \cup \Xi)^*$,$1 \leqslant i \leqslant n$, wprowadzę oznaczenie:

$$[\alpha_i]_{i=1}^{n} = \alpha_1 \alpha_2 \ldots \alpha_n.$$

Rozdział 3

Równania postaci $A(X)Y = YB(X)$

Przytoczę na początek kilka lematów wraz z potrzebnymi definicjami.

Niech e będzie równaniem. Słowo w jest *krótkie*, jeżeli $|w| \leqslant 10|e|$, oraz *długie*, jeżeli $|w| > 10|e|$. Podobnie, okres p jest *krótki*, jeżeli $p \leqslant 10|e|$ i *długi*, jeżeli $p > 10|e|$.

Niech C, D będą dwoma słowami i niech e będzie równaniem z dwoma niewiadomymi X i Y. Zdefiniuję $X_{C,D}(e)$ jako zbiór tych słów postaci $(CD)^nC$, dla których $((CD)^nC, y)$ jest rozwiązaniem e dla jakiegoś y. Natomiast $Y_{C,D}(e)$ zdefiniuję jako zbiór słów y takich, że $((CD)^nC, y)$ jest rozwiązaniem e dla jakiejś liczby nieujemnej n.

Lemat 3 ([1]) *Załóżmy, że $e : \varphi = \psi$ jest równaniem nad $\Xi = \{X, Y\}$ i C, D są dwoma słowami. Wtedy*

(i) $\bar{\#}_{X_{C,D}(e)}$ *jest stała;*

(ii) $\bar{\#}_{Y_{C,D}(e)}$ *jest stała, wykładnicza albo typu* $\mathcal{D}_1$.

Niech A, B, C będą ustalonymi słowami i niech e będzie równaniem z dwoma niewiadomymi. Zdefiniuję X_1 jako zbiór tych słów postaci A^nBA^nC, dla których (A^nBA^nC, y) jest rozwiązaniem e dla jakiegoś y. Natomiast Y_1 zdefiniuję jako zbiór słów y takich, że (A^nBA^nC, y) jest rozwiązaniem e dla jakiejś liczby nieujemnej n.

Lemat 4 ([1]) *Załóżmy, że $e : \varphi = \psi$ jest równaniem nad $\Xi = \{X, Y\}$ i X_1, Y_1 są zdefiniowane w powyższy sposób, wtedy*

(i) $\bar{\#}_{X_1}$ *jest stała;*

(ii) $\bar{\#}_{Y_1}$ *jest stała, wykładnicza albo typu* $\mathcal{D}_1$.

Zgodnie z [1] wprowadzę oznaczenie e_2 dla równań postaci $AXY = YXB$, oraz oznaczenie s_1 dla układu dwóch różnych równań typu e_2, to jest

$$s_1 : \begin{cases} AUV = VUB \\ CUV = VUD \end{cases},$$

gdzie $A, B, C, D \in \Sigma^*$, $|A| = |B|$, $|C| = |D|$ i aby równania były różne wymagane jest, aby $A \neq C$ lub $B \neq D$.

Istotne będzie badanie struktury i złożoności zbioru

$$S_1 = \{\, uvu \mid (u, v) \in \mathrm{Sol}(s_1) \,\}.$$

Lemat 5 ([1]) *Jeżeli dwa równania w s_1 są różne, wtedy istnieje stała k, taka że dla dowolnego rozwiązania (u, v) układu s_1 z v wystarczająco długim, uvu ma okres k. W szczególności $\bar{\#}_{S_1}$ jest stała. Słowa w S_1, takie że v jest krótkie, są postaci $(CD)^n Cv(CD)^n C$.*

Rozpatrzmy teraz układ równań

$$s_2 : \begin{cases} AUV = VUB \\ CDU = UDC \end{cases},$$

gdzie $A, B, C, D \in \Sigma^*$, CD pierwotne. Podobnie, jak dla s_1, istotny będzie zbiór

$$S_2 = \{\, uvu \mid (u, v) \in \mathrm{Sol}(s_2) \,\}.$$

Lemat 6 ([1]) *Jeżeli $(u, v) \in Sol(s_2)$, wtedy uvu albo ma krótki okres, albo jest postaci*

$$uvu = ((CD)^n CA)^{m+1} (CD)^{l+n-j} C. \tag{3.1}$$

Wprowadzę oznaczenie

$$A(X) = [XA_i]_{i=1}^k,$$
$$B(X) = [B_i X]_{i=1}^k.$$

W tym rozdziale przejdę do badania równań postaci

$$e : [XA_i]_{i=1}^k Y = Y [B_i X]_{i=1}^k, \tag{3.2}$$

gdzie $A_i, B_i \in \Sigma^*, 1 \leqslant i \leqslant k$

Ilie i Plandowski w pracy [1] udowodnili następujące twierdzenie

Twierdzenie 1 ([1]) *(i) $\bar{\#}_{L_X(e)} \in \{stała, \mathcal{D}_1, liniowa, wykładnicza\}$,*

(ii) $\bar{\#}_{L_Y(e)} \in \{stała, \mathcal{D}_1, \mathcal{D}_2, liniowa, wykładnicza\}$.

W tym rozdziale zbadam, które kombinacje złożoności można wykluczyć, co podsumuję w Twierdzeniu 2.

W przypadku, gdy $k = 1$, równanie (3.2) otrzymuje postać $XAY = YBX$, a to równanie może mieć złożoności

$$(X \to \text{liniowa}, Y \to \text{liniowa}),$$

jeżeli $A = PQR$ i $B = RQP$, gdzie $R, Q, P \in \Sigma^*$ i $A \neq B$ (Wynika to z Twierdzenia 16 w [1]), albo

$$(X \to \text{wykładnicza}, Y \to \text{wykładnicza}),$$

jeżeli $A = B$, ponieważ

$$\{ (x, x) \mid w \in \Sigma^* \} \subseteq \text{Sol}(XAY = YBX).$$

W przeciwnym przypadku (to znaczy gdy nie istnieją $P, Q, R \in \Sigma^*$, takie że $A = PQR$, $B = RQP$) równanie takie nie ma rozwiązań ([6]). W dalszej części będę zakładał, że $k \geqslant 2$.

Rozpatrzmy rozwiązanie $(x, y) \in \text{Sol}(A(X)Y = YB(X))$. Wtedy $A(X)$ i $B(X)$ są sprzężone przez y. Niech $\text{cycle}(aw) = wa$ dla dowolnego $a \in \Sigma$, $w \in \Sigma^*$. Zatem, korzystając z własności sprzężenia, mamy

$$A(x) = \text{cycle}^t(B(x)),$$

dla jakiegoś $0 \leqslant t < |B(x)|$. Istnieje l, $2 \leqslant l \leqslant k$, takie że albo $t = |[B_i x]_{i=1}^{l-2} B'|$, gdzie B' jest prefiksem B_{l-1} (przerwanie obrotu wypada w środku którejś stałej), albo $t = |[B_i x]_{i=1}^{l-2} B_{l-1} x'|$, gdzie x' jest prefiksem x (przerwanie obrotu wypada w środku x, czyli obrazu niewiadomej X. Zarówno w tym jak i w poprzednim przypadku zawierają się przypadki przerwania obrotu pomiędzy stałą i x, oraz pomiędzy x i stałą).

Całkowita złożoność poszczególnych języków jest to maksimum ze złożoności wszystkich składowych. (Dla ustalonego równania jest tylko stała liczba składowych, więc dodanie do maksymalnej złożoności pozostałych (mniejszych lub równych) złożoności nie może zwiększyć sumarycznej złożoności). Pary złożoności mogą być dowolnymi kombinacjami w ten sposób uzyskanych złożoności.

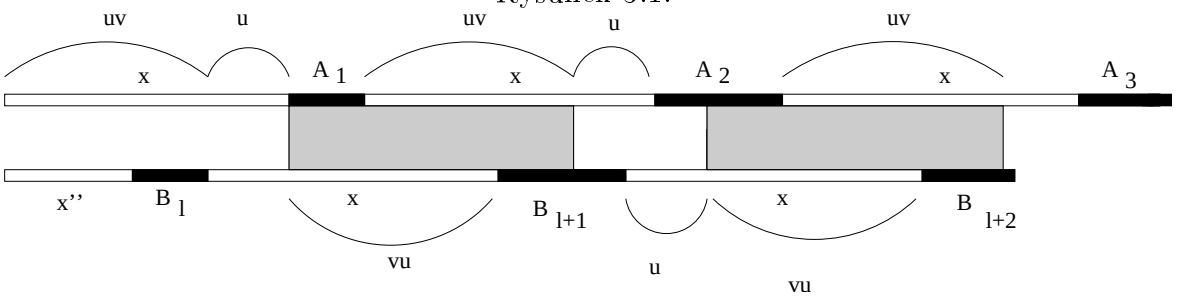
W pierwszym przypadku (przerwanie w stałej), X ma krótki okres, ponieważ zachodzi: $x \ldots = B' x \ldots$ gdzie B' jest sufiksem pewnego B_i. Zatem, na mocy Lematu 3, możliwe są następujące złożoności:

$$(X \to \text{stała}, Y \to \text{stała}),$$

$$(X \to \text{stała}, Y \to \mathcal{D}_1),$$

$$(X \to \text{stała}, Y \to \text{wykładnicza}).$$

Rysunek 3.1:



Dodając po kolei rozwiązania powstałe z wszystkich takich obrotów otrzymamy również jedną z tych par złożoności (największą, która wystąpiła).

Teraz zajmę się badaniem możliwych do uzyskania (to znaczy tych, których nie można na razie wykluczyć) par złożoności uzyskanych z obrotów cyklicznych przerwanych w obrazie niewiadomej.

Niech $t = |[B_i x]_{i=1}^{l-2} B_{l-1} x'|$ dla dowolnego $2 \leqslant l \leqslant k$ i dowolnego $x' \in \Sigma^*$ takiego, że $x = x'x'', x'' \in \Sigma^*$. Wtedy równanie z obrotem przyjmuje postać

$$[xA_i]_{i=1}^{k} = x''[B_i x]_{i=l}^{k} [B_i x]_{i=1}^{l-2} B_{l-1} x'. \tag{3.3}$$

Niech u będzie nakładającym się fragmentem pierwszego x z lewej strony i pierwszego pełnego x (nie x'') z prawej strony równania (3.3). Przypadek, w którym taki fragment nie istnieje, czyli gdy $|x''B_l| \geqslant |x|$ powoduje, że x ma krótki okres, to znaczy nie dłuższy niż $|B_l|$, więc z Lematu 3 możliwe są złożoności

$$(X \to \text{stała}, Y \to \text{stała}),$$

$$(X \to \text{stała}, Y \to \mathcal{D}_1),$$

$$(X \to \text{stała}, Y \to \text{wykładnicza}).$$

W przypadku, gdy takie u istnieje, z porównania $|x|$ pierwszych liter lewej i prawej strony równania (3.3) otrzymuję równanie:

$$x = x''B_l u. \tag{3.4}$$

Założę, że $u \leqslant \frac{|x|}{2}$. (Przypadek $u \geqslant \frac{|x|}{2}$ jest podobny). Sytuacja ta jest przedstawiona na rysunku 3.1. Jeśli u jest krótkie, to znowu x ma krótki okres i rezultat jest taki sam jak w poprzednim przypadku. Można więc w dalszej części założyć, że u jest wystarczająco długie. Założony wcześniej warunek $|u| \leqslant \frac{|x|}{2}$ prowadzi do nierówności $|u| \leqslant |x''B_l|$, ponieważ z równania (3.4) otrzymujemy $|x| = |x''B_l| + |u|$. Zatem u jest prefiksem $x''B_l$. Niech więc $x''B_l = uv$. Wtedy

$$x = uvu.$$

Z porównania nakładającej się części (prefiksu) uv drugiego x po lewej stronie i sufiksu vu pierwszego x po prawej stronie, otrzymujemy

$$A_1uv = vu\text{pref}_{|A_1|}(B_{l+1}u), \tag{3.5}$$

Jest to równanie typu e_2.

Biorąc odpowiednio długie x (długość zależna wyłącznie od równania) spowoduje się, że odpowiadające sobie wystąpienia x po lewej i po prawej stronie równania (3.3) będą na siebie zachodzić. Rozwiązania z krótkim x (zatem y ma krótki okres) można pominąć, gdyż stanowią tylko stały wkład w ogólną złożoność. Zatem w podobny sposób można utworzyć równania dla nakładających się prefiksu uv trzeciego x po lewej stronie równania z sufiksem vu drugiego x po prawej stronie. Jeżeli otrzymane równanie będzie różne od (3.5), wtedy można podzielić rozwiązania na te z długim v i z krótkim v. Gdy v jest długie, z Lematu 5 $uvu = x$ ma stały okres, więc z Lematu 3 możliwe są trzy kombinacje

$$(X \to \text{stała}, Y \to \text{stała}),$$

$$(X \to \text{stała}, Y \to \mathcal{D}_1),$$

$$(X \to \text{stała}, Y \to \text{wykładnicza}).$$

Rozpatrywanie rozwiązań z krótkim v nie wprowadzi żadnej nowej złożoności, gdyż z Lematu 5, $uvu = x$ jest postaci $x = (CD)^nCv(CD)^nC$ i, z Lematu 4 wynika, że możliwe są tylko dwie pierwsze pary złożoności z przypadku z długim v.

Ta sama sytuacja powstanie, gdy którekolwiek dwa równania otrzymane w ten sposób (porównanie nachodzących na siebie uv z $(k+1)$-szego wystąpienia x po lewej stronie równania z vu k-tego wystąpienia x po prawej stronie) są różne.

Odrębna sytuacja zachodzi, gdy wszystkie równania otrzymane w ten sposób są równe. Aby równania były równe potrzeba, aby miały równe długości, zatem

$$|A_1| = |B_l|, |A_2| = |B_{l+1}|, \ldots, |A_k| = |B_{l-1}|.$$

Możliwe są dwa przypadki.

Pierwszy przypadek $|A_1| = |A_2| = \ldots = |A_k|$. Wtedy z założenia, że wszystkie równania otrzymane z porównywania uv i vu są równe, wynika $A_1 = A_2 = \ldots = A_k$ oraz $B_1 = B_2 = \ldots = B_k$, więc równanie (3.2) redukuje się do

$$(XA_1)^kY = Y(B_1X)^k,$$

które jest równoważne równaniu $XA_1Y = YB_1X$. To równanie jest równaniem z przypadku $k = 1$, więc możliwe są złożoności

$$(X \to \text{wykładnicza}, Y \to \text{wykładnicza}),$$

$$(X \to \text{liniowa}, Y \to \text{liniowa}).$$

Należy w tym miejscu podkreślić, że obydwie wyżej uzyskane kombinacje złożoności są od razu złożonościami dla całego równania (3.2) i nie będą brały udziału w sumowaniu ze złożonościami otrzymanymi dla obrotów przerwanych w stałej. (W przeciwieństwie do pozostałych złożoności w omawianym przypadku przerwania w obrazie niewiadomej).

Drugi przypadek, gdy nie wszystkie długości stałych są równe, to znaczy $|A_i| \neq |A_j|$ dla jakiegoś $1 \leqslant i, j \leqslant k$. Bez straty ogólności można założyć, że $|A_1| < |A_2|$. Wtedy z porównania nakładających się częściowo (dlatego, że x odpowiednio długie, oraz $|B_{l+1}| = |A_2| > |A_1|$) drugiego u z drugiego x po lewej stronie równania z pierwszym u drugiego x po prawej stronie równania otrzymujemy

$$B'_{l+1}u = uA'_2, \tag{3.6}$$

gdzie B'_{l+1} jest sufiksem B_{l+1}, a A'_2 jest prefiksem A_2, obydwa o długości $|B_{l+1}| - |A_1|$. Zatem u ma krótki okres, oraz (u, v) jest rozwiązaniem układu równań postaci s_2, powiedzmy

$$\begin{cases} A_1UV = VU\text{pref}_{|A_1|}(B_{l+1}u) \\ CDU = UDC \end{cases}, \tag{3.7}$$

gdzie $A, B, C, D \in \Sigma^*$, CD pierwotne, pierwsze równanie w (3.7) to równanie (3.5). Z założeń wynika, że jest dokładnie jedno takie równanie.

Jeśli u jest wystarczająco długie, wtedy wszystkie równania dla u (tworzone analogicznie do równania (3.6) dla kolejnych x) są postaci $(CD)^kU = U(DC)^k, k \geqslant 0$. Zatem równanie (3.3) jest równoważne układowi (3.7). Mamy $x = uvu$ dla (u, v) będącego rozwiązaniem równania (3.7) i

$$y = (A(x))^p[B_ix]_{i=1}^{l-2}B_{l-1}x' \tag{3.8}$$

dla $p \geqslant 0$, ponadto $x' = x(uvB_l^{-1})^{-1} = uA$. Z (3.8) i Lematu 6 wynika, że w tym przypadku osiągana jest złożoność

$$(X \to \mathcal{D}_1, Y \to \mathcal{D}_2),$$

o ile tylko uvu jest postaci (3.1), w przeciwnym przypadku Lematu 6 (uvu ma krótki okres, zatem zastosowanie ma Lemat 3) złożoność redukuje się do

$$(X \to \text{stała}, Y \to \mathcal{D}_1).$$

W ten sposób pokazałem wszystkie możliwe złożoności dla obrotów przerwanych w jednej, ustalonej niewiadomej (wybór l). Ponieważ $2 \leqslant l \leqslant k$, zatem dla ustalonego

równania jest stała liczba l do wyboru. Aby uzyskać wszystkie możliwe złożoność sumy wszystkich tych rozwiązań należy zbadać wszystkie kombinacje (biorąc za każdym razem maksymalnązłożoność dla X, oraz maksymalna złożoność dla Y), w wyniku czego otrzymujemy następujące złożoności (pomijam na razie te końcowe, które nie biorą udziału w tych sumowaniach, gdyż osiągnięcie ich wymaga równań o szczególnej postaci o których wiadomo, jakie mają złożoności):

$$(X \to \text{stała}, Y \to \text{stała}),$$

$$(X \to \text{stała}, Y \to \mathcal{D}_1),$$

$$(X \to \text{stała}, Y \to \text{wykładnicza}),$$

$$(X \to \mathcal{D}_1, Y \to \mathcal{D}_2),$$

$$(X \to \mathcal{D}_1, Y \to \text{wykładnicza}).$$

Dodanie złożoności dla przerwanych obrotów w stałej nie spowoduje powstania innej pary złożoności. Do tej listy kombinacji należy dodać te dwie szczególne, co podsumowuje następujące twierdzenie

Twierdzenie 2 *Każda para języków definiowana przez niewiadome X i Y w równaniu postaci $[XA_i]_{i=1}^k Y = Y[B_i X]_{i=1}^k$ ma jedną z następujących złożoności:*

(i) $(X \to \textit{stała}, Y \to \textit{stała})$,

(ii) $(X \to \textit{stała}, Y \to \mathcal{D}_1)$,

(iii) $(X \to \textit{stała}, Y \to \textit{wykładnicza})$,

(iv) $(X \to \mathcal{D}_1, Y \to \mathcal{D}_2)$,

(v) $(X \to \mathcal{D}_1, Y \to \textit{wykładnicza})$,

(vi) $(X \to \textit{liniowa}, Y \to \textit{liniowa})$,

(vii) $(X \to \textit{wykładnicza}, Y \to \textit{wykładnicza})$.

Pokażę teraz konkretne równania dla niektórych złożoności. (Numeracja jak w Twierdzeniu 2)

(i) $XaXaaaY = YaaXaaX$

Obroty przerwane w stałych w sposób oczywisty prowadzą do rozwiązań gdzie x oraz y są potęgami a:

$$xaxaaa = aaxaax,$$

$$xaxaaa = axaaxa,$$

$$xaxaaa = xaaxaa.$$

Obrót przerwany w obrazie niewiadomej prowadzi do układu równań

$$\begin{cases} auv = vua \\ aauv = vuaa \end{cases},$$

z którego wynika po podstawieniu pierwszego równania do drugiego $aauv = auva$, czyli $auv = uva$, więc $u, v, x = uvu$ są potęgami a.

Zatem równanie otrzymuje postać $a^iY = Ya^i$, z czego wynika, że także y ma stałą złożoność.

(ii) $XbXaY = YXbaX$

Z obrotów przerwanych w obrazie niewiadomej dochodzimy do sprzecznosci: Pierwszy przypadek:

$$\begin{cases} buv = vub \\ cuv = vua \end{cases},$$

gdzie $c \in \Sigma$, co prowadzi do tego, że albo $|v| > 0$ i v kończy się jednocześnie na a i b, albo $|v| = 0$ i u kończy się jednocześnie na a i b. Analogicznie drugi przypadek prowadzi do sprzeczności:

$$\begin{cases} buv = vua \\ cuv = vub \end{cases}.$$

Także obrót $xbxa = axxb$ jest sprzeczny, $a \neq b$.

Z obrotu (w tym przypadku $t = 0$)

$$xbxa = xbax$$

otrzymujemy $x = a^i, i \in \mathbf{N}$. Podobnie z obrotu

$$xbxa = xxba$$

otrzymujemy $x = b^i$ i złożoności w obydwu przypadkach wynoszą stała i $\mathcal{D}_1$.

(iii) nie jest znane żadne równanie.

(iv) $XabcXcbabcY = YcbaXcbabcX$

Równanie to, wraz z dowodem złożoności znajduje się w [1].

(v) nie jest znane żadne równanie.

(vi) $XbaY = YabX$

Równanie to pochodzi z pracy [1], w której znajduje się dowód liniowej złożoności L_X oraz L_Y.

(vii) $XY = YX$

Dowolna para (w, w), $w \in \Sigma^*$ jest rozwiązaniem, zatem obydwa języki mają złożoność wykładniczą.

Rozdział 4

Długość minimalnego rozwiązania równania z jedną niewiadomą

Teraz zajmę się badaniem maksymalnej długości minimalnego rozwiązania równania z jedną niewiadomą.

Wprowadzę oznaczenia z [7]. Lewą stronę równania oznaczę przez L, a prawą stronę przez R. Dla równania z jedną niewiadomą X, która występuje taką samą liczbę razy (niech to będzie m razy) w L i w R, $l_1^x, \ldots, l_m^x$ i $r_1^x, \ldots, r_m^x$ oznaczają pozycje wystąpień X po lewej i po prawej stronie równania, w kolejności rosnącej. Zdefiniuję l_k^A i r_k^A dla słowa A i $1 \leqslant k \leqslant m$ przez

$$l_k^A = l_k^x + (k-1)(|A|-1),$$
$$r_k^A = r_k^x + (k-1)(|A|-1).$$

Oznaczę $A = h(X)$ dla dowolnego rozwiązania h.

Lemat 7 ([7]) *Dla dowolnego słowa $A \in \Sigma^*$ i dowolnego $1 \leqslant k \leqslant m$, $l_k^A - r_k^A = l_k^x - r_k^x$.*

Dowód $l_k^A - r_k^A = l_k^x + (k-1)(|A|-1) - r_k^x - (k-1)(|A|-1) = l_k^x - r_k^x$. □

Lemat 8 ([7]) *Dla dowolnego słowa $A \in \Sigma^*$ i dowolnego $1 \leqslant k \leqslant m$,*

$$h(L)[l_k^A .. l_k^A + |A| - 1] = A,$$
$$h(R)[r_k^A .. r_k^A + |A| - 1] = A.$$

Dowód Lemat udowodnię dla L, dla R dowód analogiczny. Z definicji morfizmu, $h(L)$ wygląda w następujący sposób

$$\begin{aligned} h(L) \;=\; & L[1..l_1^x - 1]AL[l_1^x + 1..l_2^x - 1]\ldots \\ & L[l_{k-1}^x + 1..l_k^x - 1]AL[l_k^x + 1..l_{k+1}^x - 1]\ldots \\ & L[l_{m-1}^x + 1..l_m^x - 1]AL[l_m^x + 1..n]. \end{aligned} \tag{4.1}$$

Ponieważ dla dowolnego słowa M, długość podsłowa $M[i..j]$ wynosi $j-i+1$, więc po zsumowaniu długości stałych do k-tego wystąpienia A otrzymujemy $((l_1^x-1)-1+1)+((l_2^x-1)-(l_1^x+1)+1)+\ldots+((l_k^x-1)-(l_{k-1}^x+1)+1)=l_k^x-k$. Po dodaniu $k-1$ razy (wszystkie przed k-tym wystąpieniem A) długości słowa A zmniejszonej o 1 (morfizm zmienia jedną literę x na słowo A) otrzymujemy $l_k^x-k+(k-1)|A|=l_k^A-1$. Zatem $h(L)[l_k^A]$ jest pozycją wystąpienia A które następuje po $L[l_k^x+1..l_k^x-1]$ po prawej stronie równania (4.1). □

Wprowadzę oznaczenie

$$d_k=|l_k^x-r_k^x|$$

dla $1\leqslant k\leqslant m$. Zachodzi lemat:

Lemat 9 ([7]) *Niech h będzie rozwiązaniem równania $L=R$, $|L|_X=|R|_X=m$. Dla $1\leqslant k\leqslant m$ i $d_k\neq 0$, jeśli $|h(X)|\geqslant d_k$, to $h(X)$ ma okres d_k.*

Dowód Bez straty ogólności można założyć, że $r_k^x<l_k^x$. Jeżeli $|A|=d_k$, z definicji d_k jest okresem A. Jeżeli $|A|>d_k$, z Lematu 7, $l_k^A=r_k^A+l+k^x-r_k^x=r_k^A+d_k<r_k^A+|A|$. Ponieważ $h(X)=A, h\in \mathrm{Sol}(L=R)$, więc h(L)[i]=h(R)[i] dla każdego $1\leqslant i\leqslant |h(L)|$. Rozważmy podciągi $h(L)$ i $h(R)$, wtedy $h(L)[l_k^A..r_k^A+|A|-1]=h(R)[l_k^A..r_k^A+|A|-1]$. Z lematu 8,

$$\begin{aligned} h(L)[l_k^A..r_k^A+|A|-1] &= A[1..|A|-(l_k^A-r_k^A)],\\ h(R)[l_k^A..r_k^A+|A|-1] &= A[1+(l_k^A-r_k^A)..|A|]. \end{aligned}$$

W ten sposób, z lematu 7,

$$A[1..|A|-(l_k^x-r_k^x)]=A[1+(l_k^x-r_k^x)..|A|],$$

co implikuje, że $l_k^x-r_k^x$ jest okresem A. □

Lemat 10 ([7]) *Niech h będzie rozwiązaniem równania $L=R$ w półgrupie wolnej z niewiadomą X, $|L|_X=|R|_X=m$, i niech p będzie okresem $h(X)$. Jeśli*

$$|h(X)|\geqslant \max_{1\leqslant k\leqslant m} d_k+p-1,$$

wtedy h' określony dla niewiadomej X przez $h'(X)=h(X)[1..|h(X)|-p]$ także jest rozwiązaniem.

Twierdzenie 3 ([7]) *Dla każdego równania w półgrupie wolnej z jedną niewiadomą długość minimalnego rozwiązania jest nie większa niż $\max_{1\leqslant k\leqslant m} d_k+\min_{1\leqslant k\leqslant m, d_k\neq 0} d_k-2$ w przypadku, gdy obydwie strony zawierają taką samą liczbę wystąpień niewiadomej. Ograniczenie jest ścisłe.*

Lemat 11 *Przykład ciągu równań dla których ograniczenie podane w Twierdzeniu 3 jest optymalne, ze stałą liczbą wystąpień niewiadomej:*

$$XXbaa(ba)^i = a(ba)^iXabX. \tag{4.2}$$

Dowód Twierdzenie 3 daje oszacowanie $4i+2$, gdyż długość maksymalnej różnicy wynosi $2i+3$, a długość minimalnej różnicy wynosi $2i+1$.

Ponieważ $baa(ba)^i \neq a(ba)^iab$, więc słowo puste nie jest rozwiązaniem tego równania.

Zatem każde rozwiązanie kończy się na a (porównanie ostatnich liter lewej i prawej strony równania) oraz zaczyna od a (porównanie pierwszych liter).

Jeśli jakiekolwiek rozwiązanie jest krótsze od $2i+1$, powiedzmy o długości $j < 2i+1$, to lewa strona równania zaczyna sie od $\mathrm{pref}_j(a(ba)^i)\mathrm{pref}_j(a(ba)^i)$, czyli zawiera podsłowo aa na pozycji j (pierwsze a to ostatnia litera pierwszego x, a drugie to pierwsza litera drugiego x). czyli $\mathrm{pref}_{j+1}(xx)$ zawiera aa. Prawa strona jednak zaczyna się od $a(ba)^i$ (które jest długości przynajmniej $j+1$) i które nie zawiera podsłowa aa, zatem lewa i prawa strona równania nie mogą być sobie równe. Zatem $|x| \geqslant 2i+1$.

Jeśli $|x| = 2i+1$, wtedy musi być $x = a(ba)^i$ (porównanie prefiksów), więc x nie jest rozwiązaniem, ponieważ:

$$a(ba)^ia(ba)^ibaa(ba)^i \neq a(ba)^ia(ba)^iaba(ba)^i.$$

Zatem $|x| > 2i+1$, można więc zauważyć, że $aa(ba)^i$ jest sufiksem x (porównanie sufiksów stron równania wykorzystujące aktualne oszacowanie minimalnej długości x). Ponieważ jednocześnie x ma prefiks $a(ba)^i$, zatem najkrótsze słowo spełniające te dwa warunki ma postać $a(ba)^ia(ba)^i$ i, jak łatwo zauważyć, jest rozwiązaniem równania:

$$a(ba)^ia(ba)^ia(ba)^ia(ba)^ibaa(ba)^i = a(ba)^ia(ba)^ia(ba)^iaba(ba)^ia(ba)^i.$$

$|a(ba)^ia(ba)^i| = 4i+2$, a zatem dokładnie tyle, ile wynosi oszacowanie z twierdzenia. □

Oznaczę odpowiednio lewą i prawą stronę równania (4.2) przez:

$$\psi_{L,i} = XXbaa(ba)^i,$$

$$\psi_{P,i} = a(ba)^iXabX.$$

Rozpatrzmy równanie

$$\psi^j_{L,i} = \psi^j_{P,i}.$$

Jest ono równoważne poprzedniemu równaniu $\psi_{L,i} = \psi_{P,i}$, zatem minimalne rozwiązanie ma długość $4i+2$. Minimalna i maksymalna odległość odpowiadających sobie

wystąpień niewiadomych się nie zmienia, zatem oszacowanie pozostaje optymalne. Liczba niewiadomych w równaniu wynosi $4j$, zatem może być dowolnie duża. Jest to więc przykład dowolnie długiego równania, z dowolnie dużą liczbą wystąpień niewiadomej, z dowolnie długim minimalnym rozwiązaniem, dla którego oszacowanie jest optymalne.

Prawdziwe jest więc twierdzenie

Twierdzenie 4 *Dla dowolnych* $i, j \in \mathbf{N}$ *minimalne rozwiązanie* $\psi^j_{L,i} = \psi^j_{P,i}$ *ma długość*

$$\max_{1 \leqslant k \leqslant m} d_k + \min_{1 \leqslant k \leqslant m, d_k \neq 0} d_k - 2,$$

co pokazuje, że granica podana w Twierdzeniu 3 jest optymalna.

Rozdział 5

Badanie skończoności zbioru rozwiązań równania z jedną niewiadomą

Udowodnię teraz następujący lemat, który posłuży do badania skończoności zbioru rozwiązań równania z jedną niewiadomą.

Lemat 12 *Niech h będzie rozwiązaniem równania $L = R$ w półgrupie wolnej z niewiadomą X, $|L|_X = |R|_X = m$, i niech p będzie okresem $h(X)$. Jeśli*

$$|h(X)| \geqslant \max_{1 \leqslant k \leqslant m} d_k + p - nwd(d_1, d_2, \ldots, d_m, p),$$

wtedy h' określony dla niewiadomej X przez $h'(X) = h(X)[1..p]h(X)$ także jest rozwiązaniem.

Dowód Z Lematu 9 wynika, że $h(X)$ ma okresy $d_1, \ldots, d_m$, oraz z warunku na długość i Lematu o okresowości (Lemat 1) wynika, że ma także okresy $nwd(d_1, p)$, $\ldots nwd(d_m, p)$. Oznaczę te liczby odpowiednio przez $p_1, \ldots, p_m$. Teraz zdefiniuję ciąg q_i, $1 \leqslant i \leqslant m$ w następujący sposób: $q_1 = p_1$, $q_n = nwd(q_{n-1}, p_n)$ dla $2 \leqslant n \leqslant m$. Stosując $m - 1$ razy Lemat o okresowości otrzymujemy, że q_m jest okresem $h(X)$. Z konstrukcji wynika także, że q_m dzieli wszystkie p_i oraz wszystkie $d_i, 1 \leqslant i \leqslant m$.

Rozważmy kolejno wszystkie nakładające się i-te obrazy niewiadomych w $h(L)$ i $h(R)$, $1 \leqslant i \leqslant m$:

$$xB = Cx,$$

gdzie x jest wystąpieniem $h(X)$, a B i C są słowami o długości d_i. Przez x' oznaczę $h'(X)$. Z faktu, że q_m dzieli d_i wynika, że

$$x[1..q_m]xB = Cx[1..q_m]x,$$

Stosując to rozumowanie $\frac{p}{q_m}$ razy, otrzymuję

$$x'B = Cx'$$

Po zastąpieniu i-tego x przez x' pozostały fragment $h(L) = h(R)$ (to znaczy nienachodzący na i-te x po lewej ani po prawej stronie) się nie zmienia, zatem x' także jest rozwiązaniem $L = R$. □

Można teraz udowodnić twierdzenie:

Twierdzenie 5 *Niech $L = R$ będzie równaniem w półgrupie wolnej z jedną niewiadomą X, takim że $|L|_X = |R|_X$. Jeżeli $L = R$ ma rozwiązanie długości przynajmniej*

$$\max_{1 \leqslant k \leqslant m} d_k + \min_{1 \leqslant k \leqslant m, d_k \neq 0} d_k - 1,$$

wtedy ma nieskończenie wiele rozwiązań. Co więcej, ograniczenie to jest optymalne.

Dowód Jeżeli wszystkie d_k są równe 0, wtedy oczywiście każde słowo jest rozwiązaniem równania, lub równanie jest sprzeczne, więc albo nie ma żadnego rozwiązania, albo jest ich nieskończenie wiele.

Załóżmy zatem, że istnieje $d_k \neq 0$. Gdyby równanie miało rozwiązanie przynajmniej tej długości oraz miało tylko skończenie wiele rozwiązań, to weźmy to najdłuższe rozwiązanie oraz jego okres, d_k. Z Lematu 12 wynika, że istnieje rozwiązanie od niego dłuższe o d_k. Sprzeczność.

Aby udowodnić optymalność, podam ciąg równań, które mają rozwiązanie o długości $(\max_{1 \leqslant k \leqslant m} d_k + \min_{1 \leqslant k \leqslant m, d_k \neq 0} d_k - 1) - 1$, lecz zbiór rozwiązań jest skończony.

Rozpatrzmy ciąg równań

$$a^i bXaX = Xba^i Xa. \tag{5.1}$$

Porównanie sufiksów ($aX = Xa$) prowadzi do wniosku, że każde rozwiązanie jest postaci a^j. Dla dowolnego rozwiązania a^j musi zatem zachodzić $a^i ba^j = a^j ba^i$ (porównanie prefiksów). Gdyby $i < j$, wtedy lewa strona na pozycji $i+1$ miałaby literę b, natomiast prawa strona literę a. Gdyby $i > j$, wtedy lewa strona równania na pozycji $j + 1$ miałaby literę a, natomiast prawa strona miałaby literę b. Musi więc zachodzić $i = j$. a^i jest (jedynym) rozwiązaniem równania (5.1):

$$a^i ba^i aa^i = a^i ba^i aa^i.$$

Tak więc istnieje rozwiązanie długości $(\max_{1 \leqslant k \leqslant m} d_k + \min_{1 \leqslant k \leqslant m, d_k \neq 0} d_k - 1) - 1 = (i + 1) + 1 - 2 = i$, oraz zbiór rozwiązań jest skończony. □

Porównajmy teraz nakładające się wystąpienia pierwszych niewiadomych w równaniu, w przypadku gdy rozwiązanie jest dłuższe od d_1.

$$xB_0 = A_0x. \tag{5.2}$$

Być może B_0 zawiera pewien prefiks x, ale jest on jednoznacznie wyznaczony przez prefiks A_0. Z zależności tej wynika, że istnieją stałe p i q, takie że

$$A_0 = (pq)^i, \tag{5.3}$$

$$B_0 = (qp)^i,$$

$$x = (pq)^j p.$$

Wynika z tego, że $|pq|$ jest okresem x. Rozpatrzmy rozwiązania odpowiednio długie (dłuższe niż ograniczenie w Lemacie 12). Z Lematu 12 wynika więc, że albo wszystkie odpowiednio długie słowa postaci $(pq)^j p$ są rozwiązaniami, albo żadne z nich, a wtedy zbiór rozwiązań jest skończony (wszystkie rozwiązania są krótsze od wspomnianego ograniczenia). Co ważne, powyższe stałe p i q mogą być znalezione w czasie liniowym ze względu na długość równania: Aby znaleźć $|pq|$, wystarczy znaleźć dwa pierwsze wystąpienia (niech to będą pos_1, pos_2) słowa B_0 w słowie A_0^3. Algorytm Knutha-Morrisa-Pratta ([9]) znajduje te wystąpienia w czasie liniowym. Długością pq jest $pos_2 - pos_1$. Długość p wynosi $pos_1 - 1$, zatem długością q jest $pos_2 - 2pos_1 + 1$. Znając długości, można znaleźć same słowa korzystając z równania (5.3):

$$p = A_0[1..pos_1 - 1],$$

$$q = A_0[pos_1..pos_2 - 1].$$

Aby zbadać skończoność zbioru wystarczy więc sprawdzić, czy jedno, odpowiednio długie słowo $x = (pq)^j p$, jest rozwiązaniem.

W przypadku, gdy pq jest pierwotne, odpowiedź na pytanie o skończoność wynika z następującgo lematu, pochodzącego z [8].

Oznaczmy stałe w równaniu: $L = A_0X \ldots, R = XB_1X \ldots$

Lemat 13 ([8]) *Niech $u, v \in \Sigma^*$ będą takie, że uv pierwotne, $|uv| \leqslant |A_0B_1|$. Jest wtedy możliwe zbadać w czasie $O(|L| + |R|)$ dla wszystkich liczb naturalnych $n \geqslant 0$, czy $x = (uv)^n u$ jest rozwiązaniem $L = R$.*

Pokażę inną metodę, działającą dla dowolnego słowa, nie tylko dla słów postaci $(uv)^n u$. Można sprawdzić, czy słowo x jest rozwiązaniem w czasie liniowym ze względu na $n = |x|$ i $|L| + |R|$ w następujący sposób: W czasie liniowym można wyliczyć tablicę reprezentującą funkcję prefiksową π (opis tworzenia tej tablicy w czasie liniowym, jako

części algorytmu Knutha-Morrisa-Pratta, znajduje się w [9]). Funkcja (tablica) ta dla słowa x długości n jest określona wzorem

$$\pi[j] = \max\{\, k \mid k < j \text{ i } x[1..k] \text{ jest prefiksem } x[1..j]\,\}, 1 \leqslant j \leqslant n.$$

Korzystając z tej tablicy można znaleźć w czasie liniowym wszystkie okresy (niewiększe od n) słowa x: $n - \pi[n]$ jest najkrótszym okresem x, $n - \pi[\pi[n]]$ jest drugim co do do długości okresem, itd. Ponieważ, z definicji, $\pi[k] < k$, zatem otrzymane w ten sposób liczby są coraz większe, dopóty, dopóki pewna iteracja nie osiągnie 0, wtedy π jest niezdefiniowane, mamy okres n i proces przerywamy. Wszystkie w ten sposób otrzymane okresy można zapamiętać w tablicy bitowej $T[1..n]$

$$T[i] = \begin{cases} 1, & \text{jeżeli } i \text{ jest okresem } x, \\ 0, & \text{w przeciwnym przypadku.} \end{cases}$$

Przy pomocy tablicy T możemy sprawdzić, czy x jest rozwiązaniem równania $L = R$. Przechodzimy równocześnie obydwie strony równania ($L = R$, ponieważ rozmiar $h(L) = h(R)$ może nie być liniowy ze względu na n) pamiętając dwa wskaźniki: dla lewej i prawej strony. Wskaźnik jest to pozycja w słowie L (odpowiednio R), lub pozcyja (w przypadku gdy wskaźnik wskazuje na X) wraz z indeksem w słowie x

Przechodzenie polega na wykonaniu pewnej ilości kroków (nie więcej niż $2n$ kroków, gdyż w każdym kroku przynajmniej jeden ze wskaźników się zwiększa, co wyniknie z następującej konstrukcji). Początkowo $j_1 = 1$ i $j_2 = 1$. Przy zwiększeniu wskaźnika, jeśli okaże się, że wskazuje on teraz na X, wtedy odpowiedni indeks i_1 lub i_2 ustawia się na 1. Wykonanie kroku jest uzależnione od tego, na co wskazują wskaźniki.

- Jeżeli obydwa wskaźniki wskazują na litery z Σ, wtedy albo przechodzimy do następnego kroku zwiększając obydwa wskaźniki, gdy są to te same litery, albo przerywamy z wynikiem "x nie jest rozwiązaniem", gdy te litery są różne.

- Jeżeli jeden ze wskaźników (powiedzmy j_1) wskazuje na literę z Σ, a drugi (powiedzmy j_2) na X z pewnym indeksem i_2, wtedy porównujemy tą literę z $x[i_2]$ i gdy są to różne litery, wtedy przerywamy z wynikiem "x nie jest rozwiązaniem", w p. p. zwiększamy wskaźnik j_1, oraz zwiększamy indeks. Jeśli zwiększony indeks jest większy od n, wtedy zwiększamy także wskaźnik j_2.

- Jeżeli obydwa wskaźniki wskazują na X, co może zajść tylko w przypadku, gdy jeden z indeksów jest równy 1 (niech to będzie i_1, przypadek, gdy obydwa indeksy są równe 1, polega na zwiększeniu obydwu wskaźników), wtedy patrzymy do tablicy $T[i_2]$, gdzie i_2 jest indeksem różnym od 1. Jeżeli na tej pozycji w T jest 0, wtedy przerywamy z wynikiem "x nie jest rozwiązaniem", w przeciwnym przypadku zwiększamy wskaźnik skojarzony z indeksem i_2, oraz zwiększamy indeks i_2 do $n - i_2 + 2$.

Jeżeli jeden ze wskaźników osiągnął $n+1$, wtedy albo przerywamy z wynikiem "x nie jest rozwiązaniem równania", gdy drugi wskaźnik wynosi mniej niż $n+1$ (ten przypadek nie zajdzie, gdy obydwie strony równania mają tyle samo wystąpień X), albo kończymy z wynikiem "x jest rozwiązaniem równania", w p. p.

Mając powyższy, liniowy algorytm na sprawdzanie, czy dane słowo jest rozwiązaniem równania, możemy udowodnić następujące twierdzenie:

Twierdzenie 6 *Dla każdego równania $L = R$ z jedną niewiadomą można sprawdzić w czasie $O(|L| + |R|)$, czy zbiór rozwiązań jest skończony.*

Dowód Jeżeli $|L|_X \neq |R|_X$, wtedy jest co najwyżej jedno rozwiązanie (rozwiązanie może być tylko jednej, ustalonej na podstawie równania, długości), zatem zbiór rozwiązań jest skończony.

Jeżeli $|L|_X = |R|_X$, wtedy z (5.2) znajdujemy słowa p i q takie, że każde rozwiązanie x musi być postaci $(pq)^i p$, bierzemy $i = \left\lceil \frac{2*|L|}{|pq|} \right\rceil + 1$, co powoduje, że $(pq)^i p$ spełnia warunek podany w Twierdzeniu 5 i sprawdzamy zgodnie z podanym algorytmem, czy $(pq)^i p$ jest rozwiązaniem. Jeśli jest, wtedy ze wspomnianego twierdzenia wynika, że jest nieskończenie wiele rozwiązań, jeśli nie, wtedy jest skończenie wiele rozwiązań. □

Spis literatury

[1] L. Ilie i W. Plandowski, Two-variable word equations, *Theoret. Informatics Appl.* **34** (2000) 467-501.

[2] G. S. Makanin, The problem of solvability of equations in a free semigroup, *Mat. Sb.* **103** (1977) 147-236. Tłumaczenie angielskie: *Math. USSR Sbornik* **32** (1977) 129-198.

[3] J. Karhumäki, F. Mignosi i W. Plandowski, The expressibility of languages and relations by word equations, *Proc. of ICALP'97.* Springer, Berlin, *Lecture Notes in Comput. Sci.* **1256** (1995) 98-109.

[4] L. Ilie, Subwords and power-free words are not expressible by word equations, *Fundamenta Informaticae* **XX** (2001) 1-9.

[5] C. Choffrut i J. Karhumäki, Combinatorics of Words, *Handbook on Formal Languages.* Springer, Berlin(1997).

[6] Yu. I. Hmelevskii, Equations in free semigroups. *Trudy Mat. Inst. Steklov* **107** (1971). Tłumaczenie angielskie: *Proc. Steklov Inst. of Mathematics* **107** (1971). Amer. Math. Soc. (1976).

[7] K. Baba, S. Tsuruta, A. Shinohara i M. Takeda, On the Length of the Minimum Solution of Word Equations in One Variable. *Proc. of MFCS 2003*, w druku.

[8] S. Eyono Obono, P. Goralcik i M. Maksimenko, Efficient Solving of the word equations in one variable. *Proc. of MFCS'94.* Springer, Berlin, *Lecture Notes in Computer Science* **841** (336-341).

[9] T. H. Cormen, C. E. Leiserson, R. L. Rivest, Wprowadzenie do algorytmów. WNT, Warszawa 1998.