39 Algorytm Makanina

39.1 Współczynnik periodyczności

Współczynnikiem periodyczności słowa w nazywamy największą liczbę naturalną n taką, że $w = up^nv$ dla pewnych słów u i v, i pewnego niepustego słowa p. Współczynnikiem periodyczności zbioru słów nazywamy maksimum współczynników periodyczności słów z tego zbioru. Współczynnikiem periodyczności rozwiązania równania w słowach nazywamy współczynnik periodyczności składowych tego rozwiązania. Nietrudno zauważyć, że jeżeli n jest współczynnikiem periodyczności słowa w, $w = up^nv$ i $p \neq \varepsilon$, to p jest słowem prostym.

Pierwszy ważny rezultat Makanina to ograniczenie współczynnika periodyczności. Nie jest jasne, dlaczego jest to ważne, ponieważ istnieją dowolnie długie słowa o współczynniku periodyczności 3 (to jest cytat, Volker Diekert, Makanin's Algoritm).

39.1.1 Nieskończone słowa o współczynniku periodyczności równym 3 i 2

Weźmy homomorfizm $\mu : \{a, b\}^* \to \{a, b\}^*$ zdefiniowany wzorami

$$\mu(a) = ab \ \text{ oraz } \ \mu(b) = ba.$$

Zauważmy, że

$$\mu^{n+1}(a) = \mu^n(\mu(a)) = \mu^n(ab) = \mu^n(a)\mu^n(b).$$

Tak więc $\mu^n(a) < \mu^{n+1}(a)$ oraz $\mid \mu^n(a) \mid = 2^n$. Będziemy rozważać nieskończyny ciąg liter a i b, którego odcinkami początkowymi są słowa postaci $\mu^n(a)$. Ten ciąg nazywamy słowem (nieskończonym) Thue'go-Morse'a.

Słowo u zachodzi na siebie w słowie w, jeżeli $w = x_1uy_1$ i $w = x_2uy_2$ dla pewnych słów takich, że $0 < \mid \mid x_1 \mid - \mid x_2 \mid \mid < \mid u \mid$.

Lemat 39.1 *W słowie w jest zachodzące na siebie podsłowo wtedy i tylko wtedy, gdy dla pewnej litery a i pewnego słowa v słowo $avava$ jest podsłowem w.*

Dowód. Przypuśćmy, że u zachodzi na siebie w słowie w. Oczywiście, $u = sz = zt$ dla pewnych niepustych s, z i t takich, że su jest podsłowem w. Niech $s = av$. Słowo z też zaczyna się literą a, więc $u = avaz'$ dla pewnego z'. Mamy też $su = avavaz'$. Wobec tego, $avava$ jest podsłowem w. Implikacja odwrotna jest oczywista. □

Lemat 39.2 *Jeżeli $x \in \{ab, ba\}^*$, to $axa \notin \{ab, ba\}^*$ oraz $bxb \notin \{ab, ba\}^*$.*

Dowód. Weźmy najkrótsze słowo takie, że $x \in \{ab, ba\}^*$ oraz $axa \in \{ab, ba\}^*$ lub $bxb \in \{ab, ba\}^*$. Przyjmijmy, że $axa \in \{ab, ba\}^*$. Oczywiście, $axa = abyba$ dla pewnego $y \in \{ab, ba\}^*$. Stąd $x = byb$ i mamy sprzeczność z definicją x. □

Lemat 39.3 *Jeżeli w słowie $\mu(w)$ jest zachodzące na siebie podsłowo, to takie podsłowo jest też w słowie w.*

Dowód. Jeżeli w słowie $\mu(w)$ jest zachodzące na siebie podsłowo, to $\mu(w) = xavavay$. Ponieważ $\mu(w)$ ma parzystą długość, więc x lub y ma parzystą długość. Załózmy, że x ma parzystą długość i $x = \mu(x')$. Gdyby v było słowem parzystej długości, to zarówno ava, jak i v należałyby do $\{ab, ba\}^*$, a to nie jest możliwe. Wobec tego, długość v jest nieparzysta. Tak więc $av = \mu(au)$ oraz $ay = \mu(au')$ dla pewnych podsłow u i u' słowa w. Mamy więc $\mu(w) = xavavay = \mu(x'auauau')$. □

Wniosek 39.4 *Dla dowolnego n, żadne słowo nie zachodzi na siebie w słowie $\mu^n(a)$.* □

Wniosek 39.5 *Dla dowolnego n, słowo $\mu^n(a)$ nie zawiera niepustego podsłowa postaci u^3. Także słowo Thue'go-Morse'a nie zawiera niepustego słowa postaci u^3.* □

Lemat 39.6 *Istnieje nieskończone słowo zbudowane z liter a, b i c, które nie zawiera niepustego podsłowa postaci u^2.*

Dowód. Słowo Thue'go-Morse'a daje się jednoznacznie przedstawić w postaci konkatenacji słów a, ab i abb. Uznajmy te słowa za nowe litery. Jeżeli otrzymane w ten sposób słowo zawiera podsłowo postaci u^2, to słowo Thue'go-Morse'a zawiera podsłowo v^2, gdzie v jest konkatenacją słów a, ab i abb. Tak więc $v = av'$ i dalszy ciąg słowa Thue'go-Morse'a (za v^2) też zaczyna się literą a. Słowo Thue'go-Morse'a zawiera więc słowo $av'av'a$, a to nie jest możliwe. □

39.1.2 Oszacowanie współczynnika periodyczności

Twierdzenie 39.7 *Jeżeli $\mathcal{E}$ jest równaniem w słowach długości d, występują w nim dwie stałe i p jest współczynnikiem periodyczności minimalnego rozwiązania $\mathcal{E}$, to $p \leqslant (2d-4)2^{1.07d}$.*

Dowód. (Właściwie idea dowodu.) Niech $X_1, \ldots, X_n$ będzie minimalnym rozwiązaniem równania $\mathcal{E}$ i niech ma współczynnik periodyczności s (minimalne rozwiązanie równania w słowach to takie, które minimalizuje sumę

$$\sum_{i=1}^{n} | X_i |$$

długości składowych). Weźmy też proste słowo P takie, że $X_i = YP^sZ$ dla pewnych i, X i Y.

Przypuśćmy, że

$$W = A_0[P^{k_i}A_i]_{i=1}^n. \tag{34}$$

Takie przedstawienie słowa W nazywamy P-podziałem, jeżeli dla $i < n$ słowo P jest sufiksem A_i, a dla $i > 0$ – prefiksem A_i.

Przedstawienie (34) jest P-stabilne, jeżeli P^2 nie jest podsłowem A_i dla $i = 0, \ldots, n$, $A_i \neq P$ z wyjątkiem $i = 0$ i $i = n$, P jest sufiksem A_i dla $i \neq n$ oraz P jest prefiksem A_i dla $i > 0$.

Zauważmy, że najdrobniejszy P-podział słowa W (podział (34), dla którego suma $n + k_1 + \ldots + k_n$ jest największa) jest też P-stabilnym przedstawieniem słowa W. Tak więc dowolne słowo W ma P-stabilne przedstawienie. Zachodzi też lemat o jednoznaczności P-stabilnych przedstawień.

Bierzemy rozwiązanie $X_1, \ldots, X_n$ i znajdujemy P-stabilne przedstawienia jego składowych. Następnie, po podstawieniu rozwiązania do równania konstruujemy P-stabilne przedstawienie stron równania. Jeżeli składowe uznamy za słowa parametryczne, to strony równania będą słowami parametrycznymi z funkcjami liniowymi jako wykładnikami. Porównując obie strony równania i korzystając z lematu o jednoznaczności otrzymujemy pewien uklad równań liniowych o niewielkich współczynnikach. Współczynnik periodyczności pomniejszony o 2 jest składową rozwiązania tego układu i to składową rozwiązania minimalnego (gdyby było mniejsze rozwiązanie, to byłoby też krótsze rozwiązanie równania w słowach, z mniejszymi wykładnikami).

Dalej korzystamy z oszacowania minimalnego rozwiązania układu równań liniowych. □

Lemat 39.8 *Rozważamy równanie*

$$x_n a\, x_n b\, x_{n-1} b\, \ldots\, x_2 b\, x_1 = a x_n\, x_{n-1}^5 b\, x_{n-2}^5 b\, \ldots\, x_1^5 b\, a^5.$$

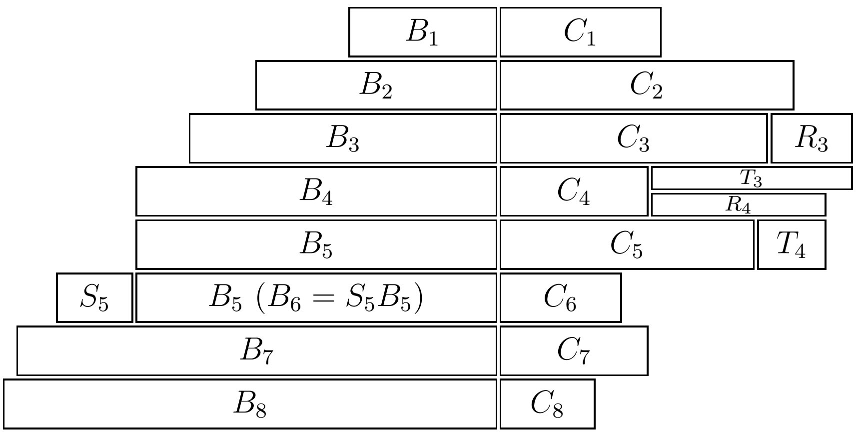
Równanie to ma dokładnie jedno rozwiązanie takie, że $x_i = a^{5^i}$. Dla dużych n współczynnik periodyczności tego rozwiązania jest większy od $2^{0.29d}$, gdzie $d = 8n+2$ jest długością podanego równania. □

39.2 Wieże domino

Trzy ciągi słów $B_1, \ldots, B_k$, $C_1, \ldots, C_k$ oraz $S_1, \ldots, S_{k-1}$ nazywamy wieżą domino wysokości k, jeżeli słowa B_i i C_i są niepuste dla $i = 1, \ldots, k$,

$$B_{i+1} = S_i B_i \text{ oraz } C_i R_i = C_{i+1} T_i$$

dla wszystkich $i = 1, \ldots, k-1$ oraz dla pewnych słów R_i i T_i.



Rysunek 3: Wieża domino

Zbiór słów $X_1, \ldots, X_n$ może być ustawiony w wieżę domino wysokości k, jeżeli istnieje wieża domino $B_1, \ldots, B_k$, $C_1, \ldots, C_k$ oraz $S_1, \ldots, S_{k-1}$ wysokości k taka, że dla $i = 1, \ldots, k$ słowo B_iC_i jest jednym ze słów $X_1, \ldots, X_n$.

Lemat 39.9 *Przypuśćmy, że zbiór $X_1, \ldots, X_n$ może być ustawiony w wieżę domino $B_1, \ldots, B_k$, $C_1, \ldots, C_k$ oraz $S_1, \ldots, S_{k-1}$ wysokości k. Jeżeli spełniony jest warunek*

$$j - i \geqslant K \Rightarrow |\, S_i \ldots S_{j-1} \,| > 0,$$

dla wszystkich i, j, $1 \leqslant i < j \leqslant k$, to pewne słowo X_t jest postaci P^sQ dla $s \geqslant \frac{k}{Kn^2} - 1$.

Dowód. Przypuśćmy, że oprócz podanych założeń spełnione są równości $K = 1$ i $n = 1$. Wtedy wszystkie słowa S_i są niepuste i zachodzą równości $S_iB_iC_i = B_{i+1}C_{i+1}T_i$ (wszystkie słowa R_i mogą być puste), czyli

$$S_iX_1 = X_1T_i$$

dla $i = 1, \ldots, k-1$. Zauważmy też, że $S_i \leqslant B_{i+1} < B_{i+1}C_{i+1} = X_1$. Każde rozwiązanie równania $sx = xt$ ma postać $s = uv$, $t = vu$ i $x = (uv)^lu$ dla pewnych słów u i v, oraz dla pewnego $l \geqslant 0$. Wobec tego, dla $i = 1, \ldots, k-1$ istnieją słowa U_i, V_i orax liczby l_i takie, że $S_i = U_iV_i$ i $X_1 = S_i^{l_i}U_i < S_i^{l_i+1}$. Ponieważ S_i jest właściwym prefiksem X_1, więc dodatkowo $l_i > 0$.

Pokażemy, że jeżeli $1 < i+1 \leqslant k-1$, to słowa S_i i S_{i+1} komutują. Mamy

$$S_{i+1}S_iB_i = S_{i+1}S_iB_i = S_{i+1}B_{i+1} = B_{i+2} < X_1.$$

W tej sytuacji wystarczy się powołać na lemat 35.6. Istnieje więc słowo proste P takie, że $S_i = P^{r_i}$ oraz $S_{i+1} = P^{r_{i+1}}$. Stosując ponownie lemat 35.6 pokazujemy, że takie słowo P nie tylko istnieje dla każdej pary słów S_i i S_{i+1}, ale jest wspólne dla wszystkich S_i, czyli równość $S_i = P^{r_i}$ są prawdziwe dla $i = 1, \ldots, k-1$.

Zachodzą też równości

$$X_1 = B_kC_k = S_{k-1}B_{k-1}C_k = S_{k-1}S_{k-2}B_{k-2}C_k = \ldots = S_{k-1}S_{k-2}\ldots S_1B_1C_k = P^{\sum_{i=1}^{k-1} r_i}B_1C_k.$$

Stąd wynika, że współczynnik periodyczności X_1 jest

$$\geqslant \sum_{i=1}^{k-1} r_i \geqslant k - 1.$$

Tak więc dowiedliśmy lemat przy dodatkowym założeniu $K = n = 1$.

Jeżeli $n > 1$, to chcielibyśmy wyrzucić część elementów wieży domino i doprowadzić do sytuacji, w której $n = 1$. Nie zawsze jest to możliwe. Jeżeli $B_1 = B_2 = B_3$, $S_1 = S_2 = \varepsilon$ oraz $C_1 = aa$, $C_2 = a$ i $C_3 = ab$, to te trzy ciągi tworzą wieżę domino, ale ciągi B_1, B_3, C_1, C_3 oraz S_1 wieży domino nie tworzą. Zawsze z wieży domino można usunąć pierwszą i ostatnią trójkę (lub kilka początkowych i kilka końcowych trójek). Można też usunąć z wieży domino dowolną trójkę B_j, C_j i S_j (zastępując jednoczeście S_{j-1} przez $S_{j-1}S_j$) lub dowolny zbiór trójek, jeżeli dla wszystkich $i = 1, \ldots, k-1$ i pewnych słów T_i zachodzą równości $C_i = C_{i+1}T_i$ (zamiast $C_iR_i = C_{i+1}T_i$).

Teraz pokażemy lemat przy założeniu, że $C_i = C_{i+1}T_i$ dla $i = 1, \ldots, k-1$ i pewnych T_i. Wtedy można dowieść, że pewne słowo X_t jest równe P^sQ dla $s \geqslant \frac{k}{Kn} - 1$. Aby to zrobić, najpierw z wieży domino usuwamy wszystkie trójki z wyjątkiem tych, których numery są postaci $1 + iK$. Po usunięciu podanych trójek otrzymujemy wieżę o wysokości $\geqslant \frac{k}{K}$. Niech X_l będzie tym słowem, które najczęściej występuje w ciągu $B_1C_1, B_{1+K}C_{1+K}, \ldots, B_{1+iK}C_{1+iK}, \ldots$ Słowo X_l występuje w tym ciągu na przynajmniej $\frac{k}{nK}$ pozycjach. Usuwając z wieży słowa B_i i C_i takie, że $B_iC_i \neq X_l$ otrzymujemy wieżę domino wysokości $\geqslant \frac{k}{nK}$. Dla tej wieży parametry K i n są równe 1. W tej sytuacji z przedstawionego rozumowania wynika, że X_l jest równe P^sQ dla $s \geqslant \frac{k}{nK} - 1$.

Weźmy teraz dowolną wieżę domino. Ciąg długości $\mid B_1 \mid, \mid B_2 \mid, \ldots, \mid B_k \mid$ jest niemalejący i jego wyrazy są mniejsze od długości najdłuższego słowa spośród $X_1, \ldots, X_n$. Liczby $\mid X_1 \mid, \ldots, \mid X_n \mid$ dzielą ciąg $\mid B_1 \mid, \mid B_2 \mid, \ldots, \mid B_k \mid$ na $\leqslant n$ odcinków. Najdłuższy odcinek $\mathcal{X}$ ma $\geqslant \frac{k}{n}$ elementów. Jeżeli z rozważanej wieży usuniemy wszystkie trójki B_i, C_i i S_i z wyjątkiem takich, że $\mid B_i \mid \in \mathcal{X}$ (usuwamy więc tylko pewne początkowe i końcowe fragmenty wieży), to otrzymamy wieżę domino o wysokości $\geqslant \frac{k}{n}$. Zauważmy, że elementy skonstruowanej w ten sposób wieży dla dowolnych i, j spełniają nierówności

$$\mid B_i \mid < \mid B_jC_j \mid. \tag{35}$$

Wieżę $B_1, \ldots, B_{k'}$, $C_1, \ldots, C_{k'}$ oraz $S_1, \ldots, S_{k'-1}$ spełniającą nierówności (35) przekształcamy teraz w wieżę $B_1, \ldots, B_{k'}$, $C'_1, \ldots, C'_{k'}$ oraz $S_1, \ldots, S_{k'-1}$ tej samej wysokości taką, że $C'_{i+1} \leqslant C'_i \leqslant C_i$ dla wszystkich i (zamiast $C'_{i+1} \leqslant C'_i$ możemy napisać $C'_{i+1}T_i = C'_i$ dla pewnego T_i). Sposób przekształcania jest dalej opisany. Jeżeli do przekształconej tak wieży zastosujemy wyżej wykazany rezultat, to otrzymamy tezę twierdzenia.

Przypuśćmy, że $C_r < C_{r+1}$ dla pewnego r. Wtedy $C_{r+1} = C_rQ$ i $B_{r+1}C_{r+1} = YQ$. Stąd słowo $Y = B_{r+1}C_r$ ma sufiks równy B_rC_r. Jeżeli są spełnione nierówności (35), to słowo Y jest dłuższe od każdego słowa B_i. Dzieki temu dla wszystkich i takich, że $B_iC_i = B_{r+1}C_{r+1}$ możemy C_i zastąpić przez C'_i spełniające równość $C'_iQ = C_i$ (także $X_j = B_{r+1}C_{r+1}$ zastępujemy przez Y). Takie zmiany powodują zmniejszenie sumy $\sum_{i=1}^{k'} \mid C_i \mid$ i nie mogą być wykonywane nieskończenie wiele razy. Po wykonaniu pewnej liczby takich poprawek otrzymamy wieżę taką, że $C_{i+1} \leqslant C_i$ dla wszystkich i. □

39.3 Równania uogólnione (graniczne)

W pracy Makanina, zgodnie z zawartą tam terminologia, zostały zdefiniowane równania uogólnione, równania te są nazywane granicznymi (boundary) lub pozycyjnymi (position). Równania graniczne wykorzystują pojęcie granicy. Intuicyjnie, granica to coś, co znajduje się między dwoma sąsiednimi literami słowa.

Równanie uogólnione (wg Makanina) składa się

1) ze skończonego zbioru granic $\{l_1, \ldots l_{g+1}\}$, zbiór granic jest uporządkowany liniowo (np. zgodnie z indeksami),

2) ze zbioru baz $\{x_1, \ldots, x_n, y_1, \ldots, y_n, c_1, \ldots, c_m\}$, bazy c_i nazywamy stałymi, pozostałe nazywamy zmiennymi, bazy x_i i y_i nazywamy (wzajemnie) dualnymi,

3) z funkcji L, R, T, funkcje L, R są określone na bazach, ich wartościami są granice, T przyporządkowuje bazom stałym litery (albo słowa nad alfabetem stałych), dla każdej bazy z zachodzi nierówność $L(z) < R(z)$,

4) ze skończonego zbioru związków granicznych (pojęcie to zostanie wyjaśnione później).

Elementy składające się na równanie uogólnione spełniają też pewne warunki, które nie zostały podane. Intuicyjnie, warunki te stwierdzają, że nie ma niepotrzebnych granic i umożliwiają szacowanie liczby granic.

Przypuśćmy, że mamy funkcję $\alpha : \{l_1, \ldots l_{g+1}\} \to A^+$. Funkcję tę rozszerzamy do (oznaczanej w ten sam sposób funcji $\alpha : \{(l_i, l_j) : 1 \leqslant i \leqslant j \leqslant g+1\} \to A^*$ przyjmując, że

$$\alpha(l_i, l_j) = \alpha(l_i) \ldots \alpha(l_{j-1}).$$

Rozwiązaniem równania uogólnionego nazywamy funkcję $\alpha : \{l_1, \ldots l_{g+1}\} \to A^+$ taką, że

1) $\alpha(L(x_i), R(x_i)) = \alpha(L(y_i), R(y_i))$ dla $i = 1, \ldots, n$,

2) $\alpha(L(c_i), R(c_i)) = T(c_i)$ dla $i = 1, \ldots, m$,

3) oraz spełniającą związki graniczne (zostanie to wyjaśnione w dalszym ciągu).

Współczynnikiem periodyczności rozwiązania α równania uogólnionego nazywamy współczynnik periodyczności słów $\alpha(L(x_i), R(x_i))$ dla $i = 1, \ldots, n$

39.4 Przekształcanie równania w równanie uogólnione

Przypuśćmy, że mamy równanie w słowach $u = w$ i jego rozwiązanie α'. Jeżeli w równaniu podstawimy składowe rozwiązania, to otrzymamy dwa identyczne słowa

$$X_1 X_2 \ldots X_a = \alpha'(u) = \alpha'(v) = Y_1 Y_2 \ldots Y_b,$$

gdzie X_i oraz Y_i są albo składowymi rozwiązania, albo stałymi równania. Rozważmy zbiór

$$\mathcal{X} = \{\varepsilon, X_1, X_1 X_2, \ldots, X_1 X_2 \ldots X_a, Y_1, Y_1 Y_2, \ldots, Y_1 Y_2 \ldots Y_b\}.$$

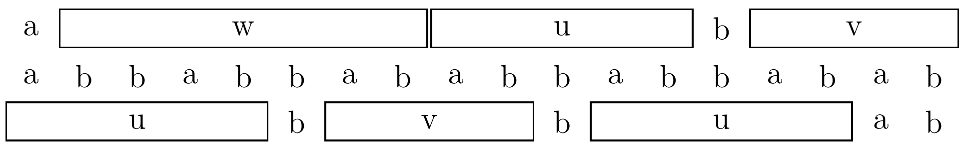
Skonstruujemy równanie uogłonione $\Gamma(u, v)$ (konstrukcja jest przedstawiona na rysunku 4). Zbiór granic tego równania jest równoliczny ze zbiorem $\mathcal{X}$ (niektóre elementy zbioru $\mathcal{X}$ mogą być wymienione dwukrotnie, $\mathcal{X}$ ma mniej elementów niż zostało wymienione w definicji). Niech l będzie różnowartościową funkcją określoną na $\mathcal{X}$ i przyjmującą wartości w zbiorze granic. Zauważmy, że zbiór $\mathcal{X}$ jest liniowo uporządkowany przez relację „być prefiksem". Przyjmujemy, że zbiór granic jest tak uporządkowany, że l jest izomorfizmem porządkowym (krótszy elementom $\mathcal{X}$ odpowadają wcześniejsze (mniejsze) granice).

Teraz określimy bazy równania $\Gamma(u, v)$. Baz stałych jest tyle, ile wystąpień stałych w równaniu $u = v$ i możemy je utożsamiać z wystąpieniami stałych. Funkcja T wystąpieniu litery a przyporządkowuje a. Przypuśćmy, że zmienna z występuje w równaniu $u = w$ na t pozycjach. Najpierw ponumerujemy różne wystąpienia z. Przypuśćmy, że z_i oznacza i-te wystąpienie z. Z każdą parą $z_i = z_{i+1}$ dla $i = 1, \ldots, t-1$ wiązemy dwie bazy zmienne x_i oraz y_i i to samo robimy z pozostałymi zmiennymi. Przyjmijmy, że w rozwiązaniu α' zmiennej z odpowiada słowo Z, ($\alpha'(z) = Z$) oraz i-te wystąpienie z w słowie $\alpha'(u)$ (lub $\alpha'(v)$) jest poprzedzone słowem P, a $i+1$-wsze – słowem Q. Wtedy słowa P, Q, PZ i QZ należą do $\mathcal{X}$. Funkcje L i R definiujemy tak, aby

$$L(x_i) = l(P), \quad R(x_i) = l(PZ), \quad L(y_i) = l(Q) \ \text{ oraz } \ R(y_i) = l(QZ).$$

To jest już pełna definicja równania $\Gamma(u, v)$ (równanie to nie zawiera związków granicznych).

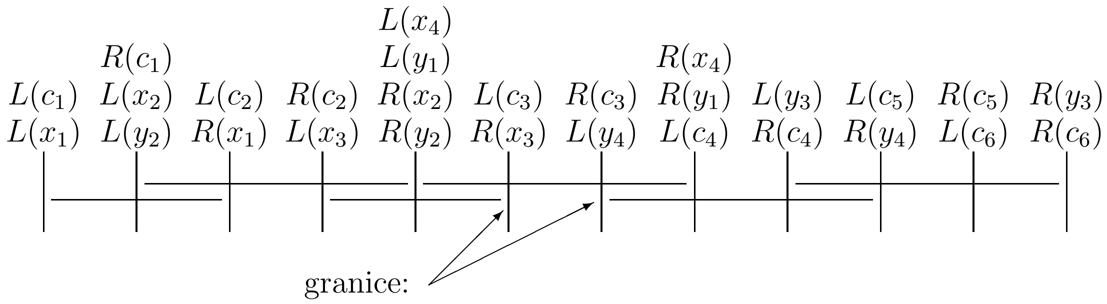
I. Równanie R postaci $awubv = ubvbuab$
i jego rozwiązanie α' takie, że $\alpha'(u) = abbab$, $\alpha'(w) = bbabbab$ i $\alpha'(v) = abab$



II. Równanie uogólnione Γ wyznaczone przez R i α' oraz jego rozwiązania α:

Bazy zmienne: $x_1, x_2, x_3, x_4, y_1, y_2, y_3, y_4$. Bazy stałe: $c_1, c_2, c_3, c_4, c_5, c_6$.

Funkcja T: $T(c_1) = a$, $T(c_2) = b$, $T(c_3) = b$, $T(c_4) = b$, $T(c_5) = a$, oraz $T(c_6) = b$.



α: a $bbab$ b ab ab b ab b ab a b

III. Rozwiązanie γ równania Γ i wyznaczone przez nie rozwiązanie γ' równania R

γ: a $bbaabba$ b ba $abbaab$ b a b $baabba$ a b

$$\gamma'(u) = abbaabba, \quad \gamma'(w) = bbaabbabba, \quad \gamma'(v) = baabbaab.$$

Rysunek 4: Równanie uogólnione odpowiadające równaniu w słowach

Teraz zdefiniujemy rozwiązanie α równania $\Gamma(u,v)$, wyznaczone przez α'. Jeżeli P i Q są dwoma kolejnymi słowami z $\mathcal{X}$ (żadne słowo X spełniające $P < X < Q$ nie należy do clx), to granicy $l(P)$ przyporządkowujemy słowo $\alpha(l(P))$ takie, że $P\alpha(l(P)) = Q$. Nietrudno zauważyć, że tak zdefiniowana funkcja α jest rozwiązaniem $\Gamma(u,v)$.

Także odwrotnie, dowolne rozwiązanie γ rownania $\Gamma(u,v)$ wyznacza rozwiązanie γ' równania $u = v$. Składową rozwiązania γ' odpowiadającą zmiennej z definiujemy w następujący sposób: bierzemy dowolne wystąpienie z, np. wystąpienie z_1, i odpowiadającą temu wystąpieniu bazę, zgodnie z wyżej stosowanymi oznaczeniami jest to baza x_1 (lub y_1) i definiujemy

$$\gamma'(z) = \gamma(L(x_1), R(x_1)).$$

Lemat 39.10 *Dla każdego równania w słowach $u = w$ istnieje równania uogólnione $\Gamma_1, \ldots, \Gamma_n$ takie, że $u = w$ ma rozwiązanie o współczynniku periodyczności s wtedy i tylko wtedy, gdy jedno z równań $\Gamma_1, \ldots, \Gamma_n$ ma rozwiązanie o współczynniku periodyczności s.*

Dowód. Można też dowieść, że jest odpowiedniość między rozwiązaniami równania $u = v$ i równań uogólnionych $\Gamma_1, \ldots, \Gamma_n$. Aby dowieść ten lemat, korzystamy z wyżej przedstawionej konstrukcji. Trzeba zauważyć, że ta konstrukcja nie wymaga znajomości rozwiązania równania $u = w$. Mając dowolne prefiksy $u' < u$ i $w' < w$ wystarczy zdecydować, które ze słów u' i w' stanie się dłuższe po podstawieniu składowych jakiegoś rozwiązania. Nietrudno zauważyć, że takie decyzje możemy podjąc tylko na skończenie wiele sposobów. □

39.5 Związki graniczne

Przypuśćmy, że mamy równanie uogólnione Γ. Związkiem granicznym nazywamy ciąg

$$(l_1, z_1, z_2, \ldots, z_t, l_2),$$

gdzie l_1 i l_2 są granicami, a $z_1, \ldots, z_t$ są bazami równania uogólnionego ($z_i \in \{x_1, \ldots, x_m, y_1, \ldots, y_m\}$), który dla pewnego $u \leqslant t$ spełnia następujące warunki:

1) $L(z_1) < l_1 < R(z_1)$ oraz $L(\bar{z}_t) < l_2 < R(\bar{z}_t)$,

2) $L(z_{i+1}) \leqslant L(\bar{z}_i)$ dla $i = 1, \ldots, u-1$ ($\bar{z}_i$ oznazca bazą dualną do z_i),

3) $L(z_{i+1}) \geqslant L(\bar{z}_i)$ dla $i = u, \ldots, t-1$,

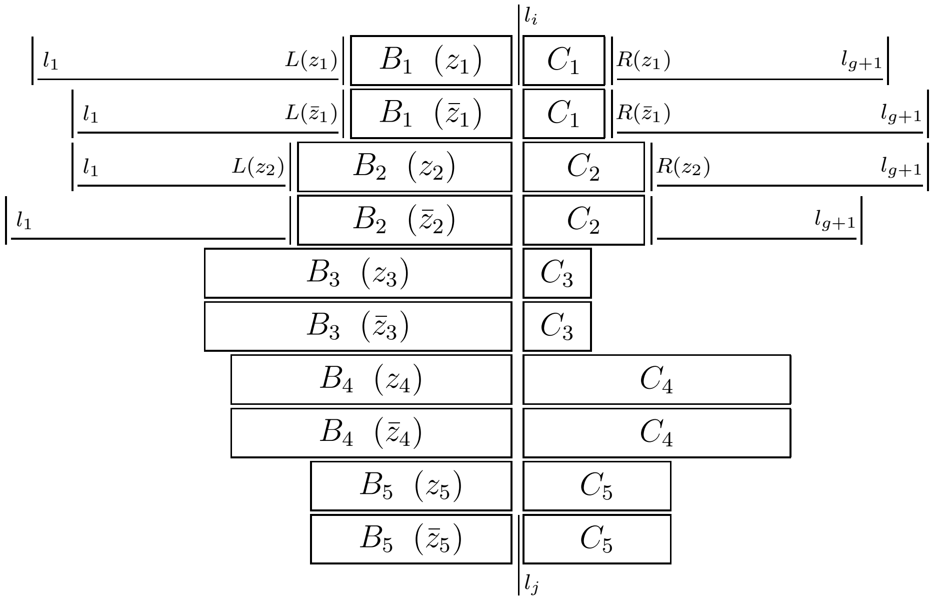
4) gramica l_2 jest wartością funkcji R.

Bazę z_t nazywamy końcową bazą związku $(l_1, z_1, z_2, \ldots, z_t, l_2)$, a liczbę t – jego długością.

Zgodnie z definicją rozwiązanie równania uogólnionego spełnia wszystkie związki graniczne z tego równania. Rozwiązanie α uogólnionego równania Γ spełnia związek graniczny $(l_1, z_1, z_2, \ldots, z_t, l_2)$ (patrz rysunek 5), jeżeli istnieją słowa $B_1, \ldots, B_t, C_1, \ldots, C_t$ takie, że

1) $\alpha(L(z_i), R(z_i)) = B_iC_i$ dla $i = 1, \ldots, t$,

2) $\alpha(L(z_1), l_1) = B_1$ oraz $\alpha(L(z_t), l_2) = B_t$,



Rysunek 5: Związek graniczny $(l_i, z_1, z_2, z_3, z_4, z_5, l_j)$ i jego rozwiązanie

3) $\alpha(L(z_{i+1}), L(\bar{z}_i))B_i = B_{i+1}$ dla $i = 1, \ldots, u-1$,

4) $\alpha(L(\bar{z}_i), L(z_{i+1}))B_{i+1} = B_i$ dla $i = u, \ldots, t-1$.

Lemat 39.11 *Jeżeli rozwiązanie uogólnionego równania spełnia związki graniczne*

$$(l_1, z_1, z_2, \ldots, z_t, l_2) \ \ i \ \ (l_1', z_1, z_2, \ldots, z_t, l_2),$$

to $l_1 = l_1'$.

Dowód. Właściwie szkic dowodu. Różnicę między dwoma granicami w uogólnionym równaniu można mierzyć liczbą granic pośrednich, a jeżeli dane jest rozwiązanie równania, to także sumą długości słów wstawianych przez to rozwiązanie między obie granice. Jeżeli rozwiązanie spełnia związek graniczny $(l_1, z_1, z_2, \ldots, z_t, l_2)$, to znając różnicę między granicami $L(\bar{z}_t)$ i l_2, oraz wzajemne położenie baz $z_1, z_2, \ldots, z_t$ i baz do nich dualnych można ustalić różnicę między granicami $L(z_1)$ i l_1. Wobec tego, granice l_1 i l_1' (ze sformułowania lematu) są tak samo położone w stosunku do $L(z_1)$ i z tego powodu są równe. □

Z powyższy lemat umożliwia dwie rzeczy. Jeżeli znamy liczbę baz równania uogólnionego i potrafimy oszacować długość związków granicznych, to – po pierwsze – potrafimy oszacować liczbę związków granicznych (patrz punkt 4 z definicji związku granicznego). Dzięki temu i dzięki warunkom wspomnianym w rozdziale 39.3 potrafimy też oszacować liczbę granic równania.

39.6 Dopuszczalne równania uogólnione

Wiemy już, że łatwo daje się badać niesprzeczność równań w słowach z jedną stałą literą (równań, w których być może wiele razy występuje tylko jedna litera). Analogiczny fakt jest prawdziwy dla równań uogólnionych i ma analogiczny dowód.

Lemat 39.12 *Istnieje algorytm, który rozwiązuje problem niesprzeczności uogólnionego równania w słowach z jedną stała.* □

Uogólnione równanie nazywamy dopuszczalnym, jeżeli występują w nim dwie litery (funkcja T przyjmuje przynajmniej dwie wartości) i po zastąpieniu wszystkich występujących w nim liter jedną wybraną otrzymujemy równanie niesprzeczne. Łatwo zauważyć, że jeżeli w równaniu niesprzecznym utożsamimy wszystkie litery, to otrzymamy równanie, które też jest niesprzeczne. Tak więc równania dopuszczalne to takie, których niesprzeczność nie da się łatwo rozstrzygnąć korzystając z algorytmu z podanego lematu. Jest też oczywiste, że istnieje algorytm, który bada, czy dane równanie uogólnione jest dopuszczalne, a jeżeli nie jest dopuszczalne, to rozstrzyga, czy jest niesprzeczne.

Lemat 39.13 *Istnieje obliczalna funkcja f taka, że jest $\leqslant f(n, m, d)$ uogólnionych równań dopuszczalnych z najwyżej $2n$ bazami zmiennymi i m bazami stałym, i w których związki graniczne mają długość nie przekraczającą d.*

Dowód. W rzeczywistości f zależy od jeszcze jednego argumentu, który jest potrzebny do szacowania liczby granic i w najważniejszych zastosowaniach lematu daje sie oszacować przez n.

Dowodzony lemat jest konsekwencją lematu 39.11. Założenia lematu 39.11 są prawdziwe dla równań dopuszczalnych po utożsamieniu występujących w nich stałych. Dzięki temu możemy oszacować za pomocą parametrów n, m i d liczbę związków granicznych w równaniu, a w konsekwencji – także liczbę granic. □

39.7 Znormalizowane równania uogólnione

Precyzyjna definicja równania znormalizowanego nie zostanie podana. Za to zauważmy, że związki graniczne można bez trudu nieistotnie wydłużać. Na przykład, rozwiązanie równania uogólnionego spełnia związek graniczny

$$(l_1, z_1, \ldots, z, \bar{z}, z, \ldots, z_t, l_2)$$

wtedy i tylko wtedy, gdy spełnia związek

$$(l_1, z_1, \ldots, z, \bar{z}, z, \bar{z}, z \ldots, \bar{z}, z, \ldots, z_t, l_2).$$

W równaniu uogólninym wystarczy więc zamieścić tylko krótszy z tych związków. W równaniu znormalizowanym znajdzie się najwyżej krótszy z tych dwóch związków.

Makanin udowodnił, że każde równanie uogólnione można znormalizować nie zwiększając parametrów tego równania (patrz lemat 39.16), a także następujący lemat:

Lemat 39.14 *Jeżeli znormalizowane równanie uogólnione ma n par baz dualnych, jest w nim związek graniczny długości d i ma rozwiązanie o współczynniku periodyczności s, to $d \leqslant 4 \cdot (s+1) \cdot n^4$.*

Dowód. Jest to prosta konsekwencja lematu 39.9 i faktu, że dla związków granicznych równań znormalizowanych w tym lemacie można przyjąć $K = 2n^2$. □

39.8 Przekształcanie równań uogólnionych, indeks rozwiązania

Przekształcanie równań uogólnionych przypomina transformacje Nielsena, a algorytm Makanina przypomina algorytm Plotkina. Algorytm Plotkina dla danej pary złożonej z równania i jego rozwiązania pozwala znaleźć podobną parę z krótszym rozwiązaniem. W przypadku algorytmu Makanina jest podobnie, ale długość rozwiązania trzeba zastąpić pojęciem indeksu. W przeciwieństwie do długości rozwiązania, indeks zależy nie tylko od rozwiązania, ale zależy także od równania (jest określony dla pary złożonej z równania uogólnionego i jego rozwiązania). Tutaj indeks nie zostanie zdefiniowany. Może zostać zdefiniowany jako liczba naturalna, w szczególności indeks można zmniejszać tylko skończenie wiele razy.

Nie zostaną także zdefiniowane wykorzystane przez Makanina przekształcenia równań uogólnionych. Przekształcenia te mają własności podane w następującym lemacie.

Lemat 39.15 *Dla dowolnego znormalizowanego dopuszczalnego równania uogólnionego Γ istnieje równanie uogólnione Γ_1 spełniające następujące warunki:*

1) *główne parametry równania, w tym liczby baz stałych i zmiennych, nie zwiększają się,*

2) *jeżeli Γ ma rozwiązanie z indeksem I i współczynnikiem periodyczności s, to Γ_1 ma rozwiązanie z indeksem mniejszym od I i współczynnikiem periodyczności mniejszym lub równym s,*

3) *jeżeli Γ_1 ma rozwiązanie, to Γ ma rozwiązanie.* □

Teraz możemy też sformułować lemat o normalizacji.

Lemat 39.16 *Dla dowolnego równania uogólnionego Γ istnieje ciąg znormalizowanych równań uogólnionych $\Gamma_1, \ldots, \Gamma_k$ taki, że*

1) *każde równanie z ciągu $\Gamma_1, \ldots, \Gamma_k$, które nie jest dopuszczalne, jest sprzeczne,*

2) *główne parametry, w tym liczby baz stałych i zmiennych, opisujące równania $\Gamma_1, \ldots, \Gamma_k$ nie przekraczają głównych parametrów równania Γ,*

3) *jeżeli Γ ma rozwiązanie z indeksem I i współczynnikiem periodyczności s, to dla pewnego i równanie Γ_i ma rozwiązanie z indeksem $\leqslant I$ i współczynnikiem periodyczności s,*

4) *jeżeli dla pewnego i równanie Γ_i ma rozwiązanie, to Γ też ma rozwiązanie.* □

39.9 Opis algorytmu Makanina

Algorytm Makanina zostanie opisany jako niedeteministyczny.

Krok 1. Dla danego równania w słowach znajdujemy liczbę s_0 ograniczającą z góry współczynnik periodyczności minimalnego rozwiązania (patrz twierdzenie 39.7),

Krok 2. dane równanie zastępujemy niedeterministycznie równaniem uogólnionym Γ (jednym z wielu skonstruowanych w lemacie 39.10, równanie to jest znormalizowane, ponieważ nie ma w nim związków granicznych),

Krok 3. ustalamy liczby n_0 par baz zmiennych i m_0 baz stałych równania Γ, inicjujemy licznik pętli podstawiając liczbę $f(n_0, m_0, 4(s_0 + 1)n_0^4)$, gdzie f jest funkcją z lematu 39.13 szacującą liczbę znormalizowanych równań uogólnionych,

Krok 4. jeżeli wartość licznika jest mniejsza od 0, to przerywamy wykonywanie algorytmu bez akceptowania, w przeciwnym razie zmniejszamy wartość licznika o 1,

Krok 5. jeżeli w Γ nie występują dwie różne stałe, to sprawdzamy, czy jest niesprzeczne korzystając z algorytmu wspomnianego w lemacie 39.12, kończymy wykonywnie algorytmu przekazując uzyskaną informację o niesprzeczności Γ (jeżeli Γ jest sprzeczne, to przerywamy wykonywanie algorytmu bez akceptowania),

Krok 6. korzystając ponownie z algorytmu z lematu 39.12 sprawdzamy, czy równanie Γ jest dopuszczalne, jeżeli nie, to jest sprzeczne i przerywamy wykonywanie algorytmu bez akceptowania,

Krok 7. teraz sprawdzamy, czy równanie ma związek graniczny o długości większej niż $4(s_0+1)n_0^4$, jeżeli tak, to przerywamy pracę algorytmu bez akceptowania,

Krok 8. w przeciwnym razie do Γ stosujemy przekształcenie opisane w dowodzie lematu 39.15, a następnie otrzymane w ten sposób równanie zastępujemy jednym z równań znormalizowanych, o których mowa w lemacie 39.16 (wynik przekształceń zapamiętujemy w zmiennej Γ),

Krok 9. wracamy do kroku 4.

39.10 Poprawność algorytmu Makanina

Twierdzenie 39.17 *Algorytm Makanina rozwiązuje problem niesprzeczności równań w słowach.*

Dowód. Najpierw pokażemy, że pozytywna odpowiedź algorytmu Makanina jest poprawna. Zauważmy, że algorytm Makanina wykorzystuje trzy konstrukcje równań uogólnionych, opisane w lematach 39.10, 39.15 i 39.16. Lematy te gwarantują, że jeżeli skonstruowane na ich podstawie równanie jest niesprzeczne, to zostało otrzymane z równania niesprzecznego. Tak więc, jeżeli podczas wykonania algorytmu Makanina pojawi się niesprzeczne równanie Γ, to wyjściowe równanie w słowach też jest niesprzeczne. Algorytm Makanina przekazuje informację, że badane równanie jest niesprzeczne tylko w kroku 5 po sprawdzeniu, czy równanie Γ jest nieprzeczne. Poprawność algorytmu Makanina wynika więc z poprawności algorytmu zastosowanego w kroku 5.

Przypuśćmy teraz, że dane jest niesprzeczne równanie w słowach. Bierzemy minimalne rozwiązanie α' tego równania. Załóżmy, że α ma współczynik periodyczności s. W kroku 2 algorytm wybiera niedeterministycznie równanie Γ. Zgodnie z lematem 39.10, równanie Γ można tak wybrać, aby miało rozwiązanie α o współczynniku periodyczności s. Z twierdzenia 39.7 wynika, że $s \leqslant s_0$. Jeżeli równanie Γ jest dopuszczalne, to w kroku 8 zostanie zastąpione (niedeterministycznie) pewnym równaniem Γ_1. Zgodnie z lematami 39.15 i 39.16 równanie Γ_1 można tak wybrać, aby miało rozwiązanie α_1 o indeksie mniejszym niż indeks α, o współczynniku periodyczności $\leqslant s \leqslant s_0$ i o liczbach baz odpowiednio $\leqslant 2n_0$ i $\leqslant m_0$. Dalej będziemy analizować sposób wykonania algorytmu Makanina, w którym krok 8 zawsze jest tak realizowany. Realizowany w taki sposób algorytm Makanina konstruuje ciąg równań $\Gamma_1, \Gamma_2, \ldots$ z ciągiem rozwiązań $\alpha_1, \alpha_2, \ldots$ o coraz mniejszych indeksach, o współczynnikach periodyczności $\leqslant s_0$, z bazami, których jest odpowiednio $\leqslant 2n_0$ i $\leqslant m_0$. Ponieważ maleją indeksy rozwiązań, więc te ciągi są skończone. Będziemy analizować, z jakiego powodu tak realizowany algorytm Makanina zakończy pracę.

Tak realizowany algorytm Makanina nie zakończy pracy ani w kroku 5, ani w kroku 6 stwierdzając sprzeczność równania, gdyż wszystkie równania Γ_i są niesprzeczne, także po utożsamieniu stałych. Nie przerwie też pracy w kroku 8, ponieważ każde z równań Γ_i jest znormalizowane i ma rozwiązanie α_i o współczynniku periodyczności $\leqslant s_0$, a zgodnie z lematem 39.14, takie równania mają związki graniczne długości$\leqslant 4(s_0 + 1)n_0^4$. Są więc tylko dwie możliwości zatrzymania algorytmu Makanina: albo występująca w nim pętla zostanie wykonana maksymalną liczbę razy, albo zakończy pracę stwierdzając niesprzeczność danego równania.

Rozważane ciągi równań $\Gamma_1, \Gamma_2, \ldots$ z rozwiązaniami $\alpha_1, \alpha_2, \ldots$ można konstruować na wiele sposobów. Wszystkie są skończone. Teraz dodatkowo założymy, że rozważamy najkrótsze takie ciągi i ciągi te mają długość t. Przypuśćmy, że algorytm Makanina analizując ciąg $\Gamma_1, \Gamma_2, \ldots, \Gamma_t$ wykonał zawartę w nim pętlę maksymalną liczbę razy, a więc $t > f(n_0, m_0, 4(s_0+1)n_0^4)$. Każde równanie w ciągu $\Gamma_1, \Gamma_2, \ldots, \Gamma_t$ ma $\leqslant n_0$ baz zmiennych, $\leqslant m_0$ baz stałych i związki graniczne o długości $\leqslant 4(s_0 + 1)n_0^4$. Na podstawie lematu 39.13, takich równań jest $\leqslant f(n_0, m_0, 4(s_0 + 1)n_0^4)$. Istnieją więc i, j, $i < j \leqslant t$ takie, że $\Gamma_i = \Gamma_j$. Oznacza to, że ciąg $\Gamma_1, \Gamma_2, \ldots, \Gamma_t$ nie jest najkrótszym, rozważanym ciągiem. Realizując algorytm Makanina mamy prawo rozważać także krótszy ciąg $\Gamma_1, \Gamma_2, \ldots \Gamma_i, \Gamma_{j+1}, \ldots, \Gamma_t$. W ten sposób uzyskaliśmy sprzeczność z wyborem ciągu $\Gamma_1, \Gamma_2, \ldots, \Gamma_t$.

Uzyskana sprzeczność świadczy o tym, że po uruchomieniu algorytmu Makanina z niesprzecznym równaniem w słowach jest taki spoób pracy algorytmu, który prowadzi do stwierdzenia, że dane równanie jest niesprzeczne. $\square$