

Równania napisów

Artur JEŹ*

Przyjrzyjmy się problemowi równań napisów¹: dla ustalonego alfabetu, np. a, b, c, \dots , piszemy równania używające liter z tegoż alfabetu oraz zmiennych, które będziemy oznaczać przez X, Y, Z, \dots . Dla uproszczenia pomyślmy o jednym równaniu, np. $aXXXX = XaYY$. Rozwiązaniem będą takie napisy, złożone z liter alfabetu, że po podstawieniu pod zmienne uzyskamy takie same napisy. W naszym przykładzie podstawienie $X = aa, Y = aaa$ jest rozwiązaniem.² Innym przykładem jest równanie $XbaYb = baaababbab$, którego rozwiązaniem jest np. $X = baaa, Y = bba$.

³ Równania tego typu rozważane były od lat 60. XX wieku, z jednej strony ze względu na związki z 10. problemem Hilberta, z drugiej ze względu na związki z teorią grup... Ale obie te historie to tematy na osobne artykuły.

Pierwszy sposób rozwiązywania tego typu równań podał G. Makanin w 1977 roku, w roku 1999 inne, dużo prostsze podejście zaproponował W. Plandowski. My przyjrzymy się dziś rozwiązaniu upraszczającemu drugie z nich, na podstawie mojego artykułu *Recompression: A Simple and Powerful Technique for Word Equations*, (Journal of the ACM, 2016 63:1, art. 4).

Zanim przystąpimy do prezentacji sposobu rozwiązywania takich równań, należałoby ustalić, do czego będziemy dążyć: chcemy podać algorytm, czyli systematyczny przepis pokazujący, jak uzyskać rozwiązanie (lub stwierdzić, że go nie ma). Przykładem algorytmu dla układu równań liniowych jest: „Powtarzaj: weź pierwsze równanie, jeśli jest niesprzeczne i zawiera zmienną, to wylicz jedną zmienną w oparciu o inne i podstaw do pozostałych równań, usuń równanie”. Przykładem algorytmu dla równania kwadratowego $ax^2 + bx + c = 0$ jest: „Wylicz $\Delta = \sqrt{b^2 - 4ac}$, jeśli $\Delta < 0$, to nie ma rozwiązań, w przeciwnym przypadku rozwiązania to $(-b \pm \sqrt{\Delta})/2a$ ”. Podamy tego typu algorytm, choć dłuższy i bardziej skomplikowany, dla równań napisów.

Nasz algorytm będzie wielokrotnie sprawdzał niezależnie kilka możliwości, np. osobno sprawdzi możliwości, że pierwsza litera podstawienia pod X to a lub b , a potem osobno, czy poza tą jedną literą coś jeszcze w podstawieniu jest, czy nie, itd. Jeśli istnieje rozwiązanie, to algorytm znajdzie je dla którejś sekwencji możliwości, dla innych być może nie znajdzie.⁴ Nie brzmi to jak dobry pomysł – w takim razie algorytm mógłby po prostu przejrzeć wszystkie możliwe podstawienia, litera po literze... Dlatego też zagwarantujemy coś jeszcze – algorytm będzie przekształcał równanie w taki sposób, że dla którejś z możliwości wiodących do rozwiązania rozważane równanie zawsze będzie miało najwyżej $8n^2$ liter i n zmiennych, gdzie n to długość równania wejściowego.⁵ W szczególności dla niektórych możliwości nie znajdziemy rozwiązania (trudno!), a inne doprowadzą nas do równania dłuższego niż $8n^2 + n$ – wtedy algorytm przestaje dalej liczyć.⁶ Takie podejście można też zrealizować inaczej, bez odwoływania się do sprawdzania wszystkich możliwości: możemy wypisać wszystkie równania długości najwyżej $8n^2 + n$, zaznaczyć, które równania możemy przekształcić w które, i sprawdzić, czy potrafimy przekształcić wejściowe równanie w jakieś bardzo proste niesprzeczne równanie, np. $a = a$ albo $X = X$.

Od teraz skupimy się na wykazaniu, że algorytm zawsze ma „dobry wybór”, tj. potrafi przekształcić niesprzeczne równanie długości nie większej niż $8n^2 + n$ w niesprzeczne równanie długości również nie większej niż $8n^2 + n$. Ważne jest też, że niezależnie od wyboru nie przekształcimy równania sprzecznego w niesprzeczne.

⁷Równanie napisów może mieć wiele rozwiązań i algorytm jest w stanie podać reprezentację wszystkich; dla uproszczenia będziemy wymagać jedynie, o ile istnieje, aby podał dowolne. W szczególności: jeśli istnieje rozwiązanie, to istnieje też takie, które używa tylko liter występujących w równaniu, choć sam alfabet może być większy – wszystkie inne litery w rozwiązaniu można zastąpić przez ustaloną literę; nowe podstawienie dalej jest rozwiązaniem.⁸ Będziemy też zakładać, że podstawienie pod zmienną jest *niepuste*, tj. w podstawieniu $X = w$ słowo w ma dodatnią długość.

Aby rozwiązać równanie napisów, będziemy *skracać* jego krótkie fragmenty: popatrzmy jeszcze raz na równanie $XbaYb = baaababbab$ (które ma rozwiązanie $X = baaa, Y = bba$): jeśli zastąpimy ba przez „świeżą”, tj. nieużywaną w równaniu,

literę c , to otrzymane równanie $XcYb = caacbc$ ma rozwiązanie $X = caa$, $Y = bc$. Łatwo też sprawdzić, że jeśli nowe równanie ma rozwiązanie, to stare też ma – wystarczy zamienić w podstawieniu każde c przez ba . Niestety, ten pomysł nie działa, jeśli w oryginalnym równaniu spróbujemy zastąpić ab przez d : po lewej stronie równania nie ma żadnego ab , w podstawieniu $X = baaa$, $Y = bba$ również nie ma, jest za to po prawej stronie równania. Problemem jest to, że ab jest częściowo w podstawieniu i częściowo poza. Parę ab , gdzie $a \neq b$, nazwiemy *złą* (dla ustalonego rozwiązania), jeśli b jest pierwszą literą podstawienia pod X oraz aX występuje w równaniu lub a jest ostatnią literą podstawienia pod Y oraz Yb występuje w równaniu (lub dla powyższych założeń XY występuje w równaniu). Parę, która nie jest zła, nazwiemy *dobrą*.⁹ Jeśli para ab , gdzie $a \neq b$, jest dobra, to skrócenie ab daje równanie, które ma odpowiadające, krótsze rozwiązanie.¹⁰ Jednocześnie, jeśli nowe równanie ma rozwiązanie, to stare też miało: wystarczy każde c zastąpić przez ab .

⁹Tu przydaje się informacja, że podstawienia pod zmienne są niepuste.

¹⁰Czyli przez kontrapozycję: jeśli stare równanie nie miało rozwiązania, to nowe też nie ma.

¹¹Tu znów istotne jest, że podstawienia pod zmienne są niepuste.

¹²A skąd wiadomo, że jest zła? Trzeba osobno sprawdzić wszystkie możliwości...

¹³Ponownie, wszystkie możliwości trzeba osobno sprawdzić.

Co do złych par, to łatwo zauważyć, że jeśli równanie ma n wystąpień zmiennych, to jest najwyżej $2n$ różnych złych par: po jednej na każdą stronę zmiennej.¹¹ Co więcej, ustaloną złą parę łatwo udobruchać (czyli uczynić dobrą): jeśli ab jest zła,¹² to dla każdej zmiennej sprawdzamy: jeśli jej podstawienie zaczyna się na b , czyli $X = bw$, to zastępujemy X przez bX (a podstawienie przez $X = w$), analogicznie, jeśli podstawienie kończy się na a , to zastępujemy X przez Xa . Jeśli w ten sposób podstawienie za X stało się puste, to usuwamy X z równania.¹³ Łatwo sprawdzić, że nowe równanie ma rozwiązanie oraz para ab jest dobra; jeśli zaś nowe równanie ma rozwiązanie, to miało je też stare równanie: wystarczy dokleić litery, które „wypchnęliśmy” ze zmiennych. W nowym równaniu możemy skrócić ab .

Złe pary to nie jedyny problem: skoro zakładamy, że w parze ab litery a i b są różne, to nie potrafimy skrócić długich powtórzeń jednej litery, np. nie wiemy, co zrobić z równaniem $aXXX = XaYY$ z pierwszego przykładu. Zamiast skracać pary, będziemy skracać *maksymalne powtórzenia*: a^ℓ jest *maksymalnym powtórzeniem* a ,¹⁴ jeśli na lewo i prawo nie ma litery a .¹⁵ Podobnie jak w przypadku par, możemy zdefiniować, co to znaczy, że litera a jest *zła* (jest zmienna X zaczynająca się na a i aX występuje w równaniu lub...) i *dobra*. Jeśli a jest dobre, to możemy równocześnie zastąpić wszystkie maksymalne wystąpienia a , powiedzmy dla każdego ℓ zastępujemy maksymalne a^ℓ przez a_ℓ ; przy czym a_ℓ to tylko wygodna nazwa dla nas, algorytm nie musi potem pamiętać, że ta litera powstała z a^ℓ . Jeśli a jest zła, to chcemy ją „udobruchać”, ale zastąpienie X przez aX może nie wystarczyć, gdyż podstawienie pod X może dalej zaczynać się na a . Zamiast tego, jeśli $X = a^\ell w a^r$, gdzie w nie zaczyna się ani nie kończy na a , to zastępujemy X przez $a^\ell X a^r$,¹⁶ odpowiednio zmieniając podstawienie pod X (teraz $X = w$). Wszystkie możliwe ℓ, r chcemy sprawdzić osobno i tu pojawia się problem: ℓ i r mogą być dowolnie duże. Wiadomo jednak, że jeśli istnieje rozwiązanie równania, to istnieje też rozwiązanie, w którym wszystkie takie ℓ, r mają najwyżej wykładniczą długość.¹⁷ To wciąż dużo, ale zamiast pisać ℓ razy a , możemy zapisać (a, ℓ) , a zapis pozycyjny ℓ ma wielomianowo wiele cyfr. I tak zaraz skrócimy wszystkie maksymalne powtórzenia a . Tak samo jak dla par możemy wykazać, że jest najwyżej $2n$ złych liter, a jeśli rozpatrzmy jednocześnie złe pary i litery, to podobnie stwierdzamy, że w sumie złych par i liter jest najwyżej $2n$.

¹⁴Oznaczamy $a^\ell = \overbrace{a \cdots a}^{\ell \text{ razy}}$, czyli ℓ razy powtórzone a .

¹⁵To uwzględnia przypadek, że na lewo lub prawo nie ma żadnej litery.

¹⁶Jeśli $X = a^\ell$, to zastępujemy X przez a^ℓ , usuwając przy okazji X .

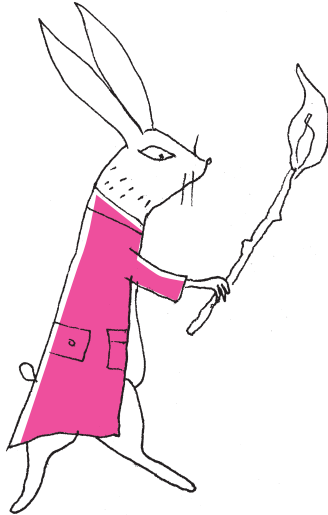
¹⁷Ten problem sprowadza się do układu liniowych równań diofantycznych. Wiadomo, że jeśli ma on rozwiązanie, to jest ono najwyżej wykładnicze. Ponownie, jest to temat na osobny artykuł...

Teraz możemy już podać wybory algorytmu, które prowadzą do rozwiązania, a równanie ma niezmiennie co najwyżej $8n^2$ liter (i najwyżej n zmiennych): dla danego równania, jeśli istnieje dobra para lub litera, to skracamy ją. Jeśli są tylko złe pary i litery, to wybieramy taką, że po jej udobruchaniu oraz skróceniu otrzymane równanie jest najkrótsze.¹⁸

¹⁸Znów, sprawdzamy wszystkie wybory, ale interesuje nas tylko ten dobry.

Przeanalizujmy algorytm dla tych możliwości: zauważmy najpierw, że jeśli równanie nie ma rozwiązania, to nieważne, co zrobimy, nowe równanie również nie będzie miało rozwiązania. Twierdzimy też, że przy właściwych wyborach równanie nigdy nie będzie miało więcej niż $8n^2$ liter,¹⁹ gdzie n to długość równania danego na wejściu: Jeśli istnieje dobra para lub litera, to po jej skróceniu nowe równanie jest krótsze niż stare. A co, jeśli są tylko złe pary i litery? Niech równanie ma m liter, nie licząc zmiennych. Różnych złych par i liter jest w sumie najwyżej $2n$. Dla każdej litery rozpatrzmy literę na prawo (lub lewo, dla ostatniej); ta sąsiadna

¹⁹Plus nie więcej niż n zmiennych.



litera może pochodzić z podstawienia pod zmienną. Razem tworzą parę albo są częścią maksymalnego bloku. Czyli dla każdego wystąpienia litery a w równaniu istnieje zła para lub litera, której wystąpienie zawiera to ustalone wystąpienie a . Skrócenie tej pary/litery usunie to ustalone wystąpienie a . Jako że jest $2n$ złych par i liter, „średnio” skrócenie usunie przynajmniej $\frac{m}{2n}$ liter oraz wprowadzi co najwyżej połowę z tej liczby (skrócenie pary usuwa dwie litery i dodaje jedną, a skrócenie bloku liter usuwa wiele liter i dodaje jedną). Tak więc istnieje takie skrócenie, które w sumie skróci równanie o przynajmniej $\frac{m}{4n}$ liter. Musimy też wykonać jedno udobruchanie, które wprowadza do równania najwyżej $2n$ liter: udobruchanie litery wprowadza długie powtórzenia tej litery, ale zaraz każde takie powtórzenie zostanie zastąpione przez jedną literę. Tak więc nowe równanie ma najwyżej m (stare litery) $- m/4n$ (usunięte przez skrócenie) $+ 2n$ (udobruchanie) liter. Oszacujmy:

$$m - m/4n + 2n = m(1 - 1/4n) + 2n \leq 8n^2(1 - 1/4n) + 2n = 8n^2,$$

jedyna nierówność wynika z ograniczenia $m \leq 8n^2$. I to jest koniec: dla podanych wyżej wyborów równanie znajdzie rozwiązanie i równania, które rozważa, mają najwyżej $8n^2$ liter (i najwyżej n zmiennych).

A czego nie wiemy o równaniach napisów? Dokładniejsza analiza algorytmu pokazuje, że jeśli istnieje rozwiązanie, to istnieje też rozwiązanie podwójnie wykładnicze. Nie jest jednak znane równanie, którego najkrótsze rozwiązanie jest ponad wykładnicze; popularna jest hipoteza, że jeśli istnieje rozwiązanie, to istnieje też rozwiązanie wykładnicze. Do równań napisów można też dodać dodatkowe warunki na rozwiązanie, np. żądać, by suma długości podstawień pod X i Y była równa długości podstawienia pod Z . Nie wiadomo, czy równania napisów z liniowymi warunkami na długości podstawień można rozwiązać.



Zadania

Przygotował Łukasz BOŻYK

M 1633. Pewne n przekątnych $2n$ -kąta foremnego przecina się w jednym punkcie, który nie jest wierzchołkiem tego wielokąta. Wykazać, że jest jego środkiem.
Rozwiązanie na str. 2

M 1634. Wierzchołki $2n$ -kąta foremnego oznaczono przez P_1, P_2, \dots, P_{2n} , niekoniecznie w tej kolejności. Udowodnić, że łamana zamknięta $P_1P_2 \dots P_{2n}$ zawiera parę odcinków równoległych.
Rozwiązanie na str. 2

M 1635. Prostokąt nazwiemy *parzystym*, jeśli każdy z jego wymiarów jest parzystą liczbą całkowitą. Kwadrat $n \times n$, gdzie n jest liczbą nieparzystą, podzielono na części, z których każda jest parzystym prostokątem lub kwadratem 1×1 . Znaleźć najmniejszą możliwą liczbę kwadratów 1×1 uzyskanych w takim podziale.
Rozwiązanie na str. 15

Przygotował Andrzej MAJHOFER

F 997. Jądro ^{238}U może ulec rozpadowi α lub samorzutnemu rozszczepieniu na dwa mniejsze jądra. W czasie $t = 1$ godzina zaobserwowano $N_f = 25$ reakcji rozszczepienia zachodzących w próbce 1 g czystego ^{238}U . Oszacuj, ile wynosi czas połowicznego zaniku dla ^{238}U ze względu na rozszczepienie? Stała Avogadro $N_A = 6,022 \cdot 10^{23}$. (Dla porównania: czas połowicznego zaniku ze względu na rozpad α wynosi $4,47 \cdot 10^9$ lat).
Rozwiązanie na str. 9

F 998. Podczas ruchu ciała w powietrzu siła oporu jest proporcjonalna do iloczynu kwadratu prędkości v ciała i pola S jego przekroju poprzecznego, prostopadłego do kierunku ruchu: $F_{\text{op}} = -kSv^2$. Wartość stałej k zależy od kształtu ciała. Siła oporu powoduje, że spadające swobodnie ciała po pewnym czasie spadania osiągają stałą prędkość spadku – prędkość graniczną. Jaki jest stosunek prędkości granicznych osiąganych przez kulki

- z tego samego materiału o masach m i $8m$?
- o tych samych rozmiarach, ale o gęstości ρ i 8ρ ?

Rozwiązanie na str. 16