Ciągi Fareya

Jakub PAWLEWICZ[*]

Jak szybko wypisać wszystkie ułamki z przedziału $[0, 1]$ o mianownikach nieprzekraczających 8, posortowane rosnąco? Będzie to tak zwany *ciąg Fareya* rzędu 8. Następująca technika działa zaskakująco dobrze. Zaczynamy od listy składającej się z dwóch ułamków: $\frac{0}{1}, \frac{1}{1}$. Następnie między każde kolejne dwa ułamki wstawiamy ułamek o liczniku będącym sumą ich liczników i mianowniku będącym sumą mianowników. Powtarzamy to tak długo, aż nie będziemy mogli wstawić już żadnego ułamka o odpowiednio małym mianowniku. I tak, między $\frac{0}{1}$ i $\frac{1}{1}$ wstawiamy $\frac{0+1}{1+1} = \frac{1}{2}$:

$$\frac{0}{1}, \frac{1}{2}, \frac{1}{1}.$$

Między $\frac{0}{1}$ i $\frac{1}{2}$ wstawiamy $\frac{0+1}{1+2} = \frac{1}{3}$, a między $\frac{1}{2}$ i $\frac{1}{1}$ wstawiamy $\frac{1+1}{2+1} = \frac{2}{3}$:

$$\frac{0}{1}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{1}{1}.$$

Między $\frac{0}{1}$ i $\frac{1}{3}$ wstawiamy $\frac{0+1}{1+3} = \frac{1}{4}$, między $\frac{1}{3}$ i $\frac{1}{2}$ wstawiamy $\frac{1+1}{3+2} = \frac{2}{5}$, między $\frac{1}{2}$ i $\frac{2}{3}$ wstawiamy $\frac{1+2}{2+3} = \frac{3}{5}$ oraz między $\frac{2}{3}$ i $\frac{1}{1}$ wstawiamy $\frac{2+1}{3+1} = \frac{3}{4}$:

$$\frac{0}{1}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{3}{4}, \frac{1}{1}.$$

Dalej między $\frac{0}{1}$ i $\frac{1}{4}$ wstawiamy $\frac{0+1}{1+4} = \frac{1}{5}$, między $\frac{1}{4}$ i $\frac{1}{3}$ wstawiamy $\frac{1+1}{4+3} = \frac{2}{7}$, między $\frac{1}{3}$ i $\frac{2}{5}$ wstawiamy $\frac{1+2}{3+5} = \frac{3}{8}$, między $\frac{2}{5}$ i $\frac{1}{2}$ wstawiamy $\frac{2+1}{5+2} = \frac{3}{7}$, między $\frac{1}{2}$ i $\frac{3}{5}$ wstawiamy $\frac{1+3}{2+5} = \frac{4}{7}$, między $\frac{3}{5}$ i $\frac{2}{3}$ wstawiamy $\frac{3+2}{5+3} = \frac{5}{8}$, między $\frac{2}{3}$ i $\frac{3}{4}$ wstawiamy $\frac{2+3}{3+4} = \frac{5}{7}$ oraz między $\frac{3}{4}$ i $\frac{1}{1}$ wstawiamy $\frac{3+1}{4+1} = \frac{4}{5}$:

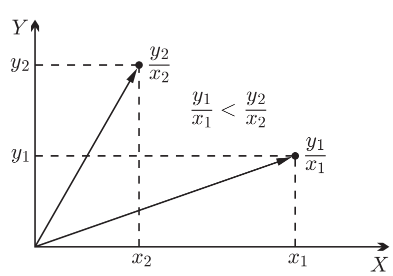
$$\frac{0}{1}, \frac{1}{5}, \frac{1}{4}, \frac{2}{7}, \frac{1}{3}, \frac{3}{8}, \frac{2}{5}, \frac{3}{7}, \frac{1}{2}, \frac{4}{7}, \frac{3}{5}, \frac{5}{8}, \frac{2}{3}, \frac{5}{7}, \frac{3}{4}, \frac{4}{5}, \frac{1}{1}.$$

Między $\frac{0}{1}$ i $\frac{1}{5}$ wstawimy jeszcze ułamki $\frac{1}{8}, \frac{1}{7}, \frac{1}{6}$ oraz między $\frac{4}{5}$ i $\frac{1}{1}$ wstawimy jeszcze $\frac{5}{6}, \frac{6}{7}, \frac{7}{8}$. Między pozostałe ułamki już nic nie wstawimy, bo suma mianowników jest większa niż 8. Ostatecznie otrzymujemy listę:

$$\frac{0}{1}, \frac{1}{8}, \frac{1}{7}, \frac{1}{6}, \frac{1}{5}, \frac{1}{4}, \frac{2}{7}, \frac{1}{3}, \frac{3}{8}, \frac{2}{5}, \frac{3}{7}, \frac{1}{2}, \frac{4}{7}, \frac{3}{5}, \frac{5}{8}, \frac{2}{3}, \frac{5}{7}, \frac{3}{4}, \frac{4}{5}, \frac{5}{6}, \frac{6}{7}, \frac{7}{8}, \frac{1}{1}.$$

Są to wszystkie szukane ułamki. Pozostaje wykazać, dlaczego ta metoda jest poprawna.

W numerze 6/2008 *Delty* pojawił się artykuł wyjaśniający tę i inne własności ciągów Fareya za pomocą okręgów Forda, odwołujący się do własności inwersji względem okręgu. W tym artykule prezentujemy inną ciekawą interpretację geometryczną ułamków Fareya, która pozwala w prosty sposób uzasadnić przedstawioną metodę ich generowania.



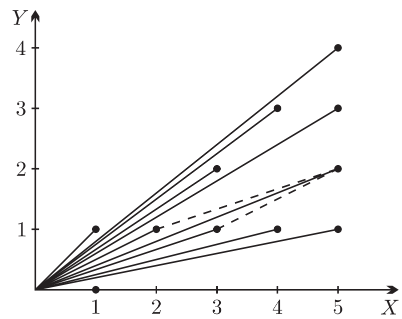
Rys. 1. Ułamki jako punkty i wektory, porównywanie ułamków.

Każdy ułamek $\frac{y}{x}$ $(x, y \geqslant 0)$ będziemy utożsamiali z punktem (x, y) lub z wektorem o początku w punkcie $(0, 0)$ i końcu w (x, y). Taka reprezentacja ma tę miłą własność, że porównywanie ułamków można wykonywać przez sprawdzanie nachylenia odpowiadających im wektorów względem osi X (rysunek 1). Wektorowi $(1, 0)$ będziemy przyporządkowywać ułamek $\frac{0}{1}$, a wektorowi $(0, 1)$ ułamek $\frac{1}{0}$, który odpowiada nieskończoności.

[*]Instytut Informatyki UW

Mówimy, że punkt $(x, y) \in \mathbb{Z}^2$ jest punktem *prymitywnym*, jeżeli $\mathrm{NWD}(x, y) = 1$. Punkty prymitywne są też nazywane punktami widocznymi ze środka układu współrzędnych. Bierze się to stąd, że jeżeli punkt (x, y) jest prymitywny, to na odcinku łączącym punkty $(0, 0)$ i (x, y) nie znajduje się żaden inny punkt o współrzędnych całkowitych. Ułamkom nieskracalnym odpowiadają punkty prymitywne.

Geometryczna interpretacja generowania ciągu Fareya rzędu n jest następująca. Na początku rysujemy wektory $(1, 0)$ i $(1, 1)$. Następnie między dwa kolejne wektory wstawiamy wektor będący ich sumą. Powtarzamy to dopóty, dopóki możemy wstawić wektor o współrzędnej x nieprzekraczającej n. W ten sposób jesteśmy w stanie znaleźć wszystkie punkty prymitywne w trójkącie ograniczonym osią X, prostą $y = x$ i prostą $x = n$.



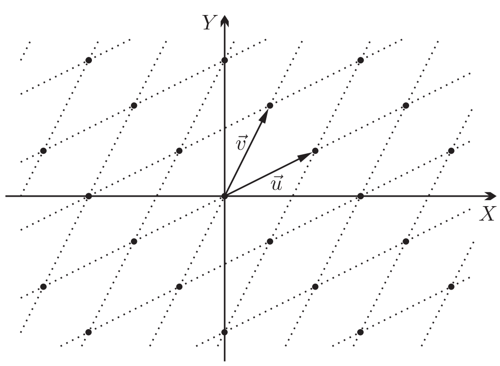
Rys. 2. Ciąg Fareya rzędu 5 w ujęciu geometrycznym; wstawienie ułamka $\frac{2}{5}$ pomiędzy $\frac{1}{3}$ i $\frac{1}{2}$.

Dlaczego to działa? Aby udowodnić poprawność tej metody, wprowadzimy kilka nowych pojęć. W poniższych definicjach zakładamy, że wektory $\vec{u}$ i $\vec{v}$ mają współrzędne nieujemne i nie są równoległe.

Kratą rozpiętą na wektorach $\vec{u}$ i $\vec{v}$ nazywamy zbiór

$$L(\vec{u}, \vec{v}) = \{a\vec{u} + b\vec{v} \mid a, b \in \mathbb{Z}\}.$$

Wektory $\vec{u}$ i $\vec{v}$ nazywamy bazą kraty.



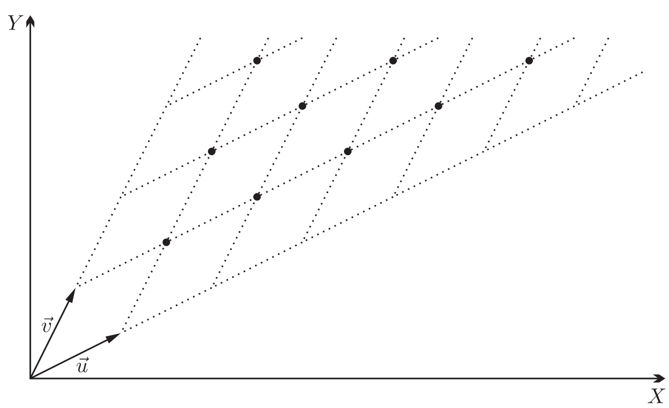
Rys. 3. Krata $L(\vec{u}, \vec{v})$.

Kratę $L(\vec{u}, \vec{v})$ nazywamy *pełną*, jeżeli zawiera wszystkie punkty kratowe (tzn. punkty o obu współrzędnych całkowitych):

$$L(\vec{u}, \vec{v}) = \mathbb{Z}^2.$$

Ćwierćkratą rozpiętą na wektorach $\vec{u}$ i $\vec{v}$ nazywamy zbiór

$$L^+(\vec{u}, \vec{v}) = \{a\vec{u} + b\vec{v} \mid a, b \in \mathbb{Z}^+\}.$$



Rys. 4. Ćwierćkrata $L^+(\vec{u}, \vec{v})$.

Ćwierćkratę $L^+(\vec{u}, \vec{v})$ nazywamy *pełną*, jeżeli składa się ze wszystkich punktów kratowych zawartych w ćwierćpłaszczyźnie wyznaczonej przez wektory $\vec{u}$ i $\vec{v}$:

$$L^+(\vec{u}, \vec{v}) = \{a\vec{u} + b\vec{v} \mid a, b \in \mathbb{R}^+\} \cap \mathbb{Z}^2.$$

Sprawdzanie pełności ćwierćkraty sprowadza się do sprawdzania pełności kraty. Wynika to z dosyć oczywistego faktu.

Fakt 1. *Krata $L(\vec{u}, \vec{v})$ jest pełna wtedy i tylko wtedy, gdy ćwierćkrata $L^+(\vec{u}, \vec{v})$ jest pełna.*

Dowód. Załóżmy, że krata $L(\vec{u}, \vec{v})$ jest pełna. Weźmy dowolny punkt kratowy (x, y) należący do ćwierćpłaszczyzny wyznaczonej przez wektory $\vec{u}$ i $\vec{v}$: $(x, y) = a\vec{u} + b\vec{v}$, przy czym $a, b \in \mathbb{R}^+$. Ponieważ (x, y) jest punktem kratowym, więc z tego, że krata $L(\vec{u}, \vec{v})$ jest pełna, wynika, że a i b są całkowite. Stąd $a, b \in \mathbb{Z}^+$, a zatem $(x, y) \in L^+(\vec{u}, \vec{v})$.

Załóżmy teraz, że ćwierćkrata $L^+(\vec{u},\vec{v})$ jest pełna. Weźmy dowolny punkt o współrzędnych całkowitych (x,y). Mamy jednoznacznie wyznaczone $a,b\in\mathbb{R}$, takie że $(x,y)=a\vec{u}+b\vec{v}$ (dlaczego?). Niech n będzie taką liczbą całkowitą, że $a+n>0$ i $b+n>0$. Wtedy punkt $(x',y')=(x,y)+n\vec{u}+n\vec{v}$ jest punktem kratowym znajdującym się wewnątrz ćwierćkraty rozpiętej na wektorach $\vec{u}$ i $\vec{v}$, a ponieważ $L^+(\vec{u},\vec{v})$ jest ćwierćkratą pełną, więc istnieją $a',b'\in\mathbb{Z}^+$, takie że $(x',y')=a'\vec{u}+b'\vec{v}$. Wynika stąd, że $a=a'-n$ i $b=b'-n$ są całkowite, więc $(x,y)\in L(\vec{u},\vec{v})$. $\square$

Ta sama krata może być opisana przez różne pary wektorów $\vec{u}$ i $\vec{v}$. W szczególności, krata pełna może mieć różne bazy. Przykładową kratą pełną jest $L((1,0),(0,1))$. Poniższy fakt pozwoli nam na tworzenie innych baz kraty pełnej.

Fakt 2. *Jeżeli $L(\vec{u},\vec{v})$ jest kratą pełną, to $L(\vec{u},\vec{u}+\vec{v})$ i $L(\vec{u}+\vec{v},\vec{v})$ również są kratami pełnymi.*

Dowód. Jeżeli $(x,y)=a\vec{u}+b\vec{v}$ i $a,b\in\mathbb{Z}$, to $(x,y)=(a-b)\vec{u}+b(\vec{u}+\vec{v})$, skąd wynika, że $L(\vec{u},\vec{u}+\vec{v})$ jest kratą pełną. Podobnie pokazujemy, że $L(\vec{u}+\vec{v},\vec{v})$ jest kratą pełną. $\square$

Poczyńmy jeszcze jedno proste spostrzeżenie.

Fakt 3. *Jeżeli $L(\vec{u},\vec{v})$ jest kratą pełną, to $\vec{u}+\vec{v}$ jest punktem prymitywnym.*

Dowód. Na półprostej wychodzącej z punktu $(0,0)$ w kierunku $\vec{u}+\vec{v}$ najbliższym punktem kratowym jest $\vec{u}+\vec{v}$, co wynika z pełności $L(\vec{u},\vec{v})$. $\square$

Teraz możemy szybko udowodnić poprawność metody generowania listy ułamków. Otóż jest ona równoważna generowaniu wszystkich punktów prymitywnych. Startujemy od listy złożonej z dwóch wektorów $(1,0)$ i $(1,1)$. Następnie wstawiamy między dwa wektory ich sumę tak długo, jak współrzędna x nie przekracza n. W dowolnym momencie lista wektorów jest posortowana według nachylenia. Niezmiennik jest następujący. Po pierwsze, wektory z listy są punktami prymitywnymi. Po drugie, każde kolejne dwa wektory są bazą pełnej ćwierćkraty. Na podstawie faktów 1, 2 i 3, wstawienie między dwa wektory ich sumy zachowuje niezmiennik. Wynika stąd, że w ten sposób wszystkie wektory reprezentują ułamki nieskracalne. Co więcej, każdy ułamek zostanie wygenerowany.

Wyjaśnijmy to drugie spostrzeżenie. Załóżmy przeciwnie, że jakiś punkt prymitywny (x,y), $x\leqslant n$ (czyli ułamek $\frac{y}{x}$) nie został wygenerowany. Wtedy punkt ten należy do jakiejś ćwierćkraty rozpiętej na pewnych dwóch kolejnych wektorach z listy. Ponieważ jest to już końcowa lista, więc współrzędna x sumy tych wektorów jest większa od n, co oznacza, że wszystkie punkty kratowe w danej ćwierćkracie mają współrzędną x większą od n, a to stanowi sprzeczność.

Rozwiązanie zadania M 1278.
Ponieważ liczba n daje z dzielenia przez 4 resztę 3, więc liczba n nie może być kwadratem liczby całkowitej. Wobec tego wszystkie dzielniki liczby n można połączyć w rozłączne pary $(d,n/d)$. Wystarczy zatem wykazać, że dla każdego dzielnika d liczby n suma $d+n/d$ jest podzielna przez 24.

Ponieważ liczby d i 24 są względnie pierwsze, więc wystarczy udowodnić, że

$$24 \mid d(d+n/d)=d^2+n.$$

Liczba $n+1$ jest podzielna przez 24, więc dowodzona podzielność sprowadza się do wykazania, że

$$24 \mid d^2-1=(d-1)(d+1).$$

Ponieważ liczby d i 24 są względnie pierwsze, więc liczba d jest nieparzysta i niepodzielna przez 3. Stąd $8 \mid (d-1)(d+1)$ (obie liczby $d-1$ oraz $d+1$ są parzyste i jedna z nich dzieli się przez 4) oraz $3 \mid (d-1)(d+1)$ (jedna z liczb $d-1$ lub $d+1$ jest podzielna przez 3). Stąd ostatecznie $24 \mid (d-1)(d+1)$.