Arytmetyka interwałowa, czyli kiedy $x \cdot x \neq x^2$

Marek GUTOWSKI

Hasło „arytmetyka interwałowa" zapewne niewiele mówi większości Czytelników. I nic w tym dziwnego, gdyż ten sposób rachowania istnieje od niedawna; dość powiedzieć, że poświęcony mu kwartalnik *Interval Computations* kończy właśnie trzeci rok istnienia. Ale do rzeczy.

Interwałem nazywamy uporządkowaną parę liczb rzeczywistych (x_1, x_2), takich że $-\infty < x_1 \leq x_2 < +\infty$. Nie należy jednak mylić obiektów zwanych interwałami z liczbami zespolonymi czy współrzędnymi punktów płaszczyzny, które też są uporządkowanymi parami liczb rzeczywistych, ale o zupełnie innych własnościach.

W dalszym ciągu będziemy oznaczać interwały dużymi literami alfabetu lub zapisywać w takiej samej formie, jak przywykliśmy przedstawiać przedziały (czyli właśnie interwały) liczbowe, np. $X = [x_1, x_2] = \{x : x_1 \leq x \leq x_2\}$, gdyż narzucającą się interpretacją geometryczną takiego tworu jest, oczywiście. odcinek położony na osi liczbowej. Szczególny, zdegenerowany, interwał, to $[x, x]$ odpowiadający zwyczajnej liczbie rzeczywistej. Cztery podstawowe działania arytmetyki interwałowej zdefiniowane są następująco (w miejsce $\circ$ możemy wstawić dowolny z symboli: $+, -, \cdot, /$):

$$(*) \qquad X \circ Y = [\min(x \circ y), \max(x \circ y)], \quad \text{gdzie } x \in X,\ y \in Y.$$

Tak więc wynikami operacji artymetycznych na interwałach są również interwały. Konkretnie wygląda to tak:

$$X + Y = [x_1, x_2] + [y_1, y_2] = [x_1 + y_1, x_2 + y_2],$$

$$X - Y = [x_1, x_2] - [y_1, y_2] = [x_1 - y_2, x_2 - y_1] \ (!),$$

$$X \cdot Y = [x_1, x_2] \cdot [y_1, y_2] =$$

$$= [\min(x_1y_1, x_1y_2, x_2y_1, x_2y_2), \max(x_1y_1, x_1y_2, x_2y_1, x_2y_2)],$$

$$X/Y = [x_1, x_2]/[y_1, y_2] =$$

$$= [\min(x_1/y_1, x_1/y_2, x_2/y_1, x_2/y_2), \max(x_1/y_1, x_1/y_2, x_2/y_1, x_2/y_2)].$$

Ostatni wynik jest dobrze określony (tzn. wynik jest też interwałem) tylko w przypadku, gdy mianownik nie zawiera zera, tzn. albo $y_1 > 0$, albo $y_2 < 0$, w pełnej analogii do zwyczajnego dzielenia. Proszę też zwrócić uwagę na różnice między dodawaniem a odejmowaniem. Mnożenie wydaje się nieco skomplikowane, ale nietrudno się przekonać, że podany wzór jest zgodny z definicją (*). Nietrudno sprawdzić, że mnożenie i dodawanie są przemienne, tzn. $X + Y = Y + X$ oraz $X \cdot Y = Y \cdot X$ dla dowolnych X i Y, tak, jak by się chciało.

Spotykają nas jednak pewne niespodzianki. Niech np.

$$A = [1, 2], \quad B = [-1, 2] \quad \text{oraz} \quad C = [2, 3].$$

Wtedy

$$A \cdot B + A \cdot C = [1, 2] \cdot [-1, 2] + [1, 2] \cdot [2, 3] = [-2, 4] + [2, 6] = [0, 10],$$

gdy tymczasem

$$A \cdot (B + C) = [1, 2] \cdot ([-1, 2] + [2, 3]) = [1, 2] \cdot [1, 5] = [1, 10].$$

Tak więc prawo rozdzielności mnożenia względem dodawania nie działa w arytmetyce interwałowej, bo $A \cdot B + A \cdot C \neq A \cdot (B + C)$. Można jednak dowieść, że zawsze, jeśli tym razem interwał rozumieć jako przedział, czyli zbiór liczb, $A \cdot (B + C) \subseteq A \cdot B + A \cdot C$.

Wiedząc, jak operować na interwałach w zakresie czterech działań arytmetycznych, można by się pokusić o badanie funkcji, których argumentami są interwały. Najprostsze z nich to takie, które w wyniku dają liczby rzeczywiste, np.

- centrum (środek) interwału: $m([x_1, x_2]) = (x_1 + x_2)/2$,
- szerokość interwału: $w([x_1, x_2]) = x_2 - x_1$,
- średnica interwału: $d([x_1, x_2])$, oznaczana też symbolicznie jako $|[x_1, x_2]| = \max(|x_1|, |x_2|)$.

Rozwiązanie zadania F 865.
Z twierdzenia o wiriale wynika, że średnia wartość czasowa wyrażenia $\frac{1}{2}r\frac{\mathrm{d}\Phi}{\mathrm{d}r}$ jest równa energii kinetycznej E_k cząstki. Podstawiając $V = ar^4$ otrzymujemy zależność między potencjałem a energią kinetyczną

$$V = \frac{1}{2}E_k\,.$$

Całkowita energia jest więc równa:

$$E_1 = V + E_k = \frac{3}{2}E_k\,.$$

Energia kinetyczna jest związana z temperaturą ciała zależnością $E_k = \frac{3}{2}\mathrm{k}T$, gdzie k jest stałą Boltzmanna. Stąd energia pojedynczego atomu wynosi $E_1 = \frac{9}{4}\mathrm{k}T$. Mnożąc energię E_1 przez liczbę Avogadro N_A oraz uwzględniając, że $\mathrm{R} = N_A\mathrm{k}$ (R to stała gazowa), otrzymujemy energię jednego mola

$$E = \frac{9}{4}\mathrm{R}T\,,$$

skąd ciepło molowe wynosi

$$C = \frac{9}{4}\mathrm{R}\,.$$

Analogiczny rachunek dla potencjału postaci $\frac{1}{r}$ prowadzi do ciepła molowego ciał stałych

$$C_0 = 3\mathrm{R}\,.$$

Porównując otrzymujemy

$$\frac{C}{C_0} = \frac{3}{4}\,.$$

Rozwiązanie zadania F 866.
Twierdzenie o wiriale łączy pochodną potencjału z energią kinetyczną cząstki poruszającej się w tym potencjale

$$r\frac{\mathrm{d}V}{\mathrm{d}r} = 2E_k\,.$$

Podstawiając wyrażenie na potencjał otrzymujemy

$$E_k = -V\,.$$

Stąd całkowita energia cząstki jest równa

$$E = V + E_k = 0\,.$$

Oznacza to, że cząstka w takim potencjale nie może wytwarzać trwałych stanów.

W naturalny sposób można też tak rozszerzyć określenia wielu znanych funkcji, aby mogły one operować na interwałach i w wyniku również dawać interwały. Robimy to formalnie tak: Niech dana będzie funkcja f określona na liczbach. Wówczas definiujemy funkcję F określoną na interwałach w sposób następujący:

$$(**) \qquad F(X) = F([x_1, x_2]) = [\inf_{x\in X}(f(x)), \sup_{x\in X}(f(x))]\,.$$

Słownie można to wyrazić tak: wynikiem działania świeżo zdefiniowanej funkcji F na danym odcinku (interwale) jako argumencie jest interwał rozciągający się od kresu dolnego do kresu górnego wartości funkcji f na tymże odcinku. A oto przykłady.

Logarytm:

$$\mathrm{LOG}([x_1, x_2]) = [\log(x_1), \log(x_2)] \quad (x_1 > 0)\,.$$

Funkcja *signum* (znak) może dać w wyniku tylko jeden z pięciu interwałów, z których trzy są zdegenerowane: $[-1,-1]$, $[-1,0]$, $[0,0]$, $[0,1]$, $[1,1]$.

Podnoszenie do kwadratu:

$$([x_1, x_2])^2 = \begin{cases} [\min(x_1^2, x_2^2), \max(x_1^2, x_2^2)], & \text{jeśli interwał } [x_1, x_2] \text{ nie zawiera zera,} \\ [0, \max(x_1^2, x_2^2)] & \text{w pozostałych przypadkach.} \end{cases}$$

Zauważmy interesujący paradoks: $X \cdot X$ nie musi być równe X^2!, tj. wynik podnoszenia interwału do kwadratu jest na ogół inny niż wynik mnożenia dwóch identycznych interwałów. Podnoszenie do kwadratu daje interwał o takiej samej lub mniejszej szerokości niż mnożenie dwóch jednakowych czynników. Różnica pojawia się wtedy, gdy interwał $[x_1, x_2]$ zawiera zero (i tylko wtedy). Paradoks jest, oczywiście, pozorny (mnożenie jest funkcją dwóch zmiennych, a operacja podnoszenia do kwadratu ma tylko jeden argument — są to więc dwie różne funkcje), ilustruje jednak, że w rachunkach interwałowych wskazana jest ostrożność.

W przypadku funkcji bardziej skomplikowanych, zwłaszcza funkcji wielu zmiennych, byłoby wielce niewygodne posługiwanie się bezpośrednio definicją (**). Mając wzór określający „zwykłą" funkcję liczbowo-liczbową konstruujemy jej rozszerzenie interwałowe posługując się regułami (*). Otrzymany w taki sposób wynik jest na ogół inny niż ten, który można by otrzymać z definicji (**). To, co otrzymamy, nazywa się funkcją inkluzywną (obejmującą, zawierającą), gdyż produkuje ona jako wyniki interwały zawierające (obejmujące) prawdziwy wynik. Sztuka polega na tym, aby znajdować możliwie dobre funkcje inkluzywne, czyli takie, które nie „zawyżają" zanadto wyniku obliczeń, tj. dają interwały o możliwie małej szerokości. Na przykład, zapis funkcji dwóch zmiennych $f(x_1, x_2) = x_1^2 + x_1x_2$ lepiej będzie wstępnie przekształcić do postaci $x_1 \cdot (x_1 + x_2)$.

Do czego może być przydatny taki sposób rachowania?

Jedno z zastosowań to obliczanie rozmaitych wielkości na podstawie niepewnych lub niedokładnych danych (np. projektując jakieś urządzenie jesteśmy zmuszeni do operowania takimi właśnie wielkościami – patrz tolerancje parametrów wykonania elementów mechanicznych lub elektrycznych).

Przykład. Obliczyć przyspieszenie grawitacyjne z okresu wahań wahadła matematycznego. Odpowiedni wzór ma postać: $\mathrm{g} = 4\pi^2 l/T^2$. Długość l wahadła wynosi 1 m, a błąd pomiaru jest nie większy od 1 mm, natomiast zmierzony czas 300 wahnięć wyniósł 602 s z błędem nie przekraczającym 0,5 s. Możemy więc powiedzieć, że $l \in [0{,}999, 1{,}001]$ m, a $T \in [2{,}005, 2{,}0083333]$ s. Posługując się definicjami (*) i (**), oraz traktując czynnik $4\pi^2$ jako interwał zdegenerowany, otrzymamy ostatecznie:

$$\mathrm{g} \in [9{,}7780815,\ 9{,}8302602]\ \mathrm{m/s^2}\,,$$

co można też zapisać jako

$$\mathrm{g} = (9{,}8041376 \pm 0{,}0260561)\ \mathrm{m/s^2}\,.$$

Nawiasem mówiąc, przytoczyliśmy wyniki w postaci „surowej", tj. w takiej, w jakiej ukazały się na ekranie kalkulatora, choć, oczywiście, nie ma to większego sensu, przynajmniej dla fizyka, który z pewnością ograniczyłby się do podania mniejszej liczby cyfr znaczących, np. $9{,}804 \pm 0{,}026$ m/s^2. Uzasadnieniem jest wyraźnie przesadzona dokładność w określeniu T.

Drugą dobrze opracowaną dziedziną jest rozwiązywanie układów równań nieliniowych. Znany jest algorytm pozwalający na znalezienie **wszystkich** rozwiązań rzeczywistych danego układu równań we wskazanym obszarze. Chodzi, oczywiście, o algorytm numeryczny. Idea jego jest prosta: załóżmy, że układ równań potrafimy zapisać w postaci $f_i(x_1, x_2, \ldots, x_k) = 0$, gdzie i numeruje równania, a $x_1, x_2, \ldots, x_k$ to niewiadome. Wskazany obszar jest wielowymiarową kostką o bokach odpowiadających minimalnym i maksymalnym wartościom poszukiwanych niewiadomych, innymi słowy – iloczynem kartezjańskim odcinków, w których poszukujemy rozwiązań w każdej ze zmiennych. Jeśli teraz skonstruujemy dobrą funkcję inkluzywną dla każdego równania, to możemy dla danej kostki rozstrzygnąć, czy może ona zawierać jakieś rozwiązanie badanego układu. (Dobra funkcja inkluzywna to taka, która w wyniku daje interwał zdegenerowany, jeżeli jej argument też jest interwałem zdegenerowanym – krótko mówiąc, daje w wyniku możliwie „małe" interwały.) Warunek jest prosty: każde równanie **musi** „przechodzić przez zero" wewnątrz danej kostki. Jeśli choć jedno z równań nie spełnia tego warunku, to w danej kostce na pewno nie ma rozwiązania. Najczęściej jednak wynikiem takiego testu będzie odpowiedź „nie wiadomo". W takim przypadku dzielimy kostkę na dwie części, tnąc prostopadle do najdłuższego boku. W kostkach „potomnych" znów przeprowadzamy prosty test przejścia przez zero, i tak dalej, aż pozostaną tylko bardzo małe kostki o rozmiarach porównywalnych z dokładnością maszynową. Jeśli funkcje, które opisują nasze równania, są ponadto różniczkowalne, to możliwe jest stwierdzenie pozytywne, że dana kostka z **pewnością** zawiera przynajmniej jedno rozwiązanie. Widać, że sposób podziału kostek na mniejsze zapewnia, że każda z nich w końcu stanie się „mała", oraz że za pomocą metody nie jesteśmy w stanie znaleźć rozwiązań zespolonych, a jedynie rzeczywiste. Ponadto, nie jest określona krotność rozwiązań, ale to zwykle jest mniej interesujące dla osób, które po prostu poszukują jakichkolwiek rozwiązań trudnych problemów.

Skoro mowa o algorytmach, które kojarzą się zwykle z komputerami, to należy dodać, że istnieją już kompilatory FORTRANu oraz PASCALa (PASCAL XSC) „znające się" na danych typu INTERVAL i mające wbudowanych kilka funkcji interwałowych. Niestety, kompilatory takie są zainstalowane, jak dotąd, wyłącznie w dużych komputerach. Użytkownicy komputerów osobistych mogą sobie zdefiniować w PASCALu swój własny typ interwałowy, na przykład tak:

```
Type interval=record
x1, x2: real
end;
```

oraz napisać odpowiednie procedury wykonujące cztery działania arytmetyczne na interwałach. Kto nie ma dostępu do komputera, może się zastanowić, traktując to, jako rozrywkę umysłową, jak skonstruować interwałowe rozszerzenie funkcji dwóch zmiennych $f(x_1, x_2) = \max(x_1, x_2)$.

Literatura:

1. G. Alefeld, J. Herzberger, *Introduction to Interval Computations*, Academic Press, New York, 1983.
2. R.B. Kearfott, *Some Tests of Generalized Bisection*, ACM Transactions on Mathematical Software, vol. 13(1987), str. 197.

Rozwiązanie zadania M 679.
Przypuśćmy, że teza zadania jest fałszywa; wtedy $2A_n$ dzieli się przez $n+2$. Mamy jednak

$$\begin{aligned} 2A_n = &2 + (2^{1993} + n^{1993}) + \\ &+ (3^{1993} + (n-1)^{1993}) + \ldots + \\ &+ ((n-1)^{1993} + 3^{1993}) + \\ &+ (n^{1993} + 2^{1993}), \end{aligned}$$

a każda z liczb postaci $j^{1993} + (n+2-j)^{1993}$, gdzie $j = 2, 3, \ldots, n$, dzieli się bez reszty przez $n+2$ – to wynika ze wzoru

$$x^{2m+1} + y^{2m+1} = (x+y)(x^{2m} - x^{2m-1}y + \ldots + y^{2m}).$$

Stąd wynika, że $2A_n$ daje z dzielenia przez $n+2$ resztę 2, a to jest sprzeczność.

Rozwiązanie zadania M 680.
Łatwo zauważyć, że po każdym kroku nie zmienia się parzystość liczby białych kul w urnie. Zatem, jeśli początkowa liczba kul białych n jest parzysta, to ostatnia kula w urnie będzie na pewno czarna, czyli szukane prawdopodobieństwo jest równe zeru. Jeśli zaś n jest nieparzyste, to ostatnia kula w urnie na pewno będzie biała – szukane prawdopodobieństwo jest równe jedności. Ostatecznie, prawdopodobieństwo jest równe $\frac{1+(-1)^{n-1}}{2}$.

Rozwiązanie zadania M 681.
Załóżmy, że d jest wspólnym dzielnikiem rozpatrywanych liczb; wtedy każda z liczb

$$\binom{n+j}{k-1} = \binom{n+j+1}{k} - \binom{n+j}{k}, \quad j = 0, 1, \ldots, k-1$$

także dzieli się przez d. Postępując dalej podobnie, stwierdzimy w końcu, że liczba

$$\binom{n}{0} = 1$$

dzieli się przez d, czyli $d = 1$. zatem największy wspólny dzielnik rozpatrywanych liczb także jest równy 1.