A. Jančařík

Početní algoritmy IV - testy prvočíselnosti

Učitel matematiky, Vol. 15 (2007), No. 4, 205–212

Persistent URL: http://dml.cz/dmlcz/150680

POČETNÍ ALGORITMY IV – TESTY PRVOČÍSELNOSTI

ANTONÍN JANČAŘÍK[1]

V předchozím článku o početních algoritmech jsme se seznámili s pojmem *výpočetní složitosti algoritmu.* Na praktických ukázkách bylo demonstrováno, že různé algoritmy provádí stejný výpočet v různém čase. Ukázali jsme si také, že některé algoritmy nemají, a to ani s využitím nejmodernější techniky, nejmenší šanci dokončit výpočet. Pokud algoritmus, který se snaží provést rozklad „velkého" čísla na prvočísla, bude potřebovat pro dokončení výpočtu asi 31 623 153 207 852 661 404 573 973 miliard let, výsledku se od něj asi nedočkáme. V tomto díle nahlédneme trochu pod pokličku *kryptografům* (odborníkům na šifrování) a odborníkům na teorii čísel a ukážeme si, jak asi fungují algoritmy, které lze efektivně použít pro ověřování prvočíselnosti a pro rozklady velkých čísel na prvočísla. Na těchto dvou matematických problémech stojí bezpečnost nejčastěji používaného algoritmu pro šifrování – RSA. Je to s podivem, ale mnoho, a to i nejnovějších a nejefektivnějších algoritmů využívá velmi jednoduché myšlenky, které vychází ze vztahů, s nimiž se seznamují žáci již na základní či střední škole.

Testy prvočíselnosti

Prvním algoritmem, se kterým se seznámíme, je test prvočíselnosti založený na malé Fermatově větě. Testy prvočíselnosti jsou navrženy tak, aby se pokusily zjistit, zda zadané číslo je číslo složené. Avšak pouze některé testy prvočíselnosti potvrdí, že testované číslo je skutečně prvočíslo. Jinými slovy, pokud test zjistí, že zadané číslo je složené, tak je jisté, že číslo není prvočíslo. Naopak, pokud test nezjistí, že číslo je složené, často u některých testů zůstává možnost, že číslo přece jen složené je.

[1]Příspěvek byl vypracován s podporou grantu GAČR 406/05/P561.

Test prvočíselnosti, se kterým se žáci běžně seznamují na základní škole – postupné testování všemi případnými děliteli – není tohoto typu. Tento test dává jednoznačnou odpověď, číslo buď je prvočíslo nebo číslo složené. Další zvláštností tohoto testu je to, že v případě složeného čísla současně nachází i jeho dělitele. Běžné testy prvočíselnosti řeší pouze otázku, zda číslo je složené, o jeho dělitelích neposkytují žádnou informaci. Z tohoto pohledu je testování všech možných dělitelů lepším testem, bohužel je však tento test velmi pomalý a proto je v „reálných" případech nepoužitelný.

Testem prvočíselnosti v našem slova smyslu by byl test, kdy žák testuje dělitelnost pouze čísly, kde zná kriteria dělitelnosti. Pokud malé číslo není dělitelné 2, 3, 5, 7 ani 11, tak bude „pravděpodobně" prvočíslo. Tento test však selhává u tak velkého množství čísel, že je naprosto nevhodný.

Prvočísla a pseudoprvočísla

Malá Fermatova věta patří mezi velmi známé matematické výsledky. Tvrzení, že pro všechna prvočísla p a všechna přirozená čísla x platí vztah

$$x^{p-1} \equiv 1 \pmod{p} \tag{1}$$

se však po dlouhou dobu hodil pouze pro příklady matematické olympiády. Převrat přinesl příchod internetu, kdy se našlo velmi praktické využití této věty v RSA algoritmu. Malá Fermatova věta však nabízí i velmi přirozený test prvočíselnosti. Chceme-li otestovat, zda nějaké číslo je prvočíslo, zvolíme si náhodně přirozené číslo a otestujeme, zda platí rovnost (1).

Ukázka 1

Chceme otestovat, zda číslo 91 je prvočíslo. Zvolíme si například $x = 3$ a ověříme, že $3^{90} \equiv 1 \pmod{91}$. Přesto, že číslo 3^{90} má při zápisu v desítkové soustavě 43 cifer, ověřit uvedený vztah není náročné. Pro výpočet totiž nepotřebujeme vědět, kolik 3^{90} přesně je, nepočítáme totiž, kolik je 3^{90}, ale jaký je zbytek 3^{90} po dělení 91. Výpočet může probíhat například takto (všechny výpočty

jsou prováděny modulo 91):

$$3^1 = 3$$
$$3^2 = 9$$
$$3^4 = 3^2 \cdot 3^2 = 9 \cdot 9 = 81 \equiv -10$$
$$3^8 = 3^4 \cdot 3^4 \equiv (-10) \cdot (-10) = 100 \equiv 9$$
$$3^{16} = 3^8 \cdot 3^8 \equiv 9 \cdot 9 = 81 \equiv (-10)$$
$$3^{32} = 3^{16} \cdot 3^{16} \equiv (-10) \cdot (-10) = 100 \equiv 9$$
$$3^{64} = 3^{32} \cdot 3^{32} \equiv 9 \cdot 9 = 81 \equiv (-10)$$
$$3^{90} = 3^{64} \cdot 3^{16} \cdot 3^8 \cdot 3^2 \equiv (-10) \cdot (-10) \cdot 9 \cdot 9 = (-90) \cdot (-90) \equiv$$
$$\equiv 1 \cdot 1 = 1$$

Zjišťujeme tedy, že číslo 91 prošlo testem prvočíselnosti, a proto o něm nemůžeme nic říct. Přesto je uvedená vlastnost natolik zajímavá, že si vysloužila samostatné pojmenování. Říkáme, že *číslo 91 je pseudoprvočíslo vzhledem k bázi 3.*

Tím jsme narazili na hlavní problém testů prvočíselnosti. V případě negativního výsledku nám neposkytují „žádnou" informaci. Výhodou testů prvočíselnosti je však to, že výpočty v nich použité probíhají velmi rychle, a proto lze test vícekrát opakovat. V případě našeho čísla 91 můžeme v dalším pokusu zvolit za x například číslo 2 a zjistit, že rovnost (1) neplatí a tudíž je číslo 91 číslem složeným. Pokud bychom měli více štěstí a zvolili si za x číslo dva již v prvním pokusu, dostali bychom odpověď již při prvním pokusu. Zde je nutné připomenout, že již jistě víme, že číslo 91 je složené, v testu však není ani nejmenší náznak toho, jak mohou vypadat jeho dělitelé.

Důležitou otázkou spojenou s uvedeným testem je to, pro kolik čísel selhává vztah (1), pokud zvolíme za p složené číslo. Jinými slovy, jakou máme šanci, že se trefíme do čísla, pro něž rovnost neplatí. Zde je odpověď velmi příznivá, pokud test (1) selhává pro jedno číslo, tak selhává minimálně pro polovinu všech čísel. To znamená, že pokud test (1) selhává, máme šanci více než 50 % odhalit toto selhání v prvním pokusu, více než 75 % ve druhém, až například více než 99,999 9 % ve dvacátém. Tedy i u velmi velkých

čísel nám stačí jen malý počet pokusů, abychom s velkou pravděpodobností ověřili, zda test selhává, tedy zda číslo je prvočíslo nebo číslo složené.

Všechny úvahy v předchozím odstavci jsou založeny na předpokladu, že test (1) selhává u složených čísel alespoň pro jedno x. Bohužel však existují složená čísla, která splňují tvrzení malé Fermatovy věty, a tudíž je výše uvedený test vždy označí za prvočísla. Tato čísla se nazývají *Carmichaelova čísla*. Carmichaelova čísla musí splňovat mnoho podmínek, a proto se vyskytují velmi zřídka, nejmenším z nich je číslo 561. Přesto, že se vyskytují „zřídka", neznamená to, že jich je málo. V roce 1992 byl publikován důkaz, že Carmichaelových čísel je nekonečně mnoho. Test prvočíselnosti založený na malé Fermatově větě tak utrpěl poměrně výrazný šrám na kráse.

Existují proto i jiné testy prvočíselnosti a další se stále hledají. Problémy z teorie čísel jasně ukazují, kolik je ve světě matematiky stále nezodpovězených otázek. Na každém kroku zde potkáváme problémy, které jsou doposud otevřené, a na svá řešení stále čekají.

Jako příklad jen uvedu, že výše uvedený test je rozpracován podrobněji v testu, který se nazývá *Miller-Rabinův*. O něm je dokázáno, že určitě funguje pro nejméně jedno x v poměrně malém intervalu, ovšem za předpokladu, že platí obecná Riemannova hypotéza. Otázka, zda obecná Riemannova hypotéza platí, je však, i přes vypsanou prémii ve výši miliónů korun, již více než sto let nevyřešena. Možná její zodpovězení čeká na některého z našich žáků.

Faktorizace

Testy prvočíselnosti, s výjimkou základního avšak velmi neefektivního prověřování všech dělitelů, nedávají odpověď na otázku, jak vypadá rozklad zadaného čísla na prvočísla. Proto se hledaly a stále hledají efektivní metody, jak nalézt dělitele zvoleného velkého lichého čísla, o kterém víme, že je složené. Ve většině případů, vzhledem k návaznosti na RSA algoritmus, předpokládáme, že zadané „velké" liché číslo n je součinem dvou prvočísel. Mnoho velmi efektivních faktorizačních algoritmů je založeno na velmi jednoduché myšlence. Cílem těchto algoritmů je nalézt dvě čísla a a

b takové, že jejich druhé mocniny mají stejný zbytek modulo n. Potom, můžeme použít rozklad $a^2 - b^2 = (a - b) \cdot (a + b) = k \cdot n$ a čísla $(a - b)$ nebo $(a + b)$ jsou soudělná s číslem n.

Nejjednodušším způsobem, jak nalézt rozklad čísla n, je následující postup, známý jako *Fermatova faktorizace*. Tento postup je velmi efektivní v případě, kdy n lze zapsat jako rozdíl druhých mocnin dvou přirozených čísel, z nichž jedno je „velmi malé". K tomu dochází například tehdy, když n lze zapsat jako součin dvou čísel, která se příliš neliší.

Podkladem pro fungování algoritmu Fermatovy faktorizace je vztah mezi přirozenými čísly, který lze použít pro faktorizace čísla $n - n = a \cdot b$ $(a > b)$ a přirozenými čísly, která lze použít pro zápis čísla n jako rozdíl čtverců, $n = t^2 - s^2$. Tento vztah je vyjádřen následujícími vzorci:

$$t = \frac{a+b}{2}\,, \qquad s = \frac{a-b}{2}\,, \qquad a = t + s\,, \qquad b = t - s.$$

Pokud s je velmi malé, tak čísla a, b se příliš neliší od $\sqrt{n}$ a můžeme postupovat tak, že postupně budeme za t volit čísla větší než $\sqrt{n}$, a to až do té doby, než nalezneme číslo t takové, že $t^2 - n$ bude druhá mocnina nějakého přirozeného čísla.

Ukázka 2

Chceme nalézt rozklad čísla 200 819 na součin dvou menších čísel. Nejprve spočítáme odmocninu z čísla $\sqrt{200\,819}$. Nyní budeme postupně volit za t čísla větší než 448. Začneme nejprve číslem 449.

$$449^2 - 200\,819 = 782,$$

což ovšem není druhá mocnina. Budeme pokračovat číslem 450.

$$450^2 - 200\,819 = 1\,681 = 41^2.$$

V tomto případě jsme tedy již uspěli a s použitím výše uvedených vztahů dostáváme, že $200\,819 = 450^2 - 41^2 = (450 - 41) \cdot (450 + 41) = 409 \cdot 491$.

Metodu Fermatovy faktorizace lze částečně zobecnit. Pokud nenalezneme t, které se liší o čtverec od n, můžeme se pokusit hledat t takové, že se liší o čtverec malého číslo od $2n$, $3n$, ...

Ukázka 3
Chceme nalézt rozklad čísla 141 467. Pokud budeme testovat čísla $\sqrt{141\,467} \cong 377$, žádnou druhou mocninu nenalezneme. Pokud však zkusíme čísla od $\sqrt{3 \cdot 141\,467} \cong 652$, rychle zjistíme, že $655^2 - 3 \cdot 141\,467 = 68^2$. Nyní spočteme největšího společného dělitele čísel $655 + 68$ a 141 467 a dostáváme, že číslo 141 467 je dělitelné číslem 241 a tedy $141\,467 = 241 \cdot 587$.

Metoda síta

Pro hledání dělitelů velkých čísel se dnes používají metody, které vycházejí z myšlenky popsané v předchozím odstavci – nalezení čísel s a t takových, že $n = t^2 - s^2$. Nehledáme však čísla s, t přímo, ale snažíme se je zkonstruovat z výsledků, které jsme obdrželi v průběhu výpočtu. Myšlenku, jak se s a t konstruují, si představíme v následující ukázce.

Ukázka 4
Chceme nalézt rozklad čísla 4 633. Postupně (modulo 4 633) zjišťujeme, že $67 \equiv -144$ a $68^2 \equiv -9$. Ani jedno z těchto čísel není čtvercem, je však zřejmé, že jejich součin čtverec již tvoří. Vyzkoušíme tedy číslo $67 \cdot 68 = 4\,556$ a dostáváme $4\,556^2 \equiv 36^2$. Nyní spočteme největší společného dělitele čísel $4\,556 + 36$ a 4 633 a dostáváme, že číslo 4 633 je dělitelné číslem 41 a tedy $4\,633 = 41 \cdot 113$.

Uvedený postup se používá tak, že se nejprve shromažďují čísla, jejichž druhé mocniny (modulo n) se dají vyjádřit pomocí předem zvolené báze prvočísel (tato báze obvykle obsahuje desítky prvočísel). Tento postup, připomínající prosívání, dal název celému algoritmu. Podle přesného postupu vyhledávání se pak algoritmy jmenují *Quadratic Sieve* (kvadratické síto), či *Number Field Sieve* (síto číselného tělesa). Do detailů zde nebudeme zacházet, protože vyžadují hlubší znalosti z teorie čísel. Po nashromáždění dostatečného množství čísel se z nich – postupem, vcházejících

z úvah použitých v ukázce 4 – sestavují naše čísla t a s. Metody sít patří mezi to nejefektivnější, co bylo pro faktorizaci čísel vynalezeno. Lze je efektivně používat i pro rozklad čísel, která mají v desítkovém zápise více než sto číslic.

Metoda Monte Carlo

Posledním algoritmem, který si v dnešním díle představíme, je *Rho metoda*, nebo jiným názvem *Metoda Monte Carlo*. Musím se přiznat, že i přesto, že na teoretické rovině vím, proč algoritmus funguje, stále mne udivuje, že fungovat může.

Postup je primitivní. Zvolíme se nějakou jednoduchou, obvykle kvadratickou, funkci f, pomocí níž generujeme řadu čísel. Tuto řadu vytváříme tak, že první číslo si zvolíme, další čísla dostáváme tak, že předchozí číslo dosadíme do f a výsledek použijeme jako další hodnotu. U každého nového čísla zjišťujeme, zda rozdíl nového čísla a některého z předchozích není soudělný s číslem n. V okamžiku, kdy najdeme soudělné číslo, nalezli jsme i dělitele čísla n. Aby celý algoritmus fungoval, a fungoval rychle, je potřeba, aby se funkce f chovala pokud možno „náhodně". Heuristickým přístupem byly vytipovány funkce, které jsou vhodnější než jiné. Přesto, že celý postup se tváří „náhodně", je překvapivě efektivní a lze jej, při vhodném naprogramování, využít i pro faktorizaci čísel, která mají desítky cifer.

Ukázka 5

Pokusíme se faktorizovat číslo 4 087 pomocí metody Monte Carlo. Za funkci f zvolíme $x^2 + x + 1$ a výchozí hodnotu zvolíme 2. Postupně (modulo 4 087) dostává hodnoty 7, 57, 3 307, 2 745, 1 343, 2 626 a 3 734. Nebudeme provádět výpočty všech největších společných dělitelů, v tomto kroku však zjišťujeme, že NSD(3 734 − − 3 307, 4 087) = 61. Tím současně nacházíme i jednoho dělitele čísla 4 087 a to číslo 61 a zjišťujeme, že $4\,087 = 61 \cdot 67$.

Závěr

V tomto čísle jsme se, na rozdíl od předchozích dílů, seznámili s algoritmy, které pravděpodobně ve výuce na základní či střední škole běžně nevyužijete. Postupy zde uvedené však mohou být

vhodným zpestřením látky v nepovinných seminářích, či výkladu pro nadané žáky např. v přípravě na matematickou olympiádu. Možná, že máte také ve třídě žáky, kteří kladou zvídavé otázky, z nichž některé s teorií čísel a kryptologií souvisí. Doufám, že tento článek také ukázal, že nejnovější výsledky současné matematiky nemusí být vždy založené pouze na velmi složitých úvahách, a proto pro „běžnou" populaci nesrozumitelné.

RNDr. Antonín Jančařík, Ph.D.
Katedra matematiky a didaktiky matematiky PdF UK
M. D. Rettigové 4
116 39 Praha 1
e-mail: antonin.jancarik@pedf.cuni.cz