PRACE NAUKOWE AKADEMII EKONOMICZNEJ WE WROCŁAWIU
Nr 1117 — 2006
Dydaktyka matematyki

Andrzej Wilkowski
Akademia Ekonomiczna we Wrocławiu

O KRZYWYCH ELIPTYCZNYCH

1. Kryptologia jest nauką o bezpiecznych sposobach przechowywania i przesyłania informacji. Wraz ze wzrostem mocy obliczeniowej komputerów i powstaniem Internetu w latach siedemdziesiątych zeszłego wieku pojawiły się kryptosystemy z kluczem publicznym. Oznacza to, że każdy użytkownik może zakodować wiadomość, a tylko niektórzy mogą ją rozszyfrować. Wszystkie te systemy bazują na funkcjach jednokierunkowych. W praktyce znaczy to, że argument x i wartość $f(x)$ przedstawione w postaci binarnej mają długość tego samego rzędu. Wówczas $f(x)$ jest łatwe do obliczenia (w czasie wielomianowym w stosunku do długości x) i, na koniec, funkcja f jest trudna do odwrócenia (to znaczy, znając wartość funkcji, trudno podać jej argument). W tym miejscu należy zaznaczyć, że w stosunku do żadnej funkcji do dziś nie udowodniono, że jest funkcją jednokierunkową. Można tylko przypuszczać, że pewne funkcje mają tę własność.

Na przykład $f(p, q) = pq$ dla dwóch liczb pierwszych p, q (jest to zadanie łatwe, możliwe do wykonania w czasie wielomianowym w stosunku do długości danych). Obliczanie wartości funkcji f^{-1} odpowiada rozkładowi na czynniki pierwsze (jak do tej pory nie jest znany algorytm umożliwiający taki rozkład w czasie wielomianowym w zależności od długości danych). Jeśli ktoś chciałby zarobić na liczbach pierwszych dość pokaźną sumę, proszę bardzo: RSA, amerykańska firma zajmująca się szyfrowaniem, oferuje 20 tys. dolarów dla śmiałka, który zdoła ustalić, jakie dwie liczby pierwsze pomnożone przez siebie dają 3107418240490043721350750035888567930037346022842727545720161948823206440518081504556346829671723286782437916272838033415471073108501919548529007337724822783525742386454014691736602477652346609 (aktualne w kwietniu 2005 roku). W latach osiemdziesiątych okazało się, że do tworzenia kryptosystemów z kluczem publicznym można używać krzywych eliptycznych, znanych także z dowodu wielkiego twierdzenia Fermata. Ogólnie można powiedzieć, że krzywe eliptyczne są

rodzajem nieosobliwych, jednorodnych krzywych algebraicznych stopnia 3. Ogranicza je część przestrzeni topologicznie równoważna torusowi. Genus takich krzywych jest równy 1 (tak samo jak genus torusa). Przejście z torusa do postaci algebraicznej krzywej umożliwia eliptyczna funkcja Weierstrassa (w tej pracy nie będziemy jej szerzej analizować). Zdefiniujmy zatem podstawowe pojęcie niniejszej pracy.

Definicja 1

Krzywą eliptyczną **E** nad ciałem **F** nazywamy krzywą zadaną równaniem postaci

$$Y^2 + a_1XY + a_3Y = X^3 + a_2X^2 + a_4X + a_6,\ a_i \in \mathbf{F}. \qquad (1)$$

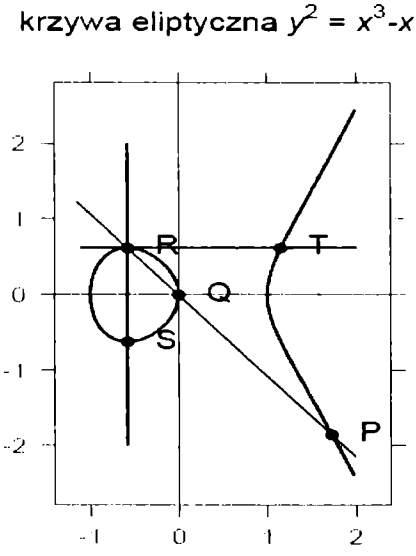
Przez **E(F)** oznaczamy zbiór złożony z punktów $(x, y) \in \mathbf{F}^2$, których współrzędne spełniają powyższe równanie, oraz „punktu w nieskończoności" oznaczanego przez O. Zbiór ten nazywamy zbiorem **F**-wymiernych punktów krzywej **E**. Aby krzywa dana równaniem (1) była krzywą eliptyczną, musi być także gładka. Oznacza to, że obie pochodne cząstkowe nie mogą jednocześnie znikać.

Zazwyczaj za ciało **F** przyjmuje się ciało liczb zespolonych, rzeczywistych, wymiernych lub ciała skończone. Gdy charakterystyka ciała jest różna od 2 i 3, wówczas równanie krzywej przybiera postać

$$Y^2 = X^3 + aX + b,\quad a, b \in \mathbf{F}. \qquad (2)$$

W takim przypadku warunek gładkości odpowiada żądaniu, aby wielomian po prawej stronie (2) nie miał pierwiastków wielokrotnych. Zachodzi to wtedy i tylko wtedy, gdy jego wyróżnik $\Delta = -16(4a^3 + 27b^2)$ jest niezerowy.

Przykład takiej krzywej nad ciałem liczb rzeczywistych $\Re$ przedstawiony jest na rys. 1.



Rys. 1. Krzywa eliptyczna

Źródło: Andrzej M. Boryszkowski (www.ipipan.gda.pl).

Określimy teraz dodawanie punktów $\Re$-wymiernych danej krzywej **E**. Niech **E** będzie krzywą eliptyczną nad ciałem liczb rzeczywistych $\Re$, zadaną równaniem (2), a P oraz Q dwoma $\Re$-wymiernymi punktami krzywej **E**, $P, Q \in \mathbf{E}(\Re)$. Punkt przeciwny do P i sumę $P + Q$ definiujemy za pomocą poniższych reguł:

- jeśli P jest punktem O w nieskończoności, to określamy $-P$ jako O,
- O jest elementem neutralnym dodawania,
- dla $P \neq O$ punkt $-P$ przeciwny do P będzie punktem symetrycznym do punktu P względem osi **X**,
- dla dowolnych trzech punktów P, Q, R krzywej eliptycznej, które leżą na jednej prostej, postulujemy, że $P + Q + R = O$.

W przypadku krzywej z powyższego rysunku mamy zatem $R + S + O = O$, stąd $R = -S$ i $S = -R$ czy $P + Q = S = R + T$ albo $Q + Q = O$.

Okazuje się, że powyższa definicja sumy zadaje na zbiorze punktów krzywej eliptycznej strukturę grupy abelowej.

Do tworzenia kryptosystemów z kluczem publicznym wykorzystuje się krzywe eliptyczne nad ciałami skończonymi F_q (w naturalny sposób przenosi się tam dodawanie). Liczba N punktów F_q-wymiernych na krzywej eliptycznej zdefiniowanej nad F_q spełnia wtedy nierówność

$$q + 1 - 2\sqrt{q} \leq N \leq q + 1 + 2\sqrt{q}\,.$$

Wynika stąd, że liczba punktów na krzywej nad ciałem skończonym jest bliska liczbie elementów w samym ciele (obecnie istnieją efektywne algorytmy wyznaczające N, gdy q ma kilkaset cyfr dziesiętnych). Funkcję jednokierunkową będącą bazą kryptosystemu określamy następująco: dane są P, $Q \in \mathbf{E}(F_q)$, znaleźć n takie, że np.= Q. Obecny stan wiedzy nie pozwala na rozwiązanie tego zadania inaczej jak tylko metodą prób i błędów. Zaletą kryptosystemów opartych na krzywych eliptycznych jest, między innymi, skrócenie kluczy, a zatem obniżenie wymagań odnośnie do pamięci i procesora, przy takim samym poziomie bezpieczeństwa jak w konwencjonalnych kryptosystemach opartych na faktoryzacji (na przykład RSA). Na obecnym etapie wiedzy o algorytmach, przy porównywalnej sile kryptograficznej, rozmiar klucza w kryptosystemie opartym na krzywych eliptycznych rośnie nieco szybciej niż pierwiastek sześcienny z rozmiaru klucza systemu konwencjonalnego. Zatem kluczom o rozmiarach 1024 i 4096 bitów (często spotykanym dla RSA) odpowiadają klucze 173- i 313-bitowe w kryptosystemach opartych na krzywych eliptycznych. Ważną cechą takich kryptosystemów jest także duża elastyczność w wyborze grupy (dla każdej potęgi q liczby pierwszej istnieje tylko jedna grupa multiplikatywna w ciele F_q, ale wiele krzywych nad tym ciałem).

2. Krzywe eliptyczne pojawiają się również przy analizie niektórych klasycznych problemów teorii liczb, na przykład w dowodzie wielkiego twierdzenia Fermata (WTF). W dalszym ciągu będziemy rozpatrywali krzywe nad ciałem liczb

wymiernych Q. Każdą taką krzywą można przedstawić w postaci równania o współczynnikach całkowitych

$$\mathbf{E}: y^2 = g(x) = x^3 + a_2x^2 + a_1x + a_0,$$

gdzie a_0, a_1, a_2 są tak dobrane, aby wyróżnik $\Delta(\mathbf{E})$ był możliwie najmniejszy ($\Delta(\mathbf{E}) = (r_1 - r_2)^2(r_1 - r_3)^2(r_2 - r_3)^2$, gdzie r_1, r_2, r_3 są pierwiastkami $g(x)$). Tak przekształcona krzywa nazywana jest modelem minimalnym krzywej eliptycznej. Dla takiej krzywej wprowadźmy kongruencję:

$$E_p : y^2 \equiv g(x) \pmod p,$$

poprzez redukcję modulo p współczynników krzywej **E** dla danej liczby pierwszej p. Widać, że taka kongruencja określa nam pewną krzywą nad ciałem skończonym F_p. Oznaczamy ją przez E_p i nazywamy redukcją modulo p krzywej eliptycznej **E**. Mówimy, że krzywa eliptyczna **E** ma dobrą redukcję (mod p), jeżeli krzywa E_p jest nieosobliwa nad ciałem F_p. W przeciwnym wypadku mamy do czynienia ze złą redukcją. Wiadomo, że jeśli p nie dzieli Δ, to E_p jest nieosobliwa, a więc i eliptyczna. Gdy zaś osobliwość krzywej E_p jest węzłem ($4a^3 + 27b^2 = 0$ oraz $a \neq 0$ w równaniu (2)), wówczas mówimy o semistabilnej redukcji krzywej eliptycznej **E**.

Definicja 2

Krzywa eliptyczna **E** jest semistabilna, jeżeli ma dobrą lub semistabilną redukcję dla każdej liczby pierwszej p.

Semistabilność oznacza, że dla każdej liczby pierwszej tylko dwa z trzech pierwiastków krzywej są w relacji kongruencji (mod p).

Ważnym parametrem krzywej eliptycznej w postaci minimalnej jest jej przewodnik (konduktor).

Definicja 3

Liczba N jest przewodnikiem krzywej **E**, gdy:

- $N = \prod_p p^{v_p}$, p – liczba pierwsza,
- $v_p = 0$, jeżeli **E** ma dobrą redukcję (mod p),
- $v_p = 1$, jeżeli **E** ma semistabilną redukcję (mod p),
- $v_p = 2 + \lambda_p$, w przeciwnym wypadku,

gdzie λ_p jest całkowite i nieujemne, tylko dla $p = 2, 3$ λ_p może być dodatnie.

Okazuje się, że p dzieli Δ wtedy i tylko wtedy, gdy p dzieli N, a krzywa eliptyczna **E** jest semistabilna, jeżeli jej przewodnik N jest bezkwadratowy.

Rozpatrzmy krzywą o równaniu

$$y^2 = x(x - A)(x + B),$$

gdzie A, B są względnie pierwszymi liczbami naturalnymi, oraz A jest podzielne przez 16. Frey powiązał tego typu krzywe z ewentualnymi rozwiązaniami równania Fermata. Mianowicie, jeśli dla liczby pierwszej $p \geq 5$ mamy $a^p + b^p = c^p$ oraz liczby naturalne a, b, c są parami względnie pierwsze i a jast parzyste, to krzywa Freya o równaniu $y^2 = x(x - a^p)(x + b^p)$ ma bardzo dziwne własności.

Rzeczywiście, jej minimalny wyróżnik to:

$$\Delta_{min} = \frac{(abc)^{2p}}{2^8}.$$

Ponieważ jest on różny od zera, więc tak zadana krzywa jest krzywą eliptyczną. Przewodnik N takiej krzywej jest bezkwadratowy, więc w przypadku gdy WTF nie zachodzi, krzywa Freya jest semistabilna. Widać także, że każdy dzielnik pierwszy tego wyróżnika pojawiałby się w jego rozkładzie na czynniki pierwsze w bardzo dużej potędze. Frey doszedł do wniosku, że tak być nie może, co było bodźcem do szukania dowodu WTF.

Istotną charakterystyką, oprócz wyróżnika i przewodnika, dla krzywych eliptycznych są liczby a_n. Definiujemy:

$$a_p = p + 1 - \left|E_p(F_p)\right|,$$

gdzie p jest liczbą pierwszą nie dzielącą Δ, a $\left|E_p(F_p)\right|$ oznacza liczbę punktów na krzywej eliptycznej **E** modulo p. Do wyznaczania tych liczb służą formy modularne.

Niech H będzie górną półpłaszczyzną zespoloną z dodanym punktem w nieskończoności i skończoną liczbą punktów na prostej $y = 0$, zwanych ostrzami (definicji nie podajemy). Określimy pojęcie formy modularnej tylko dla ciężaru 2, gdyż dalsze rozumowanie ogranicza się jedynie do tego przypadku.

Definicja 4

Formą modularną poziomu k i ciężaru drugiego nazywamy odwzorowanie

$$f : H \to C$$

$$f\left(\frac{az+b}{cz+d}\right) = (cz+d)^2 f(z),$$

gdzie $z \in H$, a, b, c, d, k są całkowite, k dzieli c, a $ad - bc = 1$.

Funkcja f jest ponadto holomorficzna w każdym punkcie zbioru H (wymaga to określenia tego pojęcia dla ostrzy).

Wymieńmy podstawowe własności form modularnych:

- zbiór form modularnych poziomu k i ciężaru 2, $M_2(k)$ jest przestrzenią liniową nad C,
- podzbiór $S_2(k)$ tego zbioru złożony z form znikających we wszystkich ostrzach jest podprzestrzenią liniową przestrzeni liniowej $M_2(k)$,
- $S_2(2)$ składa się tylko z formy zerowej,
- forma modularna ma rozwinięcie na szereg Fouriera postaci

$$f(z) = \sum_{n=0}^{\infty} c_n e^{2\pi n z}, \text{ gdy } f \in S_2(k); \text{ to wtedy mamy ponadto } c_0 = 0.$$

Jak wspomnieliśmy wcześniej, ważną charakterystyką krzywej eliptycznej są liczby a_n. Niosą one istotne informacje lokalne o danej krzywej. Istotą dowodu WTF było powiązanie tych danych lokalnych za pomocą pewnego niezmiennika globalnego. Jest to daleko idące uogólnienie znanego faktu, że każda liczba natu-

ralna jest iloczynem potęg liczb pierwszych. Okazało się mianowicie, na podstawie konkretnych obliczeń, że dla wielu specjalnych krzywych eliptycznych liczby a_p są równe współczynnikom rozwinięcia w szereg Fouriera pewnej formy modularnej. Krzywe eliptyczne o powyższej własności nazywamy **krzywymi eliptycznymi modularnymi.** Od lat pięćdziesiątych analizowana była hipoteza dotycząca takich krzywych.

Hipoteza Shimury-Taniyamy

Każda krzywa eliptyczna jest modularna.

W latach następnych Ribet udowodnił, że przy założeniu prawdziwości tej hipotezy, dla krzywych semistabilnych, krzywe Freya nie mogą istnieć, a co za tym idzie, WTF jest prawdziwe. I wreszcie w roku 1995 Wiles udowodnił hipotezę Shimury-Taniyamy dla krzywych eliptycznych semistabilnych.

Twierdzenie

Każda semistabilna krzywa eliptyczna jest modularna.

Wniosek

Wielkie Twierdzenie Fermata jest prawdziwe.

Literatura

Blake, G. Seroussi, N. Smart (2004). *Krzywe eliptyczne w kryptografii.* Wydawnictwa Naukowo-Techniczne. Warszawa.

J. Browkin (2003). *Siódmy problem milenijny: Hipoteza Birch a i Swinnertona-Dyera.* Wiadomości Matematyczne XXXIX. Polskie Towarzystwo Matematyczne. Warszawa.

N. Koblitz (1993). *Introduction to Elliptic Curves and Modular Forms.* Springer-Verlag. New York.

N. Koblitz (1995). *Wykład z teorii liczb i kryptografii.* Wydawnictwa Naukowo-Techniczne. Warszawa.

N. Koblitz (2000). *Algebraiczne aspekty kryptografii.* Wydawnictwa Naukowo-Techniczne. Warszawa.

M. Kutyłowski, W. Strothmann (1999). *Kryptografia. Teoria i praktyka zabezpieczania systemów komputerowych.* Oficyna Wydawnicza READ ME. Łódź.

P. Ribenboim (2001). *Wielkie Twierdzenie Fermata dla laików.* Wydawnictwa Naukowo-Techniczne. Warszawa.

J.H. Silverman (1994). *Advanced Topics in The Arithmetic of Elliptic Curves.* Springer-Verlag, GTM 151.

S. Singh (1999). *Tajemnica Fermata.* Wydawnictwo Prószyński i S-ka. Warszawa.

S. Singh (2001). *Księga szyfrów.* Wydawnictwo Albatros A. Kuryłowicz. Warszawa.

A. Wiles (1995). *Modular elliptic curves and Fermat's Last Theorem.* Annals of Mathematics 142.

ON ELLIPTIC CURVES

Summary

The work is dedicate to elliptic curves. In the first part one gave arithmetical properties of the group of points of this curve over the finite field. They there make possible the construction of the asymmetric cryptosystem with the public key. The second part talks over the use of elliptic curves in the proof of the Great Fermat Theorem.