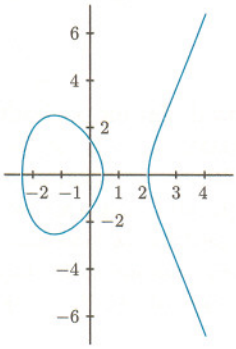
Krzywe eliptyczne

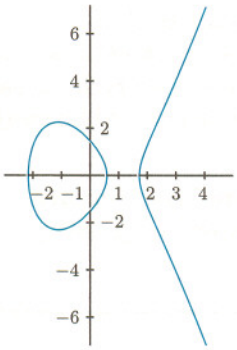
Jerzy KONARSKI

Za płaską krzywą algebraiczną stopnia n uważa się zwykle zbiór rozwiązań równania postaci $w(x,y)=0$, gdzie $w(x,y)$ jest dowolnym wielomianem dwóch zmiennych stopnia $n>0$. Na przykład równania $y-x^2=0$, $xy-1=0$ i $xy=0$ opisują krzywe stopnia 2 (parabolę, hiperbolę oraz parę prostych). Jedna z podstawowych własności płaskich krzywych algebraicznych to zależność między stopniem krzywej a liczbą jej punktów wspólnych z linią prostą. Mianowicie, dowolna prosta L albo przecina daną krzywą stopnia n w co najwyżej n punktach, albo jest w niej zawarta. Aby się o tym przekonać, wystarczy do równania krzywej $w(x,y)=0$ podstawić postać parametryczną $x=at+b$, $y=ct+d$ prostej L. Otrzymujemy równanie zmiennej t, w którym po lewej stronie stoi wielomian $w_L(t)=w(at+b,ct+d)$ stopnia nie większego niż n. Jeśli ten wielomian nie jest wielomianem zerowym, to ma co najwyżej n pierwiastków, a każdemu pierwiastkowi odpowiada punkt przecięcia prostej L z krzywą.



Rys. 1. $y^2=x^3-5x+2$.

Są trzy powody, dla których liczba punktów przecięcia może być mniejsza niż n. Po pierwsze, wielomian $w_L(t)$ może mieć stopień niższy niż n. Po drugie, nawet jeśli jest stopnia n, może się nie rozkładać na czynniki liniowe i w konsekwencji mieć mniej pierwiastków niż n. Po trzecie, może mieć tzw. pierwiastki wielokrotne. Sytuacja, w której jest mniej punktów przecięcia, sprawia pewien kłopot – wymyka się spod kontroli. W celu jej uniknięcia stosuje się następujące środki.

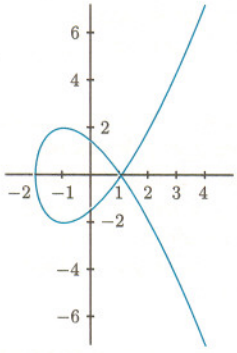


Rys. 2. $y^2=x^3-4x+2$.

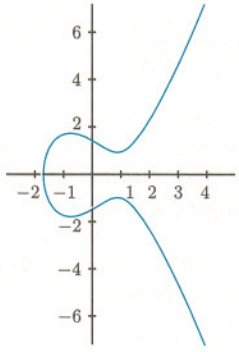
Po pierwsze, pierwiastki wielokrotne liczy się z ich krotnościami. Wtedy np. jedyny punkt wspólny paraboli $y-x^2=0$ i prostej $x=t+1$, $y=2t+1$, który odpowiada pierwiastkowi podwójnemu $t=0$ wielomianu $2t+1-(t+1)^2$, jest liczony podwójnie. Sytuacja taka ma miejsce zawsze wtedy, gdy dana prosta jest styczna do krzywej – algebraiczna definicja prostej stycznej jest następująca: prosta $x=at+b$, $y=ct+d$ jest styczną do krzywej $w(x,y)=0$ w punkcie (b,d), jeśli krotność pierwiastka $t=0$ równania $w(at+b,ct+d)=0$ jest większa niż 1 (czyli wielomian $w_L(t)$ dzieli się przez t^2).

Po drugie, zamiast zwykłej płaszczyzny rozważa się tzw. płaszczyznę rzutową, która powstaje ze zwykłej płaszczyzny przez dołączenie tzw. punktów w nieskończoności, po jednym dla każdego kierunku prostych zawartych w płaszczyźnie. Wtedy uzyskujemy dodatkowe punkty przecięcia: np. prosta „pionowa" $x=1$ przecina parabolę $y-x^2=0$ nie tylko w punkcie $(1,1)$, ale także w punkcie w nieskończoności (nazwijmy go P) odpowiadającym kierunkowi „pionowemu". Punkt P należy do prostej $x=1$ i do paraboli, bo zarówno ta prosta, jak i parabola „mają w nieskończoności kierunek pionowy". W przestrzeni rzutowej (w odpowiednio dobranym układzie współrzędnych) wielomian $w_L(t)$ będzie miał zawsze stopień n.

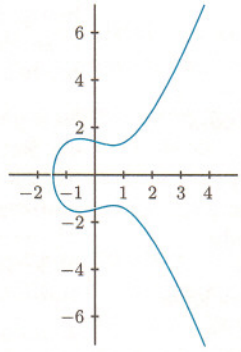
Trzecim środkiem jest stosowanie liczb zespolonych zamiast rzeczywistych jako współrzędnych punktów przestrzeni. Zachodzi twierdzenie: w zespolonej przestrzeni rzutowej dowolna prosta przecina krzywą stopnia n w dokładnie n punktach (licząc z krotnościami). Na przykład krzywa opisana równaniem $x^2+y^2=-2$ nie ma z prostą $x=y$ żadnych punktów wspólnych o współrzędnych rzeczywistych (bo sama krzywa nie ma takich punktów), ma natomiast punkty wspólne o współrzędnych zespolonych (i,i) oraz $(-i,-i)$.



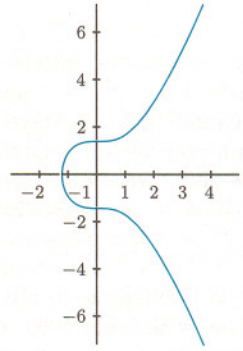
Rys. 3. $y^2=x^3-3x+2$.



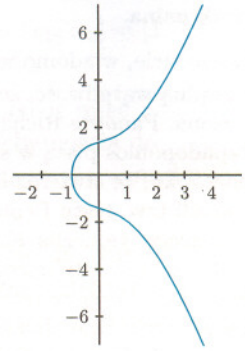
Rys. 4. $y^2=x^3-2x+2$.



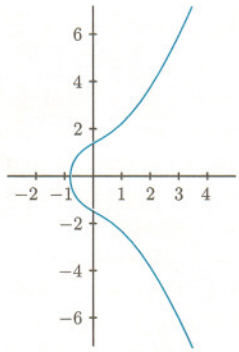
Rys. 5. $y^2=x^3-x+2$.



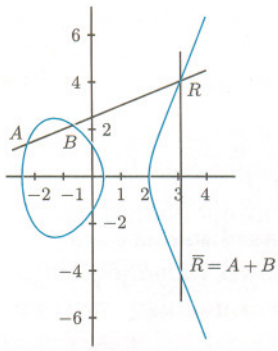
Rys. 6. $y^2=x^3+2$.



Rys. 7. $y^2=x^3+x+2$.



Rys. 8. $y^2 = x^3 + 2x + 2$.



Rys. 9. Dodawanie na krzywej $y^2 = x^3 - 5x + 2$. Elementem neutralnym jest punkt P w nieskończoności (kierunek pionowy).

Twierdzenie Mordella–Weila mówi, że grupa $E(\mathbb{Q})$ punktów $\mathbb{Q}$-wymiernych (czyli o obu współrzędnych będących liczbami wymiernymi) na krzywej eliptycznej E określonej nad ciałem liczb wymiernych jest grupą (przemienną) skończenie generowaną. Wiadomo, że każda taka grupa jest izomorficzna (może być utożsamiona) z grupą postaci $\mathbb{Z} \times \ldots \times \mathbb{Z} \times \mathbb{Z}_{m_1} \times \ldots \times \mathbb{Z}_{m_r}$, gdzie $\mathbb{Z}$ oznacza grupę liczb całkowitych z działaniem dodawania, a $\mathbb{Z}_{m_i}$ oznacza grupę złożoną z liczb całkowitych 0, 1, 2, $\ldots m_i - 1$ z działaniem dodawania modulo m_i. W przypadku grup $E(\mathbb{Q})$ liczby m_i nie mogą być dowolne: albo występuje tylko jedna i jest mniejsza od 13 i różna od 11, albo występują dwie i jedna z nich jest równa 2, a druga jest nie większa od 4. Czynników postaci $\mathbb{Z}$ może być nawet więcej niż 10, może ich też wcale nie być. Na przykład grupa punktów $\mathbb{Q}$-wymiernych krzywej opisanej równaniem $y^3 + x^3 = 1$ ma trzy elementy, a więc pasuje do powyższego opisu – jest izomorficzna z grupą $\mathbb{Z}_3$.

Uwaga: niektóre stwierdzenia wymienione powyżej, np. postać równania opisującego krzywą eliptyczną, nie są prawdziwe dla ciał charakterystyki 2 i 3.

Ciekawszy przykład otrzymamy, biorąc krzywą $x^2 - y^2 = 2$. Ze wspomnianego twierdzenia wynika, że prosta $x = y$ ma z tą krzywą (w przestrzeni rzutowej nad ciałem liczb zespolonych) dwa punkty wspólne. Sprzeczność otrzymana z podstawienia $x = y$ oznacza tylko, że nie ma punktów wspólnych na zwykłej (afinicznej) płaszczyźnie, pozostaje więc tylko możliwość, iż jedynym punktem wspólnym jest punkt w nieskończoności odpowiadający prostej $x = y$ i liczony z krotnością dwa. Znaczy to, że nasza krzywa jest styczna do prostej $x = y$ w jej punkcie w nieskończoności.

Krzywe eliptyczne to krzywe opisane w płaszczyźnie rzutowej przez wielomian nierozkładalny stopnia 3 (a więc krzywe stopnia 3, nie zawierające prostej) i spełniające dodatkowo warunek gładkości: w każdym punkcie krzywej istnieje dokładnie jedna prosta styczna do niej. Nie są gładkie np. krzywe opisane równaniami $y^2 - x^3 - x^2 = 0$ oraz $y^2 - x^3 = 0$. Pierwsza ma w punkcie $(0, 0)$ dwie styczne $y = x$ i $y = -x$, druga ma w punkcie $(0, 0)$ styczną podwójną $y = 0$. Pomijając szczegóły, wspomnimy tylko, że gładkość jest równoważna gładkości w sensie intuicyjnym (tzn. brakowi ostrzy i punktów samoprzecięcia).

Własnością wyjątkową wśród krzywych jest to, że każda krzywa eliptyczna E ma naturalną strukturę grupy przemiennej, tzn. można określić działanie dodawania punktów, które jest łączne, przemienne, ma element neutralny oraz element przeciwny dla każdego elementu. Konstrukcja tego działania dodawania opiera się na tym, że dowolna prosta przechodząca przez dwa punkty A i B, leżące na E, ma z E jeszcze dokładnie jeden punkt wspólny (wielomian $w_L(t)$ jest stopnia 3 i ma dwa pierwiastki, więc na mocy twierdzenia Bézouta musi mieć także i trzeci). Punkty liczymy, oczywiście, z krotnościami, a więc np. jeśli $A = B$, mamy na myśli prostą styczną do E w punkcie A. Najpierw wybieramy dowolny punkt $O \in E$, który będzie pełnił rolę elementu neutralnego działania. Dla każdego punktu $C \in E$ symbolem $\overline{C}$ oznaczać będziemy trzeci punkt wspólny z E prostej OC. Niech A i B będą punktami E i niech prosta AB przecina E jeszcze w punkcie R (rys. 9). Sumę $A + B$ określamy jako $\overline{R}$. Bardzo łatwe sprawdzenie przemienności ($A + B = B + A$) i neutralności zera ($A + O = A$) oraz nieco trudniejsze – istnienia elementu przeciwnego – zostawiamy Czytelnikowi. Dowód łączności (opuszczamy go ze względu na brak miejsca) wymaga już skorzystania z pewnych, nietrudnych zresztą, twierdzeń geometrii algebraicznej. W odpowiednim układzie współrzędnych każda krzywa eliptyczna jest opisana na płaszczyźnie równaniem postaci $y^2 = x^3 + ax + b$, w którym a i b są takimi liczbami, że wielomian $x^3 + ax + b$ nie ma pierwiastków wielokrotnych. Mówiąc ściślej, krzywa składa się z punktów (x, y) płaszczyzny spełniających to równanie oraz z punktu P w nieskończoności odpowiadającego kierunkowi pionowemu. Punkt P jest punktem przegięcia, tzn. styczna w tym punkcie przecina krzywą z krotnością 3. Wygodnie jest wybrać właśnie punkt P jako element neutralny działania dodawania. Wtedy, po pierwsze, dla każdego punktu R punkt $\overline{R}$ jest położony symetrycznie do R względem osi x, a po drugie, jest to element przeciwny do R (rys. 9).

Na rysunkach 1–8 przedstawione są krzywe $y^2 = x^3 + ax + b$ dla $b = 2$ i a całkowitych od -5 do 2. Dla $a = -3$ otrzymujemy krzywą z punktem samoprzecięcia, która nie jest krzywą eliptyczną. Do tej pory (nie mówiąc tego wyraźnie) rozpatrywaliśmy krzywe eliptyczne określone nad ciałem liczb rzeczywistych, tzn. takie, że współczynniki równań opisujących krzywe były liczbami rzeczywistymi. Również współrzędne wszystkich punktów na krzywej były liczbami rzeczywistymi, czyli rozpatrywaliśmy tylko tzw. punkty $\mathbb{R}$-wymierne. Otóż krzywe eliptyczne można w zasadzie bez żadnych kłopotów zdefiniować nad dowolnym ciałem, np. liczb zespolonych, wymiernych lub też jednym z ciał skończonych. Mimo że często nie przypominają krzywych jako zbiory, używa się dla nich nazwy „krzywe". Krzywe eliptyczne określone nad ciałem liczb wymiernych, a także nad ciałami skończonymi odgrywają bardzo ważną rolę w teorii liczb ze względu na różne związki z innymi pojęciami.

Krzywa eliptyczna określona nad ciałem skończonym F ma, oczywiście, tylko skończenie wiele punktów F-wymiernych (o współrzędnych z ciała F). Krzywe takie znalazły zastosowanie m.in. w kryptografii.