Uniwersytet Warszawski
Wydział Matematyki, Informatyki i Mechaniki

Michał Jaszczyk
Nr albumu: 219450

Iloczyn Weila w systemach kryptograficznych

Praca magisterska
na kierunku MATEMATYKA

Praca wykonana pod kierunkiem
dra hab. Jacka Pomykały
Instytut Matematyki

Październik 2011

Oświadczenie kierującego pracą

Potwierdzam, że niniejsza praca została przygotowana pod moim kierunkiem i kwalifikuje się do przedstawienia jej w postępowaniu o nadanie tytułu zawodowego.

Data Podpis kierującego pracą

Oświadczenie autora (autorów) pracy

Świadom odpowiedzialności prawnej oświadczam, że niniejsza praca dyplomowa została napisana przeze mnie samodzielnie i nie zawiera treści uzyskanych w sposób niezgodny z obowiązującymi przepisami.

Oświadczam również, że przedstawiona praca nie była wcześniej przedmiotem procedur związanych z uzyskaniem tytułu zawodowego w wyższej uczelni.

Oświadczam ponadto, że niniejsza wersja pracy jest identyczna z załączoną wersją elektroniczną.

Data Podpis autora (autorów) pracy

Streszczenie

W niniejszej pracy opisano iloczyn Weila, algorytm obliczający jego wartości oraz jego zastosowania w kryptografii. Praca zawiera również krótkie wprowadzenie do krzywych eliptycznych.
Iloczyn Weila to funkcja $w\colon E[n] \times E[n] \to \mathbb{K}$, która prowadzi z podgrupy n-torsyjnej $E[n]$ na krzywej eliptycznej E nad ciałem $\mathbb{K}$ (czyli podgrupy składającej się z tych punktów P na krzywej E, które spełniają zależność $nP = \mathcal{O}$) w grupę pierwiastków n-tego stopnia z jedności w ciele $\mathbb{K}$. Cechy charakterystyczne iloczynu Weila to dwuliniowość, antysymetria oraz niezdegenerowanie.
Przedstawiony w pracy algorytm obliczający wartości iloczynu Weila został zaproponowany przez Millera. Algorytm ten realizuje schemat „podwajaj-i-dodawaj" podobny do tego występującego w algorytmie szybkiego podnoszenia do n-tej potęgi. Częścią pracy jest autorska implementacja tego algorytmu.
Zastosowania iloczynu Weila zaprezentowane w pracy obejmują redukcję MOV (Menezes-Okamoto-Vanstone) umożliwiającą przeprowadzanie ataków na systemy kryptograficzne oparte na krzywych eliptycznych oraz konstrukcje kryptosystemów opartych na tożsamości, tzn. pozwalających na takie szyfrowanie i podpisywanie wiadomości, że jednym z kluczy jest dowolny ciąg bitów, np. adres poczty elektronicznej.

Słowa kluczowe

krzywe eliptyczne, iloczyn Weila, algorytm Millera, redukcja MOV, kryptografia oparta na tożsamości

Dziedzina pracy (kody wg programu Socrates-Erasmus)

11.1 Matematyka

Klasyfikacja tematyczna

94. Information and communication, circuits
94A. Communication, information
94A60. Cryptography

Tytuł pracy w języku angielskim

Weil pairing in cryptographic systems

Spis treści

Wstęp

Matematyka uważana jest za naukę oderwaną od rzeczywistości i mającą niewielki wpływ na życie ludzi. Jest jednak zupełnie odwrotnie: codziennie korzystamy z różnych zdobyczy cywilizacyjnych takich jak komputery i telefony komórkowe; przemieszczamy się za pomocą samochodów, statków i samolotów; leczymy się za pomocą nowoczesnych leków; mieszkamy w zaawansowanych technicznie budynkach. Wszystko to jest wynikiem postępu różnych nauk, a te, od nauk ścisłych aż do „niematematycznych" nauk społecznych i humanistycznych, potrzebują narzędzi matematycznych. Jesteśmy zatem zależni od matematyki w każdej dziedzinie życia, nawet jeśli nieświadomie. Z tego właśnie powodu matematyka nazywana jest królową nauk.

Różne nauki w różnym stopniu korzystają ze zdobyczy poszczególnych działów matematyki. I tak na przykład mechanika klasyczna opiera się na analizie, głównie na rachunku różniczkowym i całkowym; równania różniczkowe cząstkowe służą do opisywania i symulowania wielu zjawisk, w tym prognozowania pogody czy projektowania powierzchni aerodynamicznych (kadłubów samolotów, karoserii samochodów itp.); rachunek prawdopodobieństwa i statystyka znajdują zastosowanie w ekonomii i przy różnego rodzaju grach losowych, są także nieodzowne podczas przeprowadzania eksperymentów psychologicznych i socjologicznych; współczesne problemy fizyki związane z ogólną teorią względności oraz mechaniką kwantową formułowane są w języku zaawansowanej algebry.

Jednym z działów matematyki jest teoria liczb. Przez długi czas była ona uważana za gałąź matematyki, która, chociaż ma niemały wpływ na matematykę samą w sobie, to znajduje niewiele zastosowań poza nią. Znaczenie teorii liczb w życiu codziennym niesamowicie wzrosło wraz z pojawieniem się komputerów.

W obecnych czasach komputery wspomagają nas w wykonywaniu wielu czynności, w tym pozwalają nam komunikować się drogą elektroniczną. To spowodowało, że należało rozwiązać komputerowe odpowiedniki klasycznych problemów związanych z komunikacją: potwierdzeniem tożsamości nadawcy komunikatu oraz zapewnieniem, że komunikat zostanie odczytany jedynie przez odbiorcę. Problem zaprojektowania systemu komputerowego, który miałby te dwie cechy, dał początek współczesnej kryptografii.

Komputerowe systemy kryptograficzne bazują na fakcie, że pewne problemy są trudne obliczeniowo. Egzemplarz takiego problemu spreparowany w taki sposób, że znane jest jego rozwiązanie, stanowi „sekret" pozwalający dwóm stronom na bezpieczną komunikację. Jest to komputerowy odpowiednik kłódki i klucza, którym dysponują tylko dwie osoby. Tak się przy tym składa, że problemy, które dają podstawę systemom kryptograficznym, bardzo często związane są z teorią liczb. Przykładowo, problem rozkładu na czynniki pierwsze leży u podstaw systemu RSA, a na problemie logarytmu dyskretnego opiera się protokół Diffiego-Hellmana.

Teoria liczb i kryptografia stanowią zatem furtkę, poprzez którą pozornie abstrakcyjne obiekty matematyczne mogą trafić do codziennego (nawet jeśli nieświadomego) użytku. W ten sposób zastosowanie praktyczne znalazły m.in. krzywe eliptyczne oraz pojęcie z nimi stowarzyszone,

które jest kluczowe w niniejszej pracy: iloczyn Weila.
Iloczyn Weila to pewna dwuargumentowa operacja działająca na punktach zadanego skończonego rzędu na krzywej eliptycznej i prowadząca w zbiór pierwiastków z jedności w ciele. Dwie kluczowe cechy iloczynu Weila to niezdegenerowanie oraz dwuliniowość. W kontekście iloczynu Weila niezdegenerowanie oznacza, że jego wartościami są również pierwiastki pierwotne. Dwuliniowość zaś to odpowiednik pojęcia znanego z algebry liniowej przeniesiony na grupy abelowe. Z punktu widzenia kryptografii szczególnie dwuliniowość jest istotna, ponieważ powoduje, że pewne problemy decyzyjne na krzywych eliptycznych stają się łatwe.
Celem niniejszej pracy jest przedstawienie na przykładzie iloczynu Weila procesu wdrażania za pomocą teorii liczb i kryptografii pojęcia matematycznego do codziennego użytku oraz pokazanie, że proces ten może być względnie nieskomplikowany. Wybór iloczynu Weila na pojęcie centralne w całej pracy podyktowany jest trzema czynnikami. Po pierwsze, na podstawie iloczynu Weila udało się skonstruować wiele ciekawych kryptosystemów, które wykazują niespotykane wcześniej cechy. Po drugie, za pomocą iloczynu Weila można przeprowadzać ataki na istniejące kryptosystemy oparte na krzywych eliptycznych. Po trzecie, istnieje wydajny algorytm, który pozwala obliczać wartości iloczynu Weila, dzięki czemu wszystkie teoretyczne konstrukcje oparte na iloczynie Weila można zrealizować w praktyce. Widać więc, że jest iloczyn Weila pojęciem niezwykle interesującym z kryptograficznego punktu widzenia.
Praca składa się z pięciu rozdziałów. W rozdziale pierwszym znajduje się krótkie wprowadzenie do tematyki krzywych eliptycznych. Rozdział drugi zawiera bardziej szczegółowe studium wyjątkowo ważnego pojęcia związanego z krzywymi eliptycznymi: struktury grupy abelowej na krzywej eliptycznej. Celem rozdziału trzeciego jest zdefiniowanie i zanalizowanie własności iloczynu Weila. Rozdział czwarty zawiera omówienie algorytmu Millera. Wreszcie, rozdział piąty opisuje zastosowania iloczynu Weila w kryptografii. Ponadto, w ramach pracy powstała autorska implementacja algorytmu Millera, którą można znaleźć na dołączonej do pracy płycie kompaktowej.

Rozdział 1

Krzywe eliptyczne

Teoria krzywych eliptycznych to niezwykle interesujący dział matematyki, o czym świadczy chociażby to, że krzywe eliptyczne odgrywają kluczową rolę w dowodzie wielkiego twierdzenia Fermata. Krzywe eliptyczne na dobre zadomowiły się również w kryptografii – stanowią podstawę wielu kryptosystemów.
W rozdziale tym krótko przypomnimy podstawowe elementy teorii krzywych eliptycznych, które przedstawimy w ujęciu kryptograficznym (nie zaś bardziej ogólnym, geometryczno-algebraicznym). Ponieważ jest to tylko przypomnienie, a także ze względu na objętość przedstawianego materiału i jego niewielką trudność, wszystkie stwierdzenia w tym rozdziale podajemy bez dowodu.
Szczegółowe wprowadzenie do teorii krzywych eliptycznych można znaleźć w pracach [1] i [2] oraz w książce [3].

1.1. Definicja

Na potrzeby niniejszej pracy przyjmujemy następującą definicję krzywej eliptycznej.

Definicja 1.1.1. Dane jest ciało $\mathbb{K}$ oraz dwa jego elementy A i B. *Krzywa eliptyczna nad ciałem* $\mathbb{K}$ *o parametrach* A *i* B, oznaczana symbolem $E_{A,B}(\mathbb{K})$ (w skrócie $E(\mathbb{K})$ lub E), to zbiór składający się ze wszystkich elementów (x, y) zbioru $\mathbb{K} \times \mathbb{K}$, dla których zachodzi następująca zależność:

$$y^2 = x^3 + Ax + B \tag{1.1.2}$$

Ponadto, każda krzywa eliptyczna zawiera jeszcze jeden dodatkowy element oznaczany symbolem $\mathcal{O}$.

Ustalamy ponadto następujące określenia.

Definicja 1.1.3. Dana jest krzywa eliptyczna E. Element $\mathcal{O}$ krzywej E zwany jest *punktem w nieskończoności na krzywej* E lub jej *identycznością*. Pozostałe elementy krzywej E zwane są *punktami skończonymi na krzywej* E.

Definicja 1.1.4. Dana jest krzywa eliptyczna E. Punkty skończone P na krzywej E postaci $P = (a, 0)$ zwane są *punktami rzędu dwa na krzywej* E.

Definicja 1.1.5. Dany jest punkt skończony P na krzywej eliptycznej E postaci $P = (a, b)$. *Punkt sprzężony do punktu* P, oznaczany symbolem $\overline{P}$, to punkt skończony $(a, -b)$. Ponadto, punkt w nieskończoności na krzywej eliptycznej uznajemy za sprzężony do samego siebie.

Fakt 1.1.6. *Dla każdego punktu na krzywej eliptycznej istnieje dokładnie jeden punkt sprzężony do niego.*

Fakt 1.1.7. *Punkty rzędu dwa na krzywej eliptycznej są sprzężone do samych siebie.*

Definicja 1.1.8. Dana jest krzywa eliptyczna E. Równanie 1.1.2 zwane jest *równaniem krzywej* E. Wielomian $x^3 + Ax + B$ występujący po jego prawej stronie, oznaczany symbolem $\kappa(E)$ (w skrócie κ), zwany jest *wielomianem charakterystycznym krzywej* E.

W definicji 1.1.1 ciało $\mathbb{K}$ może być dowolne, w szczególności może być skończone lub nie oraz może mieć dowolną charakterystykę. Charakterystyka równa 2 lub 3 jest źródłem wielu trudności, np. już sama definicja krzywej eliptycznej nie jest odpowiednia w takiej sytuacji.

Uwaga 1.1.9. Zakładamy odtąd, o ile nie będzie zaznaczone inaczej, że charakterystyka ciał, nad którymi rozważamy krzywe eliptyczne, jest różna od 2 i 3.

Jest jeszcze jedno źródło trudności, którym nie będziemy zajmować się w niniejszej pracy.

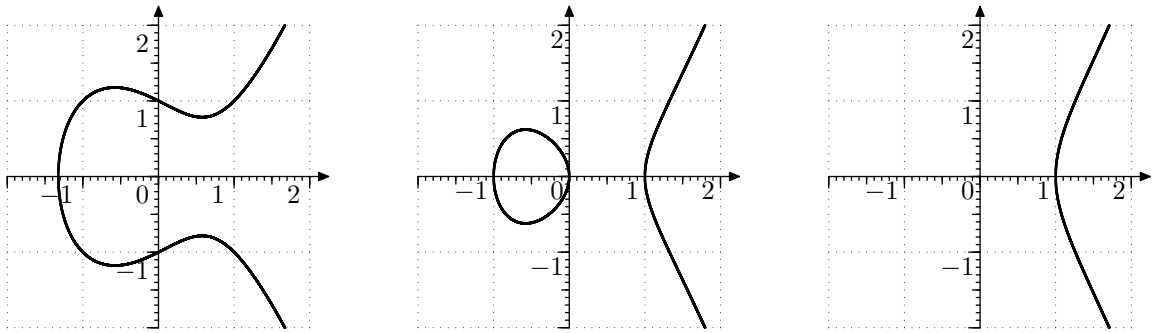
Definicja 1.1.10. Dana jest krzywa eliptyczna E nad ciałem $\mathbb{K}$. Jest ona *zdegenerowana*, jeżeli jej wielomian charakterystyczny κ ma w domknięciu algebraicznym $\overline{\mathbb{K}}$ ciała $\mathbb{K}$ pierwiastek wielokrotny. Jeżeli krzywa nie jest zdegenerowana, to mówimy, że jest *niezdegenerowana*.

Uwaga 1.1.11. Zakładamy odtąd, o ile nie będzie zaznaczone inaczej, że krzywe eliptyczne, które rozpatrujemy, są niezdegenerowane.

Twierdzenie 1.1.12. *Dane jest ciało* $\mathbb{K}$ *o charakterystyce różnej od* 2 *i* 3. *Wówczas krzywa eliptyczna* E *nad ciałem* $\mathbb{K}$ *o parametrach* A *i* B *jest zdegenerowana wtedy i tylko wtedy, gdy* $\frac{A^3}{27} + \frac{B^2}{4} = 0$.

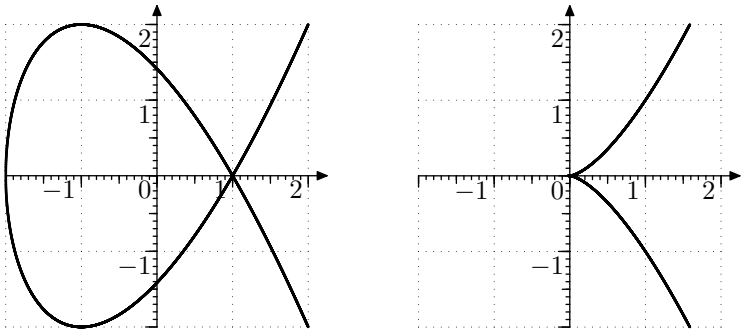
Następujące przykłady przedstawiają krzywe eliptyczne określone nad ciałem liczb rzeczywistych. Nie będziemy zajmować się takimi krzywymi, ale posłużą nam one do wyrobienia sobie „geometrycznej intuicji" na temat rozważań w dalszej części pracy.

Przykład 1.1.13. Na rysunku 1.1.14 przedstawione są fragmenty wykresów niezdegenerowanych krzywych eliptycznych nad ciałem liczb rzeczywistych. Krzywe różnią się od siebie kształtem: ilością spójnych składowych lub punktów przegięcia. Symetria wykresów względem osi odciętych jest konsekwencją faktu 1.1.6.



(a) $E_{-1,1}(\mathbb{R})$ (b) $E_{-1,0}(\mathbb{R})$ (c) $E_{0,-1}(\mathbb{R})$

Rysunek 1.1.14: Niezdegenerowane krzywe eliptyczne nad ciałem liczb rzeczywistych



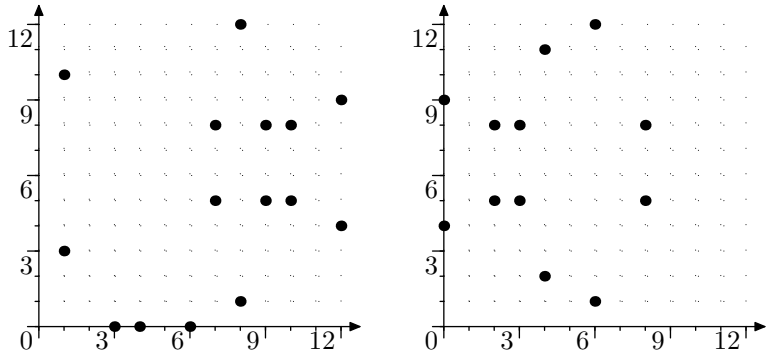
(a) $E_{-3,2}(\mathbb{R})$ (b) $E_{0,0}(\mathbb{R})$

Rysunek 1.1.16: Zdegenerowane krzywe eliptyczne nad ciałem liczb rzeczywistych

Przykład 1.1.15. Na rysunku 1.1.16 przedstawione są fragmenty wykresów zdegenerowanych krzywych eliptycznych nad ciałem liczb rzeczywistych. Przyczyną zdegenerowania jest raz podwójny, raz potrójny pierwiastek wielomianu charakterystycznego.

Dla kontrastu przedstawiamy również przykłady krzywych eliptycznych nad ciałem skończonym.

Przykład 1.1.17. Na rysunku 1.1.18 przedstawione są wykresy krzywych eliptycznych nad ciałem skończonym. Punkt w nieskończoności nie jest zaznaczony.



(a) $E_{2,6}(\mathbb{F}(13))$ (b) $E_{7,3}(\mathbb{F}(13))$

Rysunek 1.1.18: Krzywe eliptyczne nad ciałem skończonym $\mathbb{F}(13)$

1.2. Wielomiany i funkcje wymierne

Nasze rozważania w dalszej części pracy będą często dotyczyć wielomianów i funkcji wymiernych na krzywych eliptycznych, dlatego też przeanalizujemy teraz te pojęcia. Wykazują one wiele podobieństw do wielomianów i funkcji wymiernych jednej zmiennej, są też pewne różnice.

Definicja wyrażenia wielomianowego

Przyjmujemy następującą definicję wyrażenia wielomianowego na krzywej eliptycznej.

Definicja 1.2.1. Dana jest krzywa eliptyczna E nad ciałem $\mathbb{K}$. *Wyrażenie wielomianowe na krzywej E* to element pierścienia ilorazowego $\mathbb{K}[x,y]/(\kappa(E)-y^2)$. Pierścień wszystkich wyrażeń wielomianowych na krzywej E oznaczamy symbolem $\mathbb{K}[E]$.

Sens powyższej definicji jest następujący. Dowolna krzywa eliptyczna nad ciałem $\mathbb{K}$ to podzbiór zbioru $\mathbb{K}\times\mathbb{K}$ (nie licząc punktu w nieskończoności), dlatego pierwszym kandydatem na pierścień wyrażeń wielomianowych jest pierścień $\mathbb{K}[x,y]$. Współrzędne skończonych punktów krzywej spełniają równanie krzywej, dlatego też wielomiany dwóch zmiennych różniące się o wielokrotność wielomianu $\kappa-y^2$ dadzą tę samą funkcję wielomianową, niezależnie od wyboru ciała $\mathbb{K}$. Stąd pierścień $\mathbb{K}[x,y]$ dzielimy przez ideał $(\kappa-y^2)$ i uzyskujemy pierścień ilorazowy $K[x,y]/(\kappa-y^2)$.
Wyrażenie wielomianowe na krzywej eliptycznej jest w takim razie klasą abstrakcji pewnej relacji równoważności na zbiorze wielomianów dwóch zmiennych. Nie jest wygodnie myśleć o wyrażeniach wielomianowych w ten sposób. Następujący lemat pokazuje, w jaki sposób wybrać reprezentantów klas abstrakcji do dalszych rozważań.

Twierdzenie 1.2.2. *Dane jest wyrażenie wielomianowe f na krzywej eliptycznej E nad ciałem $\mathbb{K}$. Wówczas istnieją wyrażenia wielomianowe u i v z pierścienia $\mathbb{K}[x]$ takie, że $u+yv\in f$. Wyrażenia u i v są wyznaczone jednoznacznie.*

Przykład 1.2.3. Dane jest wyrażenie wielomianowe f na krzywej eliptycznej $E_{4,7}(\mathbb{F}(13))$ takie, że $x^3+y^3+x^2+y^2\in f$. Niech $u=2x^3+x^2+4x+7$ i $v=x^3+4x+7$ będą wyrażeniami wielomianowymi z pierścienia $\mathbb{F}(13)[x]$. Wówczas $u+yv\in f$.

Uwaga 1.2.4. Zbiór wyrażeń wielomianowych na krzywej eliptycznej nad ciałem $\mathbb{K}$ będziemy odtąd utożsamiać ze zbiorem złożonym z wyrażeń wielomianowych postaci $u+yv$, gdzie $u,v\in\mathbb{K}[x]$. Zbiór ten będziemy oznaczać $(1+y)\mathbb{K}[x]$.

Zauważmy, że zbiór $(1+y)\mathbb{K}[x]$ nie ma struktury pierścienia – w naturalny sposób jest w nim określone dodawanie, ale nie można wykonywać mnożenia. Nie jest to problem, ponieważ z kontekstu zawsze będzie wynikało, na jakiej krzywej rozpatrywane są wyrażenia wielomianowe. Co więcej, czasami wystarczy, że w miejsce powstałego podczas mnożenia czynnika y^2 będziemy podstawiać symbol κ i nie będzie nam potrzebna dokładna znajomość parametrów A i B krzywej.

Definicja wyrażenia wymiernego

Przyjmujemy następującą definicję wyrażenia wymiernego na krzywej eliptycznej.

Definicja 1.2.5. Dana jest krzywa eliptyczna E nad ciałem $\mathbb{K}$. *Wyrażenie wymierne na krzywej E* to element ciała ułamków pierścienia $\mathbb{K}[E]$. Ciało wszystkich wyrażeń wymiernych na krzywej E oznaczamy symbolem $\mathbb{K}(E)$.

Sens tej definicji jest taki sam, jak w przypadku każdego innego ciała ułamków: bierzemy zbiór ułamków formalnych $\frac{f}{g}$, gdzie f i g to wyrażenia wielomianowe na krzywej eliptycznej, po czym utożsamiamy ułamki $\frac{f_1}{g_1}$ oraz $\frac{f_2}{g_2}$, jeżeli zachodzi równość $f_1g_2=f_2g_1$.

Fakt 1.2.6. *Zbiór wyrażeń wielomianowych na krzywej eliptycznej można zanurzyć w zbiór wyrażeń wymiernych na tej krzywej, przypisując wyrażeniu wielomianowemu f wyrażenie wymierne $\frac{f}{1}$.*

Uwaga 1.2.7. Wyrażenia wielomianowe na krzywej eliptycznej będziemy odtąd utożsamiać z odpowiadającymi im wyrażeniami wymiernymi.

Podobnie jak wyrażenia wielomianowe, wyrażenia wymierne są klasami abstrakcji pewnej relacji równoważności, co nie jest wygodne.

Twierdzenie 1.2.8. *Dane jest wyrażenie wymierne r na krzywej eliptycznej E nad ciałem $\mathbb{K}$. Wówczas istnieją wyrażenia wymierne u i v z ciała $\mathbb{K}(x)$ takie, że $u + yv = r$. Wyrażenia u i v są wyznaczone jednoznacznie.*

Przykład 1.2.9. Dane jest wyrażenie wymierne r na krzywej eliptycznej $E_{4,7}(\mathbb{F}(13))$ takie, że $(x^2+y^2, y^3+x^3) \in r$. Niech u i v będą wyrażeniami wymiernymi z ciała $\mathbb{F}(13)(x)$ określonymi następująco:

$$\begin{aligned} u &= \frac{x^6 + x^5 + 4x^4 + 7x^3}{12x^9 + x^7 + 6x^6 + 4x^5 + x^4 + 10x^3 + 2x^2 + 10x + 8} \\ v &= \frac{12x^6 + 12x^5 + 5x^4 + 8x^3 + 3x^2 + 9x + 3}{12x^9 + x^7 + 6x^6 + 4x^5 + x^4 + 10x^3 + 2x^2 + 10x + 8} \end{aligned}$$

Wówczas $u + yv = r$.

Uwaga 1.2.10. Zbiór wyrażeń wymiernych na krzywej eliptycznej nad ciałem $\mathbb{K}$ będziemy odtąd utożsamiać ze zbiorem złożonym z wyrażeń wymiernych postaci $u + yv$, gdzie u i v to wyrażenia wymierne z ciała $\mathbb{K}(x)$. Zbiór ten będziemy oznaczać $(1 + y)\mathbb{K}(x)$.

Podobnie jak w przypadku zbioru $(1 + y)\mathbb{K}[x]$, zbiór $(1 + y)\mathbb{K}(x)$ nie ma struktury ciała – nie można mnożyć ani dzielić. Jednak zawsze albo będą dane parametry krzywej eliptycznej, albo w miejsce powstałego podczas mnożenia lub dzielenia czynnika y^2 będziemy podstawiać symbol κ.
Wyrażenia wymierne możemy również przedstawiać w postaci ilorazu elementów zbioru $(1 + y)\mathbb{K}[x]$.

Twierdzenie 1.2.11. *Dane jest wyrażenie wymierne r na krzywej eliptycznej E nad ciałem $\mathbb{K}$. Wówczas istnieją wyrażenia wielomianowe f i g ze zbioru $(1 + y)\mathbb{K}[x]$ takie, że $\frac{f}{g} = r$.*

Przykład 1.2.12. Dane jest wyrażenie wymierne r na krzywej eliptycznej $E_{4,7}(\mathbb{F}(13))$ takie, że $(x^2+y^2, y^3+x^3) \in r$. Niech $f = x^3+x^2+4x+7$ i $g = x^3+y(x^3+4x+7)$ będą wyrażeniami wielomianowymi ze zbioru $(1 + y)\mathbb{F}(13)[x]$. Wówczas $\frac{f}{g} = r$.

Funkcje wielomianowe i wymierne

Wyrażenia wielomianowe wyznaczają funkcje wielomianowe, a wyrażenia wymierne – funkcje wymierne. Wyrażenia i funkcje wielomianowe będziemy określać wspólnym mianem *wielomian*. W przypadku wyrażeń i funkcji wymiernych nie ma trzeciego określenia, którym można by je wspólnie nazwać, będziemy więc posługiwać się określeniem *funkcja wymierna*.

Uwaga 1.2.13. Jest jasne, jak obliczyć wartość funkcji wymiernej w punkcie skończonym P postaci $P = (a, b)$. Wartość tę oznaczać będziemy symbolem $r(a, b)$ lub $r(P)$.

Jedyna ewentualna trudność związana z obliczaniem wartości funkcji wymiernej w punkcie skończonym pojawia się, gdy wartość mianownika jest równa zero.

Definicja 1.2.14. Dana jest funkcja wymierna r na krzywej eliptycznej E oraz punkt skończony P na krzywej E. Funkcja r ma w punkcie P *wartość nieskończoną*, co zapisujemy jako $r(P) = \infty$, jeżeli istnieją wielomiany f i g na krzywej E takie, że $r = \frac{f}{g}$, $f(P) \neq 0$ oraz $g(P) = 0$.

Twierdzenie 1.2.15. *Dana jest funkcja wymierna r na krzywej eliptycznej E oraz punkt skończony P na tej krzywej. Niech f i g będą dowolnymi wielomianami na krzywej E takimi, że $r = \frac{f}{g}$. Wówczas jeżeli $r(P) = 0$, to $f(P) = 0$.*

Twierdzenie 1.2.16. *Dana jest funkcja wymierna r na krzywej eliptycznej E oraz punkt skończony P na tej krzywej. Niech f i g będą dowolnymi wielomianami na krzywej E takimi, że $r = \frac{f}{g}$. Wówczas jeżeli $r(P) = \infty$, to $g(P) = 0$.*

Sprzężenie i norma

Następujące dwa pojęcia są bardzo przydatne, ponieważ pozwalają sprowadzić zagadnienie dotyczące wielomianów i funkcji wymiernych na krzywej eliptycznej do przypadku wyrażeń jednej zmiennej.

Definicja 1.2.17. Dana jest funkcja wymierna r na krzywej eliptycznej E nad ciałem $\mathbb{K}$ postaci $r = u + yv$, gdzie $u, v \in \mathbb{K}(x)$. *Funkcja wymierna sprzężona do funkcji* r, oznaczana symbolem $\overline{r}$, to funkcja $u - yv$.

Definicja 1.2.18. Dana jest funkcja wymierna r na krzywej eliptycznej E nad ciałem $\mathbb{K}$ postaci $r = u + yv$, gdzie $u, v \in \mathbb{K}(x)$. *Norma funkcji wymiernej* r, oznaczana symbolem $N(r)$, to funkcja wymierna jednej zmiennej $r\overline{r} = (u + yv)(u - yv) = u^2 - y^2v^2 = u^2 - \kappa v^2$.

Przykład 1.2.19. Dany jest wielomian f na krzywej eliptycznej $E_{4,7}(\mathbb{F}(13))$ taki, że $y^3+x^3 \in r$. Wielomian f można przedstawić w postaci $f = x^3 + y(x^3 + 4x + 7)$. Wielomian sprzężony $\overline{f}$ jest równy $x^3 - y(x^3 + 4x + 7)$. Norma $N(f)$ jest równa $12x^9 + x^7 + 6x^6 + 4x^5 + x^4 + 10x^3 + 2x^2 + 10x + 8$.

Fakt 1.2.20. *Dla dowolnych funkcji wymiernych r i s na krzywej eliptycznej zachodzą zależności $\overline{rs} = \overline{r}\,\overline{s}$ oraz $N(rs) = N(r)N(s)$.*

Stopień wielomianu i funkcji wymiernej

Przyjmujemy następującą definicję stopnia wielomianu na krzywej eliptycznej.

Definicja 1.2.21. Dany jest wielomian f na krzywej eliptycznej E. *Stopień wielomianu* f, oznaczany symbolem $\deg(f)$, to stopień jego normy $N(f)$ traktowanej jak wielomian jednej zmiennej.

Przykład 1.2.22. Dany jest wielomian f na krzywej eliptycznej $E_{4,7}(\mathbb{F}(13))$ taki, że $y^3+x^3 \in f$. Jego norma została policzona w przykładzie 1.2.19, skąd widzimy, że jego stopień $\deg(f)$ jest równy 9.

Sens tej definicji jest następujący. W pierścieniu $\mathbb{K}[x]$ stopień dowolnego wielomianu wyznaczamy na podstawie stopnia wielomianu x, zaś wielomianowi x przypisujemy stopień równy 1, aby zachować związek stopnia wielomianu z m.in. ilością miejsc zerowych. Tę koncepcję przenosimy na wielomiany na krzywej eliptycznej: wielomianowi x przypisujemy stopień 2, a wielomianowi y stopień 3. Dzięki temu stopnie wielomianów występujących po obu stronach równania krzywej są równe oraz, jak się przekonamy, stopień wielomianu ma związek z ilością jego miejsc zerowych. Ponadto, do obliczania stopnia wielomianu f wybieramy z klasy abstrakcji, którą jest wielomian f, reprezentanta w postaci $u + yv$, gdzie $u, v \in \mathbb{K}[x]$.

Uwaga 1.2.23. Aby uniknąć nieporozumień, stopień wielomianu jednej zmiennej x oznaczać będziemy symbolem $\deg_x$.

Fakt 1.2.24. *Dla dowolnego wielomianu f na krzywej eliptycznej E postaci $f = u + yv$, gdzie $u, v \in \mathbb{K}[x]$, zachodzi zależność* $\deg(f) = \max(2\deg_x(u), 3 + 2\deg_x(v))$.

Dysponując stopniem wielomianu możemy określić stopień funkcji wymiernej.

Definicja 1.2.25. Dana jest funkcja wymierna r na krzywej eliptycznej E nad ciałem $\mathbb{K}$ postaci $r = \frac{f}{g}$, gdzie f i g to wielomiany na krzywej E. *Stopień funkcji wymiernej* r, oznaczany symbolem $\deg(r)$, to wielkość $\deg(f) - \deg(g)$.

Przykład 1.2.26. Dana jest funkcja wymierna r na krzywej eliptycznej $E_{4,7}(\mathbb{F}(13))$ taka, że $(x^2 + y^2, x^3 + y^3) \in r$. Stopień wielomianu $x^2 + y^2$ jest równy 6, a stopień wielomianu $x^3 + y^3$ jest równy 9. Stąd stopień funkcji r jest równy -3.

Stopień funkcji wymiernej jest dobrze określony, tzn. nie zależy od wyboru reprezentacji.

Twierdzenie 1.2.27. *Dana jest funkcja wymierna r na krzywej eliptycznej E. Wówczas dla dowolnych wielomianów f_1, g_1, f_2, g_2 na krzywej E takich, że $\frac{f_1}{g_1} = r = \frac{f_2}{g_2}$, zachodzi* $\deg(f_1) - \deg(g_1) = \deg(f_2) - \deg(g_2)$.

Na stopnie wielomianów i funkcji wymiernych na krzywej eliptycznej przenosi się zasadnicza własność znana z teorii wyrażeń jednej zmiennej.

Twierdzenie 1.2.28. *Dane są funkcje wymierne r i s na krzywej eliptycznej E. Wówczas* $\deg(rs) = \deg(r) + \deg(s)$.

Wartość funkcji wymiernej w punkcie w nieskończoności

Chcemy określić wartość funkcji wymiernej (zatem także wielomianu) w punkcie w nieskończoności. W przypadku funkcji wymiernych nad ciałem liczb rzeczywistych obliczamy po prostu granicę wartości funkcji, gdy argument dąży do nieskończoności. Uzyskana w ten sposób wartość, jeśli jest skończona, jest po prostu ilorazem współczynników stojących przy najwyższych potęgach w mianowniku i liczniku. Dzięki tej obserwacji możemy w analogiczny sposób określić wartość funkcji wymiernej w punkcie w nieskończoności.

Definicja 1.2.29. Dana jest funkcja wymierna r na krzywej eliptycznej E. Wartość funkcji r w punkcie $\mathcal{O}$ ustalamy następująco:

- jeżeli $\deg(r) < 0$, to $r(\mathcal{O}) = 0$;
- jeżeli $\deg(r) > 0$, to $r(\mathcal{O}) = \infty$;
- jeżeli $\deg(r) = 0$, to przedstawiamy funkcję r w postaci $\frac{f}{g}$, gdzie f i g to wielomiany na krzywej E i wówczas:
  - jeżeli stopnie wielomianów f i g są parzyste, to ich wiodące składniki mają postać odpowiednio ax^d i bx^d, wówczas $r(\mathcal{O}) = \frac{a}{b}$;
  - jeżeli stopnie wielomianów f i g są nieparzyste, to ich wiodące składniki mają postać odpowiednio ayx^d i byx^d, wówczas również $r(\mathcal{O}) = \frac{a}{b}$.

Przykład 1.2.30. Dana jest krzywa eliptyczna $E_{4,7}(\mathbb{F}(13))$. Wówczas $\left(\frac{x^3+y^3}{x^2+y^2}\right)(\mathcal{O}) = \infty$; $\left(\frac{x^2+y^2}{x^3+y^3}\right)(\mathcal{O}) = 0$; $\left(\frac{x^3}{2x^2+y^2}\right)(\mathcal{O}) = 1$; $\left(\frac{y^3}{2x^3+y^3}\right)(\mathcal{O}) = 1$.

Pod wieloma względami funkcje wymierne zachowują się w punkcie $\mathcal{O}$ i punktach skończonych podobnie, co pokazuje następujące twierdzenie.

Twierdzenie 1.2.31. *Dane są funkcje wymierne r i s na krzywej eliptycznej E. Jeżeli $r(\mathcal{O}) \neq \infty$ oraz $s(\mathcal{O}) \neq \infty$, to $(r + s)(\mathcal{O}) = r(\mathcal{O}) + s(\mathcal{O})$ oraz $(rs)(\mathcal{O}) = r(\mathcal{O})s(\mathcal{O})$.*

Miejsca zerowe i bieguny

Przyjmujemy następującą definicję miejsca zerowego oraz bieguna funkcji wymiernej na krzywej eliptycznej.

Definicja 1.2.32. Dana jest funkcja wymierna r na krzywej eliptycznej E. *Miejsce zerowe funkcji* r (odpowiednio, *biegun funkcji* r) to taki punkt P na krzywej E, że $r(P) = 0$ (odpowiednio, $r(P) = \infty$).

Przykład 1.2.33. Dana jest krzywa eliptyczna $E_{4,7}(\mathbb{F}(13))$. Wówczas punkt $(6, 0)$ jest biegunem funkcji $\frac{1}{y}$, a punkt $\mathcal{O}$ jest biegunem funkcji x. Punkt $(1, 5)$ jest miejscem zerowym funkcji $x - 1$, a punkt $\mathcal{O}$ jest miejscem zerowym funkcji $\frac{x}{y}$.

Fakt 1.2.34. *Jeżeli punkt skończony P jest miejscem zerowym (biegunem) funkcji wymiernej r, to punkt skończony $\overline{P}$ jest miejscem zerowym (biegunem) funkcji wymiernej $\overline{r}$.*

Wniosek 1.2.35. *Jeżeli punkt skończony P postaci $P = (a, b)$ jest miejscem zerowym wielomianu f, to wartość a jest miejscem zerowym wielomianu $N(f)$.*

Fakt 1.2.36. *Punkty rzędu dwa na krzywej eliptycznej to miejsca zerowe wielomianu charakterystycznego tej krzywej. Niezdegenerowana krzywa eliptyczna nad ciałem algebraicznie domkniętym ma dokładnie trzy punkty rzędu dwa.*

Podczas badania miejsc zerowych i biegunów funkcji wymiernych na krzywej eliptycznej będziemy chcieli uwolnić się od konieczności pilnowania czy ciało, nad którym zdefiniowana jest krzywa, jest algebraicznie domknięte.

Fakt 1.2.37. *Dane są dwa ciała $\mathbb{K}$ i $\mathbb{L}$ takie, że $\mathbb{K} \subset \mathbb{L}$ oraz parametry A i B z ciała $\mathbb{K}$. Wówczas $E_{A,B}(\mathbb{K}) \subset E_{A,B}(\mathbb{L})$.*

Fakt ten pozwala nam uwolnić się od pytania czy ciało $\mathbb{K}$ jest algebraicznie domknięte – rozważania na temat krzywej nad ciałem $\mathbb{K}$ zawsze możemy potraktować jak rozważania na temat krzywej o tych samych parametrach nad większym ciałem $\mathbb{L} = \overline{\mathbb{K}}$, która jest nadzbiorem danej krzywej.

Definicja 1.2.38. Dane są dwa ciała $\mathbb{K} \subset \mathbb{L}$ oraz parametry $A, B \in \mathbb{K}$. *Punkty $\mathbb{K}$-wymierne na krzywej $E_{A,B}(\mathbb{L})$* to te punkty krzywej $E_{A,B}(\mathbb{L})$, które są jednocześnie punktami krzywej $E_{A,B}(\mathbb{K})$.

Uwaga 1.2.39. Do końca tego rozdziału przyjmujemy, że rozpatrywane krzywe są określone nad ciałami algebraicznie domkniętymi.

Podobnie jak w przypadku funkcji wymiernych jednej zmiennej, chcemy wprowadzić pojęcie krotności miejsca zerowego i bieguna.

Twierdzenie 1.2.40. *Dana jest krzywa eliptyczna E nad ciałem $\mathbb{K}$ oraz punkt P na krzywej E. Wówczas istnieje funkcja wymierna u na krzywej E taka, że $u(P) = 0$ oraz dla każdej funkcji wymiernej r na krzywej E istnieje liczba całkowita d oraz funkcja wymierna s na krzywej E taka, że $s(P) \neq 0$, $s(P) \neq \infty$ oraz zachodzi następująca równość:*

$$r = u^d s \tag{1.2.41}$$

Liczba d nie zależy od wyboru funkcji u.

Definicja 1.2.42. Dany jest punkt P na krzywej eliptycznej E. *Unifikator w punkcie P* to dowolna funkcja wymierna u na krzywej E, której istnienie postuluje twierdzenie 1.2.40.

Przykład 1.2.43. Dla dowolnej krzywej eliptycznej E i punktu P na krzywej E unifikatorami są następujące funkcje:

- jeżeli $P = (a, 0)$, to $u = y$;
- jeżeli $P = (a, b)$, gdzie $b \neq 0$, to $u = x - a$;
- jeżeli $P = \mathcal{O}$, to $u = \frac{x}{y}$.

Unifikatory posłużą nam do przeniesienia na krzywe eliptyczne pojęcia krotności miejsca zerowego lub bieguna.

Definicja 1.2.44. Dana jest funkcja wymierna r na krzywej eliptycznej E oraz punkt P na krzywej E. Niech u będzie dowolnym unifikatorem w punkcie P. *Rząd funkcji r w punkcie P*, oznaczany symbolem $\mathrm{ord}_P(r)$, to liczba całkowita d występująca w równości 1.2.41. Ponadto:

- jeżeli $d = 0$, to funkcja r nie ma w punkcie P ani miejsca zerowego, ani bieguna;
- jeżeli $d > 0$, to mówimy, że *funkcja r ma w punkcie P d-krotne miejsce zerowe*;
- jeżeli $d < 0$, to mówimy, że *funkcja r ma w punkcie P $|d|$-krotny biegun.*

Przykład 1.2.45. Dana jest krzywa eliptyczna $E_{3,4}(\mathbb{F}(13))$. Funkcja wymierna $\frac{x^3}{y^2}$ na tej krzywej ma w punkcie $(0, 2)$ trzykrotne miejsce zerowe, a w punkcie $(12, 0)$ dwukrotny biegun. W punkcie $\mathcal{O}$ ma rząd równy 0.

Rząd funkcji wymiernej spełnia zależność podobną do tej związanej ze stopniem funkcji.

Twierdzenie 1.2.46. *Dane są funkcje wymierne r i s na krzywej eliptycznej E oraz punkt skończony P na krzywej E. Wówczas* $\mathrm{ord}_P(rs) = \mathrm{ord}_P(r) + \mathrm{ord}_P(s)$.

Możemy teraz wyrazić szereg ważnych własności wielomianów i funkcji wymiernych na krzywej eliptycznej związanych z miejscami zerowymi i biegunami.

Twierdzenie 1.2.47. *Dany jest wielomian f na krzywej eliptycznej E. Wówczas zachodzi następująca zależność:*

$$\sum_{P \in E \setminus \{\mathcal{O}\}} \mathrm{ord}_P(f) = \deg(f)$$

Wniosek 1.2.48. *Dana jest funkcja wymierna r na krzywej eliptycznej E. Wówczas zachodzi następująca zależność:*

$$\sum_{P \in E} \mathrm{ord}_P(r) = 0$$

Wniosek 1.2.49. *Funkcja wymierna na krzywej eliptycznej ma miejsce zerowe lub biegun tylko w skończonej liczbie punktów.*

Wniosek 1.2.50. *Krotność każdego miejsca zerowego lub bieguna funkcji wymiernej na krzywej eliptycznej jest skończona.*

Ze względu na definicję stopnia wielomianu na krzywej eliptycznej możemy jeszcze wysnuć następujący wniosek.

Wniosek 1.2.51. *Wielomian na krzywej eliptycznej nie może mieć stopnia równego* 1*, zatem nie może mieć jednego jednokrotnego miejsca zerowego.*

1.3. Dywizory

Rozmieszczenie miejsc zerowych i biegunów funkcji wymiernych na krzywej eliptycznej wykazuje wiele regularności. Narzędziem, które pozwoli nam je badać, są dywizory.

Definicja

Przed podaniem definicji dywizora na krzywej eliptycznej przypomnijmy, czym jest abelowa grupa wolna.

Definicja 1.3.1. Dany jest zbiór S, liczba naturalna n, parami różne elementy $s_1, \ldots, s_n$ zbioru S oraz liczby całkowite $a_1, \ldots, a_n$. *Skończona suma formalna nad zbiorem S długości n o składnikach $s_1, \ldots s_n$ i współczynnikach $a_1, \ldots, a_n$* to następujące wyrażenie:

$$\sum_{k=1}^{n} a_k \langle s_k \rangle$$

Abelowa grupa wolna generowana przez zbiór S to zbiór złożony ze wszystkich skończonych sum formalnych nad zbiorem S wyposażony w działanie polegające na formalnym dodaniu dwóch sum i pogrupowaniu składników ze względu na elementy zbioru S.

Uwaga 1.3.2. Sumę formalną nad zbiorem S będziemy często zapisywać w następującej postaci:

$$\sum_{s \in S} a(s) \langle s \rangle$$

Choć zapis tego nie sugeruje, funkcja $a \colon S \to \mathbb{Z}$ przyjmuje wartość niezerową tylko skończoną ilość razy.

Definicja 1.3.3. Dana jest krzywa eliptyczna E. *Dywizor na krzywej eliptycznej E* to element abelowej grupy wolnej generowanej przez punkty krzywej E. Zbiór wszystkich dywizorów na krzywej E oznaczamy symbolem $\mathrm{Div}(E)$.

Uwaga 1.3.4. Zgodnie z poczynioną przed chwilą uwagą 1.3.2 dywizory na krzywej eliptycznej E będziemy często zapisywać w następującej postaci:

$$\sum_{P \in E} a(P) \langle P \rangle$$

Dla dywizorów określamy stopień oraz normę.

Definicja 1.3.5. Dany jest dywizor Δ na krzywej eliptycznej E postaci $\Delta = \sum_{P \in E} a(P) \langle P \rangle$. *Stopień dywizora* Δ, oznaczany symbolem $\deg(\Delta)$, to następująca wielkość:

$$\deg(\Delta) = \sum_{P \in E} a(P)$$

Definicja 1.3.6. Dany jest dywizor Δ na krzywej eliptycznej E postaci $\Delta = \sum_{P \in E} a(P) \langle P \rangle$. *Norma dywizora* Δ, oznaczana symbolem $|\Delta|$, to następująca wielkość:

$$|\Delta| = \sum_{P \in E \setminus \{\mathcal{O}\}} |a(P)|$$

Przykład 1.3.7. Dana jest krzywa eliptyczna $E_{3,4}(\mathbb{F}(13))$. Dywizor $3 \langle (0,2) \rangle - 2 \langle (12,0) \rangle$ na tej krzywej ma stopień równy 1 i normę równą 5.

Dywizory i funkcje wymierne

Przydatność dywizorów polega na tym, że można za ich pomocą reprezentować informacje o wszystkich miejscach zerowych i biegunach funkcji wymiernej oraz ich krotnościach.

Definicja 1.3.8. Dana jest funkcja wymierna r na krzywej eliptycznej E. *Dywizor funkcji* r, oznaczany symbolem $\mathrm{div}(r)$, to dywizor na krzywej E określony następująco:

$$\mathrm{div}(r) = \sum_{P \in E} \mathrm{ord}_P(r) \langle P \rangle$$

Przykład 1.3.9. Dana jest krzywa eliptyczna $E_{3,4}(\mathbb{F}(13^2))$. Funkcja wymierna $\frac{x^3}{y^2}$ na tej krzywej ma dywizor równy $3\langle(0,2)\rangle + 3\langle(0,11)\rangle - 2\langle(12,0)\rangle - 2\langle(\alpha,0)\rangle - 2\langle(\beta,0)\rangle$, gdzie α i β to pierwiastki wielomianu $x^2 - x + 4$ w ciele $\mathbb{F}(13^2)$.

Okazuje się, że dywizor niosący informacje o miejscach zerowych i biegunach funkcji wymiernej wyznacza ją niemalże jednoznacznie.

Fakt 1.3.10. *Dwie funkcje wymierne na krzywej eliptycznej różniące się o czynnik stały mają taki sam dywizor.*

Twierdzenie 1.3.11. *Dane są funkcje wymierne r i s na krzywej eliptycznej E. Zależność $\mathrm{div}(r) = \mathrm{div}(s)$ zachodzi wtedy i tylko wtedy, gdy iloraz $\frac{r}{s}$ jest stały i niezerowy.*

Wniosek 1.3.12. *Funkcja wymierna na krzywej eliptycznej nie ma miejsca zerowego ani bieguna wtedy i tylko wtedy, gdy jest stała i niezerowa.*

Wniosek 1.3.13. *Dywizor na krzywej eliptycznej wyznacza funkcję wymierną z dokładnością do stałego niezerowego czynnika.*

Rozdział 2

Grupy na krzywych eliptycznych

Niezwykłą cechą krzywych eliptycznych jest fakt, że można na nich zadać niebanalną strukturę grupową. Grupy na krzywych eliptycznych są niezmiernie istotne w matematyce, również w kryptografii. Z punktu widzenia niniejszej pracy grupy na krzywych eliptycznych są o tyle ważne, że struktura grupowa krzywej eliptycznej pojawia się w definicji iloczynu Weila.
Dlatego też rozdział ten w całości poświęcony jest dokładnemu zdefiniowaniu i zbadaniu własności operacji grupowej na punktach krzywej eliptycznej. Operacja ta wydaje się dosyć nieintuicyjna, dlatego najpierw przeanalizujemy kilka innych pojęć, które motywują taką właśnie definicję działania grupowego na krzywej.
Rozdział ten, podobnie jak rozdział poprzedni, stanowi jedynie przypomnienie wiadomości o grupie na krzywej eliptycznej. Jednak ze względu na duże znaczenie grup na krzywych eliptycznych, przedstawimy dowody podawanych stwierdzeń.

2.1. Linie

Aby zdefiniować działanie grupowe na krzywej eliptycznej, będziemy wielokrotnie posługiwać się pewną specyficzną klasą wielomianów. Są to odpowiedniki wielomianów liniowych i linii prostych.

Definicja

Przyjmujemy następującą definicję linii na krzywej eliptycznej.

Definicja 2.1.1. Dana jest krzywa eliptyczna E nad ciałem $\mathbb{K}$ oraz trzy elementy a, b i c ciała $\mathbb{K}$. *Linia na krzywej eliptycznej E o współczynnikach a, b i c* to wielomian $ax + by + c$. Jeżeli $b = 0$, to mówimy, że linia l jest *pionowa*. Jeżeli $a = 0$ i $b = 0$, to mówimy, że linia l jest *zdegenerowana*.

Definicja 2.1.2. Dana jest linia l na krzywej eliptycznej E oraz punkt P na krzywej E. *Linia l przecina krzywą E w punkcie P*, jeżeli $l(P) = 0$.

Definicja 2.1.3. Dana jest linia l na krzywej eliptycznej E oraz punkt $P \in E$. *Linia l jest styczna do krzywej E w punkcie P*, jeżeli $\mathrm{ord}_P(l) > 1$.

Przykład 2.1.4. Dana jest krzywa eliptyczna $E_{3,4}(\mathbb{F}(13))$. Linia $l_1 = 2x + 3y + 6$ na tej krzywej przecina ją w punktach $(0, 11)$, $(5, 12)$ i $(7, 2)$. Nie jest ona styczna do krzywej w żadnym z tych punktów. Linia pionowa $l_2 = x + 1$ przecina krzywą w punkcie $(12, 0)$ i jest w tym punkcie styczna.

Własności

Następujące własności są to po prostu podane wcześniej fakty zastosowane do przypadku linii.

Fakt 2.1.5. *Dana jest niezdegenerowana linia l na krzywej eliptycznej E postaci $l = ax + by + c$. Jeżeli $b \neq 0$, to* $\deg(l) = 3$*, w przeciwnym razie* $\deg(l) = 2$.

Fakt 2.1.6. *Dana jest niezdegenerowana linia l na krzywej eliptycznej E postaci $l = ax + by + c$. Jeżeli $b \neq 0$, to* $|\mathrm{div}(l)| = 3$*, w przeciwnym razie* $|\mathrm{div}(l)| = 2$.

Będziemy mieć dużo do czynienia z liniami, dlatego musimy uściślić wyniki, które przedstawia fakt 2.1.6.

Fakt 2.1.7. *Dana jest niezdegenerowana linia l na krzywej eliptycznej E. Niech P, Q i R będą trzema parami różnymi punktami skończonymi krzywej E. Wówczas dywizor* $\mathrm{div}(l)$ *może mieć jedną z następujących postaci:*

- $\mathrm{div}(l) = \langle P\rangle + \langle Q\rangle + \langle R\rangle - 3\langle \mathcal{O}\rangle$;
- $\mathrm{div}(l) = 2\langle P\rangle + \langle Q\rangle - 3\langle \mathcal{O}\rangle$;
- $\mathrm{div}(l) = 3\langle P\rangle - 3\langle \mathcal{O}\rangle$;
- $\mathrm{div}(l) = \langle P\rangle + \langle Q\rangle - 2\langle \mathcal{O}\rangle$;
- $\mathrm{div}(l) = 2\langle P\rangle - 2\langle \mathcal{O}\rangle$.

Przykład 2.1.8. Dana jest krzywa eliptyczna $E_{2,3}(\mathbb{F}(13))$. Wówczas dywizory następujących linii nad tą krzywą wyczerpują wszystkie przypadki wymienione w fakcie 2.1.7:

$$\begin{aligned}
\mathrm{div}(5x + 12y + 4) &= \langle(0,4)\rangle + \langle(3,6)\rangle + \langle(9,10)\rangle - 3\langle\mathcal{O}\rangle \\
\mathrm{div}(3x + y + 9) &= 2\langle(0,4)\rangle + \langle(9,3)\rangle - 3\langle\mathcal{O}\rangle \\
\mathrm{div}(3x + y + 11) &= 3\langle(3,6)\rangle - 3\langle\mathcal{O}\rangle \\
\mathrm{div}(x) &= \langle(0,4)\rangle + \langle(0,9)\rangle - 2\langle\mathcal{O}\rangle \\
\mathrm{div}(x + 1) &= 2\langle(12,0)\rangle - 2\langle\mathcal{O}\rangle
\end{aligned}$$

Uwaga 2.1.9. Składniki dywizorów wymienionych w fakcie 2.1.7 są pogrupowane. Czasami nie będziemy ich grupować, tylko zapisywać w jednej z następujących postaci:

- $\mathrm{div}(l) = \langle P\rangle + \langle Q\rangle + \langle R\rangle - 3\langle \mathcal{O}\rangle$ – jest to najbardziej ogólna forma, może zdarzyć się, że punkty P, Q oraz R nie są parami różne lub że jeden z nich nie jest skończony;
- $\mathrm{div}(l) = \langle P\rangle + \left\langle \overline{P}\right\rangle - 2\langle \mathcal{O}\rangle$ – jest to ogólna postać linii pionowych, może zdarzyć się, że $P = \overline{P}$.

Następujące twierdzenia pokazują, w jaki sposób linie przecinają się z krzywymi eliptycznymi.

Twierdzenie 2.1.10. *Dane są dwa różne punkty skończone P i Q na krzywej eliptycznej E nad ciałem $\mathbb{K}$. Wówczas przez punkty P i Q przechodzi jedna (z dokładnością do czynnika stałego) linia.*

Dowód. Oznaczmy $P = (a, b)$, $Q = (c, d)$, gdzie $a, b, c, d \in \mathbb{K}$ i rozważmy dwa przypadki.

1. Jeżeli $P = \overline{Q}$, to rozpatrujemy linię:

$$l(x, y) = x - a \tag{2.1.11}$$

   Jak nietrudno sprawdzić, $l(P) = 0$ oraz $l(Q) = 0$, ponieważ $c = a$.

2. Jeżeli $P \neq \overline{Q}$, to rozpatrujemy linię:

$$l(x, y) = \left(\frac{d-b}{c-a}\right)(x - a) - (y - b) \tag{2.1.12}$$

   Ponownie łatwo sprawdzić, że $l(P) = 0$ oraz $l(Q) = 0$. Zauważmy też, że iloraz $\frac{d-b}{c-a}$ jest dobrze określony, ponieważ $c \neq a$.

W obu przypadkach można pokazać, że inne linie przechodzące przez punkty P i Q różnią się od linii l o czynnik stały. Wystarczy rozpatrzyć odpowiedni układ równań liniowych na współczynniki linii: jedno równanie ma postać $l(P) = 0$, a drugie – $l(Q) = 0$. Są to dwa równania liniowe trzech zmiennych, przestrzeń rozwiązań jest więc jednowymiarowa. Widać też, że wszystkie rozwiązania są proporcjonalne, ponieważ układ równań jest jednorodny. □

Twierdzenie 2.1.13. *Dane są dwa różne punkty skończone P i Q na krzywej eliptycznej E nad ciałem $\mathbb{K}$. Niech l będzie linią przechodzącą przez punkty P i Q. Wówczas:*

- *jeżeli $P = \overline{Q}$, to* $\operatorname{div}(l) = \langle P\rangle + \langle Q\rangle - 2\langle \mathcal{O}\rangle$*;*
- *jeżeli $P \neq \overline{Q}$, to istnieje punkt skończony R na krzywej E taki, że* $\operatorname{div}(l) = \langle P\rangle + \langle Q\rangle + \langle R\rangle - 3\langle \mathcal{O}\rangle$*, może przy tym zdarzyć się, że $R = P$ lub $R = Q$.*

Dowód. Rozważmy oba przypadki.

1. Jeżeli $P = \overline{Q}$, to na podstawie twierdzeń 2.1.10 i 1.2.47 oraz faktu 2.1.5 widzimy, że linia l ma dwa różne miejsca zerowe lub jedno podwójne. Drugi przypadek jest niemożliwy, ponieważ punkty P i Q są jej miejscami zerowymi i są różne. Zatem $\operatorname{div}(l) = \langle P\rangle + \langle Q\rangle - 2\langle \mathcal{O}\rangle$.

2. Jeżeli $P \neq \overline{Q}$, to ponownie z twierdzeń 2.1.10 i 1.2.47 oraz faktu 2.1.5 widzimy, że linia l ma trzy różne miejsca zerowe, jedno pojedyncze i jedno podwójne lub jedno potrójne. Ostatni przypadek znów jest niemożliwy. W pozostałych przypadkach można dobrać punkt skończony R (być może równy P lub Q) taki, że $\operatorname{div}(l) = \langle P\rangle + \langle Q\rangle + \langle R\rangle - 3\langle \mathcal{O}\rangle$.

□

Twierdzenie 2.1.14. *Dane są dwa różne, niesprzężone do siebie punkty skończone P i Q na krzywej eliptycznej E nad ciałem $\mathbb{K}$ postaci $P = (a, b)$ i $Q = (c, d)$, gdzie $a, b, c, d \in \mathbb{K}$. Niech l będzie linią przechodzącą przez punkty P i Q. Niech R będzie punktem skończonym na krzywej E postaci $R = (e, f)$, gdzie $e, f \in \mathbb{K}$, takim, że* $\operatorname{div}(l) = \langle P\rangle + \langle Q\rangle + \langle R\rangle - 3\langle \mathcal{O}\rangle$*. Wówczas:*

$$e = \left(\frac{d-b}{c-a}\right)^2 - a - c \tag{2.1.15}$$

$$f = b + \left(\frac{d-b}{c-a}\right)(e - a) \tag{2.1.16}$$

Dowód. Rozważmy normę $N(l)$ linii l. Jest ona równa:

$$\begin{aligned} N(l) = l\bar{l} &= \left(\left(\frac{d-b}{c-a}\right)(x-a) - (y-b)\right)\overline{\left(\left(\frac{d-b}{c-a}\right)(x-a) - (y-b)\right)} \\ &= \left(\left(\frac{d-b}{c-a}\right)(x-a) + b\right)^2 - x^3 - Ax - B \\ &= -x^3 + \left(\frac{d-b}{c-a}\right)^2 x^2 + \cdots \end{aligned}$$

Na mocy wniosku 1.2.35 jej miejsca zerowe to a, c oraz e. Współrzędną e znajdujemy ze wzoru Viete'a na sumę pierwiastków:

$$\begin{aligned} a + c + e &= \left(\frac{d-b}{c-a}\right)^2 \\ e &= \left(\frac{d-b}{c-a}\right)^2 - a - c \end{aligned}$$

Współrzędną f obliczamy z równania $l(R) = 0$:

$$\begin{aligned} \left(\frac{d-b}{c-a}\right)(e-a) - (f-b) &= 0 \\ f &= b + \left(\frac{d-b}{c-a}\right)(e-a) \end{aligned}$$

□

Twierdzenie 2.1.17. *Dany jest punkt skończony P na krzywej eliptycznej E nad ciałem $\mathbb{K}$. Wówczas przez punkt P przechodzi jedna (z dokładnością do czynnika stałego) linia styczna do krzywej E w punkcie P.*

Dowód. Oznaczmy $P = (a, b)$, gdzie $a, b \in \mathbb{K}$ i rozważmy dwa przypadki.

1. Jeżeli P jest punktem rzędu dwa, tzn. $P = \overline{P}$ i $b = 0$, to rozpatrujemy linię:

$$l(x, y) = x - a \tag{2.1.18}$$

   Ma ona stopień równy 2, a jej dywizor, zgodnie z uwagą 2.1.9, ma postać $\langle P\rangle + \left\langle \overline{P}\right\rangle - 2\langle\mathcal{O}\rangle$. Skoro P jest punktem rzędu dwa, to dywizor ten jest równy $2\langle P\rangle - 2\langle\mathcal{O}\rangle$, zatem linia l jest styczna do krzywej w punkcie P.

2. Jeżeli P nie jest punktem rzędu dwa, to rozpatrujemy linię:

$$l(x, y) = \left(\frac{3a^2 + A}{2b}\right)(x-a) - (y-b) \tag{2.1.19}$$

   Linia l ma w punkcie P miejsce zerowe, zatem $\mathrm{ord}_P(l) > 0$. Pokażemy, że $\mathrm{ord}_P(l) > 1$. W tym celu wykażemy, że jeśli $l = us$, gdzie u jest unifikatorem w punkcie P, a s jest funkcją wymierną, to $s(P) = 0$. Ponieważ punkt P nie jest podwójny, bierzemy

unifikator $u(x,y) = x - a$. Wyznaczamy funkcję wymierną s.

$$\begin{aligned} s = \frac{l}{u} &= \frac{\left(\frac{3a^2+A}{2b}\right)(x-a) - (y-b)}{x-a} \\ &= \frac{3a^2+A}{2b} - \frac{y-b}{x-a} \\ &= \frac{3a^2+A}{2b} - \frac{y^2-b^2}{(x-a)(y+b)} \\ &= \frac{3a^2+A}{2b} - \frac{x^3+Ax+B-a^3-Aa-B}{(x-a)(y+b)} \\ &= \frac{3a^2+A}{2b} - \frac{(x-a)(x^2+ax+a^2+A)}{(x-a)(y+b)} \\ &= \frac{3a^2+A}{2b} - \frac{x^2+ax+a^2+A}{y+b} \end{aligned}$$

Widzimy teraz, że $s(P) = 0$, Zatem $\mathrm{ord}_P(l) > 1$, czyli linia l jest styczna do krzywej w punkcie P.

Podobnie jak w dowodzie twierdzenia 2.1.10, to, że inne linie styczne do krzywej w punkcie P różnią się od linii l o czynnik stały, wynika z rozpatrzenia odpowiedniego układu równań liniowych. Jedno równanie ma postać $l(P) = 0$, a drugie – $s(P) = (\frac{l}{u})(P) = 0$ (można je przekształcić do postaci równania liniowego na współczynniki linii). □

Twierdzenie 2.1.20. *Dany jest punkt skończony P na krzywej eliptycznej E nad ciałem $\mathbb{K}$. Niech l będzie linią styczną do krzywej E w punkcie P. Wówczas:*

- *jeśli $P = \overline{P}$, to* $\mathrm{div}(l) = 2\langle P\rangle - 2\langle\mathcal{O}\rangle$*;*
- *jeśli $P \neq \overline{P}$, to* $\mathrm{div}(l) = 2\langle P\rangle + \langle Q\rangle - 3\langle\mathcal{O}\rangle$*, przy czym punkt Q jest skończony i może zdarzyć się, że $Q = P$.*

Dowód. Rozważmy dwa przypadki.

1. Jeżeli $P = \overline{P}$, to na podstawie twierdzeń 2.1.17 i 1.2.47 oraz faktu 2.1.5 widzimy, że linia l ma dwa różne miejsca zerowe lub jedno podwójne. Pierwszy przypadek jest niemożliwy, ponieważ punkt P jest jej podwójnym miejscem zerowym. Zatem $\mathrm{div}(l) = 2\langle P\rangle - 2\langle\mathcal{O}\rangle$.
2. Jeżeli $P \neq \overline{P}$, to ponownie z twierdzeń 2.1.17 i 1.2.47 oraz faktu 2.1.5 widzimy, że linia l ma trzy różne miejsca zerowe, jedno pojedyncze i jedno podwójne lub jedno potrójne. Pierwszy przypadek znów jest niemożliwy. W pozostałych przypadkach można dobrać punkt skończony Q (być może równy P) taki, że $\mathrm{div}\, l = 2\langle P\rangle + \langle Q\rangle - 3\langle\mathcal{O}\rangle$.

□

Twierdzenie 2.1.21. *Dany jest punkt skończony P na krzywej eliptycznej E nad ciałem $\mathbb{K}$ postaci $P = (a,b)$, gdzie $a, b \in \mathbb{K}$, niebędący punktem rzędu dwa. Niech l będzie linią styczną do krzywej E w punkcie P. Niech Q będzie punktem skończonym krzywej postaci $Q = (c,d)$, gdzie $c, d \in \mathbb{K}$, takim, że* $\mathrm{div}(l) = 2\langle P\rangle + \langle Q\rangle - 3\langle\mathcal{O}\rangle$*. Wówczas:*

$$c = \left(\frac{3a^2+A}{2b}\right)^2 - 2a \tag{2.1.22}$$

$$d = b + \left(\frac{3a^2+A}{2b}\right)(c-a) \tag{2.1.23}$$

Dowód. Rozważmy normę $N(l)$ linii l. Jest ona równa:

$$\begin{aligned} N(l) = l\bar{l} &= \left(\left(\frac{3a^2+A}{2b}\right)(x-a)-(y-b)\right)\overline{\left(\left(\frac{3a^2+A}{2b}\right)(x-a)-(y-b)\right)} \\ &= \left(\left(\frac{3a^2+A}{2b}\right)(x-a)+b\right)^2 - x^3 - Ax - B \\ &= -x^3 + \left(\frac{3a^2+A}{2b}\right)^2 x^2 + \cdots \end{aligned}$$

Na mocy wniosku 1.2.35 jej miejsca zerowe to a (podwójne) oraz c. Współrzędną c znajdujemy ze wzorów Viete'a na sumę pierwiastków:

$$\begin{aligned} 2a + c &= \left(\frac{3a^2+A}{2b}\right)^2 \\ c &= \left(\frac{3a^2+A}{2b}\right)^2 - 2a \end{aligned}$$

Współrzędną d obliczamy z równania $l(Q) = 0$:

$$\begin{aligned} \left(\frac{3a^2+A}{2b}\right)(c-a)-(d-b) &= 0 \\ d &= b + \left(\frac{3a^2+A}{2b}\right)(c-a) \end{aligned}$$

□

Przykład 2.1.24. Dana jest krzywa eliptyczna $E_{3,4}(\mathbb{F}(13))$. Przez punkty $(3,1)$ i $(7,2)$ na tej krzywej przechodzi linia $10x + 12y + 10$, której trzecim miejscem zerowym jest punkt $(12,0)$. Przez punkt $(11,4)$ na tej krzywej przechodzi linia styczna $10x + 12y + 11$, której drugim miejscem zerowym jest punkt $(0,11)$.

Niezwykle istotny jest fakt, że współczynniki linii przechodzącej przez dwa punkty lub stycznej w punkcie, jak i współrzędne dodatkowego punktu przecięcia takiej linii z krzywą eliptyczną dają się wyrazić jako nieskomplikowane wyrażenia wymierne.

Fakt 2.1.25. *Dana jest krzywa eliptyczna E nad ciałem $\mathbb{L}$ o parametrach z ciała $\mathbb{K} \subset \mathbb{L}$ oraz dwa różne punkty skończone P i Q na krzywej E. Niech l będzie linią przechodzącą przez punkty P i Q, a R będzie trzecim punktem przecięcia linii l z krzywą E, tzn.* $\operatorname{div}(l) = \langle P\rangle + \langle Q\rangle + \langle R\rangle - 3\langle\mathcal{O}\rangle$. *Wówczas, jeżeli punkty P i Q są $\mathbb{K}$-wymierne, to również punkt R jest $\mathbb{K}$-wymierny, a współczynniki linii l są elementami ciała $\mathbb{K}$.*

Fakt 2.1.26. *Dana jest krzywa eliptyczna E nad ciałem $\mathbb{L}$ o parametrach z ciała $\mathbb{K} \subset \mathbb{L}$ oraz punkt skończony P na krzywej E niebędący punktem rzędu dwa. Niech l będzie linią styczną do krzywej E w punkcie P, a Q będzie drugim punktem przecięcia linii l z krzywą E, tzn.* $\operatorname{div}(l) = 2\langle P\rangle + \langle Q\rangle - 3\langle\mathcal{O}\rangle$. *Wówczas, jeżeli punkt P jest $\mathbb{K}$-wymierny, to również punkt Q jest $\mathbb{K}$-wymierny, a współczynniki linii l są elementami ciała $\mathbb{K}$.*

Istota tych faktów polega na tym, że ciało $\mathbb{K}$ nie musi być algebraicznie domknięte, a mimo to zawsze da się wyznaczyć współczynniki linii oraz dodatkowego punktu przecięcia linii z krzywą. A priori nie było to oczywiste – uzyskane wzory mogły być bardziej skomplikowane (np. wymagać pierwiastkowania) i konkretne wartości niezbędne do wyrażenia współczynników i współrzędnych znajdowałyby się dopiero w domknięciu algebraicznym $\overline{\mathbb{K}}$ ciała $\mathbb{K}$.

2.2. Szczególne grupy dywizorów

Zbadamy teraz kilka ważnych grup dywizorów na krzywej eliptycznej. Analiza tych grup doprowadzi nas do definicji działania grupowego na krzywej eliptycznej.

Definicje grup

Największą rozpatrywaną grupą jest oczywiście grupa złożona ze wszystkich dywizorów na krzywej eliptycznej.

Definicja 2.2.1. Dana jest krzywa eliptyczna E. *Grupa dywizorów na krzywej* E, oznaczana symbolem $\mathrm{Div}(E)$, jest to abelowa grupa wolna generowana przez punkty krzywej E.

Spośród wszystkich dywizorów na krzywej eliptycznej wyróżnioną pozycję zajmują te, które są dywizorami funkcji wymiernych.

Definicja 2.2.2. Dany jest dywizor Δ na krzywej eliptycznej E. Dywizor Δ jest *główny*, jeżeli jest on dywizorem pewnej funkcji wymiernej na krzywej E.

Następujące twierdzenie pokazuje, że dywizory główne tworzą grupę.

Twierdzenie 2.2.3. *Dane są funkcje wymierne r oraz s na krzywej eliptycznej E. Wówczas zachodzi zależność* $\mathrm{div}(rs) = \mathrm{div}(r) + \mathrm{div}(s)$.

Definicja 2.2.4. Dana jest krzywa eliptyczna E. *Grupa dywizorów głównych na krzywej* E, oznaczana symbolem $\mathrm{Prin}(E)$, to podgrupa grupy $\mathrm{Div}(E)$ składająca się ze wszystkich dywizorów głównych na krzywej E.

Grupy $\mathrm{Div}(E)$ oraz $\mathrm{Prin}(E)$ są abelowe, zatem istnieje ich grupa ilorazowa.

Definicja 2.2.5. Dana jest krzywa eliptyczna E. *Grupa Picarda na krzywej* E, oznaczana symbolem $\mathrm{Pic}(E)$, to grupa ilorazowa $\mathrm{Div}(E)/\mathrm{Prin}(E)$.

Dwa dywizory Δ_1 i Δ_2 należą do tej samej warstwy będącej elementem grupy $\mathrm{Pic}(E)$ wtedy i tylko wtedy, gdy ich różnica jest dywizorem głównym. Fakt ten prowadzi nas do następującej definicji.

Definicja 2.2.6. Dane są dywizory Δ_1 i Δ_2 na krzywej eliptycznej E. Dywizory Δ_1 i Δ_2 są *równoważne*, co oznaczamy symbolem $\Delta_1 \sim \Delta_2$, jeżeli ich różnica $\Delta_1 - \Delta_2$ jest dywizorem głównym.

Wniosek 1.2.48 daje częściową charakteryzację dywizorów głównych.

Fakt 2.2.7. *Jeżeli Δ jest dywizorem głównym, to* $\deg(\Delta) = 0$.

To prowadzi nas do definicji kolejnej wartej uwagi grupy.

Definicja 2.2.8. Dana jest krzywa eliptyczna E. *Grupa dywizorów stopnia zero na krzywej* E, oznaczana symbolem $\mathrm{Div}^0(E)$, to podgrupa grupy $\mathrm{Div}(E)$ składająca się ze wszystkich dywizorów stopnia zero na krzywej E.

Zachodzi następujący ciąg inkluzji: $\mathrm{Prin}(E) \subset \mathrm{Div}^0(E) \subset \mathrm{Div}(E)$. Zbadajmy wobec tego grupy ilorazowe. Grupa $\mathrm{Div}(E)/\mathrm{Div}^0(E)$ nie jest specjalnie interesująca, ponieważ jest izomorficzna z grupą liczb całkowitych $\mathbb{Z}$. Za to grupa $\mathrm{Div}^0(E)/\mathrm{Prin}(E)$ będzie dla nas bardzo ważna.

Definicja 2.2.9. Dana jest krzywa eliptyczna E. *Zerowa grupa Picarda na krzywej* E, oznaczana symbolem $\mathrm{Pic}^0(E)$, to grupa ilorazowa $\mathrm{Div}^0(E)/\mathrm{Prin}(E)$.

Liniowe redukcje dywizorów

Konstrukcja różnych grup na krzywej eliptycznej opiera się na równoważności dywizorów. Dwa dywizory są równoważne, gdy różnią się o dywizor główny, czyli o dywizor pewnej funkcji wymiernej. Jak wskazać funkcję wymierną, która jest „różnicą" między dwoma dywizorami? Następujące dwa twierdzenia pokazują, że taką funkcję wymierną można skonstruować, mnożąc i dzieląc odpowiednie linie.

Twierdzenie 2.2.10. *Dany jest dywizor Δ na krzywej eliptycznej E. Wówczas istnieje dywizor $\tilde{\Delta}$ na krzywej E postaci $\tilde{\Delta} = \langle P\rangle + n\langle\mathcal{O}\rangle$, gdzie $P \in E$ i $n \in \mathbb{Z}$, taki, że $\Delta \sim \tilde{\Delta}$ oraz* $\deg(\Delta) = \deg(\tilde{\Delta})$.

Dowód. Oznaczmy $\Delta = \sum_{P\in E} a(P)\langle P\rangle$. Mogą zajść następujące przypadki:

1. dywizor Δ zawiera dwa składniki $a(P)\langle P\rangle$ i $a(Q)\langle Q\rangle$, gdzie P i Q to dwa różne punkty skończone krzywej, a współczynniki $a(P)$ i $a(Q)$ są niezerowe oraz tego samego znaku;
2. dywizor Δ zawiera składnik $a(P)\langle P\rangle$, gdzie P punkt skończony krzywej, a $|a(P)| > 1$;
3. dywizor Δ ma postać $\langle P\rangle - \langle Q\rangle + n\langle\mathcal{O}\rangle$, gdzie P i Q to punkty skończone krzywej;
4. dywizor Δ ma postać $\pm\langle P\rangle + n\langle\mathcal{O}\rangle$ lub $n\langle\mathcal{O}\rangle$, gdzie P to punkt skończony krzywej.

Pokażemy, że w każdym przypadku z wyjątkiem ostatniego możemy wskazać dywizor Δ' taki, że $\Delta \sim \Delta'$, $\deg(\Delta) = \deg(\Delta')$ oraz $|\Delta| > |\Delta'|$.

1. Dla ustalenia uwagi przyjmijmy, że współczynniki $a(P)$ oraz $a(Q)$ są dodatnie. Rozważmy dywizor $\Delta' = \Delta - \mathrm{div}(l)$, gdzie l to linia przechodząca przez punkty P oraz Q, ma ona dywizor $\mathrm{div}(l)$ równy $\langle P\rangle + \langle Q\rangle + \langle R\rangle - 3\langle\mathcal{O}\rangle$. Może zdarzyć się, że punkt R jest równy P, Q lub $\mathcal{O}$. W każdym przypadku łatwo sprawdzić, że $|\Delta| > |\Delta'|$.
2. Dla ustalenia uwagi przyjmijmy, że współczynnik $a(P)$ jest dodatni. Rozważmy dywizor $\Delta' = \Delta - \mathrm{div}(l)$, gdzie l jest linią styczną do krzywej w punkcie P, ma ona dywizor $\mathrm{div}(l)$ równy $2\langle P\rangle + \langle R\rangle - 3\langle\mathcal{O}\rangle$. Również w tym przypadku $|\Delta| > |\Delta'|$, nawet jeśli punkt R jest równy P lub $\mathcal{O}$.
3. Rozważmy linię l przechodzącą przez punkty P oraz $\overline{P}$. Ma ona dywizor $\mathrm{div}(l)$ równy $\langle P\rangle + \left\langle\overline{P}\right\rangle - 2\langle\mathcal{O}\rangle$. Rozważamy dywizor $\Delta' = \Delta - \mathrm{div}(l) = -\left\langle\overline{P}\right\rangle - \langle Q\rangle + n'\langle\mathcal{O}\rangle$, do którego stosujemy rozumowanie opisane w punkcie pierwszym lub drugim (zależnie od tego, czy $\overline{P} = Q$, czy nie) i otrzymujemy dywizor Δ'', który ma normę mniejszą niż Δ. Zwróćmy uwagę, że rozumowanie w tym kroku jest poprawne również wtedy, gdy P jest punktem podwójnym – wówczas zamiast linii przechodzącej przez punkty P oraz $\overline{P}$ bierzemy linię styczną do krzywej w punkcie P.

Teraz konstruujemy ciąg dywizorów $\Delta_0, \Delta_1, \Delta_2, \ldots$ taki, że $\Delta_0 = \Delta$, a dywizor Δ_{i+1} otrzymujemy poprzez zastosowanie do dywizora Δ_i jednej z opisanych redukcji. Widać, że wszystkie dywizory w ciągu są sobie równoważne oraz mają taki sam stopień. Co więcej, ciąg ten jest skończony, bo normy kolejnych dywizorów maleją, a nie mogą tego robić w nieskończoność – są to liczby naturalne. Ostatni dywizor w ciągu Δ_k musi więc mieć postać $\pm\langle P\rangle + n\langle\mathcal{O}\rangle$ lub $n\langle\mathcal{O}\rangle$, gdyż w przeciwnym razie można by kontynuować redukowanie.
Jeżeli Δ_k jest postaci $\langle P\rangle + n\langle\mathcal{O}\rangle$, to jest to szukany dywizor. Jeżeli Δ_k ma postać $n\langle\mathcal{O}\rangle$, to również jest to szukany dywizor – zapisujemy go w postaci $\Delta_k = \langle\mathcal{O}\rangle + (n-1)\langle\mathcal{O}\rangle$. Jeśli zaś Δ_k jest postaci $-\langle P\rangle + n\langle\mathcal{O}\rangle$, to szukanym dywizorem jest dywizor $\Delta_k + \mathrm{div}(l)$, gdzie l to linia przechodząca przez punkty P oraz $\overline{P}$ (lub styczna do krzywej w punkcie P, jeżeli jest to punkt podwójny). Linia l ma dywizor $\langle P\rangle + \left\langle\overline{P}\right\rangle - 2\langle\mathcal{O}\rangle$, zatem $\Delta_k + \mathrm{div}(l) = \left\langle\overline{P}\right\rangle + (n-2)\langle\mathcal{O}\rangle$. □

Lemat 2.2.11. *Dane są punkty P i Q na krzywej eliptycznej E. Jeżeli $\langle P\rangle \sim \langle Q\rangle$, to $P = Q$.*

Dowód. Rozważmy najpierw przypadek, gdy jeden z punktów P i Q jest równy $\mathcal{O}$. Dla ustalenia uwagi przyjmijmy, że jest to punkt Q. Zatem $\langle P\rangle - \langle\mathcal{O}\rangle \sim 0$, czyli istnieje funkcja wymierna r, której dywizorem jest $\langle P\rangle - \langle\mathcal{O}\rangle$. Z kształtu dywizora widzimy, że jeżeli $P \neq \mathcal{O}$, to funkcja r jest wielomianem, który ma jedno jednokrotne miejsce zerowe w punkcie P. Na mocy wniosku 1.2.51 nie jest to możliwe, zatem $P = \mathcal{O}$.
Przyjmijmy więc, że punkty P i Q są skończone i weźmy funkcję wymierną r, której dywizorem jest $\langle P\rangle - \langle Q\rangle$. Oznaczmy $P = (a, b)$ i rozważmy dwa przypadki.

1. $P = \overline{P}$. Unifikatorem w punkcie P jest funkcja $u(x, y) = y$. Widzimy, że $\mathrm{ord}_P(r) = 1$, zatem $r = us$, gdzie $s(P) \neq 0$. Wiemy, że $\mathrm{div}(u) = \langle P\rangle + \langle P_1\rangle + \langle P_2\rangle - 3\langle\mathcal{O}\rangle$, gdzie P_1 i P_2 to pozostałe punkty podwójne, skąd otrzymujemy, że $\mathrm{div}(\frac{1}{s}) = \langle Q\rangle + \langle P_1\rangle + \langle P_2\rangle - 3\langle\mathcal{O}\rangle$. Funkcja $\frac{1}{s}$ jest zatem linią. Linia przechodząca przez punkty P_1 i P_2 musi przechodzić również przez trzeci punkt podwójny, czyli przez P. Zatem $Q = P$.

2. $P \neq \overline{P}$. Unifikatorem w punkcie P jest teraz funkcja $u(x, y) = x - a$. Po analogicznych rachunkach otrzymujemy $\mathrm{div}(\frac{1}{s}) = \langle Q\rangle + \left\langle \overline{P}\right\rangle - 2\langle\mathcal{O}\rangle$. Funkcja $\frac{1}{s}$ jest tym razem linią pionową, skąd $\overline{Q} = \overline{P}$, czyli $Q = P$.

□

Twierdzenie 2.2.12. *Dany jest dywizor zerowego stopnia Δ na krzywej eliptycznej E. Wówczas istnieje dokładnie jeden dywizor $\tilde{\Delta}$ na krzywej E postaci $\tilde{\Delta} = \langle P\rangle - \langle\mathcal{O}\rangle$, gdzie $P \in E$, taki, że $\Delta \sim \tilde{\Delta}$ oraz $\deg(\Delta) = \deg(\tilde{\Delta})$.*

Dowód. Zastosujmy twierdzenie 2.2.10. Otrzymujemy, że istnieje dywizor $\tilde{\Delta}$ postaci $\langle P\rangle + n\langle\mathcal{O}\rangle$, który jest równoważny dywizorowi Δ oraz ma taki sam stopień. Skoro $\deg(\tilde{\Delta}) = \deg(\Delta) = 0$, to musi zachodzić $n = -1$, czyli $\tilde{\Delta} = \langle P\rangle - \langle\mathcal{O}\rangle$. Jednoznaczność otrzymujemy udowodnionego przed chwilą lematu 2.2.11. □

Przykład 2.2.13. Dana jest krzywa eliptyczna $E_{3,4}(\mathbb{F}(13))$. Redukcja dywizora $\langle(0, 2)\rangle + \langle(5, 1)\rangle - \langle(7, 11)\rangle - \langle(12, 0)\rangle + \langle\mathcal{O}\rangle$ przebiega następująco.

1. Redukujemy za pomocą linii $5x + 12y + 10$ i otrzymujemy dywizor $-2\langle(7, 11)\rangle - \langle(12, 0)\rangle + 4\langle\mathcal{O}\rangle$.

2. Redukujemy za pomocą linii $8x+12y+7$ i otrzymujemy dywizor $\langle(11, 4)\rangle - \langle(12, 0)\rangle + \langle\mathcal{O}\rangle$.

3. Redukujemy za pomocą linii $x+2$, a następnie za pomocą linii $4x+10y+4$ i otrzymujemy dywizor $\langle(6, 2)\rangle$.

Zerowa grupa Picarda

Dzięki twierdzeniom 2.2.10 oraz 2.2.12 możemy szczegółowo zanalizować zerową grupę Picarda.

Twierdzenie 2.2.14. *Dana jest krzywa eliptyczna E. Niech $[\Delta]$ oznacza klasę abstrakcji dywizora Δ w grupie $\mathrm{Pic}^0(E)$. Niech funkcja $\pi\colon E \to \mathrm{Pic}^0(E)$ będzie określona wzorem:*

$$\pi(P) = [\langle P\rangle - \langle\mathcal{O}\rangle] \tag{2.2.15}$$

Wówczas funkcja π jest bijekcją.

Dowód. Funkcja π jest różnowartościowa. Jeżeli $\pi(P) = \pi(Q)$, to znaczy, że dywizory $\langle P\rangle - \langle\mathcal{O}\rangle$ oraz $\langle Q\rangle - \langle\mathcal{O}\rangle$ są w tej samej warstwie, skąd $\langle P\rangle - \langle Q\rangle \sim 0$. Z lematu 2.2.11 dostajemy $P = Q$.
Funkcja π jest „na". Weźmy dowolną warstwę będącą elementem zerowej grupy Picarda i wybierzmy z niej dowolny dywizor Δ. Zgodnie z twierdzeniem 2.2.10 jest on równoważny pewnemu dywizorowi postaci $\langle P\rangle - \langle\mathcal{O}\rangle$. Wówczas warstwa $\pi(P)$ to właśnie warstwa zawierająca dywizor Δ. □

Wniosek 2.2.16. *Zerową grupę Picarda możemy utożsamić ze zbiorem punktów krzywej eliptycznej przyporządkowując punktowi P warstwę zawierającą dywizor $\langle P\rangle - \langle\mathcal{O}\rangle$.*

Zbadajmy, który punkt jest reprezentantem warstwy, która jest sumą dwóch warstw reprezentowanych przez zadane punkty.

Fakt 2.2.17. *Dodawanie dwóch elementów zerowej grupy Picarda, które są warstwami zawierającymi dywizory $\langle P\rangle - \langle\mathcal{O}\rangle$ oraz $\langle Q\rangle - \langle\mathcal{O}\rangle$, przeprowadzamy następująco.*

1. *Jeżeli $P = \mathcal{O}$ (odpowiednio $Q = \mathcal{O}$), to wynik dodawania jest równy warstwie zawierającej dywizor $\langle Q\rangle - \langle\mathcal{O}\rangle$ (odpowiednio $\langle P\rangle - \langle\mathcal{O}\rangle$).*

2. *W przeciwnym razie dodajemy dywizory $\langle P\rangle - \langle\mathcal{O}\rangle$ oraz $\langle Q\rangle - \langle\mathcal{O}\rangle$ zgodnie z działaniem w grupie $\mathrm{Div}^0(E)$. Otrzymujemy:*

$$\langle P\rangle + \langle Q\rangle - 2\langle\mathcal{O}\rangle$$

3. *Zgodnie z twierdzeniem 2.2.12 redukujemy wynik za pomocą linii l_1 przechodzącej przez punkty P i Q (lub stycznej, jeśli $P = Q$). Otrzymujemy:*

$$\begin{aligned}\langle P\rangle + \langle Q\rangle - 2\langle\mathcal{O}\rangle - \mathrm{div}(l_1) &= \langle P\rangle + \langle Q\rangle - 2\langle\mathcal{O}\rangle - \langle P\rangle - \langle Q\rangle - \langle R\rangle + 3\langle\mathcal{O}\rangle \\ &= -\langle R\rangle + \langle\mathcal{O}\rangle\end{aligned}$$

4. *Jeżeli $P = \overline{Q}$, to $R = \mathcal{O}$ i wynik dodawania to warstwa zawierająca dywizor $\langle\mathcal{O}\rangle - \langle\mathcal{O}\rangle$.*

5. *W przeciwnym razie wykonujemy jeszcze jedną redukcję za pomocą linii l_2 przechodzącej przez punkty R oraz $\overline{R}$ (lub stycznej, gdy $R = \overline{R}$). Otrzymujemy:*

$$\begin{aligned}-\langle R\rangle + \langle\mathcal{O}\rangle + \mathrm{div}(l_2) &= -\langle R\rangle + \langle\mathcal{O}\rangle + \langle R\rangle + \left\langle\overline{R}\right\rangle - 2\langle\mathcal{O}\rangle \\ &= \left\langle\overline{R}\right\rangle - \langle\mathcal{O}\rangle\end{aligned}$$

Wynikiem jest warstwa zawierająca dywizor $\left\langle\overline{R}\right\rangle - \langle\mathcal{O}\rangle$.

2.3. Definicja

Zdefiniujemy teraz działanie grupowe na zbiorze punktów krzywej eliptycznej.

Definicja 2.3.1. Dana jest krzywa eliptyczna E nad ciałem $\mathbb{K}$. *Grupa na krzywej eliptycznej E* to grupa składająca się ze zbioru punktów krzywej E oraz działania, oznaczanego symbolem $+$, określonego w następujący sposób.

1. Elementem neutralnym działania jest punkt $\mathcal{O}$.

$$P + \mathcal{O} = P = \mathcal{O} + P \tag{2.3.2}$$

2. Elementem przeciwnym do punktu $P \neq \mathcal{O}$, oznaczanym symbolem $-P$, jest punkt $\overline{P}$.

$$P + \overline{P} = \mathcal{O} = \overline{P} + P \tag{2.3.3}$$

3. Suma dwóch punktów skończonych P i Q (przy czym $Q \neq \overline{P}$) postaci $P = (a, b)$ i $Q = (c, d)$, gdzie $a, b, c, d \in \mathbb{K}$, to punkt R postaci $R = (e, f)$, którego współrzędne są określone następująco:

$$e = \lambda^2 - a - c \tag{2.3.4}$$
$$f = -\lambda(e - a) - b \tag{2.3.5}$$

Współczynnik λ jest określony następująco:

(a) jeżeli $a \neq c$, to:

$$\lambda = \frac{d - b}{c - a} \tag{2.3.6}$$

(b) jeżeli $a = c$, to:

$$\lambda = \frac{3a^2 + A}{2b} \tag{2.3.7}$$

Uwaga 2.3.8. Nie wprowadzamy nowego oznaczenia na grupę na krzywej eliptycznej – będziemy ją oznaczać tak samo, jak oznaczamy krzywe.

Uwaga 2.3.9. W podanej definicji w punkcie 3. suma dwóch punktów P i Q nie jest określona, gdy $Q = \overline{P}$ lub gdy jeden z punktów jest nieskończony. Wynik dodawania jest wówczas określony na podstawie zależności podanych w punkcie 1. lub 2.

Uwaga 2.3.10. Przypadek opisany w punkcie 3b. zachodzi wtedy, gdy próbujemy dodać punkt skończony niebędący punktem rzędu dwa do samego siebie.

Uwaga 2.3.11. Różne przypadki występujące w definicji 2.3.1 obrazuje rysunek 2.3.12.

Fakt 2.3.13. *Jeżeli punkty P, Q i R leżą na jednej linii, to $P + Q + R = \mathcal{O}$.*

Fakt ten daje intuicję, na czym tak naprawdę polega działanie na punktach krzywej eliptycznej. Aby dodać punkty P i Q, przeprowadzamy przez nie linię, odnajdujemy trzeci punkt R, w którym przecina się ona z krzywą, po czym za wynik dodawania uznajemy punkt $\overline{R}$. Podobieństwo do działania w zerowej grupie Picarda nie jest przypadkowe.

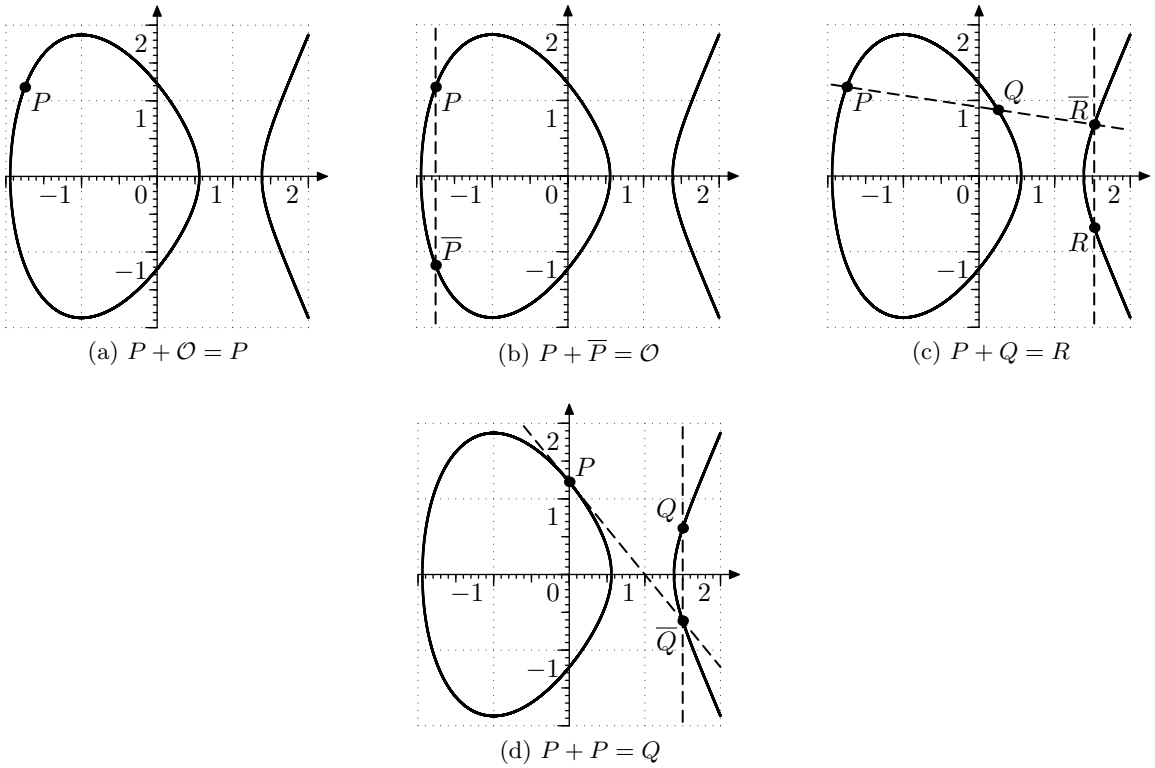
Twierdzenie 2.3.14. *Dana jest krzywa eliptyczna E. Niech π będzie bijekcją określoną wzorem 2.2.15. Wówczas dla dowolnych punktów P i Q na krzywej E zachodzą następujące zależności:*

- $\pi(P + Q) = \pi(P) + \pi(Q)$;
- $\pi^{-1}\left([\langle P\rangle - \langle\mathcal{O}\rangle] + [\langle Q\rangle - \langle\mathcal{O}\rangle]\right) = \pi^{-1}([\langle P\rangle - \langle\mathcal{O}\rangle]) + \pi^{-1}([\langle Q\rangle - \langle\mathcal{O}\rangle])$.

Dowód. Natychmiastowy na podstawie twierdzeń 2.2.10, 2.2.12 i 2.2.14, wniosku 2.2.16 oraz faktu 2.2.17. □

Wniosek 2.3.15. *Zbiór punktów na krzywej eliptycznej E wraz z operacją dodawania punktów tworzą grupę. Jest to grupa abelowa izomorficzna z grupą* $\mathrm{Pic}^0(E)$.

Udało się nam uniknąć sprawdzania praw grupowych bezpośrednio z podanej definicji (co byłoby dosyć żmudne), a przy tym zrozumieliśmy powód, dla którego działanie na krzywej eliptycznej jest zdefiniowane właśnie w taki sposób. Jest to o tyle cenne, że w literaturze fachowej, lecz niekoniecznie matematycznej, działanie grupowe na krzywej eliptycznej często definiuje się po prostu poprzez podanie odpowiednich wzorów, bez wskazania motywacji.



(a) $P + \mathcal{O} = P$ (b) $P + \overline{P} = \mathcal{O}$ (c) $P + Q = R$

(d) $P + P = Q$

Rysunek 2.3.12: Dodawanie punktów na krzywej eliptycznej $E_{-3,1\frac{1}{2}}(\mathbb{R})$

2.4. Własności

Przytoczymy teraz bez dowodu kilka najważniejszych własności grup na krzywych eliptycznych.
Przede wszystkim, ze względu na charakter wzorów występujących w definicji działania grupowego, ciało, nad którym krzywe rozpatrujemy, nie musi być algebraicznie domknięte.

Fakt 2.4.1. *Dana jest krzywa eliptyczna E nad ciałem $\mathbb{L}$ o parametrach z ciała $\mathbb{K} \subset \mathbb{L}$. Wówczas suma dwóch punktów $\mathbb{K}$-wymiernych na krzywej E jest punktem $\mathbb{K}$-wymiernym.*

Jest jasne, że w przypadku ciał algebraicznie domkniętych krzywa eliptyczna ma nieskończenie wiele elementów. W przypadku ciał skończonych sytuacja wygląda zupełnie inaczej.

Twierdzenie 2.4.2 (Hasse)**.** *Rząd grupy na krzywej eliptycznej E nad ciałem skończonym $\mathbb{GF}(q)$, spełnia następującą nierówność:*

$$q + 1 - 2\sqrt{q} \leqslant |E| \leqslant q + 1 + 2\sqrt{q} \tag{2.4.3}$$

Dowód tego twierdzenia można znaleźć np. w pracy [1], zaś intuicyjne wyjaśnienie jest następujące. Wielomian charakterystyczny κ rzadko przyjmuje tę samą wartość więcej niż raz, zatem możemy uznać, że jest on „prawie" permutacją zbioru $\mathbb{GF}(q)$. Stąd dla mniej-więcej połowy elementów a z ciała $\mathbb{GF}(q)$ można z wartości $\kappa(a)$ wyciągnąć pierwiastek kwadratowy. Zatem krzywa eliptyczna $E(\mathbb{GF}(q))$ ma około $2\frac{q}{2}$ punktów skończonych. Doliczając punkt w nieskończoności otrzymujemy około $q + 1$ punktów.
Choć nie ma jawnego wzoru na rząd grupy na krzywej eliptycznej nad ciałem skończonym, następujące twierdzenie, którego dowód można odnaleźć w książce [4], charakteryzuje ogólną strukturę takiej grupy.

Twierdzenie 2.4.4. *Grupa na krzywej eliptycznej E nad ciałem $\mathbb{GF}(q)$ jest izomorficzna z grupą $(\mathbb{Z}/m\mathbb{Z}) \times (\mathbb{Z}/n\mathbb{Z})$, przy czym $n \mid \gcd(m, q-1)$.*

Na koniec wspomnijmy o algorytmie, który pozwala dosyć efektywnie policzyć rząd grupy na krzywej nad ciałem skończonym. Został on opublikowany po raz pierwszy w pracy [11].

Twierdzenie 2.4.5 (Schoof)**.** *Istnieje algorytm, który pozwala obliczyć rząd grupy na krzywej eliptycznej E nad ciałem $\mathbb{GF}(q)$ za pomocą $O(\log^8 q)$ operacji na bitach.*

Rozdział 3

Iloczyn Weila

Iloczyn Weila to działanie dwuargumentowe określone na pewnej specyficznej podgrupie grupy na krzywej eliptycznej. Jego najważniejsze właściwości to dwuliniowość (a dokładnie odpowiednik tego pojęcia przeniesiony na grupy), charakterystyczna przeciwdziedzina – grupa pierwiastków z jedności – oraz niezdegenerowanie polegające na tym, że iloczyn Weila jest funkcją „na".
W tym rozdziale podamy definicję iloczynu Weila (poprzedzoną niewielką ilością informacji wstępnych) i przeanalizujemy jego podstawowe własności. Podamy również alternatywną definicję, która, jak zobaczymy w kolejnym rozdziale, stanowi lepszą podstawę do obliczeń.

3.1. Definicja

Mimo tego, że wprowadziliśmy już wiele elementów teorii krzywych eliptycznych, wciąż jeszcze brakuje nam pewnych wiadomości, aby móc zdefiniować iloczyn Weila. Ze względu na ograniczoną objętość pracy przedstawimy je teraz w bardzo okrojonym ujęciu.

Podgrupy n-torsyjne

Zaczniemy od zdefiniowania podgrup grupy na krzywej eliptycznej, które stanowią dziedzinę iloczynu Weila.

Definicja 3.1.1. Dana jest krzywa eliptyczna E oraz liczba całkowita n. *Mnożenie punktu na krzywej E przez liczbę całkowitą n* to funkcja $[n]\colon E \to E$ określona następująco:

$$[n](P) = \begin{cases} \underbrace{P + P + \cdots + P}_{n \text{ razy}} & \text{gdy } n > 0, \\ \mathcal{O} & \text{gdy } n = 0, \\ -[-n](P) & \text{gdy } n < 0. \end{cases} \tag{3.1.2}$$

Wyrażenie $[n](P)$ zapisujemy skrótowo jako nP.

Użycie określenia „mnożenie" oraz skrótowego zapisu nP jest zmotywowane przez własności funkcji $[n]$, które przypominają cechy operacji mnożenia liczb całkowitych.

Fakt 3.1.3. *Dla dowolnych punktów P i Q na krzywej eliptycznej E oraz liczb całkowitych*

m i n zachodzą następujące zależności:

$$\begin{aligned} m(P+Q) &= mP+mQ \\ (m+n)P &= mP+nP \\ m(-P) &= -(mP) \\ (-m)P &= -(mP) \end{aligned}$$

Jesteśmy zainteresowani takimi punktami P krzywej, które spełniają równanie $nP = \mathcal{O}$. Jak nietrudno stwierdzić na podstawie podanych własności mnożenia, tworzą one grupę.

Definicja 3.1.4. Dana jest krzywa eliptyczna E oraz liczba naturalna n. *Podgrupa n-torsyjna na krzywej E*, oznaczana symbolem $E[n]$, to podgrupa grupy na krzywej E złożona ze wszystkich punktów P na krzywej E, które spełniają zależność:

$$nP = \mathcal{O} \tag{3.1.5}$$

Elementy grupy $E[n]$ nazywamy *punktami rzędu n na krzywej E.*

Przykład 3.1.6. Rozważmy krzywą E_1 nad ciałem $\mathbb{F}(17)$ o parametrach 16 i 0. Wszystkie punkty na tej krzywej są rzędu cztery, zatem $E_1 = E_1[4]$. Można sprawdzić, że grupa na krzywej E_1 jest generowana przez punkty $(5,1)$ i $(13,5)$ oraz że ma strukturę $(\mathbb{Z}/4\mathbb{Z})\times(\mathbb{Z}/4\mathbb{Z})$. Ponadto, podgrupa 2-torsyjna na krzywej E_1 ma strukturę $(\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$ i składa się z czterech punktów: $\mathcal{O}$, $(0,0)$, $(1,0)$ i $(16,0)$.
Weźmy teraz krzywą E_2 nad ciałem $\mathbb{F}(19)$ o parametrach 14 i 12. Jej podgrupa 3-torsyjna ma strukturę $(\mathbb{Z}/3\mathbb{Z}) \times (\mathbb{Z}/3\mathbb{Z})$ i jest generowana przez punkty $(3,10)$ i $(13,15)$.

Podajemy bez dowodu kilka podstawowych własności podgrup n-torsyjnych.

Fakt 3.1.7. *Dana jest krzywa eliptyczna E oraz liczby naturalne m i n. Jeżeli $m \mid n$, to $E[m] \subset E[n]$.*

Twierdzenie 3.1.8. *Dana jest krzywa eliptyczna E nad ciałem algebraicznie domkniętym $\mathbb{K}$ oraz liczba naturalna n. Jeżeli $\mathrm{char}(\mathbb{K}) = 0$ lub $\gcd(n, \mathrm{char}(\mathbb{K})) = 1$, to podgrupa n-torsyjna $E[n]$ ma rząd równy n^2 i jest izomorficzna z grupą $(\mathbb{Z}/n\mathbb{Z}) \times (\mathbb{Z}/n\mathbb{Z})$.*

Wniosek 3.1.9. *Dana jest krzywa eliptyczna E nad ciałem algebraicznie domkniętym. Podgrupa $E[1]$ jest trywialna – składa się tylko z punktu $\mathcal{O}$. Podgrupa $E[2]$ składa się z czterech punktów: punktu $\mathcal{O}$ oraz trzech punktów rzędu dwa.*

Uwaga 3.1.10. Zakładamy od tej pory, że rozważane podgrupy n-torsyjne to zawsze mają strukturę $(\mathbb{Z}/n\mathbb{Z}) \times (\mathbb{Z}/n\mathbb{Z})$.

Pewne szczególne dywizory

Elementy podgrupy n-torsyjnej to rozwiązania równania $nQ = \mathcal{O}$. Rozważymy teraz bardziej ogólne równanie $nQ = P$. Doprowadzi nas to do definicji pewnego rodzaju dywizorów, które pojawiają się w definicji iloczynu Weila.

Twierdzenie 3.1.11. *Dany jest punkt P rzędu n na krzywej eliptycznej E. Wówczas:*

- *istnieje taki punkt Q_0 rzędu n^2 na krzywej E, że $nQ_0 = P$;*
- *każdy punkt Q spełniający równanie $nQ = P$ można przedstawić w postaci $Q = Q_0 + R$, gdzie $R \in E[n]$.*

Dowód. Zgodnie z twierdzeniem 3.1.8 grupy $E[n]$ i $E[n^2]$ są izomorficzne odpowiednio z grupami $(\mathbb{Z}/n\mathbb{Z}) \times (\mathbb{Z}/n\mathbb{Z})$ i $(\mathbb{Z}/n^2\mathbb{Z}) \times (\mathbb{Z}/n^2\mathbb{Z})$. Wystarczy teraz przeanalizować sposób, w jaki grupa $(\mathbb{Z}/n\mathbb{Z}) \times (\mathbb{Z}/n\mathbb{Z})$ jest zanurzona w grupie $(\mathbb{Z}/n^2\mathbb{Z}) \times (\mathbb{Z}/n^2\mathbb{Z})$. □

Wniosek 3.1.12. *Jest n^2 punktów Q na krzywej eliptycznej E spełniających równanie $P = nQ$, gdzie $P \in E[n]$. Zbiór tych punktów można przedstawić w postaci $\{Q_0 + R \mid R \in E[n]\}$, gdzie Q_0 jest dowolnym takim punktem.*

Rozważania te motywują definicję „dzielenia" punktu. Takie dzielenie jest podobne do pierwiastkowania liczb zespolonych – jest niejednoznaczne, ale zbiór wszystkich możliwych wyników dzielenia przejawia pewną strukturę. Zbiór ten będziemy reprezentować za pomocą dywizora.

Definicja 3.1.13. Dana jest krzywa eliptyczna E oraz liczba całkowita n. *Dzielenie punktu na krzywej E przez liczbę całkowitą n* to funkcja $[n]^{-1}\colon E[n] \to \mathrm{Div}(E)$ określona następująco:

$$[n]^{-1}(P) = \sum_{nQ=P} \langle Q \rangle \tag{3.1.14}$$

Przykład 3.1.15. Niech $(0,0)$ będzie punktem na krzywej eliptycznej $E_{16,0}(\mathbb{F}(17))$. Wówczas:

$$[2]^{-1}((0,0)) = \langle (4,3) \rangle + \langle (4,14) \rangle + \langle (13,5) \rangle + \langle (13,12) \rangle$$

Zauważmy też, że zbiór $\{(4,3),(4,14),(13,5),(13,12)\}$ można przedstawić w postaci $\{(4,3) + R \mid R \in E_{16,0}(\mathbb{F}(17))[2]\}$.

Następujące twierdzenie przedstawia własność tego rodzaju dywizorów, która będzie nas najbardziej interesować.

Twierdzenie 3.1.16. *Dany jest punkt P rzędu n na krzywej eliptycznej E. Wówczas zachodzi następująca zależność:*

$$[n]^{-1}(P) \sim n^2 \langle \mathcal{O} \rangle$$

Dowód tego twierdzenia poprzedzimy bardzo użytecznym lematem.

Lemat 3.1.17. *Dane są punkty P i Q na krzywej eliptycznej E. Wówczas dywizor $\langle P+Q \rangle - \langle P \rangle - \langle Q \rangle + \langle \mathcal{O} \rangle$ jest główny.*

Dowód. Lemat zachodzi, jeżeli $P = Q$, $P = -Q$, $P = \mathcal{O}$ lub $Q = \mathcal{O}$. W przeciwnym przypadku rozważmy funkcję $\frac{r}{s}$, gdzie r to linia przechodząca przez punkty $(P+Q)$ i $-(P+Q)$, a s to linia przechodząca przez punkty P, Q i $-(P+Q)$. Jak nietrudno sprawdzić, $\mathrm{div}(\frac{r}{s}) = \langle P+Q \rangle - \langle P \rangle - \langle Q \rangle + \langle \mathcal{O} \rangle$. □

Wniosek 3.1.18. *Dane są punkty P i Q na krzywej eliptycznej E. Wówczas $\langle P \rangle + \langle Q \rangle \sim \langle P+Q \rangle + \langle \mathcal{O} \rangle$.*

Zauważmy, że z pomocą tego lematu i płynącego z niego wniosku możemy sformułować metodę redukcji dywizorów podobną do twierdzenia 2.2.10. Ograniczymy się jednak do zastosowania tej metody do przypadku dywizorów postaci $[n]^{-1}(P)$.

Dowód twierdzenia 3.1.16. Niech α i β oznaczają generatory grupy $E[n]$. Wybierzmy dowolny punkt $Q \in E[n^2]$ taki, że $P = nQ$ Zgodnie z wnioskiem 3.1.12 dywizor $[n]^{-1}(P)$ możemy zapisać w postaci:

$$[n]^{-1}(P) = \sum_{k=0}^{n-1} \sum_{l=0}^{n-1} \langle Q + k\alpha + l\beta \rangle$$

Teraz wielokrotnie zastosujemy wniosek 3.1.18, aby zamienić sumę dywizorów na dywizor sumy. Najpierw $n-1$ razy redukujemy wewnętrzną sumę i otrzymujemy:

$$\begin{aligned}
\sum_{l=0}^{n} \langle Q + k\alpha + l\beta \rangle &= \langle Q + k\alpha \rangle + \langle Q + k\alpha + \beta \rangle + \sum_{l=2}^{n-1} \langle Q + k\alpha + l\beta \rangle \\
&\sim \langle \mathcal{O} \rangle + \langle 2Q + 2k\alpha + \beta \rangle + \langle Q + k\alpha + 2\beta \rangle + \sum_{l=3}^{n-1} \langle Q + k\alpha + l\beta \rangle \\
&\sim 2\langle \mathcal{O} \rangle + \langle 3Q + 3k\alpha + 3\beta \rangle + \langle Q + k\alpha + 3\beta \rangle + \sum_{l=4}^{n-1} \langle Q + k\alpha + l\beta \rangle \\
&\sim 3\langle \mathcal{O} \rangle + \langle 4Q + 4k\alpha + 6\beta \rangle + \langle Q + k\alpha + 4\beta \rangle + \sum_{l=5}^{n-1} \langle Q + k\alpha + l\beta \rangle \\
&\sim \dots \\
&\sim (n-1)\langle \mathcal{O} \rangle + \left\langle nQ + nk\alpha + \binom{n}{2}\beta \right\rangle
\end{aligned}$$

Wynik ten podstawiamy do zewnętrznej sumy, po czym w analogiczny sposób redukujemy ją $n-1$ razy. Otrzymujemy:

$$\begin{aligned}
\sum_{k=0}^{n} (n-1)\langle \mathcal{O} \rangle + \left\langle nQ + nk\alpha + \binom{n}{2}\beta \right\rangle &\sim (n^2-1)\langle \mathcal{O} \rangle + \left\langle n^2 Q + n\binom{n}{2}\alpha + n\binom{n}{2}\beta \right\rangle \\
&\sim (n^2-1)\langle \mathcal{O} \rangle + \left\langle n^2 Q \right\rangle \\
&\sim n^2 \langle \mathcal{O} \rangle
\end{aligned}$$

□

Wniosek 3.1.19. *Dywizor $[n]^{-1}(P) - [n]^{-1}(\mathcal{O})$ jest główny.*

Iloczyn Weila

Dysponujemy już wszystkimi niezbędnymi informacjami, aby móc podać definicję iloczynu Weila.

Definicja 3.1.20. Dana jest krzywa eliptyczna E nad ciałem $\mathbb{K}$ oraz liczba całkowita n. Niech P będzie dowolnym punktem na krzywej E. Niech f_P będzie funkcją wymierną na krzywej E określoną z dokładnością do niezerowego czynnika stałego poprzez podanie jej dywizora:

$$\operatorname{div}(f_P) = [n]^{-1}(P) - [n]^{-1}(\mathcal{O}) \tag{3.1.21}$$

Niech funkcja $t_Q \colon E \to E$ będzie określona następująco:

$$t_Q(R) = R + Q \tag{3.1.22}$$

Iloczyn Weila to funkcja $w\colon E[n] \times E[n] \to \mathbb{K}$ określona następująco:

$$w(P,Q) = \left(\frac{f_P \circ t_Q}{f_P}\right)(\mathcal{O}) \tag{3.1.23}$$

Uwaga 3.1.24. Zauważmy, że podana definicja określa nie jedną funkcję, lecz całą rodzinę funkcji – po jednej dla każdej liczby całkowitej n. Dalsze rozważania ograniczamy do tych liczb n, które są dodatnie i dla których grupa $E[n]$ ma strukturę $(\mathbb{Z}/n\mathbb{Z}) \times (\mathbb{Z}/n\mathbb{Z})$.

Uwaga 3.1.25. Przyjmujemy od tej pory, że zawsze dana jest pewna liczba całkowita n oraz odpowiadający jej iloczyn Weila $w\colon E[n] \times E[n] \to \mathbb{K}$.

Sprawdźmy teraz, że iloczyn Weila jest dobrze określony.

Lemat 3.1.26. *Wartość iloczynu Weila nie zależy od wyboru funkcji f_P.*

Dowód. Na mocy wniosku 1.3.13 dwie funkcje wymierne f_P i $\tilde{f_P}$ o takim samym dywizorze różnią się o stały niezerowy czynnik, zatem przyjmijmy, że $f_P = c\tilde{f_P}$. Podstawiamy tę zależność do wzoru 3.1.23 i otrzymujemy:

$$\begin{aligned}
\frac{f_P \circ t_Q}{f_P}(\mathcal{O}) &= \frac{(c\tilde{f_P}) \circ t_Q}{c\tilde{f_P}}(\mathcal{O}) \\
&= \frac{c(\tilde{f_P} \circ t_Q)}{c\tilde{f_P}}(\mathcal{O}) \\
&= \frac{\tilde{f_P} \circ t_Q}{\tilde{f_P}}(\mathcal{O})
\end{aligned}$$

Widzimy stąd, że wartość iloczynu Weila pozostanie taka sama niezależnie od tego, jakiej funkcji użyjemy, ponieważ wszystkie one różnią się o czynnik stały, który pojawia się zarówno w liczniku, jak i w mianowniku. □

Lemat 3.1.27. *Funkcje f_P i $f_P \circ t_Q$ mają taki sam dywizor.*

Dowód. Niech $R \in E[n^2]$ będzie dowolnym punktem takim, że $nR = P$. Zauważmy, że $f_P(S) = 0$ wtedy i tylko wtedy, gdy $(f_P \circ t_Q)(S - Q) = 0$. Podobnie, $f_P(S) = \infty$ wtedy i tylko wtedy, gdy $(f_P \circ t_Q)(S - Q) = \infty$. Wszystkie miejsca zerowe i bieguny funkcji f_P są jednokrotne. Dywizor funkcji $f_P \circ t_Q$ możemy w takim razie zapisać w następującej postaci:

$$\operatorname{div}(f_P \circ t_Q) = \sum_{S \in E[n]} \langle R + S - Q \rangle - \sum_{S \in E[n]} \langle S - Q \rangle$$

Wykonujemy „zamianę zmiennych” – w miejsce zmiennej S podstawiamy zmienną $S' = S - Q$. Zauważmy, że zmienna S' również będzie przebiegać po zbiorze $E[n]$. Otrzymujemy:

$$\begin{aligned}
\operatorname{div}(f_P \circ t_Q) &= \sum_{S' \in E[n]} \langle R + S' \rangle - \sum_{S' \in E[n]} \langle S' \rangle \\
&= \operatorname{div}(f_P)
\end{aligned}$$

□

Wniosek 3.1.28. *Funkcja $\frac{f_P \circ t_Q}{f_P}$ jest stała i niezerowa.*

Dowód. Stosujemy twierdzenie 2.2.3 i udowodniony przed chwilą lemat do obliczenia dywizora funkcji $\frac{f_P \circ t_Q}{f_P}$ i otrzymujemy:

$$\operatorname{div}\left(\frac{f_P \circ t_Q}{f_P}\right) = \operatorname{div}(f_P \circ t_Q) - \operatorname{div}(f_P) = 0$$

Funkcja mająca zerowy dywizor jest stała i niezerowa, zgodnie z wnioskiem 1.3.12. □

Uwaga 3.1.29. Wybór punktu $\mathcal{O}$ jako argumentu funkcji $\frac{f_P \circ t_Q}{f_P}$ we wzorze 3.1.23 był dowolny, bo funkcja ta i tak jest stała. Dlatego też czasem będziemy nadużywać notacji i zapisywać iloczyn Weila jako:

$$w(P, Q) = \frac{f_P \circ t_Q}{f_P}$$

Uwaga 3.1.30. Iloczyn Weila nie jest funkcją stałą – współczynnik c w tożsamości $f_P \circ t_Q = cf_P$ może zmieniać się w zależności od punktu Q mimo tego, że funkcja $f_P \circ t_Q$ zawsze ma taki sam dywizor jak funkcja f_P.

Twierdzenie 3.1.31. *Iloczyn Weila $w(P, Q)$ jest dobrze określony, tzn. jego wartość zależy tylko od punktów P i Q.*

Dowód. Natychmiastowy na podstawie lematu 3.1.26 i wniosku 3.1.28. □

Przykład 3.1.32. Policzymy wartość iloczynu Weila dla punktów $P = (1, 0)$ i $Q = (16, 0)$ rzędu dwa na krzywej eliptycznej $E_{16,0}(\mathbb{F}(17))$.
Zacznijmy od znalezienia funkcji f_P. W tym celu obliczmy dywizory $[2]^{-1}(P)$ i $[2]^{-1}(\mathcal{O})$:

$$\begin{aligned} [2]^{-1}(P) &= \langle(7, 8)\rangle + \langle(7, 9)\rangle + \langle(12, 4)\rangle + \langle(12, 13)\rangle \\ [2]^{-1}(\mathcal{O}) &= \langle\mathcal{O}\rangle + \langle(0, 0)\rangle + \langle(1, 0)\rangle + \langle(16, 0)\rangle \end{aligned}$$

Zauważmy teraz, że wielomian $(x-7)(x-12)$ ma dywizor równy $[2]^{-1}(P) - 4\langle\mathcal{O}\rangle$, a wielomian y ma dywizor równy $[2]^{-1}(\mathcal{O}) - 4\langle\mathcal{O}\rangle$. Iloraz tych dwóch wielomianów ma szukany dywizor, przyjmijmy zatem, że funkcja f_P jest równa $\frac{(x-7)(x-12)}{y}$.
Teraz wystarczy wybrać taki punkt R, żeby wartości $f_P(R + Q)$ i $f_P(R)$ były skończone i niezerowe. Weźmy $R = (10, 2)$. Wtedy $R + Q = (10, 15)$, zatem $f_P(R + Q) = -4$. Z kolei $f_P(R) = 4$. Otrzymujemy $w(P, Q) = 16$.

3.2. Własności

Udowodnimy teraz szereg podstawowych własności iloczynu Weila.

Pierwiastki z jedności

Zaczniemy od pokazania, że wartości, jakie przyjmuje iloczyn Weila, to pierwiastki z jedynki.

Twierdzenie 3.2.1. *Dane są punkty P i Q rzędu n na krzywej eliptycznej E. Wówczas $w(P, Q)^n = 1$.*

Dowód. Rozważmy następujące wyrażenie:

$$f_P \circ \underbrace{t_Q \circ \cdots \circ t_Q}_{n \text{ razy}}$$

Składanie funkcji jest łączne, zatem wyrażenie to można ponawiasować na wiele sposobów. Rozważmy dwa szczególne nawiasowania.

1. Nawiasujemy od lewej do prawej:

$$\begin{aligned} f_P \circ \underbrace{t_Q \circ \cdots \circ t_Q}_{n \text{ razy}} &= (\underbrace{(\cdots((f_P \circ t_Q) \circ t_Q}_{n-1 \text{ razy}}) \circ \cdots) \circ t_Q) \\ &= (\underbrace{(\cdots(((w(P,Q)f_P) \circ t_Q) \circ t_Q}_{n-2 \text{ razy}}) \circ \cdots) \circ t_Q) \\ &= w(P,Q)(\underbrace{(\cdots((f_P \circ t_Q) \circ t_Q}_{n-2 \text{ razy}}) \circ \cdots) \circ t_Q) \\ &= \ldots \\ &= w(P,Q)^n f_P \end{aligned}$$

2. Nawiasujemy od prawej do lewej:

$$\begin{aligned} f_P \circ \underbrace{t_Q \circ \cdots \circ t_Q}_{n \text{ razy}} &= (f_P \circ \underbrace{(t_Q \circ \cdots \circ (t_Q}_{n-1 \text{ razy}} \circ t_Q) \cdots)) \\ &= (f_P \circ \underbrace{(t_Q \circ \cdots \circ (t_Q}_{n-2 \text{ razy}} \circ t_{2Q}) \cdots)) \\ &= \ldots \\ &= f_P \circ t_{nQ} \\ &= f_P \end{aligned}$$

Z porównania wyników obu nawiasowań widzimy, że $w(P,Q)^n = 1$. □

Odwzorowanie dwuliniowe

Wykażemy teraz własność iloczynu Weila, która jest jednym z ważnych powodów jego użyteczności. Własnością tą jest dwuliniowość.

Definicja 3.2.2. Dana jest grupa addytywna $\mathbb{G}_1$ oraz grupa multiplikatywna $\mathbb{G}_2$. Odwzorowanie $b\colon \mathbb{G}_1 \times \mathbb{G}_1 \to \mathbb{G}_2$ jest *dwuliniowe*, jeżeli dla dowolnych elementów P, P_1, P_2, Q, Q_1 i Q_2 z grupy $\mathbb{G}_1$ zachodzą następujące zależności:

- $b(P_1 + P_2, Q) = b(P_1, Q)b(P_2, Q)$;
- $b(P, Q_1 + Q_2) = b(P, Q_1)b(P, Q_2)$.

Sprawdźmy, że iloczyn Weila spełnia obie wymagane tożsamości.

Lemat 3.2.3. *Dane są punkty P_1, P_2 i Q rzędu n na krzywej eliptycznej E. Wówczas zachodzi następująca zależność:*

$$w(P_1 + P_2, Q) = w(P_1, Q)w(P_2, Q) \tag{3.2.4}$$

Dowód. Policzmy dywizor funkcji $\frac{f_{P_1+P_2}}{f_{P_1} f_{P_2}}$. Otrzymujemy:

$$\operatorname{div}\left(\frac{f_{P_1+P_2}}{f_{P_1} f_{P_2}}\right) = [n]^{-1}(P_1 + P_2) - [n]^{-1}(P_1) - [n]^{-1}(P_2) + [n]^{-1}(\mathcal{O})$$

Niech r będzie dowolną funkcją wymierną na krzywej E, której dywizor jest równy $\langle P_1 + P_2\rangle - \langle P_1\rangle - \langle P_2\rangle + \langle\mathcal{O}\rangle$. Rozważmy funkcję $r \circ [n]$. Funkcja ta ma miejsce zerowe w każdym takim punkcie R, że $nR = P_1 + P_2$ lub $nR = \mathcal{O}$ oraz biegun w każdym takim punkcie R, że $nR = P_1$

lub $nR = P_2$. Po sprawdzeniu krotności miejsc zerowych i biegunów oraz uwzględnieniu faktu, mogą się one nałożyć i „wzmocnić" lub „znieść", otrzymujemy:

$$\mathrm{div}(r \circ [n]) = [n]^{-1}(P_1 + P_2) - [n]^{-1}(P_1) - [n]^{-1}(P_2) + [n]^{-1}(\mathcal{O})$$

Zatem $\frac{f_{P_1+P_2}}{f_{P_1} f_{P_2}} = c(r \circ [n])$. Zauważmy teraz, że funkcja $r \circ [n]$ jest niezmiennicza ze względu na przesunięcia o element $R \in E[n]$:

$$\begin{aligned}((r \circ [n]) \circ t_R)(S) &= r([n](t_R(S))) \\ &= r([n](S + R)) \\ &= r(nS + nR) \\ &= r(nS) \\ &= (r \circ [n])(S)\end{aligned}$$

Zatem funkcja $\frac{f_{P_1+P_2}}{f_{P_1} f_{P_2}}$ również jest niezmiennicza ze względu na takie przesunięcia. Wykorzystamy ten fakt, aby przekształcić wzór na wartość $w(P_1 + P_2, Q)$:

$$\begin{aligned}w(P_1 + P_2, Q) &= \frac{f_{P_1+P_2} \circ t_Q}{f_{P_1+P_2}} \\ &= \frac{f_{P_1+P_2} \circ t_Q}{f_{P_1+P_2}} \frac{f_{P_1} f_{P_2}}{f_{P_1} f_{P_2}} \\ &= \frac{f_{P_1+P_2} \circ t_Q}{f_{P_1} f_{P_2}} \frac{f_{P_1} f_{P_2}}{f_{P_1+P_2}} \\ &= \frac{f_{P_1+P_2} \circ t_Q}{f_{P_1} f_{P_2}} \left(\left(\frac{f_{P_1} f_{P_2}}{f_{P_1+P_2}} \right) \circ t_Q \right) \\ &= \frac{f_{P_1+P_2} \circ t_Q}{f_{P_1} f_{P_2}} \frac{(f_{P_1} \circ t_Q)(f_{P_2} \circ t_Q)}{f_{P_1+P_2} \circ t_Q} \\ &= \frac{(f_{P_1} \circ t_Q)(f_{P_2} \circ t_Q)}{f_{P_1} f_{P_2}} \\ &= w(P_1, Q) w(P_2, Q)\end{aligned}$$

□

Lemat 3.2.5. *Dane są punkty P, Q_1 i Q_2 rzędu n na krzywej eliptycznej E. Wówczas zachodzi następująca zależność:*

$$w(P, Q_1 + Q_2) = w(P, Q_1) w(P, Q_2) \tag{3.2.6}$$

Dowód. Tym razem wystarczy zamienić $t_{Q_1+Q_2}$ na $t_{Q_1} \circ t_{Q_2}$ we wzorze na $w(P, Q_1 + Q_2)$:

$$\begin{aligned}w(P, Q_1 + Q_2) &= \frac{f_P \circ t_{Q_1+Q_2}}{f_P} \\ &= \frac{f_P \circ (t_{Q_1} \circ t_{Q_2})}{f_P} \\ &= \frac{(f_P \circ t_{Q_1}) \circ t_{Q_2}}{f_P} \\ &= \frac{((w(P, Q_1) f_P) \circ t_{Q_2})}{f_P} \\ &= w(P, Q_1) \frac{f_P \circ t_{Q_2}}{f_P} \\ &= w(P, Q_1) w(P, Q_2)\end{aligned}$$

□

Wniosek 3.2.7. *Iloczyn Weila jest odwzorowaniem dwuliniowym.*

Wartości dla szczególnych argumentów

Obliczymy teraz wartość iloczynu Weila dla kilku specyficznych kombinacji punktów P i Q.

Twierdzenie 3.2.8. *Dany jest punkt P rzędu n na krzywej eliptycznej E. Wówczas zachodzi zależność $w(P,P) = 1$.*

Dowód. Niech R będzie dowolnym punktem takim, że $nR = P$. Przypomnijmy, że dywizor funkcji f_P można przedstawić w następującej postaci:

$$\operatorname{div}(f_P) = \sum_{S \in E[n]} \langle R + S \rangle - \sum_{S \in E[n]} \langle S \rangle$$

Jeśli zbadamy miejsca zerowe i bieguny funkcji $f_P \circ t_{kR}$, gdzie $k = 0, 1, \ldots, n-1$, to otrzymamy, że jej dywizor jest równy:

$$\begin{aligned} \operatorname{div}(f_P \circ t_{kR}) &= \sum_{S \in E[n]} \langle (1-k)R + S \rangle - \sum_{S \in E[n]} \langle -kR + S \rangle \\ &= [n]^{-1}((1-k)P) - [n]^{-1}((-k)P) \end{aligned}$$

Określmy funkcję $\widehat{f_P}$ następująco:

$$\widehat{f_P} = \prod_{k=0}^{n-1} f_P \circ t_{kR}$$

Jej dywizor jest równy:

$$\operatorname{div}\left(\widehat{f_P}\right) = \sum_{k=0}^{n-1} [n]^{-1}((1-k)P) - [n]^{-1}((-k)P)$$

Jest to suma, która składa się teleskopowo, skąd otrzymujemy:

$$\begin{aligned} \operatorname{div}\left(\widehat{f_P}\right) &= [n]^{-1}(P) - [n]^{-1}((1-n)P) \\ &= [n]^{-1}(P) - [n]^{-1}(P) \\ &= 0 \end{aligned}$$

Funkcja $\widehat{f_P}$ jest zatem stała, w szczególności $\widehat{f_P} = \widehat{f_P} \circ t_R$. Po rozpisaniu tej równości dostajemy:

$$\begin{aligned} \widehat{f_P} &= \widehat{f_P} \circ t_R \\ \prod_{k=0}^{n-1} f_P \circ t_{kR} &= \left(\prod_{k=0}^{n-1} f_P \circ t_{kR}\right) \circ t_R \\ \prod_{k=0}^{n-1} f_P \circ t_{kR} &= \prod_{k=1}^{n} f_P \circ t_{kR} \\ f_P &= f_P \circ t_{nR} \\ f_P &= f_P \circ t_P \end{aligned}$$

Zatem $w(P,P) = \frac{f_P \circ t_P}{f_P} = \frac{f_P}{f_P} = 1$. □

Wynik ten uogólniamy na przypadek punktów „współliniowych". Pojęcie współliniowości bierze się stąd, że grupa $E[n]$ jest izomorficzna z grupą $(\mathbb{Z}/n\mathbb{Z}) \times (\mathbb{Z}/n\mathbb{Z})$, która jest wolnym modułem wymiaru dwa nad pierścieniem $\mathbb{Z}/n\mathbb{Z}$.

Twierdzenie 3.2.9. *Dane są punkty P i Q rzędu n na krzywej eliptycznej E. Jeżeli istnieje punkt R rzędu n na krzywej E oraz liczby całkowite k i l takie, że $P = kR$ i $Q = lR$, to zachodzi zależność $w(P, Q) = 1$.*

Dowód. Wykorzystamy dwuliniowość oraz to, że $w(R, R) = 1$:

$$\begin{aligned} w(P,Q) &= w(kR, lR) \\ &= w(R, lR)^k \\ &= w(R,R)^{kl} \\ &= 1 \end{aligned}$$

□

Następujące twierdzenie jest w pewnym sensie odwrotne do poprzedniego wyniku.

Twierdzenie 3.2.10. *Dany jest punkt P rzędu n na krzywej eliptycznej E. Jeżeli $w(P, Q) = 1$ dla każdego punktu Q rzędu n na krzywej E, to zachodzi zależność $P = \mathcal{O}$.*

W dowodzie tego twierdzenia kluczową rolę odegra następujący lemat.

Lemat 3.2.11. *Dana jest funkcja wymierna r na krzywej eliptycznej E. Jeżeli funkcja r jest niezmiennicza ze względu na przesunięcie o dowolny punkt Q rzędu n na krzywej E, tzn. $r = r \circ t_Q$, to funkcję r można przedstawić w postaci $r = s \circ [n]$, gdzie s jest pewną funkcją wymierną na krzywej E.*

Nie podamy dowodu tego lematu, gdyż wykracza on poza zakres niniejszej pracy.

Dowód twierdzenia 3.2.10. Z założenia mamy $f_P \circ t_Q = f_P$, więc na mocy lematu $f_P = s \circ [n]$. Porównując dywizor funkcji f_P z kształtem funkcji $s \circ [n]$ otrzymujemy, że $\mathrm{div}(s) = \langle P \rangle - \langle \mathcal{O} \rangle$. Korzystamy z lematu 2.2.11 i dostajemy $P = \mathcal{O}$. □

Niezdegenerowanie

Mogłoby się zdarzyć, że skomplikowana definicja doprowadziła nas do funkcji trywialnej. Pokażemy teraz, że tak nie jest.

Twierdzenie 3.2.12. *Dane są punkty P i Q rzędu n na krzywej eliptycznej E. Jeżeli punkty P i Q są generatorami grupy $E[n]$, to wartość $w(P, Q)$ jest pierwiastkiem pierwotnym n-tego stopnia z jedności.*

Dowód. Skoro punkty P i Q są generatorami grupy $E[n]$, to $\mathrm{ord}(P) = n$ oraz $\mathrm{ord}(Q) = n$. Dalej, każdy punkt $R \in E[n]$ można przedstawić w postaci $R = kP + lQ$. Niech $\mathrm{ord}(w(P, Q)) = m$, tzn. m jest najmniejszą liczbą całkowitą dodatnią taką, że $w(P, Q)^m = 1$. Obliczmy wartość $w(mP, R)$:

$$\begin{aligned} w(mP, R) &= w(mP, kP + lQ) \\ &= w(P,P)^{mk} w(P,Q)^{ml} \\ &= 1 \end{aligned}$$

Z twierdzenia 3.2.10 otrzymujemy, że $mP = \mathcal{O}$. Stąd $n \mid m$, bo n jest rzędem punktu P w grupie $E[n]$. Z drugiej strony, zawsze mamy $w(P, Q)^n = 1$, zatem $m \mid n$. Ostatecznie otrzymujemy $n = m$. □

Wniosek 3.2.13. *Wartością iloczynu Weila może być dowolny pierwiastek n-tego stopnia z jedności.*

Antysymetria

Z dwuliniowości łatwo pokazać, że iloczyn Weila jest antysymetryczny.

Twierdzenie 3.2.14. *Dane są punkty P i Q rzędu n na krzywej eliptycznej E. Wówczas zachodzi zależność $w(P,Q) = w(Q,P)^{-1}$.*

Dowód. Korzystając z dwuliniowości rozpiszmy wartość $w(P+Q, P+Q)$:

$$\begin{aligned} 1 &= w(P+Q, P+Q) \\ &= w(P,P)w(P,Q)w(Q,P)w(Q,Q) \\ &= w(P,Q)w(Q,P) \end{aligned}$$

□

3.3. Definicja alternatywna

Na podstawie podanej definicji iloczynu Weila udało nam się udowodnić szereg jego własności. Definicja ta nie nadaje się jednak do tego, aby skonstruować na jej podstawie wydajny algorytm obliczający wartości iloczynu Weila. Dlatego podamy teraz alternatywną definicję, która lepiej nada się do celów obliczeniowych.

Definicja

Alternatywna definicja iloczynu Weila jest następująca.

Definicja 3.3.1. Dana jest krzywa eliptyczna E nad ciałem $\mathbb{K}$ oraz liczba całkowita n. Niech P i Q będą dowolnymi punktami na krzywej E. Niech f'_P i f'_Q będą funkcjami wymiernymi na krzywej E określonymi z dokładnością do niezerowego czynnika stałego poprzez podanie ich dywizorów:

$$\begin{aligned} \operatorname{div}(f'_P) &= n\langle P\rangle - n\langle \mathcal{O}\rangle \\ \operatorname{div}(f'_Q) &= n\langle Q\rangle - n\langle \mathcal{O}\rangle \end{aligned}$$

Alternatywna postać iloczynu Weila to funkcja $w' \colon E[n] \to \mathbb{K}$ określona następująco:

$$w'(P,Q) = (-1)^n \frac{f'_P(Q)}{f'_Q(P)} \frac{f'_Q}{f'_P}(\mathcal{O}) \tag{3.3.2}$$

Uwaga 3.3.3. Funkcje f'_P i f'_Q są wyznaczone z dokładnością do stałej, ale nie wpływa to na wartość funkcji $w'(P,Q)$, zatem jest ona dobrze określona.

Symbol lokalny i prawo wzajemności Weila

Aby móc udowodnić równoważność obu definicji, wprowadzimy tzw. symbol lokalny i posłużymy się ważną zależnością, którą spełnia: prawem wzajemności Weila.

Definicja 3.3.4. Dany jest punkt P na krzywej eliptycznej E. Niech r i s będą dwoma niezerowymi funkcjami wymiernymi na krzywej E. *Symbol lokalny* $\langle r, s\rangle_P$ określony jest następująco:

$$\langle r, s\rangle_P = (-1)^{\mathrm{ord}_P(r)\,\mathrm{ord}_P(s)} \left(\frac{r^{\mathrm{ord}_P(s)}}{s^{\mathrm{ord}_P(r)}}\right)(P) \tag{3.3.5}$$

Podstawowe własności symbolu lokalnego podsumowuje następujące twierdzenie.

Twierdzenie 3.3.6. *Dany jest punkt P na krzywej eliptycznej E. Niech r, r_1, r_2, s, s_1 i s_2 będą dowolnymi niezerowymi funkcjami wymiernymi na krzywej E. Wówczas zachodzą następujące zależności:*

- $\langle r, s\rangle_P \neq 0$;
- *jeżeli funkcje r i s nie mają w punkcie P ani miejsca zerowego, ani bieguna, to* $\langle r, s\rangle_P = 1$;
- $\langle -r, r\rangle_P = 1$;
- $\langle 1 - r, r\rangle_P = 1$;
- $\langle r, s\rangle_P \langle s, r\rangle_P = 1$;
- $\langle r_1 r_2, s\rangle_P = \langle r_1, s\rangle_P \langle r_2, s\rangle_P$;
- $\langle r, s_1 s_2\rangle_P = \langle r, s_1\rangle_P \langle r, s_2\rangle_P$;
- *jeżeli* $\mathrm{ord}_P(r) = 0$, *to* $\langle r, s\rangle_P = r(P)^{\mathrm{ord}_P(s)}$;
- *jeżeli* $\mathrm{ord}_P(s) = 0$, *to* $\langle r, s\rangle_P = \frac{1}{s(P)^{\mathrm{ord}_P(r)}}$;
- *jeżeli* $\mathrm{ord}_P(1 - s) > 0$ *oraz* $\mathrm{ord}_P(r) \neq 0$, *to* $\langle r, s\rangle_P = 1$.

Nietrywialną własnością symbolu lokalnego jest następująca zależność.

Twierdzenie 3.3.7 (Prawo wzajemności Weila)**.** *Dane są niezerowe funkcje wymierne r i s na krzywej eliptycznej E. Wówczas zachodzi następująca zależność:*

$$\prod_{P \in E} \langle r, s\rangle_P = 1 \tag{3.3.8}$$

Dowód prawa wzajemności Weila znacznie wykracza poza zakres niniejszej pracy, dlatego nie podajemy go.

Równoważność obu definicji

Udowodnimy teraz, że obie definicje iloczynu Weila są równoważne.

Twierdzenie 3.3.9. *Dane są punkty P i Q rzędu n na krzywej eliptycznej E nad ciałem $\mathbb{K}$. Wówczas zachodzi następująca zależność:*

$$w(P, Q) = w'(P, Q) \tag{3.3.10}$$

Dowód. Zakładamy, że $P \neq Q$, $P \neq \mathcal{O}$, $Q \neq \mathcal{O}$, gdyż w tych przypadkach łatwo sprawdzić, że żądana równość zachodzi.
Niech f_P oraz f_Q będą funkcjami wymiernymi na krzywej E określonymi w definicji 3.1.20. Poprzez analizę krotności miejsc zerowych i biegunów możemy sprawdzić, że następujące dywizory są równe:

$$\begin{aligned} \operatorname{div}({f_P}^n) &= \operatorname{div}(f'_P \circ [n]) \\ \operatorname{div}({f_Q}^n) &= \operatorname{div}(f'_Q \circ [n]) \end{aligned}$$

Przyjmijmy więc, że funkcje f'_P i f'_Q są wybrane tak, aby zachodziły następujące równości:

$$\begin{aligned} {f_P}^n &= f'_P \circ [n] \\ {f_Q}^n &= f'_Q \circ [n] \end{aligned}$$

Dalej, przyjmijmy, że $P', Q' \in E$ są takimi punktami, że $P = nP'$ oraz $Q = nQ'$. Rozważmy następujący dywizor:

$$(n-1)\langle Q' \rangle + \langle Q' - Q \rangle - n \langle \mathcal{O} \rangle$$

Stosując lemat 3.1.17 sprawdzamy, że jest to dywizor główny. Niech więc r będzie funkcją wymierną na krzywej E o takim dywizorze:

$$\operatorname{div}(r) = (n-1)\langle Q' \rangle + \langle Q' - Q \rangle - n \langle \mathcal{O} \rangle$$

Stosujemy uogólnione prawo wzajemności Weila do funkcji f_P i r:

$$\prod_{S \in E} \langle f_P, r \rangle_S = 1 \tag{3.3.11}$$

Ze względu na własności symbolu lokalnego dla większości punktów S będzie zachodzić $\langle f_P, r \rangle_S = 1$. Jedyne punkty, dla których może być inaczej, to Q', $Q' - Q$ oraz miejsca zerowe i bieguny funkcji f_P. Policzymy wartość symbolu lokalnego dla tych punktów.
Zacznijmy od punktu Q'. W tym punkcie zachodzi $\operatorname{ord}_{Q'}(f_P) = 0$ oraz $\operatorname{ord}_{Q'}(r) = n-1$, więc z własności symbolu lokalnego otrzymujemy:

$$\langle f_P, r \rangle_{Q'} = f_P(Q')^{n-1}$$

Teraz rozważamy punkt $Q' - Q$. Tym razem $\operatorname{ord}_{Q'-Q}(f_P) = 0$ oraz $\operatorname{ord}_{Q'-Q}(r) = 1$. Dostajemy:

$$\langle f_P, r \rangle_{Q'-Q} = f_P(Q' - Q)$$

Wymnażamy obie równości:

$$\begin{aligned} \langle f_P, r \rangle_{Q'} \langle f_P, r \rangle_{Q'-Q} &= f_P(Q')^{n-1} f_P(Q - Q') \\ &= \frac{f_P(Q' - Q)}{f_P(Q')} f_P(Q')^n \\ &= \frac{f_P}{f_P \circ t_Q}(Q' - Q) f_P(Q')^n \\ &= w(P,Q)^{-1} f_P(Q')^n \\ &= w(P,Q)^{-1}(f'_P \circ [n])(Q') \\ &= w(P,Q)^{-1} f'_P(Q) \end{aligned}$$

Zanim obliczymy wartość symbolu lokalnego dla miejsc zerowych i biegunów funkcji f_P, rozważmy następującą funkcję pomocniczą $\widehat{r}$:

$$\widehat{r}(T) = \prod_{S \in E[n]} r(S+T)$$

Jej dywizor jest równy:

$$\operatorname{div}(\widehat{r}) = \sum_{S \in E[n]} \operatorname{div}(r \circ t_S)$$

Zauważmy, że zachodzi:

$$\operatorname{div}(r \circ t_S) = (n-1)\langle Q' - S\rangle + \langle Q' - Q - S\rangle - n\langle -S\rangle$$

Stąd otrzymujemy:

$$\begin{aligned}
\operatorname{div}(\widehat{r}) &= \sum_{S \in E[n]} \operatorname{div}(r \circ t_S) \\
&= \sum_{S \in E[n]} (n-1)\langle Q' - S\rangle + \langle Q' - Q - S\rangle - n\langle -S\rangle \\
&= (n-1)\sum_{S \in E[n]} \langle Q' + S\rangle + \sum_{S \in E[n]} \langle Q' + S\rangle - n\sum_{S \in E[n]} \langle S\rangle \\
&= n[n]^{-1}(Q) - n[n]^{-1}(\mathcal{O}) \\
&= \operatorname{div}({f_Q}^n) \\
&= \operatorname{div}(f'_Q \circ [n])
\end{aligned}$$

Wybraliśmy funkcję r dowolnie, więc przyjmijmy teraz, że została wybrana tak, żeby zachodziła następująca zależność:

$$\widehat{r} = f'_Q \circ [n]$$

Obliczymy teraz fragment iloczynu 3.3.11, który odpowiada miejscom zerowym funkcji f_P. Dla każdego miejsca zerowego S funkcji f_P mamy $\operatorname{ord}_S(f_P) = 1$ oraz $\operatorname{ord}_S(r) = 0$. Zatem:

$$\begin{aligned}
\prod_{\operatorname{ord}_S(f_P)>0} \langle f_P, r\rangle_S &= \prod_{\operatorname{ord}_S(f_P)>0} \frac{1}{r(S)} \\
&= \prod_{S \in E[n]} \frac{1}{r(P'+S)} \\
&= \frac{1}{\widehat{r}(P')} \\
&= \frac{1}{(f'_Q \circ [n])(P')} \\
&= \frac{1}{f'_Q(P)}
\end{aligned}$$

Podobnie postępujemy z fragmentem iloczynu 3.3.11, który odpowiada biegunom funkcji f_P.

Pamiętajmy przy tym, że punkt $\mathcal{O}$ jest również biegunem stopnia n funkcji r.

$$\begin{aligned}
\prod_{\mathrm{ord}_S(f_P)<0} \langle f_P, r\rangle_S &= \langle f_P, r\rangle_{\mathcal{O}} \prod_{S\in E[n]\setminus\{\mathcal{O}\}} r(S) \\
&= (-1)^n \frac{{f_P}^{-n}}{r^{-1}}(\mathcal{O}) \prod_{S\in E[n]\setminus\{\mathcal{O}\}} r(S) \\
&= (-1)^n \frac{{f_P}^{-n}}{\widehat{r}^{-1}}(\mathcal{O}) \\
&= (-1)^n \frac{\widehat{r}}{{f_P}^{n}}(\mathcal{O}) \\
&= (-1)^n \frac{f'_Q \circ [n]}{f'_P \circ [n]}(\mathcal{O}) \\
&= (-1)^n \frac{f'_Q}{f'_P}(\mathcal{O})
\end{aligned}$$

Rozpiszmy jeszcze raz wzór 3.3.11:

$$\begin{aligned}
\prod_{S\in E} \langle f_P, r\rangle_S &= \langle f_P, r\rangle_{Q'} \langle f_P, r\rangle_{Q'-Q} \prod_{\mathrm{ord}_S(f_P)>0} \langle f_P, r\rangle_S \prod_{\mathrm{ord}_S(f_P)<0} \langle f_P, r\rangle_S \\
&= w(P,Q)^{-1} f'_P(Q) \frac{1}{f'_Q(P)} (-1)^n \frac{f'_Q}{f'_P}(\mathcal{O})
\end{aligned}$$

Zgodnie z prawem wzajemności Weila przyrównujemy wynik do jedynki i otrzymujemy:

$$\begin{aligned}
1 &= w(P,Q)^{-1} f'_P(Q) \frac{1}{f'_Q(P)} (-1)^n \frac{f'_Q}{f'_P}(\mathcal{O}) \\
w(P,Q) &= (-1)^n \frac{f'_P(Q)}{f'_Q(P)} \frac{f'_Q}{f'_P}(\mathcal{O}) \\
w(P,Q) &= w'(P,Q)
\end{aligned}$$

□

Rozdział 4

Implementacja iloczynu Weila

Znaczenie iloczynu Weila w kryptografii polega na tym, że istnieje wydajny algorytm, który oblicza jego wartości. Dzięki temu różnego rodzaju konstrukcje teoretyczne – systemy i ataki kryptograficzne – mogą zostać przełożone na algorytmy i ich implementacje, które mogą być zastosowane w praktyce.
W rozdziale tym zaprezentujemy wspomniany algorytm. Został on opracowany przez Millera i opublikowany w pracy [5]. Opiszemy również implementację tego algorytmu, która jest częścią niniejszej pracy.

4.1. Podstawowy wzór

Algorytm będzie obliczał wartości iloczynu Weila na podstawie definicji alternatywnej. Niestety, fragmentem wzoru 3.3.2 jest wyrażenie $\frac{f'_Q}{f'_P}(\mathcal{O})$, które zawiera elementy niewygodne z obliczeniowego punktu widzenia: miejsca zerowe, bieguny i punkt w nieskończoności. Dlatego „przesuniemy" licznik i mianownik we wzorze 3.3.2 tak, aby operować tylko na wartościach skończonych.

Definicja 4.1.1. Dane są punkty P i Q rzędu n na krzywej eliptycznej E nad ciałem $\mathbb{K}$. Niech R i S będą dowolnymi punktami na krzywej E takimi, że punkty P, Q, R, S, $P+R$, $Q+S$ i $\mathcal{O}$ są parami różne. Niech f''_P i f''_Q będą funkcjami wymiernymi na krzywej E określonymi z dokładnością do niezerowego czynnika stałego poprzez następujące dywizory:

$$\begin{aligned} \operatorname{div}(f''_P) &= n\langle P+R\rangle - n\langle R\rangle \\ \operatorname{div}(f''_Q) &= n\langle Q+S\rangle - n\langle S\rangle \end{aligned}$$

Druga alternatywna postać iloczynu Weila to funkcja $w''(P,Q)\colon E[n] \to \mathbb{K}$ określona następująco:

$$w''(P,Q) = (-1)^n \frac{f''_P(Q+S)}{f''_P(S)} \frac{f''_Q(R)}{f''_Q(P+R)} \tag{4.1.2}$$

Uwaga 4.1.3. Podobnie jak w przypadku funkcji f'_P i f'_Q z definicji 3.3.1, funkcje f''_P i f''_Q w definicji 4.1.1 są określone z dokładnością do czynnika stałego. Nie wpływa to na wartość funkcji $w''(P,Q)$, zatem jest ona dobrze określona.

Uwaga 4.1.4. Ze względu na wybór punktów R i S żaden z czynników występujących we wzorze 4.1.2 nie jest zerem ani nieskończonością.

Pokażemy teraz, że funkcje $w'(P,Q)$ i $w''(P,Q)$ są równe. Głównym narzędziem użytym w dowodzie będzie prawo wzajemności Weila.

Twierdzenie 4.1.5. *Dane są punkty P i Q rzędu n na krzywej eliptycznej E. Wówczas zachodzi następująca zależność:*

$$w'(P,Q) = w''(P,Q) \tag{4.1.6}$$

Dowód. Z postaci dywizorów określających funkcje f''_P i f''_Q łatwo zauważyć, że $f''_P = f'_P \circ t_R$ oraz $f''_Q = f'_Q \circ t_S$. Jest jeszcze inna zależność łącząca te funkcje.
Z lematu 3.1.17 wiemy, że dywizor $\langle P+R\rangle - \langle P\rangle - \langle R\rangle + \langle\mathcal{O}\rangle$ jest główny, podobnie dywizor $\langle Q+S\rangle - \langle Q\rangle - \langle S\rangle + \langle\mathcal{O}\rangle$. Niech więc funkcje wymierne r_P i r_Q będą określone poprzez te dywizory:

$$\begin{aligned}\operatorname{div}(r_P) &= \langle P+R\rangle - \langle P\rangle - \langle R\rangle + \langle\mathcal{O}\rangle \\ \operatorname{div}(r_Q) &= \langle Q+S\rangle - \langle Q\rangle - \langle S\rangle + \langle\mathcal{O}\rangle\end{aligned}$$

Można teraz sprawdzić, że $f'''_P = f'_P r_P^n$ oraz $f'''_Q = f'_Q r_Q^n$.
Zastosujmy prawo wzajemności Weila do funkcji f'_P i r_Q. Wykorzystamy przy tym własności symbolu lokalnego oraz to, że punkty P, $Q+S$, Q, S i $\mathcal{O}$ są parami różne. Otrzymujemy:

$$\begin{aligned}1 &= \prod_{T\in E} \langle f'_P, r_Q\rangle_T \\ &= \langle f'_P, r_Q\rangle_{Q+S} \langle f'_P, r_Q\rangle_Q \langle f'_P, r_Q\rangle_S \langle f'_P, r_Q\rangle_P \langle f'_P, r_Q\rangle_{\mathcal{O}} \\ &= \frac{1}{r_Q^n(P)} \frac{f'_P(Q+S)}{f'_P(Q) f'_P(S)} (-1)^{-n} \left(\frac{f'_P}{r_Q^{-n}}\right)(\mathcal{O})\end{aligned} \tag{4.1.7}$$

W analogiczny sposób stosujemy prawo wzajemności Weila do funkcji r_P i f'_Q:

$$\begin{aligned}1 &= \prod_{T\in E} \left\langle r_P, f'_Q\right\rangle_T \\ &= \left\langle r_P, f'_Q\right\rangle_{P+R} \left\langle r_P, f'_Q\right\rangle_P \left\langle r_P, f'_Q\right\rangle_R \left\langle r_P, f'_Q\right\rangle_Q \left\langle r_P, f'_Q\right\rangle_{\mathcal{O}} \\ &= r_P^n(Q) \frac{f'_Q(P) f'_Q(R)}{f'_Q(P+R)} (-1)^{-n} \left(\frac{r_P^{-n}}{f'_Q}\right)(\mathcal{O})\end{aligned} \tag{4.1.8}$$

Stosujemy prawo wzajemności Weila jeszcze raz do funkcji r_P i r_Q:

$$\begin{aligned}1 &= \prod_{T\in E} \langle r_P, r_Q\rangle_T \\ &= \langle r_P, r_Q\rangle_{P+R} \langle r_P, r_Q\rangle_P \langle r_P, r_Q\rangle_R \\ &\quad \langle r_P, r_Q\rangle_{Q+S} \langle r_P, r_Q\rangle_Q \langle r_P, r_Q\rangle_S \\ &\quad \langle r_P, r_Q\rangle_{\mathcal{O}} \\ &= \frac{r_P(Q+S)}{r_P(S)} \frac{r_Q(R)}{r_Q(P+R)} \frac{r_P(Q)}{r_Q(P)} \left(\frac{r_P}{r_Q}\right)(\mathcal{O})\end{aligned} \tag{4.1.9}$$

Ostatnią równość podnosimy do n-tej potęgi i otrzymujemy:

$$1 = \frac{r_P^n(Q+S)}{r_P^n(S)} \frac{r_Q^n(R)}{r_Q^n(P+R)} \frac{r_P^n(Q)}{r_Q^n(P)} \left(\frac{r_P^n}{r_Q^n}\right)(\mathcal{O}) \tag{4.1.10}$$

Równości 4.1.7, 4.1.8 i 4.1.10 domnażamy do wzoru 3.3.2. Otrzymujemy:

$$\begin{aligned}
w'(P,Q) &= \frac{f'_P(Q)}{f'_Q(P)}(-1)^n \left(\frac{f'_Q}{f'_P}\right)(\mathcal{O}) \\
&\quad \frac{1}{r_Q^n(P)}\frac{f'_P(Q+S)}{f'_P(Q)f'_P(S)}(-1)^{-n}\left(\frac{f'_P}{r_Q^{-n}}\right)(\mathcal{O}) \\
&\quad r_P^n(Q)\frac{f'_Q(P)f'_Q(R)}{f'_Q(P+R)}(-1)^{-n}\left(\frac{r_P^{-n}}{f'_Q}\right)(\mathcal{O}) \\
&\quad \frac{r_P^n(Q+S)}{r_P^n(S)}\frac{r_Q^n(R)}{r_Q^n(P+R)}\frac{r_P^n(Q)}{r_Q^n(P)}\left(\frac{r_P^n}{r_Q^n}\right)(\mathcal{O}) \\
&= \frac{f'_P(Q+S)r_P^n(Q+S)}{f_P(S)r_P^n(S)}\frac{f'_Q(R)r_Q^n(R)}{f'_Q(P+R)r_Q^n(P+R)}(-1)^n\left(\frac{f'_Pf'_Qr_P^nr_P^{-n}}{f'_Qf'_Pr_Q^nr_Q^{-n}}\right)(\mathcal{O}) \\
&= (-1)^n\frac{f''_P(Q+S)}{f''_P(S)}\frac{f''_Q}{f''_Q(P+R)} \\
&= w''(P,Q)
\end{aligned}$$

□

4.2. Algorytm Millera

Przedstawimy teraz pseudokod algorytmu Millera, który oblicza wartości iloczynu Weila na podstawie wzoru 4.1.2. Najpierw opiszemy zasady obowiązujące przy zapisywaniu pseudokodu, a następnie podamy treść procedur pomocniczych oraz właściwego algorytmu.

Uwaga 4.2.1. Aby uniknąć niejasności przy zapisywaniu pseudokodu, przyjmujemy następujące ustalenia:

- pseudokod będziemy zapisywać w formie podobnej do tej stosowanej w książce [17];
- operacje arytmetyczne na liczbach całkowitych ($+$, $-$, $\cdot$, **div**, **mod**) uznajemy za elementarne;
- porównania liczb całkowitych ($=$, $\neq$, $<$, $\leqslant$, $>$, $\geqslant$) uznajemy za elementarne;
- operacje arytmetyczne na elementach ciał skończonych ($+$, $-$, $\cdot$, $/$) uznajemy za elementarne;
- porównania elementów ciał skończonych ($=$, $\neq$) uznajemy za elementarne;
- porównania punktów na krzywej eliptycznej ($=$, $\neq$) uznajemy za elementarne;
- ustalamy, że notacja $p[o]$ oznacza odczytanie cechy p obiektu o, w szczególności:
  - *zero*$[K]$, *one*$[K]$ itd. oznaczają odpowiednio zero, jedynkę itd. w ciele K;
  - $A[E]$, $B[E]$, *field*$[E]$, *identity*$[E]$ oznaczają odpowiednio parametry krzywej eliptycznej E, ciało, nad którym krzywa E jest zdefiniowana i punkt w nieskończoności krzywej E;
  - $x[P]$ i $y[P]$ oznaczają współrzędne skończonego punktu P na krzywej eliptycznej;

– $a[l]$ $b[l]$ i $c[l]$ oznaczają odpowiednio współczynnik stojący przy zmiennej x, współczynnik stojący przy zmiennej y i wyraz wolny linii l na krzywej eliptycznej;

- zakładamy, że dysponujemy następującymi procedurami:

  – RANDOM-INTEGER(n), która wybiera losowo z rozkładem jednostajnym i przekazuje jako wynik liczbę całkowitą a z przedziału $[0; n)$, tzn. $0 \leqslant a < n$;

  – RANDOM-FINITE-FIELD-ELEMENT(K), która wybiera losowo z rozkładem jednostajnym i przekazuje jako wynik element ciała skończonego K;

  – FINITE-FIELD-ELEMENT-SQUARE-ROOT(K, a), która oblicza i przekazuje jako wynik dowolny pierwiastek kwadratowy elementu $a \in K$ lub stałą ERROR, jeśli taki pierwiastek nie istnieje;

  – CURVE-FINITE-POINT(E, a, b), która konstruuje i przekazuje jako wynik punkt skończony na krzywej eliptycznej E o współrzędnych a i b;

  – CURVE-POINT-CONJUGATE(E, P), która oblicza i przekazuje jako wynik punkt sprzężony do punktu P na krzywej eliptycznej E;

  – LINE-ON-CURVE(E, a, b, c), która konstruuje i przekazuje jako wynik linię na krzywej eliptycznej E o współczynnikach a, b i c.

Ponieważ wybraliśmy dosyć ubogi zestaw operacji elementarnych, przed opisem procedur związanych bezpośrednio z algorytmem Millera musimy w postaci pseudokodu przedstawić ogólne algorytmy związane z krzywymi eliptycznymi. Część z nich to ujęte w postaci pseudokodu wzory wyprowadzone wcześniej w pracy.
Rozpoczniemy od procedury obliczającej wartość linii w punkcie krzywej eliptycznej. Jest ona trywialna, ale nie uznaliśmy jej za operację elementarną, dlatego wypada ją zdefiniować.

Algorytm 4.2.2. Dana jest linia l oraz punkt skończony P na krzywej eliptycznej E. Następująca procedura na podstawie wartości E, l i P oblicza i przekazuje jako wynik wartość $l(P)$:

LINE-VALUE-AT-CURVE-FINITE-POINT(E, l, P)
1 **assert** $P \neq \mathit{identity}[E]$
2 **return** $a[l] \cdot x[P] + b[l] \cdot y[P] + c[l]$

Teraz podamy procedury wyznaczające linie przechodzące przez zadane punkty. Najpierw rozważymy przypadki szczególne, dla których wyprowadziliśmy wcześniej odpowiednie wzory.

Algorytm 4.2.3. Dany jest punkt P na krzywej eliptycznej E. Następująca procedura na podstawie wartości E i P oblicza i przekazuje jako wynik linię pionową przechodzącą przez punkt P:

VERTICAL-LINE-THROUGH-CURVE-POINT(E, P)
1 $K \leftarrow \mathit{field}(E)$
2 **if** $P = \mathit{identity}[E]$
3 **then return** LINE-ON-CURVE($E, \mathit{zero}[K], \mathit{zero}[K], \mathit{one}[K]$)
4 **else** $a \leftarrow x[P]$
5 **return** LINE-ON-CURVE($E, \mathit{one}[K], \mathit{zero}[K], -a$)

Algorytm 4.2.4. Dane są dwa różne punkty skończone P i Q na krzywej eliptycznej E. Następująca procedura na podstawie wartości E, P i Q oblicza i przekazuje jako wynik linię przechodzącą przez punkty P i Q:

LINE-THROUGH-DIFFERENT-CURVE-FINITE-POINTS(E, P, Q)

```
 1  assert P ≠ identity[E] and P ≠ identity[E]
 2  assert P ≠ Q
 3  K ← field[E]
 4  a ← x[P]
 5  b ← y[P]
 6  c ← x[Q]
 7  d ← y[Q]
 8  if a = c
 9      then return LINE-ON-CURVE(E, one[K], zero[K], −a)
10      else  λ ← (d − b)/(c − a)
11            return LINE-ON-CURVE(E, λ, − one[K], −(λ · a − b))
```

Algorytm 4.2.5. Dany jest punkt skończony P na krzywej eliptycznej E. Następująca procedura na podstawie wartości E i P oblicza i przekazuje jako wynik linię styczną do krzywej E przechodzącą przez punkt P:

TANGENT-LINE-THROUGH-CURVE-FINITE-POINT(E, P)

```
1  assert P ≠ identity[E]
2  K ← field[E]
3  a ← x[P]
4  b ← y[P]
5  if b = zero[K]
6     then return LINE-ON-CURVE(E, one[K], zero[K], −a)
7     else  λ ← (three[K] · a · a + A[E])/(two[K] · b)
8           return LINE-ON-CURVE(E, λ, − one[K], −(λ · a − b))
```

Podane procedury zbierzemy teraz w jedną całość: procedurę, która wyznacza linię przechodzącą przez dowolne dwa punkty krzywej.

Algorytm 4.2.6. Dane są punkty P i Q na krzywej eliptycznej E. Następująca procedura na podstawie wartości E, P i Q oblicza i przekazuje jako wynik linię przechodzącą przez punkty P i Q (lub styczną do krzywej w punkcie P, gdy $P = Q$):

LINE-THROUGH-CURVE-POINTS(E, P, Q)

```
 1  K ← field[E]
 2  if P = identity[E] and Q = identity[E]
 3     then return LINE-ON-CURVE(E, zero[K], zero[K], one[K])
 4  if P = identity[E]
 5     then return VERTICAL-LINE-THROUGH-CURVE-POINT(E, Q)
 6  if Q = identity[E]
 7     then return VERTICAL-LINE-THROUGH-CURVE-POINT(E, P)
 8  if P ≠ Q
 9     then return LINE-THROUGH-DIFFERENT-CURVE-FINITE-POINTS(E, P, Q)
10     else  return TANGENT-LINE-THROUGH-CURVE-FINITE-POINT(E, P)
```

Kolejne ogólne procedury to dodawanie dwóch punktów krzywej oraz mnożenie punktu przez liczbę całkowitą.

Algorytm 4.2.7. Dane są punkty P i Q na krzywej eliptycznej E. Następująca procedura na podstawie wartości E, P i Q oblicza i przekazuje jako wynik punkt $P + Q$:

```
ADD-CURVE-POINTS(E, P, Q)
 1  K ← field[E]
 2  if P = identity[E]
 3      then return Q
 4  if Q = identity[E]
 5      then return P
 6  if P = CURVE-POINT-CONJUGATE(E, Q)
 7      then return identity[E]
 8  a ← x[P]
 9  b ← y[P]
10  c ← x[Q]
11  d ← y[Q]
12  if P ≠ Q
13      then λ ← (d − b)/(c − a)
14      else  λ ← (three[K] · a · a + A[E])/(two[K] · b)
15  e ← λ · λ − a − c
16  f ← −λ · (e − a) − b
17  return CURVE-FINITE-POINT(E, e, f)
```

Algorytm 4.2.8. Dany jest punkt P na krzywej eliptycznej E oraz liczba całkowita n. Następująca procedura na podstawie wartości E, P i n oblicza i przekazuje jako wynik punkt nP:

```
MULTIPLY-CURVE-POINT(E, P, n)
 1  K ← field[E]
 2  if P = identity[E]
 3     then return identity[E]
 4  if y[P] = zero[K]
 5     then if n mod 2 = 0
 6             then return identity[E]
 7             else  return P
 8  if n = 0
 9     then return identity[E]
10  if n > 0
11     then m ← n
12     else  m ← −n
13  R ← identity[E]
14  while m > 0
15        do if m mod 2 ≠ 0
16             then R ← ADD-CURVE-POINTS(R, P)
17           P ← ADD-CURVE-POINTS(P, P)
18           m ← m div 2
19  if n > 0
20     then return R
21     else  return CURVE-POINT-CONJUGATE(E, R)
```

Ostatnią ogólną procedurą jest losowanie punktu na krzywej eliptycznej.

Algorytm 4.2.9. Dana jest krzywa eliptyczna E nad ciałem skończonym. Następująca procedura na podstawie wartości E wybiera losowo z rozkładem jednostajnym i przekazuje jako wynik punkt skończony na krzywej E:

```
RANDOM-CURVE-FINITE-POINT(E)
 1  K ← field[E]
 2  while TRUE
 3        do a ← RANDOM-FINITE-FIELD-ELEMENT(K)
 4           d ← a · a · a + A[E] · a + B[E]
 5           if d = zero[K]
 6              then if RANDOM-INTEGER(2) = 0
 7                      then return CURVE-FINITE-POINT(E, a, zero[K])
 8              else  b ← FINITE-FIELD-ELEMENT-SQUARE-ROOT(K, d)
 9                    if b ≠ ERROR
10                       then if RANDOM-INTEGER(2) = 0
11                               then return CURVE-FINITE-POINT(E, a, b)
12                               else  return CURVE-FINITE-POINT(E, a, −b)
```

Możemy teraz przejść do opisania algorytmu Millera.

Algorytm 4.2.10. Dane są punkty P i Q rzędu n na krzywej eliptycznej E nad ciałem skończonym. Następujące procedury na podstawie wartości E, n, P i Q obliczają i przekazują jako wynik wartość $w(P, Q)$:

```
COMBINE-PARTIAL-VALUES(E, A, U, V, u, v)
1  K ← field[E]
2  g ← LINE-THROUGH-CURVE-POINTS(E, U, V)
3  h ← VERTICAL-LINE-THROUGH-CURVE-POINT(ADD-CURVE-POINTS(E, U, V))
4  s ← LINE-VALUE-AT-CURVE-FINITE-POINT(E, g, A)
5  t ← LINE-VALUE-AT-CURVE-FINITE-POINT(E, h, A)
6  if s = zero[K] or t = zero[K]
7     then return ERROR
8  return u · v · (s/t)
```

```
COMPUTE-VALUE(E, n, P, R, A)
 1  K ← field[E]
 2  g ← LINE-THROUGH-CURVE-POINTS(E, P, R)
 3  h ← VERTICAL-LINE-THROUGH-CURVE-POINT(E, ADD-CURVE-POINTS(E, P, R))
 4  s ← LINE-VALUE-AT-CURVE-FINITE-POINT(E, g, A)
 5  t ← LINE-VALUE-AT-CURVE-FINITE-POINT(E, h, A)
 6  if s = zero[K] or t = zero[K]
 7     then return ERROR
 8  U ← identity[E]
 9  V ← P
10  u ← one[K]
11  v ← t/s
12  while n > 0
13        do if v = ERROR
14              then return ERROR
15           if n mod 2 ≠ 0
16              then u ← COMBINE-PARTIAL-VALUES(E, A, U, V, u, v)
17                   U ← ADD-CURVE-POINTS(U, V)
18                   if u = ERROR
19                      then return ERROR
20           v ← COMBINE-PARTIAL-VALUES(E, A, V, V, v, v)
21           V ← ADD-CURVE-POINTS(E, V, V)
22           n ← n div 2
23  return u
```

```
WEIL-PAIRING(E, n, P, Q)
 1  assert MULTIPLY-CURVE-POINT(E, P, n) = identity[E]
 2  assert MULTIPLY-CURVE-POINT(E, Q, n) = identity[E]
 3  K ← field[E]
 4  if P = Q or P = identity[E] or Q = identity[E]
 5     then return one[K]
 6  while TRUE
 7        do R ← RANDOM-CURVE-FINITE-POINT(E)
 8           S ← RANDOM-CURVE-FINITE-POINT(E)
 9           R' ← ADD-CURVE-POINTS(E, P, R)
10           S' ← ADD-CURVE-POINTS(E, Q, S)
11           if R' ≠ identity[E] and S' ≠ identity[E]
12              then a ← COMPUTE-VALUE(E, n, P, R, S')
13                   b ← COMPUTE-VALUE(E, n, P, R, S)
14                   c ← COMPUTE-VALUE(E, n, Q, S, R')
15                   d ← COMPUTE-VALUE(E, n, Q, S, R)
16                   if a ≠ ERROR and b ≠ ERROR and c ≠ ERROR and d ≠ ERROR
17                      then return (a/b) · (d/c)
```

4.3. Dowód poprawności i oszacowanie złożoności

Zajmiemy się teraz analizą algorytmu Millera: udowodnimy jego poprawność oraz oszacujemy złożoność czasową i pamięciową. Analizę poprzedzimy krótkim wyjaśnieniem idei stojącej za algorytmem.

Wyjaśnienie

Podstawę, na której opiera się algorytm Millera, stanowi następująca definicja i płynące z niej wnioski.

Definicja 4.3.1. Dany jest punkt P rzędu n na krzywej eliptycznej E. Niech R będzie dowolnym punktem na krzywej E. *Cząstkowe wartości iloczynu Weila* to rodzina funkcji wymiernych $r_{P,R}^{(m)}$ na krzywej E, gdzie $m = 0, 1, \ldots, n$, określona z dokładnością do niezerowego czynnika stałego poprzez następujące dywizory:

$$\operatorname{div}(r_{P,R}^{(m)}) = m\,\langle P + R\rangle - m\,\langle R\rangle - \langle mP\rangle + \langle \mathcal{O}\rangle \tag{4.3.2}$$

Następujący fakt motywuje nazwę oraz pokazuje związek cząstkowych wartości iloczynu Weila z samym iloczynem Weila.

Fakt 4.3.3. *Zachodzi następująca zależność:*

$$\operatorname{div}(f_P'') = \operatorname{div}(r_{P,R}^{(n)}) \tag{4.3.4}$$

Cząstkowe wartości iloczynu Weila mają następującą kluczową własność.

Lemat 4.3.5. *Dane są cząstkowe wartości iloczynu Weila $r_{P,R}^{(m)}$ oraz liczby naturalne k i l takie, że $0 \leqslant k, l \leqslant n$ i $0 \leqslant k+l \leqslant n$. Niech g będzie linią przechodzącą przez punkty kP i lP, a h będzie linią pionową przechodzącą przez punkt $(k+l)P$. Wówczas zachodzi następująca zależność:*

$$\operatorname{div}\left(r_{P,R}^{(k+l)}\right) = \operatorname{div}\left(r_{P,R}^{(k)}\right) + \operatorname{div}\left(r_{P,R}^{(l)}\right) + \operatorname{div}(g) - \operatorname{div}(h) \tag{4.3.6}$$

Dowód. Wystarczy sprawdzić dywizory obu stron równości:

$$\begin{aligned}
\operatorname{div}\left(r_{P,R}^{(k+l)}\right) &= (k+l)\langle P+R\rangle - (k+l)\langle R\rangle - \langle (k+l)P\rangle + \langle \mathcal{O}\rangle \\
&= k\langle P+R\rangle - k\langle R\rangle - \langle kP\rangle + \langle \mathcal{O}\rangle + \\
&\quad l\langle P+R\rangle - l\langle R\rangle - \langle lP\rangle + \langle \mathcal{O}\rangle + \\
&\quad \langle kP\rangle + \langle lP\rangle - \langle (k+l)P\rangle - \langle \mathcal{O}\rangle \\
&= \operatorname{div}\left(r_{P,R}^{(k)}\right) + \operatorname{div}\left(r_{P,R}^{(l)}\right) + \langle kP\rangle + \langle lP\rangle + \langle -(k+l)P\rangle - 3\langle \mathcal{O}\rangle - \\
&\quad \langle -(k+l)P\rangle - \langle (k+l)P\rangle + 2\langle \mathcal{O}\rangle \\
&= \operatorname{div}\left(r_{P,R}^{(k)}\right) + \operatorname{div}\left(r_{P,R}^{(l)}\right) + \operatorname{div}(g) - \operatorname{div}(h)
\end{aligned}$$

□

Wniosek 4.3.7. *Zachodzi następująca zależność:*

$$r_{P,R}^{(k+l)} = r_{P,R}^{(k)} r_{P,R}^{(l)} \frac{g}{h} \tag{4.3.8}$$

Widzimy teraz, na czym opiera się algorytm Millera. Lemat 4.3.5 i płynący z niego wniosek 4.3.7 sugerują algorytm typu „podwajaj-i-dodawaj", który pozwala obliczać wartości funkcji $r_{P,R}^{(n)}$.

Poprawność

Pokażemy teraz, że algorytm 4.2.10 jest poprawny, tzn. że rzeczywiście oblicza wartości iloczynu Weila i że ma własność stopu.

Lemat 4.3.9. *Procedura* COMBINE-PARTIAL-VALUES *na podstawie wartości E, A, $U = kP$, $V = lP$, $u = r_{P,R}^{(k)}(A)$ i $v = r_{P,R}^{(l)}(A)$ oblicza i przekazuje jako wynik wartość $r_{P,R}^{(k+l)}(A)$.*

Dowód. Jest to jasne – procedura ta jest bezpośrednim przełożeniem na pseudokod wniosku 4.3.7. □

Uwaga 4.3.10. Jest możliwe, że procedura Procedura COMBINE-FUNCTION-VALUES przekaże jako wynik specjalną stałą ERROR. Sytuację tę omówimy niebawem.

Lemat 4.3.11. *Procedura* COMPUTE-VALUE *w algorytmie 4.2.10 na podstawie wartości E, n, P, R i A oblicza i przekazuje jako wynik wartość $r_{P,R}^{(n)}(A)$, o ile nie kończy się błędem (tzn. nie przekazuje jako wyniku stałej* ERROR*).*

Dowód. Procedura ta jest typowym algorytmem typu „podwajaj-i-dodawaj". Jego poprawność łatwo udowodnić przez indukcję. Będziemy rozpatrywać zapis liczby n w postaci dwójkowej:

$$n = \sum_{k=0}^{\lfloor \log_2 n \rfloor} n_k 2^k$$

Jak łatwo sprawdzić, przed rozpoczęciem pętli **while** zmienna u ma wartość $r_{P,R}^{(0)}(A)$, a zmienna v ma wartość $r_{P,R}^{(1)}(A)$.

Po zakończeniu k-tego przebiegu pętli zmienna u ma wartość $r_{P,R}^{(m_k)}(A)$, gdzie $m_k = \sum_{l=0}^{k} n_l 2^l$, a zmienna v ma wartość $r_{P,R}^{(2^k)}(A)$.

Widać stąd, że po zakończeniu $(\lfloor \log_2 n \rfloor + 1)$-szego kroku pętli zmienna u będzie miała żądaną wartość $r_{P,R}^{(n)}$. □

Uwaga 4.3.12. Istotą procedury COMPUTE-VALUE jest to, że pozwala ona obliczyć wartość funkcji $r_{P,R}^{(n)}$ w zadanym punkcie A bez konieczności obliczania wyrażenia wymiernego określającego funkcję $r_{P,R}^{(n)}$. Obliczenie tego wyrażenia byłoby możliwe – wystarczy w procedurze COMBINE-PARTIAL-VALUES zastąpić mnożenie elementów ciała skończonego formalnym mnożeniem wyrażeń wymiernych. Jest to jednak mniej wydajne podejście.

Uwaga 4.3.13. Procedury COMBINE-PARTIAL-VALUES i COMPUTE-VALUE mogą zakończyć się błędem, tzn. przekazać jako wynik stałą ERROR. Sytuacja ta występuje wtedy, gdy pewne wartości wyliczone w tych procedurach są równe zero. Kontynuowanie obliczeń w tej sytuacji nie ma sensu, ponieważ doprowadziłoby to do mnożenia lub dzielenia przez zero i nie byłoby wówczas możliwe poprawne obliczenie wartości iloczynu Weila, ponieważ są one niezerowe. Aby rozwiązać ten problem, należałoby potencjalne mnożenie lub dzielenie przez zero zastąpić śledzeniem krotności pojawiających się zer i biegunów (co jest dosyć żmudne) lub zamienić operacje arytmetyczne na elementach ciała skończonego na operacje arytmetyczne na wyrażeniach wymiernych nad tym ciałem (co nie jest wydajne).

Twierdzenie 4.3.14. *Jeżeli procedura* WEIL-PAIRING *w algorytmie 4.2.10 zatrzymuje się, to jako wynik przekazuje wartość* $w(P, Q)$.

Dowód. Jest to jasne. Jeżeli żadna z wartości a, b, c, d obliczanych w tej procedurze nie jest równa stałej ERROR, to widzimy, że wyrażenie $(a/b) \cdot (d/c)$ przekazywane przez procedurę faktycznie jest równe $w(P, Q)$, co wynika z poprawności procedury COMPUTE-VALUE (lemat 4.3.11), faktu 4.3.3 i wzoru 4.1.2. □

Twierdzenie 4.3.15. *Procedura* WEIL-PAIRING *w algorytmie 4.2.10 prawie na pewno kończy się.*

Dowód. Nawet jeżeli punkty R i S wylosowane w procedurze WEIL-PAIRING są takie, że punkty P, R, $P+R$, Q, S, $Q+S$ i $\mathcal{O}$ są parami różne i wyrażenie 4.1.2 jest dobrze określone, to procedura COMPUTE-VALUE wciąż może zakończyć się błędem. Dzieje się tak dlatego, że podczas obliczania wartości COMPUTE-VALUE(E, n, P, R, A), gdzie $A = S$ lub $A = Q + S$, punkt A może pokryć się z jednym z punktów występujących w procedurach COMBINE-PARTIAL-VALUES i COMPUTE-VALUE. Gdyby przeprowadzać obliczenia symbolicznie, tzn. konstruować wyrażenia wymierne określające funkcje $r_{P,R}^{(m)}$, to obliczone ostatecznie wyrażenie $r_{P,R}^{(n)}$ nie miałoby w punkcie S ani $Q + S$ miejsca zerowego ani bieguna i można by obliczyć wartość $r_{P,R}^{(n)}(A)$. Procedura COMPUTE-VALUE nie prowadzi jednak obliczeń symbolicznych, więc może wystąpić błąd opisany w uwadze 4.3.13.
Obliczenie wartości COMPUTE-VALUE(E, n, P, R, A) powiedzie się, jeżeli punkt A nie pokryje się z żadnym z punktów R, $P, 2P, 4P, 8P, \ldots, 2^{\lfloor \log_2 n \rfloor}P$, $T_1, T_2, T_3, \ldots, T_{\lfloor \log_2 n \rfloor}$, gdzie $T_k = \left(\sum_{l=0}^{k} n_l 2^l\right) P$. Punktów tych jest łącznie $O(\log n)$.
Wniosek stąd, że obliczenie każdej z wartości a, b, c, d w jednym przebiegu pętli **while** w procedurze WEIL-PAIRING zakończy się błędem z prawdopodobieństwem niewiększym niż $O(\frac{\log n}{q})$, gdzie q to rozmiar ciała skończonego, nad którym dana jest krzywa E.
Przyjmijmy teraz, że $q = p^e$. Jeżeli prawdopodobieństwo wystąpienia błędu jest zbyt duże, możemy losować punkty R i S z odpowiednio większego ciała o rozmiarze $q' = p^{e'}$, gdzie $e' = ef$ i $f > 1$. Dla odpowiednio dużej wartości q' prawdopodobieństwo wystąpienia błędu stanie się mniejsze niż $\frac{1}{2}$, co będzie oznaczać, że procedura WEIL-PAIRING prawie na pewno kończy się, a oczekiwana liczba przebiegów pętli **while** jest stała. □

Złożoność czasowa i pamięciowa

Oszacujemy teraz złożoność asymptotyczną algorytmu Millera. Złożoność czasową będziemy mierzyć ilością niezbędnych operacji na bitach, a pamięciową – ilością wymaganych dodatkowych bitów pamięci.

Uwaga 4.3.16. Złożoność asymptotyczną będziemy mierzyć w zależności od dwóch naturalnych w tej sytuacji parametrów: rozmiaru q ciała skończonego, nad którym dana jest krzywa eliptyczna oraz liczby n oznaczającej rząd punktów w podgrupie torsyjnej, w której rozpatrujemy iloczyn Weila.

Rozpoczniemy od ustalenia złożoności operacji elementarnych i procedur, które uznaliśmy za dane.

Uwaga 4.3.17. Złożoność asymptotyczna operacji uznanych za elementarne jest następująca:

- addytywne operacje arytmetyczne na liczbach całkowitych wymagają $O(\log n)$ operacji na bitach i $O(\log n)$ bitów pamięci;
- multiplikatywne operacje arytmetyczne na liczbach całkowitych wymagają $O(\log^2 n)$ operacji na bitach i $O(\log n)$ bitów pamięci;
- porównania liczb całkowitych wymagają $O(\log n)$ operacji na bitach i $O(1)$ bitów pamięci;
- addytywne operacje arytmetyczne na elementach ciała skończonego wymagają $O(\log q)$ operacji na bitach i $O(\log q)$ bitów pamięci;
- multiplikatywne operacje arytmetyczne na elementach ciała skończonego wymagają $O(\log^2 q)$ operacji na bitach i $O(\log q)$ bitów pamięci;
- porównania elementów ciała skończonego wymagają $O(\log q)$ operacji na bitach i $O(1)$ bitów pamięci;
- porównania punktów na krzywej eliptycznej wymagają $O(\log q)$ operacji na bitach i $O(1)$ bitów pamięci;
- operacje typu $p[o]$, czyli odczytanie cechy p obiektu o, wymagają $O(1)$ operacji na bitach i $O(1)$ bitów pamięci.

Uwaga 4.3.18. Uzależnienie złożoności operacji na liczbach całkowitych od parametru n wynika stąd, że we wszystkich procedurach wykorzystywanych w algorytmie Millera nie pojawiają się liczby większe niż n.

Uwaga 4.3.19. Złożoność asymptotyczna procedur uznanych za dane jest następująca:

- wykonanie procedury Random-Integer wymaga $O(\log n)$ operacji na bitach i $O(\log n)$ bitów pamięci;
- wykonanie procedury Random-Finite-Field-Element wymaga $O(\log q)$ operacji na bitach i $O(\log q)$ bitów pamięci;
- wykonanie procedury Finite-Field-Square-Root wymaga średnio $O(\log^3 q)$ operacji na bitach i $O(\log q)$ bitów pamięci;
- wykonanie procedury Curve-Finite-Point wymaga $O(\log q)$ operacji na bitach i $O(\log q)$ bitów pamięci;

- wykonanie procedury CURVE-POINT-CONJUGATE wymaga $O(\log q)$ operacji na bitach i $O(\log q)$ bitów pamięci;
- wykonanie procedury LINE-ON-CURVE wymaga $O(\log q)$ operacji na bitach i $O(\log q)$ bitów pamięci;

Podamy teraz złożoność ogólnych procedur zaprezentowanych w tym rozdziale.

Fakt 4.3.20. *Wykonanie procedury* LINE-VALUE-AT-FINITE-POINT *wymaga* $O(\log^2 q)$ *operacji na bitach i* $O(\log q)$ *bitów pamięci.*

Fakt 4.3.21. *Wykonanie procedury* VERTICAL-LINE-THROUGH-CURVE-POINT *wymaga* $O(\log q)$ *operacji na bitach i* $O(\log q)$ *bitów pamięci.*

Fakt 4.3.22. *Wykonanie procedury* LINE-THROUGH-DIFFERENT-CURVE-FINITE-POINTS *wymaga* $O(\log^2 q)$ *operacji na bitach i* $O(\log q)$ *bitów pamięci.*

Fakt 4.3.23. *Wykonanie procedury* TANGENT-LINE-CURVE-FINITE-POINT *wymaga* $O(\log^2 q)$ *operacji na bitach i* $O(\log q)$ *bitów pamięci.*

Fakt 4.3.24. *Wykonanie procedury* LINE-THROUGH-CURVE-POINTS *wymaga* $O(\log^2 q)$ *operacji na bitach i* $O(\log q)$ *bitów pamięci.*

Fakt 4.3.25. *Wykonanie procedury* ADD-CURVE-POINTS *wymaga* $O(\log^2 q)$ *operacji na bitach i* $O(\log q)$ *bitów pamięci.*

Lemat 4.3.26. *Wykonanie procedury* MULTIPLY-CURVE-POINTS *wymaga* $O(\log n \log^2 q)$ *operacji na bitach i* $O(\log n + \log q)$ *bitów pamięci.*

Dowód. Procedura ta realizuje schemat „podwajaj-i-dodawaj". Pojedynczy przebieg pętli **while** wymaga $O(\log^2 q)$ operacji na bitach, a przebiegów takich jest $O(\log n)$. Procedura potrzebuje dodatkowej pamięci na stałą ilość punktów krzywej i liczb całkowitych. □

Lemat 4.3.27. *Wykonanie procedury* RANDOM-CURVE-FINITE-POINT *wymaga średnio* $O(\log^3 q)$ *operacji na bitach i* $O(\log q)$ *bitów pamięci.*

Dowód. Procedura ta jest zrandomizowana. Obliczenie wartości d w jednym przebiegu pętli wymaga $O(\log^2 q)$ operacji na bitach. Wykonanie pierwiastkowania, aby obliczyć wartość b, wymaga średnio $O(\log^3 q)$ operacji na bitach. Tak więc pojedynczy przebieg pętli **while** wymaga średnio $O(\log^3 q)$ operacji na bitach.
Policzmy oczekiwaną liczbę przebiegów potrzebną do znalezienia punktu krzywej. Krzywa składa się z $q \pm O(\sqrt{q})$ punktów skończonych. Jeśli pominąć punkty rzędu dwa, to punkty skończone krzywej można połączyć w pary postaci punkt i punkt sprzężony. Z tego wniosek, że dla $\frac{q}{2} \pm O(\sqrt{q})$ elementów ciała $\mathbb{K}$ wartość d będzie kwadratem pewnego elementu ciała. Prawdopodobieństwo zakończenia się pętli w danym kroku wynosi zatem $\frac{1}{2} \pm O(\frac{1}{\sqrt{q}})$. Stąd procedura ta prawie na pewno kończy się, a oczekiwana liczba przebiegów pętli **while** wynosi 2, zatem oczekiwana liczba wymaganych operacji na bitach to średnio $O(\log^3 q)$. Ponadto, procedura potrzebuje stałej ilości zmiennych pomocniczych, w których przechowywane są elementy ciała, stąd wymaga $O(\log q)$ bitów pamięci. □

Przejdziemy teraz do obliczenia złożoności asymptotycznej algorytmu Millera.

Fakt 4.3.28. *Wykonanie procedury* COMBINE-PARTIAL-VALUES *wymaga* $O(\log^2 q)$ *operacji na bitach i* $O(\log q)$ *bitów pamięci.*

Twierdzenie 4.3.29. *Wykonanie procedury* COMPUTE-VALUE *wymaga* $O(\log n \log^2 q)$ *operacji na bitach i* $O(\log n + \log q)$ *bitów pamięci.*

Dowód. Obliczenia wstępne przed pętlą **while** mają złożoność czasową $O(\log^2 q)$. Każdy krok pętli **while** również ma złożoność czasową $O(\log^2 q)$. Procedura ta realizuje schemat „podwajaj-i-dodawaj", przebiegów pętli jest maksymalnie $O(\log n)$. Stąd łączna złożoność czasowa to $O(\log n \log^2 q)$. Procedura potrzebuje stałą ilość zmiennych pomocniczych, które przechowują liczby całkowite i elementy ciała, zatem potrzebuje $O(\log n + \log q)$ bitów pamięci. □

Wniosek 4.3.30. *Wykonanie jednego przebiegu pętli w procedurze* WEIL-PAIRING *wymaga* $O((\log q + \log n) \log^2 q)$ *operacji na bitach i* $O(\log q + \log n)$ *bitów pamięci.*

Uwaga 4.3.31. Oczekiwana liczba przebiegów pętli w procedurze WEIL-PAIRING zależy od relacji między wielkością parametrów q i n.
Jeżeli $q \geqslant n$, to oczekiwana liczba przebiegów pętli jest stała. Prawdopodobieństwo wystąpienia błędu w jednym przebiegu pętli wynosi $O(\frac{\log n}{q}) = O(\frac{\log q}{q}) < \frac{1}{2}$, zatem oczekiwana liczba przebiegów pętli jest stała.
Jeżeli $q < n$, to przeprowadzamy obliczenia w ciele $\mathbb{GF}(q')$, a nie w ciele $\mathbb{GF}(q)$. Wielkość q' należy dobrać tak, aby zachodziła zależność $q' \geqslant n$. Niech $q' = q^f$. Wystarczy teraz przyjąć, że $f = \lceil \log_2 n \rceil$. Wówczas $q' = q^f > 2^f = n$. Oczekiwana liczba przebiegów pętli jest wtedy stała, a jeden przebieg wymaga $O((\log q' + \log n) \log^2 q') = O(\log^3 q \log^3 n)$ operacji na bitach i $O(\log q' + \log n) = Q(\log q \log n)$ bitów pamięci.

4.4. Opis implementacji

Częścią niniejszej pracy jest implementacja opisanego w tym rozdziale algorytmu Millera, którą teraz krótko omówimy.

Język programowania

Do implementacji został użyty język Python [18]. Podsumujmy krótko jego najważniejsze cechy.

- Python ma charakterystyczną składnię. Najbardziej rzuca się w oczy sposób oznaczania bloków kodu (tzn. ciał pętli, treści procedur, gałęzi instrukcji warunkowych itd.) – blok kodu jest „wcięty", tzn. zawiera większą liczbę białych znaków na początkach linii niż kod otaczający.
- Python jest językiem interpretowanym, tzn. instrukcje pythonowe nie są kompilowane do języka niższego poziomu (np. do kodu maszynowego), lecz każdorazowo parsowane i wykonywane bezpośrednio przez „interpreter".
- Python pozwala stosować wiele paradygmatów programowania: imperatywny, obiektowy i funkcyjny (częściowo).
- Python jest językiem dynamicznie typowanym, tzn. wszelkiego rodzaju kontrola typów (np. sprawdzanie, że argumentami operacji dodawania są liczby) jest wykonywana w trakcie interpretowania programu.
- Python dostarcza mechanizm automatycznego zarządzania pamięcią oparty o zliczanie referencji.

- Python dysponuje niezmiernie bogatą biblioteką standardową, która oprócz usług dostępnych we wszystkich językach programowania (dostęp do plików, niskopoziomowy dostęp do sieci itp.) dostarcza mechanizmy służące do prowadzenia wysokopoziomowej komunikacji sieciowej (protokoły HTTP, SMTP, FTP itp.); obsługi plików skompresowanych (TAR, GZIP, BZIP2); tworzenia interfejsów graficznych (Tk); prowadzenia komunikacji międzyprocesowej; przetwarzania danych multimedialnych itd.

- Python może być uruchomiony na wielu systemach operacyjnych: wszystkich nowszych wersjach systemu Windows firmy Microsoft, praktycznie dowolnej dystrybucji systemu Linux, a także na wielu innych systemach uniksowych, w tym na Mac OS X firmy Apple.

- Nad rozwojem języka Python czuwa jego twórca Guido van Rossum oraz społeczność skupiona wokół fundacji Python Software Foundation, dzięki czemu proces wytwarzania kolejnych wersji języka jest otwarty, przejrzysty i dostępny dla osób postronnych.

- Python udostępniany jest na licencji Python License, która jest zgodna z wymogami OSI, jest zatem licencją wolną. Oznacza to, że z Pythona można korzystać bezpłatnie w praktycznie dowolnych zastosowaniach, w tym akademickich i komercyjnych.

Wybór języka Python do zaimplementowania algorytmu Millera podyktowany był następującymi względami:

- Składnia języka Python jest niezwykle podobna do zastosowanego w pracy sposobu zapisywania pseudokodu. Dzięki temu wszystkie przedstawione algorytmy można było bez wysiłku przetłumaczyć na składnię Pythona.

- Elementem biblioteki standardowej Pythona jest implementacja „długich" liczb całkowitych, tzn. liczb całkowitych o dowolnej ilości bitów. Dzięki temu implementacja procedur była prostsza.

- Dynamiczne typowanie, automatyczne zarządzanie pamięcią oraz brak konieczności kompilacji pozwalają na bardzo łatwe pisanie prototypów w Pythonie. Dzięki temu implementacja algorytmu Millera powstała bardzo szybko.

Struktura implementacji

Pełna komputerowa implementacja algorytmu Millera obejmuje nie tylko przedstawione w tym rozdziale algorytmy, ale także pewną ilość dodatkowych elementów. Całość można podzielić na „warstwy", czyli takie części, że każda kolejna jest zależna od poprzedniej.

1. Pierwszą warstwą jest implementacja liczb całkowitych dowolnej precyzji. Koncepcją implementacji jest system pozycyjny, w którym podstawę stanowi potęga liczby dwa, np. 2^{32} lub 2^{64}. Implementacji „długich" liczb całkowitych dostarcza biblioteka standardowa Pythona.

2. Drugą warstwą jest implementacja ciał skończonych. Operacje na elementach skończonych są zaimplementowane wyłącznie za pomocą operacji na „długich" liczbach całkowitych oraz standardowych konstrukcji językowych Pythona. Przyjęto obiektowy model reprezentacji danych. Warstwę tę można podzielić na mniejsze warstwy.

    - Implementacja pierścienia liczb całkowitych polega na opakowaniu dostępnych w Pythonie „długich" liczb całkowitych w postać obiektowej.

- Implementacja pierścienia ilorazowego nad zadanym pierścieniem bazowym polega na zaimplementowaniu w pierścieniu bazowym operacji dzielenia z resztą oraz rozszerzonego algorytmu Euklidesa. Operacje w pierścieniu ilorazowym polegają na wykonywaniu operacji w pierścieniu bazowym modulo zadany element. Odwrotności w pierścieniu ilorazowym obliczane są za pomocą algorytmu Euklidesa.
- Implementacja pierścienia wielomianów nad zadanym pierścieniem bazowym polega na reprezentowaniu wielomianów za pomocą tablic elementów pierścienia bazowego.

Ciało skończone $\mathbb{F}(p)$ możemy przedstawić jako pierścień ilorazowy pierścienia liczb całkowitych podzielonego przez p. Ciało skończone $\mathbb{GF}(q)$, gdzie $q = p^e$, przedstawiamy jako pierścień ilorazowy pierścienia wielomianów nad ciałem $\mathbb{F}(p)$ podzielony przez wielomian stopnia e nierozkładalny nad ciałem $\mathbb{F}(p)$.

3. Trzecią warstwę stanowi implementacja prostego modelu obiektowego reprezentującego krzywe eliptyczne oraz punkty i linie na krzywych. Warstwa ta dostarcza jedynie reprezentację danych, nie udostępnia natomiast żadnych operacji na tych danych.

4. Czwartą i ostatnią warstwę stanowi implementacja procedur przedstawionych w tym rozdziale. Tę warstwę również można podzielić na części.

   - Implementacja akcesorów (np. $x[P]$, $identity[E]$, $a[l]$) polega ona na odpowiednim odwoływaniu się do modelu obiektowego z warstwy trzeciej.
   - Implementacja procedur uznanych za dane, w tym konstruktorów (np. Curve-Finite-Point, Line-On-Curve), które korzystają z modelu obiektowego z warstwy trzeciej oraz algorytmów (np. Finite-Field-Element-Square-Root).
   - Implementacja ogólnych procedur przedstawionych w tym rozdziale (np. Line-Through-Curve-Points, Add-Curve-Points).
   - Implementacja algorytmu Millera, w tym procedury Weil-Pairing.

Wydajność

Wydajność otrzymanej implementacji algorytmu Millera jest niska. Fakt ten jest wywołany kilkoma czynnikami.
Po pierwsze, zastosowany język Python oprócz wszystkich wymienionych zalet ma też jedną zasadniczą wadę – nie jest szybki. Przyczyną tego stanu rzeczy jest przede wszystkim to, że jest to język interpretowany. W zależności od zastosowań może on być kilka do kilkudziesięciu razy wolniejszy od języków kompilowanych, np. od C++.
Po drugie, sposób zastosowany przy implementacji ciał skończonych nie jest wydajny. Przedstawienie ciał skończonych jako pierścieni ilorazowych powoduje, że wykonanie pojedynczej operacji arytmetycznej w ciele skończonym przekłada się na wiele operacji w pierścieniach bazowych, a w konsekwencji na jeszcze więcej operacji na liczbach całkowitych. Mimo niskiej wydajności ten sposób reprezentacji został zastosowany, ponieważ jest przejrzysty i ogólny.
Po trzecie, sam algorytm Millera w postaci podanej w niniejszej pracy charakteryzuje się dużą stałą w złożoności asymptotycznej. Uważna analiza pozwala wyeliminować losowość z algorytmu (por. [5]).
W związku z powyższym należy uznać, że stanowiąca część niniejszej pracy implementacja algorytmu Millera ma charakter poglądowy i nie nadaje się do zastosowań praktycznych.

System Sage

Narzędziem, które okazało się niezwykle pomocne podczas implementacji, jest system Sage [20]. Jest to wolny odpowiednik systemów takich jak Mathematica czy Maple. Oto niektóre z funkcji, które udostępnia:

- przeprowadzanie obliczeń symbolicznych;
- symboliczne i numeryczne rozwiązywanie różnego rodzaju równań i układów równań;
- wykonywanie operacji arytmetycznych na elementach różnych obiektów algebraicznych – ciał, pierścieni itp.;
- implementacja wielu algorytmów teorioliczbowych i kryptograficznych.

Rozdział 5

Zastosowania iloczynu Weila

Iloczyn Weila jest narzędziem używanym w kryptografii wtedy, gdy rozpatrywane są zagadnienia związane z krzywymi eliptycznymi. Jak każde narzędzie, można go wykorzystać na dwa sposoby: sposób „dobry" polega na tworzeniu kryptosystemów z pomocą iloczynu Weila, zaś sposób „zły" polega na przeprowadzaniu ataków na nie.
W rozdziale tym przedstawimy zastosowania iloczynu Weila obu rodzajów. Omówimy jeden atak kryptograficzny oraz grupę kryptosystemów, które mają pewną specyficzną cechę wspólną.

5.1. Superosobliwe krzywe eliptyczne

Opiszemy najpierw pewien szczególny rodzaj krzywych eliptycznych, które dobrze nadają do się zastosowań związanych z iloczynem Weila.

Definicja 5.1.1. Dana jest liczba pierwsza p taka, że $p \equiv 2 \pmod 3$. *Superosobliwa krzywa eliptyczna nad ciałem* $\mathbb{F}(p)$ to krzywa $E_{0,1}(\mathbb{F}(p))$.

Uwaga 5.1.2. Krzywe superosobliwe definiuje się tak naprawdę w zupełnie inny, bardziej ogólny sposób. Podana definicja to jedynie przykład pewnej rodziny krzywych, które są superosobliwe w sensie ogólnej definicji. Inną rodziną krzywych superosobliwych są krzywe postaci $E_{1,0}(\mathbb{F}(p))$, gdzie $p \equiv 3 \pmod 4$.

Pierwszą interesującą własnością krzywych superosobliwych jest ich struktura grupowa.

Twierdzenie 5.1.3. *Krzywa superosobliwa* $E_{0,1}(\mathbb{F}(p))$ *składa się z* $p+1$ *punktów.*

Dowód. Rozważmy równanie $x^3 = c$ w ciele $\mathbb{F}(p)$. Weźmy dowolne dwa elementy $a, b \in \mathbb{F}(p)$ takie, że $a^3 = c$ i $b^3 = c$. Gdy $c = 0$, to $a = b = 0$. Gdy $c \neq 0$, zapisujemy elementy a i b jako potęgi generatora: $a = g^k$, $b = g^l$. Wiemy, że $a^3 = b^3$, skąd dostajemy $g^{3(k-l)} = 1$. Zatem $p - 1 \mid 3(k - l)$. Ponieważ liczba $p - 1$ jest względnie pierwsza z 3, otrzymujemy $p - 1 \mid k - l$. Stąd $a = b$.
Z rozważań tych wynika następujący wniosek: funkcja $x^3 + 1$ jest bijekcją w zbiorze $\mathbb{F}(p)$, w jej obrazie znajdują się wszystkie reszty kwadratowe, zatem krzywa superosobliwa $E_{0,1}(\mathbb{F}(p))$ zawiera dokładnie $p + 1$ punktów. □

Twierdzenie 5.1.4. *Grupa na krzywej superosobliwej* $E_{0,1}(\mathbb{F}(p))$ *jest cykliczna.*

Dowód. Zgodnie z twierdzeniem 2.4.4 krzywa E jest izomorficzna z grupą $(\mathbb{Z}/k\mathbb{Z}) \times (\mathbb{Z}/l\mathbb{Z})$, gdzie $kl = p + 1$ oraz $l \mid \gcd(k, p - 1)$. Skoro $k \mid p + 1$, to $l \mid \gcd(p + 1, p - 1)$, zatem $l = 1$

lub $l = 2$. W pierwszym przypadku twierdzenie jest udowodnione. Pokażemy, że drugi jest niemożliwy.
Przyjmijmy, że $l = 2$. Jeżeli $2 \nmid k$, to grupa $(\mathbb{Z}/k\mathbb{Z}) \times (\mathbb{Z}/l\mathbb{Z})$ jest izomorficzna z żądaną grupą $\mathbb{Z}/(p+1)\mathbb{Z}$. Natomiast jeśli $2 \mid k$, to na krzywej E są dwa punkty rzędu dwa, a więc wielomian $x^3 + 1$ ma w ciele $\mathbb{F}(p)$ dwa miejsca zerowe. To jest niemożliwe, bo wielomian ten jest bijekcją w zbiorze $\mathbb{F}(p)$, jak zauważyliśmy wcześniej. □

W kontekście iloczynu Weila krzywe superosobliwe są ważne dlatego, że łatwo wskazywać ich podgrupy n-torsyjne.

Twierdzenie 5.1.5. *Dana jest liczba pierwsza p taka, że $p \equiv 2 \pmod 3$ oraz liczba naturalna n będąca dzielnikiem liczby $p + 1$. Wówczas podgrupa n-torsyjna krzywej eliptycznej $E_{0,1}(\mathbb{GF}(p^2))$ jest izomorficzna z grupą $(\mathbb{Z}/n\mathbb{Z}) \times (\mathbb{Z}/n\mathbb{Z})$.*

Dowód. Ponieważ $3 \mid p^2 - 1$, w ciele $\mathbb{GF}(p^2)$ istnieje nietrywialny pierwiastek trzeciego stopnia z jedności. Oznaczmy go przez ξ. Wówczas automorfizmem krzywej $E_{0,1}(\mathbb{GF}(p^2))$ jest odwzorowanie $\phi\colon E_{0,1}(\mathbb{F}(p^2)) \to E_{0,1}(\mathbb{GF}(p^2))$ określone następująco:

$$\phi(x, y) = (\xi x, y) \tag{5.1.6}$$

Niech P będzie takim punktem krzywej $E_{0,1}(\mathbb{F}(p))$, że $nP = \mathcal{O}$. Punkt $\phi(P)$ również ma rząd n. Zauważmy przy tym, że $\phi(P) \notin E_{0,1}(\mathbb{F}(p))$, bo $\xi \notin \mathbb{F}(p)$. Wynika stąd już, że podgrupa n-torsyjna krzywej $E_{0,1}(\mathbb{GF}(p^2))$ ma żądaną postać, a jej generatorami są punkty P i $\phi(P)$. □

5.2. Redukcja MOV

Zagadnieniem, od którego rozpoczniemy omawianie zastosowań iloczynu Weila, jest twierdzenie udowodnione przez Menezesa, Okamoto i Vanstone'a [6], zwane w skrócie „redukcją MOV".
Redukcja MOV daje podstawy teoretyczne do przeprowadzenia pewnego specyficznego ataku na kryptosystemy oparte na krzywych eliptycznych. Dzięki algorytmowi Millera teoria może zmienić się w praktykę, dlatego odkrycie redukcji MOV spowodowało, że niebezpieczne stało się wykorzystywanie systemów kryptograficznych opartych na superosobliwych krzywych eliptycznych. Jest to o tyle niefortunne, że krzywe superosobliwe wyjątkowo dobrze nadają się do realizacji komputerowej i wiązano z nimi nadzieje stworzenie praktycznych, wydajnych kryptosystemów.
Redukcja MOV związana jest z problemem logarytmu dyskretnego oraz protokołem Diffiego-Hellmana, dlatego w pierwszej kolejności podamy definicje obu problemów.

Problem 5.2.1 (Logarytm dyskretny)**.** Dana jest grupa cykliczna $\mathbb{G}$, jej generator g oraz jej element a. Znaleźć liczbę całkowitą $k \in \mathbb{Z}$ taką, że:

$$g^k = a \tag{5.2.2}$$

Uwaga 5.2.3. Zwyczajowo problem logarytmu dyskretnego opisuje się, stosując zapis multiplikatywny działania grupowego. W przypadku zapisu addytywnego równanie 5.2.2 przybiera następującą postać:

$$kg = a \tag{5.2.4}$$

Uwaga 5.2.5. Problem logarytmu dyskretnego można rozpatrywać w przypadku dowolnej grupy. Należy wówczas rozpatrywać jej podgrupę (g) generowaną przez pewien jej element g.

Uwaga 5.2.6. Trudność problemu logarytmu dyskretnego w danej grupie zależy od sposobu reprezentacji elementów tej grupy za pomocą ciągów bitów. I tak na przykład problem logarytmu dyskretnego w grupie $\mathbb{Z}/(p-1)\mathbb{Z}$ jest łatwy, zaś w grupie multiplikatywnej ciała $\mathbb{F}(p)$, która jest przecież izomorficzna z grupą $\mathbb{Z}/(p-1)\mathbb{Z}$, jest obecnie uznawany za trudny.

Problemem spokrewnionym z problemem logarytmu dyskretnego jest zagadnienie złamania protokołu Diffiego-Hellmana, zwane w skrócie „problemem Diffiego-Hellmana".

Problem 5.2.7 (Protokół Diffiego-Hellmana)**.** Dana jest grupa cykliczna $\mathbb{G}$, jej generator g oraz elementy a i b. Niech k i l będą takimi liczbami całkowitymi, że $a = g^k$ i $b = g^l$. Znaleźć (znając tylko wartości g, a i b) element $c \in \mathbb{G}$ taki, że:

$$c = g^{kl} \tag{5.2.8}$$

Uwaga 5.2.9. Podobnie jak w przypadku problemu logarytmu dyskretnego, możemy stosować zapis addytywny (szukamy wówczas wartości klg) oraz rozpatrywać problem w dowolnej grupie.

Pokrewieństwo obu problemów polega na tym, że jeden można zredukować do drugiego.

Twierdzenie 5.2.10. *Problem Diffiego-Hellmana w grupie $\mathbb{G}$ można w czasie wielomianowym w sposób deterministyczny zredukować do problemu logarytmu dyskretnego w grupie $\mathbb{G}$.*

Dowód. Redukcja jest bardzo prosta. Jeżeli dysponujemy algorytmem rozwiązującym problem logarytmu dyskretnego w grupie $\mathbb{G}$, to postępujemy następująco:

1. na podstawie wartości $\mathbb{G}$, g i a obliczamy wartość k;
2. na podstawie wartości $\mathbb{G}$, g i b obliczamy wartość l;
3. obliczamy kl;
4. obliczamy g^{kl}.

Jak widać, aby rozwiązać egzemplarz problemu Diffiego-Hellmana wystarczy dwa razy zastosować rozwiązanie problemu logarytmu dyskretnego. □

Czy problem Diffiego-Hellmana można rozwiązać inaczej? Zagadnienie to jest o tyle istotne, że na trudności problemu Diffiego-Hellmana opiera się wiele kryptosystemów, których implementacje są wykorzystywane na co dzień, m.in. w protokole SSL używanym w sieci Internet. Wydaje się, że oba problemy są równoważne (por. [10]), a to oznacza trudność problemu Diffiego-Hellmana, a zatem bezpieczeństwo używanych kryptosystemów.
Oba podane problemy zostały przedstawione w wersji obliczeniowej. W teorii złożoności obliczeniowej często rozpatruje się wersje decyzyjne problemów. Problem Diffiego-Hellmana przybiera wówczas następującą postać.

Problem 5.2.11. Dana jest grupa $\mathbb{G}$, jej generator g oraz jej elementy a, b i c. Niech k, l i m będą takimi liczbami całkowitymi, że $a = g^k$, $b = g^l$ i $c = g^m$. Stwierdzić (znając tylko wartości g, a, b i c), czy zachodzi następująca zależność:

$$kl \equiv m \pmod{|\mathbb{G}|} \tag{5.2.12}$$

W przypadku krzywych eliptycznych obliczeniowa i decyzyjna wersja problemu Diffiego-Hellmana przybierają następującą postać.

Problem 5.2.13. Dana jest krzywa eliptyczna E nad ciałem skończonym, punkt G rzędu n na krzywej E oraz punkty P i Q będące wielokrotnościami punktu G. Niech k i l będą takimi liczbami całkowitymi, że $kG = P$ i $lG = Q$. Znaleźć (znając tylko wartości G, P i Q) punkt R na krzywej E taki, że:

$$R = klG \tag{5.2.14}$$

Problem 5.2.15. Dana jest krzywa eliptyczna E nad ciałem skończonym, punkt G rzędu n na krzywej E oraz punkty P, Q i R będące wielokrotnościami punktu G. Niech k, l i m będą takimi liczbami całkowitymi, że $kG = P$, $lG = Q$ i $mG = R$. Stwierdzić (znając tylko wartości G, P, Q i R), czy zachodzi następująca zależność:

$$kl \equiv m \pmod{n} \tag{5.2.16}$$

Czy problemy Diffiego-Hellmana na krzywej eliptycznej są trudne? Okazuje się, że iloczyn Weila ma wpływ na tę kwestię.

Twierdzenie 5.2.17 (Redukcja MOV)**.** *Obliczeniowy problem logarytmu dyskretnego na krzywej eliptycznej $E(\mathbb{GF}(p^e))$ można w czasie wielomianowym w sposób deterministyczny zredukować do obliczeniowego problemu logarytmu dyskretnego w grupie multiplikatywnej ciała $\mathbb{GF}(p^{ef})$.*

Dowód. Niech punkt G rzędu n na krzywej E oraz punkt P będący wielokrotnością punktu G będą instancją obliczeniowego problemu logarytmu dyskretnego na krzywej E. Rozważmy rozszerzenie $\mathbb{GF}(p^{ef})$ ciała $\mathbb{GF}(p^e)$ dostatecznie duże, aby istniał punkt $H \in E(\mathbb{GF}(p^{ef})$ taki, że $\mathrm{ord}(H) = n$ oraz $H \notin (G)$. Jest to możliwe, gdy $p \nmid n$. Wartość $w(G, H)$ jest wówczas pierwiastkiem pierwotnym n-tego stopnia z jedności. Oznaczmy $\mu = w(G, H)$, $P = kG$ i policzmy wartość $w(P, H)$:

$$\begin{aligned} w(P, H) &= w(kG, H) \\ &= w(G, H)^k \\ &= \mu^k \end{aligned}$$

Elementy μ i $w(P, H)$ stanowią zatem egzemplarz problemu logarytmu dyskretnego w grupie multiplikatywnej ciała $\mathbb{GF}(p^{ef})$, który ma takie samo rozwiązanie, jak egzemplarz pierwotnego problemu. □

Wniosek 5.2.18. *Obliczeniowy problem Diffiego-Hellmana na krzywej eliptycznej $E(\mathbb{GF}(p^e))$ jest nietrudniejszy od obliczeniowego problemu logarytmu dyskretnego w grupie multiplikatywnej ciała $\mathbb{GF}(p^{ef})$.*

Twierdzenie 5.2.19. *Decyzyjny problem Diffiego-Hellmana na krzywej superosobliwej $E_{0,1}(\mathbb{F}(p))$ jest łatwy.*

Dowód. Niech punkt G rzędu n na krzywej E oraz punkty P, Q i R będące wielokrotnościami punktu G będą instancją decyzyjnego problemu Diffiego-Hellmana na krzywej E. Oznaczmy $P = kG$, $Q = lG$ i $R = mG$. Oznaczmy $H = \phi(G)$, gdzie ϕ jest automorfizmem określonym równaniem 5.1.6. Punkty G i H są generatorami grupy $E[n]$, zatem wartość $w(G, H)$ jest pierwiastkiem pierwotnym n-tego stopnia z jedności. Ponadto, zachodzą następujące zależności:

$$\begin{aligned} w(P, f(Q)) &= w(G, H)^{kl} \\ w(R, f(G)) &= w(G, H)^m \end{aligned}$$

Zależność $kl \equiv m \pmod{n}$ zachodzi wtedy i tylko wtedy, gdy $w(P, f(Q)) = w(R, f(G))$, bo wartość $w(G, H)$ jest pierwiastkiem pierwotnym. □

Uwaga 5.2.20. Kluczową rolę w dowodzie odgrywa fakt, że istnieje nietrywialny, łatwo obliczalny automorfizm grupy $E[n]$. Rozumowanie to można uogólnić na wszystkie krzywe, dla których jesteśmy w stanie wskazać analogiczny automorfizm.

5.3. Szyfrowanie oparte na tożsamości

Opiszemy teraz system szyfrujący opracowany przez Boneha i Franklina [7]. Iloczyn Weila odgrywa w tym systemie kluczową rolę.
System pozwala szyfrować wiadomości w sposób asymetryczny, tzn. występują w nim klucze publiczne do szyfrowania wiadomości oraz klucze prywatne do ich odczytywania. W skład systemu wchodzą cztery algorytmy, których przeznaczenie jest następujące.

- Algorytm tworzenia parametrów systemu jest wykorzystywany przez zarządcę systemu raz, podczas tworzenia instancji systemu. Wynikiem jego działania są parametry systemu, które zarządca powinien udostępnić wszystkim użytkownikom chcącym korzystać z systemu oraz klucz główny, którego nie powinien ujawniać.

- Algorytm tworzenia klucza prywatnego jest wykorzystywany przez zarządcę wtedy, gdy nowy użytkownik chce rozpocząć korzystanie z systemu. Za pomocą tego algorytmu zarządca tworzy dla nowego użytkownika klucz prywatny, którego użytkownik nie powinien ujawniać. Klucz prywatny jest wyznaczany jednoznacznie na podstawie parametrów systemu, klucza głównego i dowolnie wybranego ciągu bitów. Ciąg ten będzie identyfikatorem nowego użytkownika, który inni użytkownicy będą wykorzystywać, aby wysyłać do niego zaszyfrowane wiadomości.

- Wiadomości są szyfrowane za pomocą algorytmu wykorzystującego parametry systemu i identyfikator odbiorcy. Dodatkowo, w procesie szyfrowania używany jest losowo wybrany parametr, dlatego w wyniku wielokrotnego zaszyfrowania tej samej wiadomości można otrzymać różne kryptogramy.

- Kryptogramy odczytywane są za pomocą algorytmu wykorzystującego parametry systemu i klucz prywatny odbiorcy.

Jak widać, kluczową cechą odróżniającą ten system od innych rozwiązań jest możliwość użycia dowolnego ciągu bitów jako klucza publicznego. O systemie mającym tę cechę mówimy, że jest „oparty na tożsamości". Takie podejście, jak zaraz zobaczymy, oferuje możliwości, których nie dają inne kryptosystemy.

Motywacja

Przedstawiamy dwa typy problemów, których rozwiązanie za pomocą kryptosystemów nieopierających się na tożsamości jest albo niemożliwe, albo bardzo trudne. Natomiast łatwo jest rozwiązać je za pomocą systemu Boneha-Franklina.

- Dana jest grupa użytkowników (np. instytucja państwowa lub korporacja), którzy chcą przesyłać między sobą zaszyfrowane wiadomości. Zastosowanie w tym celu systemu Boneha-Franklina daje następujące korzyści.

  – Użytkownicy nie muszą przechowywać kluczy publicznych innych użytkowników. Ułatwia to nawiązanie korespondencji z nową osobą, bo łatwiej jest przekazać

innym kanałem (np. za pomocą wizytówki) swój identyfikator (np. adres poczty elektronicznej) niż swój klucz publiczny. Dalej, nie ma ryzyka utracenia listy kluczy publicznych innych użytkowników. Listę taką zawsze można odtworzyć, ale może być to bardziej żmudne niż odtworzenie listy adresów poczty elektronicznej.

- Użytkownicy nie muszą przechowywać nawet swojego klucza prywatnego – zawsze można ponownie poprosić zarządcę systemu o jego wygenerowanie (uprzednio potwierdzając swoją tożsamość innymi metodami). Utrata klucza prywatnego w innych systemach może spowodować nieodwracalne konsekwencje.
- Można wysyłać wiadomości do przyszłych członków grupy, dla których nie został jeszcze utworzony klucz prywatny. Wystarczy, że znany jest identyfikator, którym będą się posługiwać.
- Łatwo jest wykluczać użytkowników z grupy. Wystarczy, że kluczem publicznym będzie ciąg bitów będący wynikiem połączenia identyfikatora odbiorcy z rokiem (odpowiednio, rokiem i miesiącem) wysłania wiadomości. W ten sposób każdy członek grupy musi co rok (odpowiednio, co miesiąc) poprosić zarządcę o wygenerowanie nowego klucza prywatnego, który będzie ważny przez kolejny rok (odpowiednio, miesiąc).
- Można wysyłać wiadomości, które będą mogły być odczytane dopiero w przyszłości. Wystarczy zmodyfikować poprzedni pomysł: zamiast bieżącej daty można do identyfikatora odbiorcy dokleić datę przyszłą, która określa, kiedy wiadomość będzie mogła być odczytana.
- Członkom grupy można nadawać poziomy uprawnień do odczytywania wiadomości. W tym celu kluczem publicznym powinien być ciąg bitów składający się z identyfikatora odbiorcy, daty oraz nazwy uprawnienia, które jest wymagane do odczytu nadawanej wiadomości. Jest to uogólnienie poprzednich dwóch pomysłów.

Ponieważ nie trzeba przechowywać kluczy publicznych, szyfrowanie elektronicznej korespondencji staje się bardziej dostępne dla przeciętnego użytkownika poczty elektronicznej. Zauważmy przy tym, że nierozwiązany pozostaje problem przesyłania zaszyfrowanych wiadomości między członkami różnych grup, np. między użytkownikami dwóch różnych serwerów poczty elektronicznej. Dlatego też nie można do końca zrezygnować z infrastruktury klucza publicznego.

Zauważmy, że zastosowanie systemu Boneha i Franklina w opisany sposób prowadzi do następującego problemu: zarządca systemu może odczytać dowolną zaszyfrowaną wiadomość. Z kryptograficznego punktu widzenia jest to zjawisko zdecydowanie niepożądane, jednak obecnie praktykuje się właśnie takie rozwiązania: administrator serwera poczty elektronicznej ma dostęp do wszystkich wiadomości.

- Przypuśćmy, że pewien użytkownik systemu opartego na infrastrukturze klucza publicznego chce zabezpieczyć się przed ujawnieniem swojego klucza prywatnego. Sytuacja taka może mieć miejsce, jeżeli będzie przechowywał klucz prywatny na niezaufanych urządzeniach (np. laptopie lub telefonie komórkowym, które mogą zostać skradzione) lub powierzy go niezaufanym osobom (np. swoim asystentom, których zadaniem jest pomoc w prowadzeniu korespondencji). W tym celu może wykorzystać system Bonehа-Franklina. Powinien utworzyć własną instancję systemu i poprosić wszystkich, którzy wysyłają do niego wiadomości, żeby szyfrowali je kluczem, który powstaje z połączenia daty wysłania wiadomości z kategorią tematyczną wiadomości. Opisane problemy można wówczas rozwiązać następująco.

– Użytkownik ma dostęp do całej swojej korespondencji, ponieważ jest w posiadaniu klucza głównego i może z jego pomocą odszyfrować dowolną wiadomość.

– Swoim asystentom może wydać klucze prywatne, które pozwalają na odczytanie wiadomości jedynie z tej kategorii tematycznej, za którą są odpowiedzialni.

– Na zagrożone kradzieżą urządzenie użytkownik może nagrać klucze prywatne pozwalające na odczytywanie wiadomości wysłanych tylko w zadanym okresie czasu (np. jeden tydzień). Jeżeli urządzenie zostanie skradzione, złodziej nie będzie w stanie odczytać wiadomości wysłanych po upływie zadanego okresu.

Szczegółowy opis działania systemu

Przedstawimy teraz działanie wszystkich algorytmów wchodzących w skład systemu Boneha-Franklina. Opiszemy również rodzaje danych pojawiających się w systemie.

Algorytm 5.3.1 (Tworzenie parametrów systemu)**.** Następujący algorytm wybiera i przekazuje jako wynik parametry instancji systemu Boneha-Franklina i jej klucz główny.

BF-Create-Instance()

1 Wybieramy liczbę pierwszą q.
2 Wybieramy grupy cykliczne $\mathbb{G}_1$ i $\mathbb{G}_2$ rzędu q.
3 Wybieramy odwzorowanie dwuliniowe $b\colon \mathbb{G}_1 \times \mathbb{G}_1 \to G_2$.
4 Wybieramy liczbę naturalną m.
5 Ustalamy, że przestrzenią wiadomości $\mathcal{M}$ jest zbiór $\{0,1\}^m$.
6 Ustalamy, że przestrzenią kryptogramów $\mathcal{C}$ jest zbiór $\mathbb{G}_1^\star \times \mathcal{M}$.
7 Ustalamy, że przestrzenią identyfikatorów użytkowników $\mathcal{I}$ jest zbiór $\{0,1\}^\star$.
8 Wybieramy funkcje haszujące $H_1\colon \mathcal{I} \to \mathbb{G}_1^\star$ i $H_2\colon \mathbb{G}_2 \to \mathcal{M}$.
9 Wybieramy generator P grupy $\mathbb{G}_1$.
10 Wybieramy element s z grupy $(\mathbb{Z}/q\mathbb{Z})^\star$.
11 Obliczamy wartość $Q = sP$.
12 Parametry systemu to krotka $\langle q, \mathbb{G}_1, \mathbb{G}_2, b, m, H_1, H_2, P, Q\rangle$.
13 Klucz główny to element s.

Algorytm 5.3.2 (Tworzenie klucza prywatnego)**.** Dany jest identyfikator ID. Następujący algorytm na podstawie wartości ID oraz klucza głównego s instancji systemu Boneha-Franklina i jej parametrów oblicza i przekazuje jako wynik klucz prywatny odpowiadający identyfikatorowi ID.

BF-Create-Private-Key(ID, s, $\langle q, \mathbb{G}_1, \mathbb{G}_2, b, m, H_1, H_2, P, Q\rangle$)

1 Obliczamy wartość $R = H_1(\text{ID})$.
2 Obliczamy wartość $S = sR$.
3 Klucz prywatny to punkt S.

Algorytm 5.3.3 (Szyfrowanie wiadomości)**.** Dana jest wiadomość M i identyfikator adresata ID. Następujący algorytm na podstawie wartości M i ID oraz parametrów instancji systemu Boneha-Franklina oblicza i przekazuje jako wynik kryptogram odpowiadający wiadomości M.

BF-ENCRYPT$(M, \text{ID}, \langle q, \mathbb{G}_1, \mathbb{G}_2, b, m, H_1, H_2, P, Q\rangle)$

1 Obliczamy wartość $R = H_1(\text{ID})$.
2 Wybieramy element r z grupy $(\mathbb{Z}/q\mathbb{Z})^\star$.
3 Obliczamy wartość $U = rP$.
4 Obliczamy wartość $V = M$ **xor** $b(R, Q)^r$.
5 Kryptogram to para $\langle U, V\rangle$.

Algorytm 5.3.4 (Odczytywanie kryptogramu)**.** Dany jest kryptogram C w postaci $C = \langle U, V\rangle$ i klucz prywatny S odpowiadający identyfikatorowi ID. Następujący algorytm na podstawie wartości C i S oraz parametrów instancji systemu Boneha-Franklina oblicza i przekazuje jako wynik wiadomość odpowiadającą kryptogramowi C.

BF-DECRYPT$(\langle U, V\rangle, S, \langle q, \mathbb{G}_1, \mathbb{G}_2, b, m, H_1, H_2, P, Q\rangle)$

1 Obliczamy wartość $M = V$ **xor** $b(S, U)$.
2 Wiadomość to wartość M.

Uwaga 5.3.5. Zauważmy, że nie sprecyzowaliśmy, jaka jest struktura grup $\mathbb{G}_1$ i $\mathbb{G}_2$. Kwestię tę omówimy za chwilę.

Pokażmy przede wszystkim, że w systemie Boneha-Franklina prawidłowo działa szyfrowanie i odczytywanie wiadomości.

Twierdzenie 5.3.6. *Dana jest wiadomość M, identyfikator adresata* ID*, odpowiadający mu klucz prywatny S, i parametry $\mathcal{P}$ instancji systemu Boneha-Franklina. Wówczas zachodzi następująca zależność:*

$$\text{BF-DECRYPT}(\text{BF-ENCRYPT}(M, \text{ID}, \mathcal{P}), S, \mathcal{P}) = M$$

Dowód. Wiadomość M jest szyfrowana symetrycznie za pomocą operacji **xor**. Wystarczy zatem sprawdzić, że klucze użyte przy szyfrowaniu i odczytywaniu wiadomości są takie same. Istotnie:

$$\begin{aligned} b(S, U) &= b(sR, U) \\ &= b(R, U)^s \\ &= b(R, rP)^s \\ &= b(R, P)^{sr} \\ &= b(R, sP)^r \\ &= b(R, Q)^r \end{aligned}$$

□

Sednem sposobu szyfrowania w systemie Boneha-Franklina jest wykorzystanie prostego algorytmu symetrycznego do szyfrowania wiadomości, stosowanie w tym algorytmie kluczy jednorazowych oraz sprytny sposób przekazania informacji o użytym kluczu. Sekretną informacją potrzebną do odczytania wiadomości jest losowa wartość r, którą nadawca „rozdziela" na dwie części. Jedynie odbiorca jest w stanie połączyć je z powrotem w całość i odczytać wiadomość. Złączenie to odbywa się za pomocą operacji dwuliniowej b. Sposób, w jaki sekret jest rozdzielony i przekazany w częściach, koncepcyjnie przypomina działanie protokołu Diffiego-Hellmana.
Konkretna realizacja systemu Boneha-Franklina wymaga wybrania konkretnych grup $\mathbb{G}_1$ i $\mathbb{G}_2$ oraz odwzorowania dwuliniowego b. W swojej pracy Boneh i Franklin opisują następującą realizację systemu opartą na superosobliwych krzywych eliptycznych.

- Podstawą systemu jest superosobliwa krzywa eliptyczna $E_{0,1}(\mathbb{F}(p))$, przy czym $q \mid p+1$. Zauważmy, że w tej sytuacji $E[q] \subset E_{0,1}(\mathbb{GF}(p^2))$.
- Rolę grupy $\mathbb{G}_1$ pełni grupa generowana przez punkt P rzędu q na krzywej E.
- Grupa $\mathbb{G}_2$ to grupa pierwiastków q-tego stopnia z jedności w ciele $\mathbb{GF}(p^2)$.
- Niech ϕ będzie funkcją określoną równaniem 5.1.6. Odwzorowanie dwuliniowe b używane w systemie jest określone następująco:

$$b(P,Q) = w(P, \phi(Q)) \tag{5.3.7}$$

Uwaga 5.3.8. Zauważmy, że iloczyn Weila jest zdegenerowany na krzywej $E_{0,1}(\mathbb{F}(p))$, dlatego w definicji funkcji b jeden z argumentów jest zmodyfikowany za pomocą automorfizmu ϕ. Można sprawdzić, że uzyskane w ten sposób odwzorowanie jest dwuliniowe.

W pracy Boneha i Franklina można znaleźć analizę bezpieczeństwa systemu. Zaznaczmy, że aby uzyskać gwarancje bezpieczeństwa na tyle rozsądne, żeby można było korzystać z systemu w praktyce, należy zastosować zmodyfikowaną wersję opisanego tutaj schematu. Opis tej modyfikacji można znaleźć w pracy [9].

5.4. Podpisy cyfrowe oparte na tożsamości

Zagadnieniem blisko spokrewnionym z szyfrowaniem są podpisy cyfrowe. Szyfrowanie gwarantuje, że wiadomość elektroniczną odczyta tylko wybrany przez nadawcę odbiorca, zaś podpisy cyfrowe pozwalają odbiorcy stwierdzić, czy wiadomość jest autentyczna.
W przypadku niektórych kryptosystemów szyfrowanie i wystawianie podpisów cyfrowych jest tak samo skomplikowane. Przykładowo, w systemie RSA operacje szyfrowania i odczytywania wiadomości są przemienne, dzięki czemu podpis cyfrowy można wystawić poprzez zaszyfrowanie wiadomości kluczem prywatnym, a nie publicznym.
W przypadku systemów opartych na tożsamości, które wykorzystują odwzorowania dwuliniowe, sytuacja nie jest aż taka prosta. System Boneha i Franklina jest swego rodzaju kamieniem milowym w swojej dziedzinie – trudno wskazać bardziej popularny system szyfrujący oparty na tożsamości. Niestety, system ten nie pozwala na wystawianie podpisów cyfrowych.
Udało się opracować wiele różnych rozwiązań korzystających z odwzorowań dwuliniowych, dzięki którym można wystawiać podpisy cyfrowe oparte na tożsamości, jednak o żadnym z nich nie można powiedzieć, że jest ono tak istotne dla kryptografii, jak system Boneha i Franklina.
Omówimy teraz jedno z takich rozwiązań opracowane przez Yi [8]. Wybór tego konkretnego rozwiązania na obiekt naszych rozważań jest podyktowany tym, że rozwiązanie to jest najbliższe systemowi Boneha-Franklina. Dzięki występowaniu wielu elementów wspólnych oba systemy można potraktować niemalże jak części jednego większego systemu.

Motywacja

Podpis cyfrowy oparty na tożsamości można zweryfikować bez znajomości klucza publicznego domniemanego nadawcy. Bardzo często tożsamość domniemanego nadawcy jest wskazana w treści weryfikowanej wiadomości. Dzięki temu weryfikowanie podpisów jest bardzo łatwe.
Rozwiązanie takie może być przydatne w dużych instytucjach, w których w obiegu pozostaje wiele dokumentów. Często jest tak, że dokumenty krążą między osobami, które nie miały wcześniej ze sobą styczności i nie wymieniły się swoimi kluczami publicznymi. Systemy oparte

na infrastrukturze klucza publicznego prowadzą centralny rejestr kluczy publicznych wszystkich członków organizacji. W rozwiązaniu korzystającym z podpisów opartych na tożsamości nie występuje punkt centralny, zatem jest ono bardziej wydajne i skalowalne.
Systemy podpisów cyfrowych opartych na tożsamości mogą też wpłynąć na popularyzację kryptografii w systemach światowej poczty elektronicznej, ponieważ znacząco obniżają trud i skomplikowanie czynności niezbędnych do sprawdzenia autentyczności wiadomości.

Szczegółowy opis działania systemu

Podobnie jak w przypadku systemu Boneha-Franklina, w skład systemu wchodzą cztery algorytmy: tworzenie parametrów systemu, tworzenie klucza prywatnego, podpisywanie wiadomości i weryfikowanie podpisu.

Algorytm 5.4.1 (Tworzenie parametrów systemu)**.** Następujący algorytm wybiera i przekazuje jako wynik parametry instancji systemu Yi i jej klucz główny.

YI-CREATE-INSTANCE()

1 Wybieramy liczby pierwsze p i q takie, że $p + 1 = 12q$.
2 Ustalamy, że $\mathbb{G}_1$ oznacza podgrupę rzędu q krzywej $E_{0,1}(\mathbb{F}(p))$.
3 Ustalamy, że $\mathbb{G}_2$ oznacza grupę pierwiastków q-tego stopnia w ciele $\mathbb{GF}(p^2)$.
4 Ustalamy, że odwzorowanie dwuliniowe $b\colon \mathbb{G}_1 \times \mathbb{G}_1 \to \mathbb{G}_2$ jest określone wzorem 5.3.7.
5 Wybieramy liczbę naturalną m.
6 Ustalamy, że przestrzenią wiadomości $\mathcal{M}$ jest zbiór $\{0,1\}^m$.
7 Ustalamy, że przestrzenią podpisów $\mathcal{S}$ jest zbiór $\mathbb{F}(p) \times \mathbb{F}(p)$.
8 Ustalamy, że przestrzenią identyfikatorów użytkowników $\mathcal{I}$ jest zbiór $\{0,1\}^\star$.
9 Wybieramy funkcje haszujące $H_1\colon \mathcal{I} \times \mathbb{Z} \to \mathbb{F}(p)$ i $H_2\colon \mathcal{M} \times \mathbb{G}_1^\star \to \mathbb{Z}$.
10 Wybieramy generator P grupy $\mathbb{G}_1$.
11 Wybieramy element s z grupy $(\mathbb{Z}/q\mathbb{Z})^\star$.
12 Obliczamy wartość $Q = sP$.
13 Parametry systemu to krotka $\langle p, q, \mathbb{G}_1, \mathbb{G}_2, b, m, H_1, H_2, P, Q\rangle$.
14 Klucz główny to element s.

Algorytm 5.4.2 (Tworzenie klucza prywatnego)**.** Dany jest identyfikator podpisującego ID. Następujący algorytm na podstawie wartości ID oraz klucza głównego s instancji systemu Yi i jej parametrów oblicza i przekazuje jako wynik klucz prywatny odpowiadający identyfikatorowi ID.

YI-CREATE-PRIVATE-KEY(ID, s, $\langle p, q, \mathbb{G}_1, \mathbb{G}_2, b, m, H_1, H_2, P, Q\rangle$)

1 Przyjmujemy $k = 0$.
2 Obliczamy $a = H_1(\text{ID}, k)^3 + 1$.
3 Jeżeli $a^{\frac{p-1}{2}} \neq 1$, to powiększamy wartość k o 1 i wracamy do kroku 2.
4 Obliczamy $b = a^{\frac{p+1}{4}}$.
5 Jeżeli $b > -b \pmod p$, to zmieniamy znak wartości b.
6 Obliczamy $U = 12(a, b)$.
7 Obliczamy $S = sU$.
8 Klucz prywatny to punkt S.

Algorytm 5.4.3 (Podpisywanie wiadomości)**.** Dana jest wiadomość M i klucz prywatny podpisującego S. Następujący algorytm na podstawie wartości M i S oraz parametrów instancji systemu Yi oblicza i przekazuje jako wynik podpis wiadomości M odpowiadający kluczowi S.

YI-SIGN$(M, S, \langle p, q, \mathbb{G}_1, \mathbb{G}_2, b, m, H_1, H_2, P, Q\rangle)$

1 Wybieramy element r z grupy $(\mathbb{Z}/q\mathbb{Z})^\star$.
2 Obliczamy $R = rP$.
3 Jeżeli $y(R) \geqslant -y(R) \pmod p$, to obliczamy $T = rQ + H_2(M, R)S$.
4 W przeciwnym razie obliczamy $T = -rQ + H_2(M, -R)S$.
5 Podpis to para $\langle x(R), x(T)\rangle$.

Algorytm 5.4.4 (Weryfikowanie podpisu)**.** Dana jest wiadomość M, podpis V w postaci $V = \langle a, c\rangle$ i identyfikator podpisującego ID. Następujący algorytm na podstawie wartości M, V i ID oraz parametrów instancji systemu Yi stwierdza, czy podpis V jest autentyczny, tzn. został utworzony dla wiadomości M przez podpisującego dysponującego kluczem prywatnym odpowiadającym identyfikatorowi ID.

YI-VERIFY$(M, \langle a, c\rangle, \text{ID}, \langle p, q, \mathbb{G}_1, \mathbb{G}_2, b, m, H_1, H_2, P, Q\rangle)$

1 Obliczamy $b = (a^3 + 1)^{\frac{p+1}{4}}$.
2 Obliczamy $d = (c^3 + 1)^{\frac{p+1}{4}}$.
3 Jeżeli $b \geqslant -b \pmod p$, to przyjmujemy $R' = (a, b)$.
4 W przeciwnym razie przyjmujemy $R' = (a, -b)$.
5 Przyjmujemy $T' = (c, d)$.
6 Obliczamy wartość U tak samo jak w algorytmie tworzenia klucza prywatnego.
7 Obliczamy $u = b(T', P)$.
8 Obliczamy $v = b(R' + H_2(M, R')U, Q)$.
9 Jeżeli $u = v$ lub $u = v^{-1}$, to podpis jest autentyczny.

Zauważmy, że system Yi nie jest aż tak ogólny, jak system Boneha i Franklina – jest oparty na superosobliwej krzywej eliptycznej, nie zaś na dowolnej grupie. Ponadto, w zaproponowanej przez Boneha i Franklina konkretnej realizacji systemu liczba q mogła być dowolnym dzielnikiem liczby $p+1$. W systemie Yi dodatkowo musi być spełniony warunek $12q = p+1$. Obostrzenia te wprowadzone są po to, żeby można było łatwo pierwiastkować w ciele $\mathbb{F}(p)$.

Uwaga 5.4.5. Jeżeli $p \equiv 3 \pmod 4$, to element a z grupy $(\mathbb{Z}/p\mathbb{Z})^\star$ jest resztą kwadratową wtedy i tylko wtedy, gdy $a^{\frac{p-1}{2}} \equiv 1 \pmod p$. Ponadto, jeżeli element a jest resztą kwadratową, to wartość $a^{\frac{p+1}{4}}$ jest jego pierwiastkiem.

Uwaga 5.4.6. Zmienna k w procedurze YI-CREATE-PRIVATE-KEY jest powiększana tak długo, aż wartość $H_1(\text{ID}, k)^3 + 1$ będzie resztą kwadratową. Prawdopodobieństwo zajścia tej sytuacji w jednym kroku wynosi $\frac{1}{2}$, zatem algorytm generowania klucza prywatnego prawie na pewno kończy się sukcesem.

Sprawdźmy teraz, że weryfikacja autentycznego podpisu faktycznie kończy się stwierdzeniem jego autentyczności.

Twierdzenie 5.4.7. *Dana jest wiadomość M, identyfikator nadawcy* ID*, odpowiadający mu klucz prywatny S i parametry $\mathcal{P}$ instancji systemu Yi. Wówczas zachodzi następująca zależność:*

$$\text{YI-VERIFY}(M, \text{YI-SIGN}(M, S, \mathcal{P}), \text{ID}, \mathcal{P}) = \text{TRUE}$$

Dowód. Zauważmy, że punkt T' obliczany w procedurze YI-VERIFY jest równy punktowi T (obliczanemu w procedurze YI-SIGN lub punktowi $-T$. Podobnie jest z punktem R'. Rozważmy dwa przypadki.

1. Jeżeli $y(R) \geqslant -y(R) \pmod p$, to $R' = R$ i $T' = \pm T$. Wówczas:

$$\begin{aligned} u &= b(T', P) \\ &= b(\pm T', P) \\ &= b(T, P)^{\pm 1} \\ &= b(rQ + H_2(M, R)S, P)^{\pm 1} \\ &= b(rsP + sH_2(M, R)U, P)^{\pm 1} \\ &= b(rP + H_2(M, R)U, P)^{\pm s} \\ &= b(R + H_2(M, R)U, Q)^{\pm 1} \\ &= b(R' + H_2(M, R')U, Q)^{\pm 1} \\ &= v^{\pm 1} \end{aligned}$$

2. Jeżeli $y(R) < -y(R) \pmod p$, to $R' = -R$ i $T' = \pm T$. Wówczas:

$$\begin{aligned} u &= b(T', P) \\ &= b(\pm T', P) \\ &= b(T, P)^{\pm 1} \\ &= b(-rQ + H_2(M, -R)S, P)^{\pm 1} \\ &= b(-rsP + sH_2(M, -R)U, P)^{\pm 1} \\ &= b(-rP + H_2(M, -R)U, P)^{\pm s} \\ &= b(-R + H_2(M, -R)U, Q)^{\pm 1} \\ &= b(R' + H_2(M, R')U, Q)^{\pm 1} \\ &= v^{\pm 1} \end{aligned}$$

□

We wspomnianej wcześniej pracy Yi znajduje się analiza bezpieczeństwa opisanego systemu.

Podsumowanie

Cel wyznaczony na początku pracy został osiągnięty. Zobaczyliśmy na przykładzie iloczynu Weila, że pozornie abstrakcyjne matematyczne pojęcie może odnaleźć drogę do zastosowania w praktyce i że droga ta nie musi być ani długa, ani trudna: kryptosystemy wykorzystujące iloczyn Weila nie odbiegają poziomem skomplikowania od najpopularniejszych obecnie kryptosystemów, a implementacja algorytmu Millera będąca częścią pracy powstała szybko, nie nastręczała większych trudności, a powstały kod źródłowy jest przejrzysty.
Kryptosystemy oparte na iloczynie Weila nie są jeszcze tak rozpowszechnione, jak np. kryptosystem RSA czy protokół Diffiego-Hellmana. Za adekwatne kryterium popularności można uznać pytanie, czy istnieje kryptosystem oparty na iloczynie Weila taki, że istnieje jego wolna, otwartoźródłowa implementacja, która jest częścią jednej z wiodących dystrybucji systemu Linuks, lub taki, który jest objęty jednym ze standardów regulujących działanie sieci Internet (np. RFC [19]). W chwili pisania niniejszej pracy odpowiedź na to pytanie jest negatywna. Aby ten stan rzeczy zmienił się, należy podjąć następujące kroki:

- udoskonalić implementację algorytmu Millera;
- zmodyfikować sam algorytm tak, aby wyeliminować z niego losowość;
- zaproponować takie modyfikacje standardów internetowych powiązanych z szyfrowaniem, infrastrukturą kluczy publicznych i pocztą elektroniczną, które umożliwią użycie kryptosystemów opartych na iloczynie Weila na szeroką skalę;
- zaimplementować kryptosystemy oparte na iloczynie Weila w popularnych otwartoźródłowych bibliotekach szyfrujących oraz serwerach i przeglądarkach poczty elektronicznej.

Powyższe kierunki dalszych prac to zagadnienia typowo praktyczne. Są również możliwe dalsze badania teoretyczne:

- konstruowanie kolejnych systemów kryptograficznych na podstawie iloczynu Weila;
- konstruowanie systemów kryptograficznych na podstawie odwzorowań podobnych do iloczynu Weila, np. na podstawie odwzorowania Tate'a;
- przeprowadzanie ataków kryptograficznych za pomocą iloczynu Weila;
- analiza odporności istniejących systemów kryptograficznych opartych na krzywych eliptycznych na potencjalny atak oparty na iloczynie Weila;
- projektowanie systemów kryptograficznych odpornych na ataki wykorzystujące iloczyn Weila.

Miejmy nadzieję, że iloczyn Weila, opierające się na nim kryptosystemy oraz wszelkie inne ciekawe zdobycze kryptografii trafią kiedyś do codziennego użytku, a dzięki nim matematyka będzie postrzegana jako bardziej przydatna i stanie się bardziej lubiana.

Bibliografia

[1] Leonard Charlap, David Robbins. *An Elementary Introduction to Elliptic Curves.* CRD Expository Report 31, 1988.

[2] Leonard Charlap, Raymond Coley. *An Elementary Introduction to Elliptic Curves II.* CCR Expository Report 34, 1990.

[3] Joseph Silverman. *The Arithmetic of Elliptic Curves.* Springer-Verlag, 1994.

[4] Michael Tsfasman, Serge Vlăduţ, Dmitry Nogin. *Algebraic Geometry Codes.* American Mathematical Society, 2007.

[5] Victor Miller. *Short Programs for Functions on Curves.* IBM, 1986.

[6] Alfred Menezes, Tatsuaki Okamoto, Scott Vanstone. *Reducing Elliptic Curve Logarithms in a Finite Field.* IEEE Transactions on Information Theory, 1993.

[7] Dan Boneh, Matthew Franklin. *Identity-based Encryption from the Weil Pairing.* Advances in Cryptology – CRYPTO 2001, 2001.

[8] Xun Yi. *An Identity-Based Signature Scheme from the Weil Pairing.* IEEE, 2003.

[9] Eiichiro Fujisaki, Tatsuaki Okamoto. *Secure Integration of Asymmetric and Symmetric Encryption Schemes.* Advances in Cryptology – CRYPTO '99, 1999.

[10] Ueli Maurer. *Towards the Equivalence of Breaking the Diffie-Hellman Protocol and Computing Discrete Logarithms.* Advances in Cryptology – CRYPTO '94, 1994.

[11] René Schoof. *Counting Points on Elliptic Curves over Finite Fields.* Journal de Théorie des Nombres de Bordeaux, 1995.

[12] Leonard Adleman, Kenneth Manders, Gary Miller. *On Taking Roots in Finite Field.* 18th Annual Symposium on Foundations of Computer Science, 1977.

[13] Michael Rabin. *Probabilistic Algorithms in Finite Fields.* SIAM Journal on Computing, 1980.

[14] Neal Koblitz. *Elliptic Curve Cryptosystems.* American Mathematical Society, 1987.

[15] Ronald Rivest, Adi Shamir, Leonard Adleman. *A Method for Obtaining Digital Signatures and Public-Key Cryptosystems.* Communications of the ACM, 1978.

[16] Whitfield Diffie, Martin Hellman. *New Directions in Cryptography.* IEEE Transactions on Information Theory, 1976.

[17] Thomas Cormen, Charles Leiserson, Ronald Rivest, Clifford Stein. *Wprowadzenie do Algorytmów*. Wydawnictwa Naukowo-Techniczne, 2005.

[18] Guido van Rossum et al. *Python Programming Language*. `http://www.python.org/`, 1990-2011.

[19] Steve Crocker, Steve Carr, Jeff Rulifson et al. *Requests for Comments*. `http://tools.ietf.org/rfc/`, 1969-2011.

[20] William Stein et al. *Sage*. `http://sagemath.org/`, 2005-2011.