Elliptische Kurven I

Franz Lemmermeyer
lemmerm@mpim-bomm.mpg.de

4. September 1999

4.9.1999

Inhaltsverzeichnis

Kapitel 1
Einführung

Dieses Kapitel dient im wesentlichen der Motivation; die Abschnitte 1.4 und 1.5 sind für das Verständnis der restlichen Kapitel nicht unbedingt erforderlich.

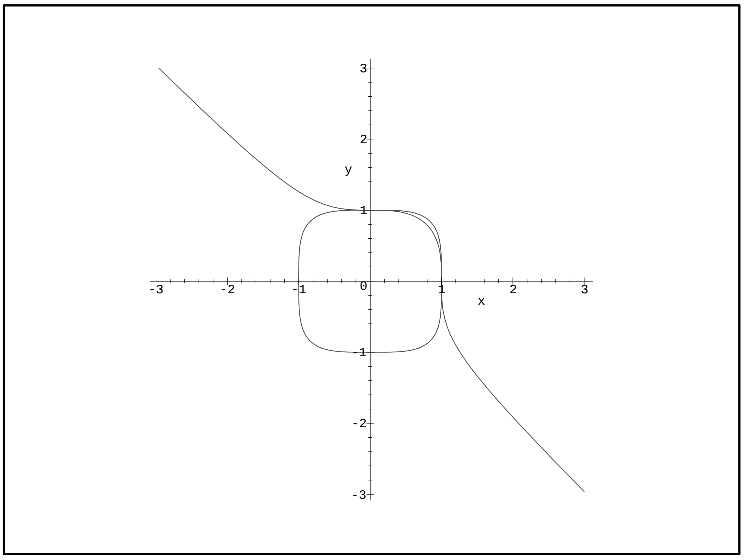
1.1 Was sind elliptische Kurven?

Eine elliptische Kurve ist eine eindimensionale glatte projektive Varietät vom Geschlecht 1 mit einem rationalen Punkt.

Diese hübsche Definition hat den Nachteil, daß man eine ganze Menge an algebraischer Geometrie benötigt, um sie zu verstehen. Viel leichter ist es, Beispiele für elliptische Kurven zu geben. Beginnen wir mit der Gleichung $x^3 + y^3 = 6z^3$, wobei wir für x, y, z ganze Zahlen zulassen. Von dieser elliptischen Kurve hat Legendre behauptet bewiesen zu haben, daß sie keine nichttrivialen Lösungen besitzt, und Mathematiker wie Lucas, Pépin oder Dudeney haben eine solche Lösung angegeben: $17^3 + 37^3 = 6 \cdot 21^3$. Der Sunday Telegraph in London veranstaltet jährlich ein Neujahrsquiz; 1995 waren zwei der Fragen die folgenden

- Solve the equation $A^3/B^3 + C^3/D^3 = 6$, where A, B, C, D are all positive whole numbers below 100.
- A special question with a £450 prize. Either give a second solution to the above equation where the four variables are all whole numbers

ABBILDUNG 1.1. FERMATKURVEN $x^3 + y^3 = 1$ UND $x^4 + y^4 = 1$



above 100 (A, B and C, D relatively prime), or demonstrate that no such second solution can exist.

Bekannter ist natürlich die elliptische Kurve $x^3 + y^3 = z^3$, von der Fermat bewiesen haben will, daß sie keine nichttrivialen Lösungen in $\mathbb{Z}$ besitzt. Euler und Gauß haben später gezeigt, daß Fermat recht hatte. Abbildung 1.1 zeigt die Fermatkurven $x^n + y^n = z^n$ für $n = 3$ und $n = 4$.

Diese Beispiele beantworten die Frage, was elliptische Kurven denn nun sind, selbstverständlich nicht, und es ist nicht klar, warum man $x^3 + y^3 = z^3$ eine elliptische Kurve nennt, $x^4+y^4 = z^4$ dagegen nicht. Unsere Antwort wird sein, daß wir uns elliptische Kurven durch eine Gleichung gegeben denken, die man die Weierstraßsche Normalform nennt und die über $\mathbb{Q}$ so aussieht: $y^2 = x^3 + ax + b$ mit rationalen Koeffizienten a, b, wobei man noch voraussetzt, daß das Polynom $x^3 + ax + b$ keine mehrfache Nullstelle besitzt. Diese Normalform ist natürlich nicht vom Himmel gefallen, sondern kommt aus der Funktionentheorie, also der komplexen Analysis, genauer gesagt aus der Theorie elliptischer Funktionen. Diese wiederum sind entstanden durch Umkehrung elliptischer Integrale, und diese verdanken ihren Namen dem Versuch, den Umfang von Ellipsen zu berechnen.

Es erscheint mir daher angebracht, den historischen Weg von Diophant über Newton bis zum Gruppengesetz auf elliptischen Kurven einerseits, sowie von der Lemniskate Fagnano's über Eulers Beiträge bis hin zu den revolutionären Arbeiten Abels über elliptische Funktionen andererseits etwas zu beleuchten. Nach einem kurzen Überblick über das, was die moderne Funktionentheorie zu diesem Thema zu sagen hat, werden wir dann schon bei der Weierstraßschen Normalform und dem Gruppengesetz auf elliptischen Kurven über $\mathbb{C}$ angelangt sein.

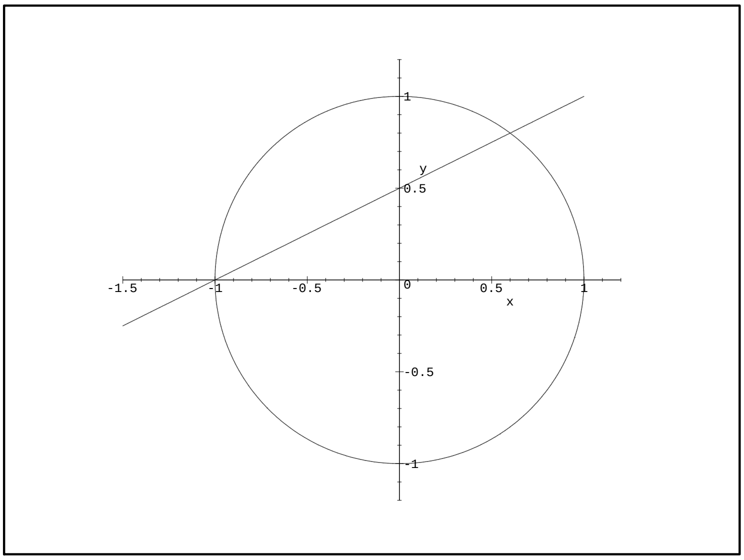
1.2 Diophant und Newton

Die Theorie der elliptischen Kurven ist Teil der arithmetischen Geometrie und liegt damit zwischen algebraischer Geometrie und klassischer Zahlentheorie. Die arithmetische Geometrie beschäftigt sich im wesentlichen mit dem Studium rationaler Punkte (solche, deren Koordinaten rationale Zahlen sind) auf geometrischen Objekten wie projektiven Varietäten, z.B. glatten Kurven; insbesondere interessiert man sich für die Beschreibung rationaler Punkte auf Kurven vom Geschlecht $g = 1$, nämlich elliptischen Kurven. Die Kurven vom Geschlecht $g = 1$ nehmen dabei eine Sonderstellung ein: der Fall $g = 0$ ist im wesentlichen trivial, der Fall $g \geq 2$ dagegen außerordentlich schwierig.

Die Sekanten-Methode

Ein Beispiel für eine Kurve vom Geschlecht 0 ist der Einheitskreis $C : x^2 + y^2 = 1$. Die Bestimmung der rationalen Punkte auf dieser Kurve ist seit dem Altertum bekannt: ist nämlich $x = \frac{a}{c}$, $y = \frac{b}{c}$ ein solcher, so ist (a, b, c) ein pythagoräisches Tripel (eine Lösung der Gleichung $a^2 + b^2 = c^2$), und die Umkehrung ist ebenfalls richtig. Ein geometrischer Zugang zur Lösung des Problems ist der folgende, der bereits Diophant bekannt war (man beachte aber, daß Diophant nur arithmetisch, nicht geometrisch argumentieren konnte: Koordinaten, d.h. die analytische Geometrie, sind eine Erfindung der Neuzeit und mit dem Namen Descartes verknüpft; ob zu recht oder zu unrecht, scheint umstritten zu sein): offenbar ist $P = (-1, 0)$ ein rationaler Punkt auf C; ist Q ein weiterer, so wird die Gerade PQ durch eine Gleichung mit *rationalen* Koeffizienten beschrieben, und sie schneidet die y-Achse (ebenfalls eine "rationale" Gerade) in einem Punkt R, der als Schnittpunkt zweier rationaler Geraden rationale Koeffizienten haben muß. Umgekehrt kann man sich irgendeinen rationalen Punkt $R = (0, t)$, $t \in \mathbb{Q}$,

Abbildung 1.2. Einheitskreis und Gerade $y = \frac{1}{2}(x+1)$



auf der y-Achse heraussuchen und die Gerade PR mit C schneiden; der von P verschiedene Schnittpunkt Q ist dann ebenfalls rational, und jeder rationale Punkt $Q \neq P$ kann offenbar auf diese Weise erhalten werden.

Natürlich lassen sich diese Überlegungen explizit durchführen: die Gerade durch P und $(0,t)$ hat die Gleichung $y = t(x+1)$; Schneiden mit C liefert $1 - x^2 = y^2 = t^2(x+1)^2$. Die Lösung $x = -1$ entspricht dem Schnittpunkt P, sodaß Kürzen von $x+1$ den zweiten Schnittpunkt Q liefert: $1 - x = t^2(1+x)$. Der Rest ist klar: Q hat x-Koordinate $x = \frac{1-t^2}{1+t^2}$, und Einsetzen in die Geradengleichung gibt $y = t(x+1) = \frac{2t}{1+t^2}$. Wir haben gezeigt:

Proposition 1.1. *Die von $(-1,0)$ verschiedenen rationalen Punkte auf $C : x^2 + y^2 = 1$ werden durch*

$$x = \frac{1-t^2}{1+t^2}, \quad y = \frac{2t}{1+t^2} \tag{1.1}$$

parametrisiert. Insbesondere sind die pythagoreischen Tripel gegeben durch $(1-t^2, 2t, 1+t^2)$.

Mit anderen Worten: die rationalen Punkte auf dem Einheitskreis sind nicht viel aufregender als diejenigen auf der Geraden $x = 0$. Die Methode Diophants dagegen, mit der dieses Ergebnis gewonnen wurde, ist Gold wert.

Sie funktioniert nämlich allgemein für Kegelschnitte, das sind die Nulstellenmengen quadratischer Gleichungen der Form

$$ax^2 + bxy + cy^2 + dx + ey + f = 0, \tag{1.2}$$

welche einen rationalen Punkt P besitzen. In der Regel werden durch solche Gleichungen Kegelschnitte beschrieben, also Kreise, Ellipsen, Hyperbeln, Parablen etc.; allerdings kann es auch vorkommen, daß (1.2) das Produkt zweier Geraden ist (solche Kurven nennt man – wie könnte es anders sein – reduzibel).

Besitzt (1.2) keinen rationalen Punkt, ist die Beschreibung aller rationalen Punkte natürlich auch nicht schwer; mit dem Hasse-Prinzip hat man sogar ein schnelles Verfahren, um festzustellen, ob Gleichungen der Form (1.2) einen rationalen Punkt besitzen oder nicht. Für Diophants Verfahren wähle man irgendeine rationale Gerade; betrachte deren rationale Punkte R, und untersuche die Schnittpunkte der Geraden PR mit der Kurve (1.2).

Übung. Man bestimme alle rationalen Punkte auf $C : x^2 + y^2 = 2$. (Bemerkung: das Aussehen der Parametrisierung wird natürlich davon abhängen, welche Gerade (oder welchen rationalen Punkt) man wählt. Es wird daher verschiedene richtige Lösungen geben).

Übung. Man bestimme alle rationalen Punkte auf $C : x^2 - y^2 = 1$.

Übung. Man zeige, daß $C : x^2 + y^2 = 3$ keinen rationalen Punkt besitzt.

Das angesprochene Hassesche Lokal-Global-Prinzip werden wir zwar nicht beweisen (was nicht schwer wäre, aber etwas Zeit kostet), soll aber wegen seiner Bedeutung wenigstens formuliert werden:

Satz 1.2. *Eine Kurve*

$$ax^2 + bxy + cy^2 = 0 \tag{1.3}$$

mit rationalen Koeffizienten $a, b, c \in \mathbb{Q}$, $ac \neq 0$, hat genau dann nichttriviale rationale Lösungen, wenn sie solche in jeder Komplettierung $\mathbb{Q}_p$ von $\mathbb{Q}$ (einschließlich $\mathbb{Q}_\infty = \mathbb{R}$) besitzt.

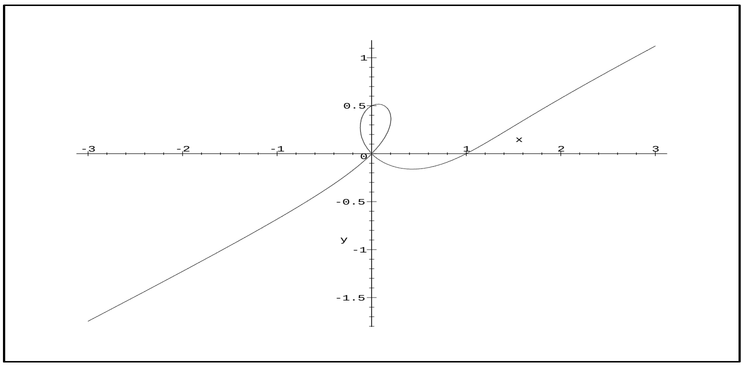
Dabei heißt die Lösung $(0, 0)$ von (1.3) trivial. Entsprechende Aussagen gelten für (1.2), da man die linearen und konstanten Terme wegtransformieren kann. Zum Glück muß man hier nicht wirklich jede Komplettierung $\mathbb{Q}_p$ betrachten: transformiert man die Gleichung so, daß die Koeffizienten a, b, c ganze Zahlen werden, so genügt es, die $\mathbb{Q}_p$ mit $p \mid 2abc$ zu betrachten; es ist also nur endlich viel zu tun.

Wie sieht es nun bei kubischen Kurven aus (solche, in denen x und y maximal zur dritten Potenz auftreten)? Das Problem ist hier, daß ein Schnittpunkt PR wie oben die Kurve im allgemeinen in zwei Punkten Q_1 und Q_2 schneidet, die zwar einer quadratischen Gleichung mit rationalen Koeffizienten genügen, aber selbst nicht rational zu sein brauchen.

Es gibt aber "Ausnahmekurven", in denen das Verfahren doch funktioniert: ist nämlich C eine kubische Kurve, die einen rationalen "Doppelpunkt" P enthält (d.h. die Kurve schneidet sich selbst in P), dann funktioniert das obige Verfahren problemlos.

Als Beispiel betrachten wir die Kurve $C : x^2 - y^2 = (x - 2y)(x^2 + y^2)$. Das Diagramm 1.3 zeigt, daß der rationale Punkt $P = (0, 0)$ tatsächlich

ABBILDUNG 1.3. $x^2 - y^2 = (x - 2y)(x^2 + y^2)$

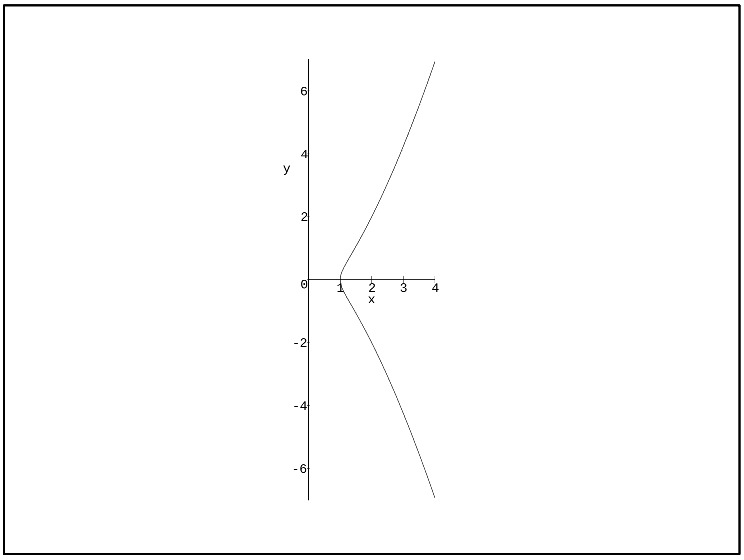


ein Doppelpunkt ist. Wählen wir $G : y = 1$ als rationale Gerade, so haben die Geraden durch P und die rationalen Punkte $R = (1, t)$ auf G die Gleichungen $x = ty$. Schneiden mit C gibt $y^2(t^2 - 1) = y^3(t - 2)(t^2 + 1)$; die doppelte Nullstelle $y = 0$ kommt von P, und Abdividieren von y^2 liefert $y = \frac{t^2-1}{(t-2)(t^2+1)}$. Die x-Koordinate ist, wie die Geradengleichung zeigt, gleich $x = ty$. Man sieht auch hier, daß solche Parametrisierungen rationaler Punkte im allgemeinen nicht alle Punkte beschreiben: solche mit $y = 0$ und $x \neq 0$ nämlich können nicht auf einer Geraden durch den Ursprung liegen; hier gibt es genau einen solchen: $y = 0$ gibt $x^2 = x^3$, und außer von $x = 0$ wird diese Gleichung auch von $x = 1$ gelöst:

Proposition 1.3. *Die Parametrisierung*

$$x = \frac{t^3 - t}{(t-2)(t^2+1)}, \quad y = \frac{t^2 - 1}{(t-2)(t^2+1)}$$

Abbildung 1.4. Newtonscher Knoten $y^2 = x^2(x-1)$



der kubischen Kurve $C : x^2 - y^2 = (x - 2y)(x^2 + y^2)$ *liefert alle rationalen Punkte außer* $P = (0, 0)$ *und* $(1, 0)$.

Kubische Kurven mit Doppelpunkt heißen *singulär*; wie wir noch sehen werden, gibt es aber noch andere Singularitäten auf kubischen Kurven als Doppelpunkte, z.B. sogenannte Spitzen. Die genaue Untersuchung singulärer kubischer Kurven werden wir im nächsten Kapitel durchführen.

Übung. Parametrisiere den Newtonschen Knoten $y^2 = x^2(x-1)$ durch Projektion auf die Gerade $x = -1$.

Übung. Parametrisiere das Cartesische Blatt $x^3 + y^3 = 3xy$.

Übung. Parametrisiere das dreiblättrige Kleeblatt $(x^2 + y^2)^2 + 3x^2y = y^3$.

Übung. Parametrisiere die Kurve $y^2 = x^3 + x^2 - x - 1$.

Übung. Parametrisiere die Kissoide des Diocles $y^2(a + x) = (a - x)^3$.

Übung. Parametrisiere die Kurve $(x^2 + y^2)^n = x^{2n-1}$.

Die Tangenten-Methode

Was kann man nun tun, wenn die kubische Kurve C nicht singulär ist? Wenn man gar keinen rationalen Punkt auf C kennt, vorläufig gar nichts.

Abbildung 1.5. Cartesisches Blatt $x^3 + y^3 = 3xy$

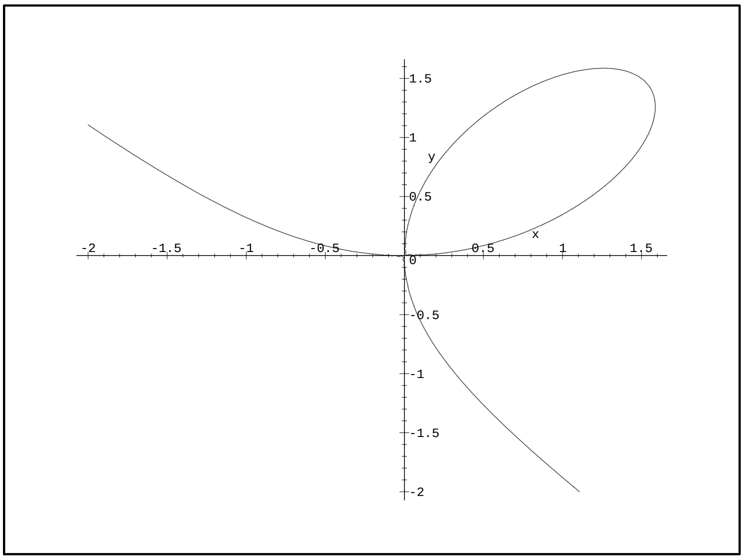


Abbildung 1.6. Dreiblättriges Kleeblatt $(x^2 + y^2)^2 + 3x^2y = y^3$

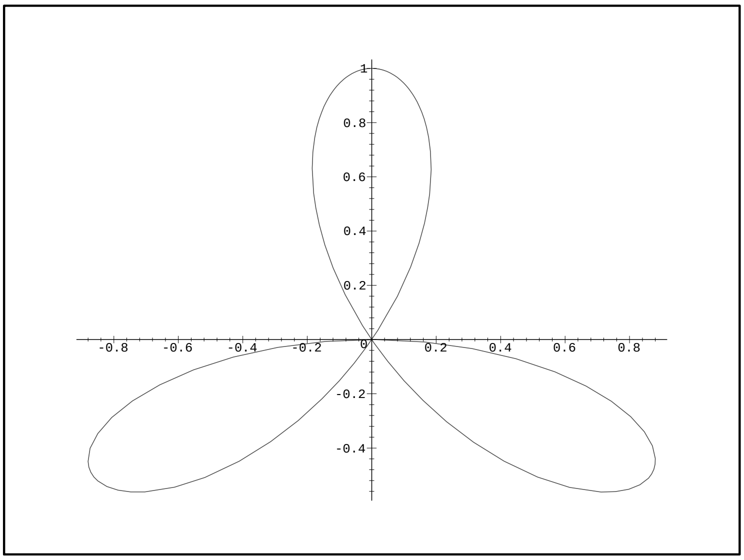
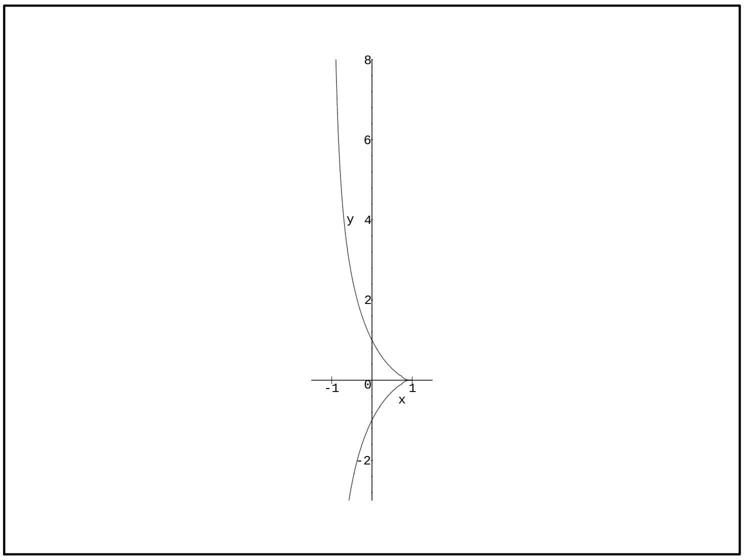


Abbildung 1.7. Kissoide des Diocles $y^2(1+x) = (1-x)^3$



Ansonsten geht man vor wie folgt: P braucht ja kein Doppelpunkt der Kurve zu sein, es genügt, wenn die Gerade durch P mit C eine "zweifache Nullstelle" besitzt, d.h. wenn wir durch P eine Tangente an C legen (und Newton hat bekanntlich gezeigt, wie man Tangenten findet[1]). Ist nämlich T eine Tangente an C in P, so dürfen wir i.a. erwarten, daß T mit der kubischen Kurve einen dritten Schnittpunkt Q hat. Ist die Tangentensteigung rational, so wird Q ebenfalls rationale Koordinaten besitzen, und wir haben eine Methode, mit der man aus einem rationalen Punkt auf C weitere konstruieren kann.

Als Beispiel wählen wir die Kurve $C : x^3+y^3 = 9$. Diese hat den rationalen Punkt $P = (2,1)$. Eine Gerade durch P hat die Gleichung $y-1 = m(x-2)$, wo m die Steigung bezeichnet. Diese soll gleich der Steigung der Kurve in P sein, die sich durch implizite Differentiation leicht ergibt: Ableiten nach x der Kurvengleichung ergibt $3x^2 + 3y^2y' = 0$, also $y' = -x^2/y^2$ und $m = -4$. Schneiden mit C liefert $9 - x^3 = (-4(x-2)+1)^3$ (das Ausmultiplizieren ist eine Tätigkeit (Tätlichkeit?), die oft bestraft wird; das läßt sich auch hier wieder schön sehen), d.h. $x^3 - 9 = 64(x-2)^3 - 48(x-2)^2 + 12(x-2) - 1$. Bringt man die -1 auf die linke Seite, steht dort $x^3 - 8 = (x-2)(x^2 +$

[1]Es ist deswegen erstaunlich, daß Newton, obwohl er die Sekanten-Methode gekannt und benutzt hat, *nichts* über die Tangenten-Methode sagt. Diese taucht erstmals bei Lagrange auf.

$2x + 4)$; durch Abdividieren der uninteressanten Nullstelle $x = 2$ finden wir $x^2 + 2x + 4 = 64(x-2)^2 - 48(x-2) + 12$. Wieder bringen wir die 12 auf die linke Seite und erhalten dort $x^2 + 2x - 8 = (x-2)(x+4)$. Daß der Faktor $x - 2$ sich zweimal kürzen läßt, ist ein gutes Zeichen: die Tangente muß ja eine doppelte Nullstelle in P besitzen. Diesmal folgt nach Abdividieren $x + 4 = 64(x-2) - 48$, also $63x = 180$ und $x = \frac{20}{7}$. Einsetzen in die Geradengleichung gibt schließlich $y = -\frac{17}{7}$ und damit $Q = (\frac{20}{7}, -\frac{17}{7})$.

Damit haben wir fast schon das Gruppengesetz auf dieser Kurve entdeckt: bis auf eine kleine Modifikation ist nämlich $Q = P + P$. Ganz entsprechend kann man jetzt weiter machen und entweder die Tangente in Q betrachten (was dann im wesentlichen auf $Q + Q = 4P$ führt), oder aber die Gerade PQ (dies gibt dann analog einen Punkt, der fast $P + Q$ heißt), und erhält so laufend neue rationale Punkte auf C.

Schreibt man die Kurve in der homogenen Form $x^3 + y^3 = 9z^3$, so hat P die Koordinaten $(2, 1, 1)$, und das Gruppengesetz ermöglicht die Berechnung der Vielfachen von P:

$$\begin{aligned}
2P &= (-17, 20, 7)\\
3P &= (919, -271, 438)\\
4P &= (-36520, 188479, 90391)\\
5P &= (169748279, -152542262, 53023559)\\
6P &= (415280564497, 676702467503, 348671682660)\\
7P &= (-14541760311678322, 14546930068742329, 714352239600649)\\
8P &= (1243617733990094836481, 487267171714352336560,\\
&\qquad 609623835676137297449)
\end{aligned}$$

An dieser Stelle drängen sich vielleicht einige Fragen auf: zuerst fragt man sich natürlich, ob die Sekanten-Tangenten-Methode immer funktioniert, also immer einen wohldefinierten Punkt auf der gegebenen kubischen Kurve liefert. Das ist nicht so, wie das Beispiel der Sekante PP' mit $P = (1, 2)$ und $P' = (2, 1)$ auf der Kurve $x^3 + y^3 = 9$ zeigt. Zweitens: falls das Verfahren funktioniert, gibt es dann Punkte, die nur endlich viele neue Punkte liefern? Solche Punkte (Hurwitz nannte sie “Ausnahmepunkte”; neben ihm hat insbesondere Beppo Levi solche Punkte untersucht) gibt es tatsächlich, und wir widmen diesen Punkten das ganze Kapitel 3.

Dann ist da die Frage, ob man z.B. in unserem Beispiel $C : x^3 + y^3 = 9$ durch obige Konstruktion unendlich viele verschiedene rationale Punkte auf C erhält. Dies ist in der Tat der Fall (und ließe sich “von Hand” nachweisen, indem man zeigt, daß Zähler und Nenner bei fortgesetzter Addition von

P immer größer werden). Wir werden das aber erst machen, wenn wir in Kapitel 4 den Begriff der Höhen auf elliptischen Kurven eingeführt haben.

Weiter mag man sich fragen, ob es auf jeder kubischen Kurve C *endlich viele* Punkte $P_1, \ldots, P_r$ gibt, sodaß sich *jeder* rationale Punkt auf C in endlich vielen Schritten mit obigen Methoden aus diesen P_j konstruieren läßt. Dies ist in der Tat so, wie Poincaré ohne Beweis angenommen und Mordell später bewiesen hat. Der entsprechende Satz heißt heute Satz von Mordell-Weil, weil Weil (Washington hat unlängst einen Artikel geschrieben, in dem "converges to two, too" vorkommt; ich darf das also auch) Mordells Beweis von $\mathbb{Q}$ auf beliebige algebraische Zahlkörper (und von elliptischen Kurven auf abelsche Varietäten) erweitert hat; dieser Satz wird – in einem Spezialfall, da uns für den allgemeinen die algebraische Zahlentheorie fehlt – ebenfalls in Kapitel 4 bewiesen werden.

Eine letzte Frage ist, ob diese Konstruktion neuer Punkte sich als "Addition" verstehen läßt: man könnte den durch die Tangente an P definierten Punkt Q als $P + P$ bezeichnen, den durch die Gerade PQ definierten als $P+Q$ etc.; die Antwort hierauf ist "fast ja": man sieht einerseits ein, daß man auf diese Weise sicher kein neutrales Element bekommt, also einen Punkt, dessen "Addition" immer zum Ausgangspunkt zurückführt. Andererseits ist die obige Definition nur ein klein wenig zu modifizieren, um zum Erfolg zu kommen; wir werden dies in Kapitel 2 tun, sobald wir projektive Ebenenen bereitgestellt haben.

Übung. Sei $P = (r, s)$ ein rationaler Punkt auf $C : x^3 + y^3 = a$ und a keine dritte Potenz in $\mathbb{Q}$. Bestimme die Tangente an C in P, sowie den dritten Schnittpunkt Q dieser Tangente mit C. (Als Zwischenergebnis für die x-Koordinate erhält man $x = r\frac{r^6+3r^3s^3+2s^6}{r^6-s^6}$; man erinnere sich an dieser Stelle aber an die binomischen Formeln oder an den euklidischen Algorithmus.) Kann es vorkommen, daß $P = Q$ ist?

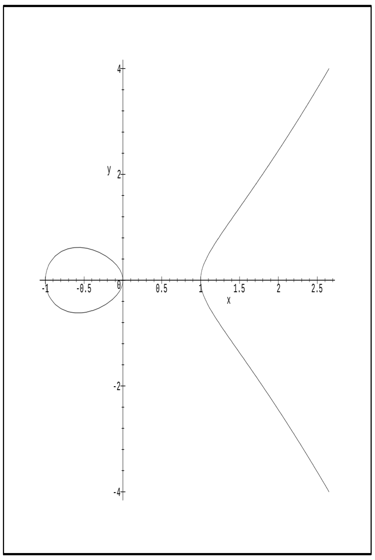
Übung. Seien $P = (r, s)$ und $Q = (t, u)$ rationale Punkte auf $C : x^3+y^3 = a$ und a keine dritte Potenz in $\mathbb{Q}$. Bestimme die Sekante PQ und den dritten Schnittpunkt R dieser Sekante mit C.

Übung. Man finde auf der Kurve $C : x^3 + y^3 = 6$ weitere rationale Punkte. Wer einen Computer mit Langzahlarithmetik hat, löse die Neujahrsfrage des London Telegraph; die korrekte Antwort lautet

$$x = 1498088000358117387964077872464225368637808093957571271237$$
$$y = 1659187585671832817045260251600163696204266708036135112763$$
$$z = 1097408669115641639274297227729214734500292503382977739220$$

4.9.1999

Abbildung 1.8. Elliptische Kurve in Weierstrass-Form



Übung. Man finde einige rationale Punkte auf $C : x^3 + y^3 = 7$.

Übung. (Diophant) Sei $0 < d = u^3 - v^3$ für positive $u, v \in \mathbb{Q}$. Zeige, daß $d = x^3 + y^3$ mit positiven $x, y \in \mathbb{Q}$ ist.

Übung. Was passiert, wenn man die Sekantenmethode auf die beiden Punkte $(1, 2)$ und $(2, 1)$ der Kurve $x^3 + y^3 = 9$ anwendet?

Übung. Sei $P = (r, s)$, $s \neq 0$, rationaler Punkt auf $E : y^2 = x^3 + ax + b$ mit $a, b \in \mathbb{Q}$. Man rechne nach, daß die Tangente an E in P die Steigung $m = \frac{a^3r^2}{2s}$ besitzt und zeige, daß der Schnittpunkt $Q \neq P$ der Tangente mit E die Koordinaten (x, y) mit $x = m^2 - 2r$ und $y = m(x - r) + s$ besitzt.

Übung. (Euler) Seien $a, b, c \in \mathbb{Z}$ nicht durch die Primzahl p teilbar. Man zeige, daß dann $ax^3 + bpy^3 + cp^2z^3 = 0$ keine Lösung in $\mathbb{Z} \setminus \{0\}$ besitzt.

Übung. (Hurwitz) Sei $\phi(x, y, z) \in \mathbb{Z}[x, y, z]$ eine kubische Form (also homogen vom Grad 3: jeder Term hat Gesamtgrad 3). Zeige, daß die diophantische Gleichung $x^3 + 2y^3 + 4z^3 + 9\phi(x, y, z) = 0$ keine nichttrivialen Lösungen in $\mathbb{Z}$ hat.

4.9.1999

1.3 À Quoi Bon?

Wozu sind elliptische Kurven gut? Bis in die 70er Jahre unseres Jahrhunderts hätten viele Mathematiker diese Frage als falsch gestellt zurückgewiesen; schließlich geht es ihnen in erster Linie nicht um den praktischen Nutzen. Seit nicht ganz 20 Jahren aber haben sie eine bessere Antwort zur Verfügung: damals hat H.W. Lenstra entdeckt, daß man mit elliptischen Kurven Primfaktoren großer Zahlen finden kann. Das hat zu einer Explosion des Interesses an elliptischen Kurven geführt, und wenig später gab es elliptische Primzahltests, elliptische Methoden zur Berechnung von Quadratwurzeln modulo p, sowie elliptische Kryptosysteme.

Elliptische Kurven haben auch in der algebraischen Zahlentheorie Anwendung gefunden, z.B. in der Konstruktion quadratischer Zahlkörper mit großem 3-Rang oder von euklidischen Zahlkörpern mit großem Grad. Die Entwicklung von Algorithmen für die Arithmetik elliptischer Kurven hat auch dazu geführt, daß früher als praktisch unlösbar geltende diophantische Gleichungen wie $y^2 = x^3 - 999$ inzwischen sogar von einem Computer gelöst werden können (die ganzzahligen Lösungen sind $(10, \pm 1)$, $(12, \pm 27)$, $(40, 251)$, $(147, \pm 1782)$, $(174, \pm 2295)$ und $(22\,480, 3\,370\,501)$).

Klassenzahlprobleme

Auch an der Lösung des Gaußschen Klassenzahlproblems waren elliptische Kurven maßgeblich beteiligt. Ein elementarer Aspekt läßt sich wie folgt beschreiben: Euler entdeckte 1772, daß das Polynom $f(x) = x^2 - x + 41$ für $x = 1, 2, \ldots, 40$ nur Primzahlwerte annimmt. Ähnliches gilt für die Polynome $f(x) = x^2 - x + m$ in der folgenden Tabelle: sie liefern Primzahlen für $x = 1, \ldots, m$.

m	3	5	11	17	41
d	-11	-19	-43	-67	-163

In der zweiten Zeile steht dabei die Diskriminante $d = 1 - 4m$ von $f(x)$. Was hat es mit diesen Werten von d auf sich? Betrachten wir dazu die Gleichung

$$x^2 - dy^2 = 4p, \tag{1.4}$$

wo p eine Primzahl ist. Falls diese Gleichung eine Lösung besitzt, muß sicher $p \nmid xy$ sein; modulo p ergibt sich dann $d \equiv (x/y)^2 \bmod p$. d.h. d ist notwendig quadratischer Rest modulo p. Sei nun umgekehrt p eine ungerade Primzahl mit $(d/p) = +1$; ist dann (1.4) mit $x, y \in \mathbb{Z}$ lösbar? Leider nicht: es ist z.B. $(-23/3) = +1$, aber offenbar $12 \neq x^2 + 23y^2$. Tatsächlich gilt diese Umkehrung für $|d| \geq 11$ genau dann, wenn d in obiger Tabelle vorkommt. Daß die Umkehrung für diese Werte richtig ist, ist mit den Methoden der algebraischen Zahlentheorie ganz einfach zu zeigen. Daß es aber kein $d < -163$ mit dieser Eigenschaft mehr gibt, ist eine Vermutung von Gauß gewesen, die im wesentlichen zuerst von Heegner mittels der Theorie der komplexen Multiplikation (also der Theorie solcher elliptischer Kurven, die eine Zusatzeigenschaft besitzen; wir werden dazu später mehr sagen können) bewiesen, wenn auch nicht sofort akzeptiert worden.

In diesem Zusammenhang ist auch folgende Beobachtung relevant:

d	$\exp(\pi\sqrt{-d})$
-43	884736743.99977746603490666...
-67	147197952743.99999866245422450...
-163	262537412640768743.99999999999925007...

Wie man sehen kann, liegen die Werte $e^{\pi\sqrt{-d}}$ für große $|d|$ sehr nahe bei ganzen Zahlen. Noch deutlicher wird es, wenn wir von diesen Zahlen 744 abziehen und die dritte Wurzel ziehen:

d	$(\exp(\pi\sqrt{-d}) - 744)^{1/3}$
-11	31.99809333222744098975227354...
-19	95.99999195891694508468060476...
-43	959.99999999991951173137537734...
-67	5279.9999999999998400738235224...
-163	640319.99999999999999999999999939...

Die Erklärung für diese numerischen Kuriositäten benutzt eine ganze Menge Technik: aus der Theorie der komplexen Multiplikation ist bekannt, daß eine bestimmte "Modulfunktion", nämlich

$$j(q) = \frac{1}{q} + 744 + 196\,884q + 21\,493\,760q^2 + \ldots,$$

für alle d, für die die oben besprochene Umkehrung gilt, an der Stelle $q = -e^{\pi i\sqrt{d}}$ eine ganze Zahlen als Funktionswert hat (sogar eine dritte Potenz einer ganzen Zahl). Da q für große $|d|$ winzig klein ist, wird sich $|\frac{1}{q} + 744|$ nur wenig von der ganzen Zahl $j(q)^{1/3}$ unterscheiden.

4.9.1999

Damit nicht genug: die ganze Zahl x, die von $\sqrt[3]{e^{\pi\sqrt{-d}}} - 744$ approximiert wird, genügt der Gleichung $x^3 + 1728 = -dy^2$ für ein $y \in \mathbb{N}$. Man sehe und staune:

$$\begin{aligned} 11 \cdot 56^2 &= 32^3 + 1728 \\ 19 \cdot 216^2 &= 96^3 + 1728 \\ 43 \cdot 4536^2 &= 960^3 + 1728 \\ 67 \cdot 4\,6872^2 &= 5280^3 + 1728 \\ 163 \cdot 401\,33016^2 &= 640320^3 + 1728 \end{aligned}$$

Eine sehr gute Einführung in die Theorie von $j(q)$ findet man in Serre [Ser]. Darüberhinaus ist bei Springer ein neues Buch [KK] über elliptische Funktionen und Modulformen erschienen, das eine ganze Menge Stoff enthält. Die Darstellung in Cox [Co1] ist ebenfalls exzellent, setzt jedoch algebraische Zahlentheorie voraus.

Monster und Mondschein

Die meisten endlichen Gruppen, die man zu Beginn des Studiums kennenlernt, sind abelsch: $\mathbb{Z}/m\mathbb{Z}$ und $(\mathbb{Z}/m\mathbb{Z})^\times$, oder allgemeiner die Gruppen $(\mathbb{F}_q, +)$ bzw. $(\mathbb{F}_q^\times, \cdot)$ in endlichen Körpern sind Beispiele. Nichtabelsche endliche Gruppen lassen sich aber leicht konstruieren: man nehme z.B. die symmetrischen Gruppen S_n aller Permutationen von $n \geq 3$ Elementen oder die $\mathrm{GL}_n(\mathbb{F}_p)$, also die Gruppe der invertierbaren $n \times n$-Matrizen mit Einträgen aus $\mathbb{F}_p$ (mit $n > 1$).

Während in abelschen Gruppen immer $aba^{-1} = b$ ist, gilt dies in nichtabelschen Gruppen natürlich nicht mehr; schlimmer noch: ist G eine endliche Gruppe mit Untergruppe U, so brauchen die Elemente gug^{-1} (mit $g \in G$ und $u \in U$) nicht mehr in U zu liegen. Tun sie dies doch, nennt man U eine normale Untergruppe von G (oder kurz Normalteiler). Die Untergruppen $\{1\}$ und G sind immer normal; falls G außer diesen beiden trivialen Normalteilern keine weiteren besitzt, nennt man G einfach. Beispielsweise sind zyklische Gruppen $\mathbb{Z}/p\mathbb{Z}$ von Primzahlordnung p einfach, weil $\{1\}$ und G ihre einzigen Untergruppen sind. Weniger triviale Beispiele nichteinfacher Gruppen sind die alternierenden Gruppen A_n für $n \geq 5$ (A_n ist die Untergruppe der geraden Permutationen in S_n; deren Einfachheit ist der Grund dafür, daß sich Gleichungen vom Grad ≥ 5 i.a. nicht mit Wurzeloperationen auflösen lassen).

Das Hauptproblem der Gruppentheorie bis in die 80er Jahre war die Klassifikation aller endlichen einfachen Gruppen, genauer der Satz, daß folgende Liste von endlichen einfachen Gruppen vollständig ist:

- zyklische Gruppen von Primzahlordnung;
- alternierende Gruppen A_n mit $n \geq 5$;
- die klassischen linearen Gruppen $\mathrm{PSL}(n,q)$, $\mathrm{PSU}(n,q)$, $\mathrm{PSp}(2n,q)$ und $\mathrm{P}\Omega^{\varepsilon}(n,q)$;
- Ausnahmegruppen vom Lie-Typ ${}^3D_4(q)$, $E_6(q)$, ${}^2E_6(q)$, $E_7(q)$, $E_8(q)$, $F_4(q)$, ${}^2F_4(2^n)'$, $G_2(q)$, ${}^2G_2(3^n)$ und ${}^2B_2(2^n)$;
- sporadische einfache Gruppen M_{11}, M_{12}, M_{22}, M_{23}, M_{24} (Mathieu-Gruppen); J_1, J_2, J_3, J_4 (Janko-Gruppen); Co_1, Co_2, Co_3 (Conway-Gruppen); HS; Mc; Suz (Co_1–'Babies'), Fi_{22}, Fi_{23}, Fi'_{24} (Fischer-Gruppen); $F_1 = M$ (das Monster), F_2, F_3, F_5, F_7 (Monster–'Babies'); Ru; Ly; ON.

Die größte unter den sporadischen (also zu keiner Familie gehörenden) einfachen Gruppen ist das "Monster", dessen Existenz 1973 von Fischer und Griess vorhergesagt und 1980 bestätigt wurde; das Monster hat Ordnung

$$8080,17424,79451,28758,86459,90496,17107,57005,75436,80000,00000$$
$$= \quad 2^{46} \cdot 3^{20} \cdot 5^9 \cdot 7^6 \cdot 11^2 \cdot 13^3 \cdot 17 \cdot 19 \cdot 23 \cdot 29 \cdot 31 \cdot 41 \cdot 47 \cdot 59 \cdot 71.$$

Nun besitzt jede endliche Gruppe G sogenannte "Darstellungen", das sind Gruppenhomomorphismen $\rho : G \longrightarrow \mathrm{Aut}\,(V)$ in die Automorphismengruppe endlichdimensionaler $\mathbb{C}$-Vektorräume. Wenn eine Darstellung ρ die Eigenschaft hat, daß es außer 0 und V keine invarianten Teilräume gibt, dann heißt ρ irreduzibel. Die Funktion $\chi : G \longrightarrow \mathbb{C}$, die jedem $g \in G$ die Spur der Matrix $\rho(g)$ zuordnet, heißt der Charakter der Darstellung. Die Anzahl der irreduziblen Charaktere jeder endlichen Gruppe ist endlich; man hat nun die Dimensionen der zu solchen irreduziblen Charakteren des Monsters gehörenden Vektorräume V berechnet und gefunden, daß die kleinsten Dimension $d_1 = 1$, $d_2 = 196883$, $d_3 = 21296876$ sind. Vergleicht man das mit der q-Entwicklung $j(q) = \frac{1}{q} + \sum j_n q^n$ der j-Invariante, so stellt man fest, daß $j_1 = d_1 + d_2$ und $j_2 = d_1 + d_2 + d_3$ gilt. Zufall? Nein, dahinter verbirgt sich inzwischen eine ganze Forschungsrichtung, die von Conway und Norton 1979 auf den Namen Mondschein getauft wurde und sich inzwischen bis in die Stringtheorie hineinverzweigt hat. Näheres findet man auf [MP].

4.9.1999

Ein Wort noch zur Klassifikation: der Umfang des Beweises wird auf ca. 10 000 Seiten geschätzt, und derzeit ist eine Buchreihe geplant, die diesen Beweis aufarbeiten soll (die ersten Bände sind bereits erschienen). Andererseits scheint es bei dem Teilproblem der Klassifikation der quasi-dünnen einfachen Gruppen so zu sein, daß ein Beweis wohl angekündigt, aber nie richtig veröffentlicht wurde. Inzwischen hat man aber anscheinend jemanden gefunden, der sich dieser Sache annehmen will.

Fermats Grosser Satz

Schließlich kann ich es mir selbstverständlich nicht entgehen lassen zu bemerken, daß die berühmt-berüchtigte Fermatsche Vermutung, wonach die Gleichung $x^n + y^n = z^n$ für $n \geq 3$ in natürlichen Zahlen nicht lösbar ist, im Jahre 1994 von Wiles ebenfalls mit Hilfe von elliptischen Kurven gelöst worden ist, und zwar durch den Beweis eines großen Teils der Vermutung von Taniyama-Shimura, die im wesentlichen aussagt, daß elliptische Kurven über $\mathbb{Q}$ eine tiefliegende Struktur besitzen, die von hochsymmetrischen Funktionen der oberen komplexen Halbebene (sogenannten modularen Funktionen) herrühren. Ganz grob geht der Beweis so: Man nimmt an, das Tripel (a, b, c) mit $abc \neq 0$ sei eine Lösung der Fermatgleichung $X^p + Y^p + Z^p = 0$ für $p \geq 7$. Eine der drei Zahlen a, b, c ist notwendig gerade, sagen wir b, und eine ist $\equiv 3 \bmod 4$, sagen wir a. Dann betrachten wir die elliptische Kurve $E_{a,b,c} : y^2 = x(x - a^p)(x + b^p)$. Man zeigt leicht, daß $E_{a,b,c}$ semistabil ist (was das bedeutet, werden wir später sehen). Nun hat G. Frey Anfang der 80er Jahre vermutet, daß die Kurve $E_{a,b,c}$ nicht modular sein kann (was modular ist, werden wir auch noch erklären). Serre hat daraufhin genau untersucht, was man beweisen müßte, um dies zu zeigen, und Ken Ribet hat 1986 eben dies getan. Nun gab es aber eine Vermutung von Taniyama-Shimura, die besagte, daß alle elliptischen Kurven mit Koeffizienten aus $\mathbb{Q}$ modular sein sollten; man konnte Ribets Resultat also so formulieren: die Taniyama-Shimura-Vermutung impliziert FLT. Tatsächlich würde es genügen, die Taniyama-Shimura-Vermutung nur für semistabile elliptische Kurven zu beweisen – und genau das hat Wiles getan.

Zum Beweis der Fermatschen Vermutung gibt es inzwischen eine ganze Reihe hervorragender Artikel. Für eine erste Einführung sind Cox [Co2] und Gouvea [Gou] empfehlenswert. Wer wissen will, was wirklich hinter dem Beweis steckt, sollte ein oder zwei Blicke in [CSS] werfen.

4.9.1999

Anwendungen in der Kryptographie

Wer sich für kryptographische Anwendungen der Theorie elliptischer Kurven interessiert, sollte unbedingt in Menezes [Men] und Koblitz [Kob] schauen.

1.4 Die Lemniskate, das AGM und π

Seien F_1 und F_2 Punkte in der euklidischen Ebene mit Abstand $\overline{F_1F_2} = 2c$. Die Menge aller Punkte P mit der Eigenschaft $\overline{PF_1} \cdot \overline{PF_2} = c^2$ nennt man eine Lemniskate.

Übung. Wählt man $F_1 = (-c, 0)$ und $F_2 = (c, 0)$, so erhält man die Gleichung $(x^2 + y^2)^2 = 2c^2(x^2 - y^2)$. Umrechnen in Polarkoordinaten ergibt $r^2 = 2c^2 \cos 2\theta$.

Übung. Betrachte die Lemniskate mit $2c^2 = 1$. Zeige, daß sie von den Gleichungen $2x^2 = r^2 + r^4$ und $2y^2 = r^2 - r^4$ parametrisiert wird. Rechne nach, daß $\dot{x}^2 + \dot{y}^2 = (1 - r^4)^{-1}$ gilt, und daß die Bogenlänge der Lemniskate im ersten Quadranten gleich

$$\int_0^1 \frac{dr}{\sqrt{1 - r^4}}$$

ist.

Im folgenden werden wir das Integral

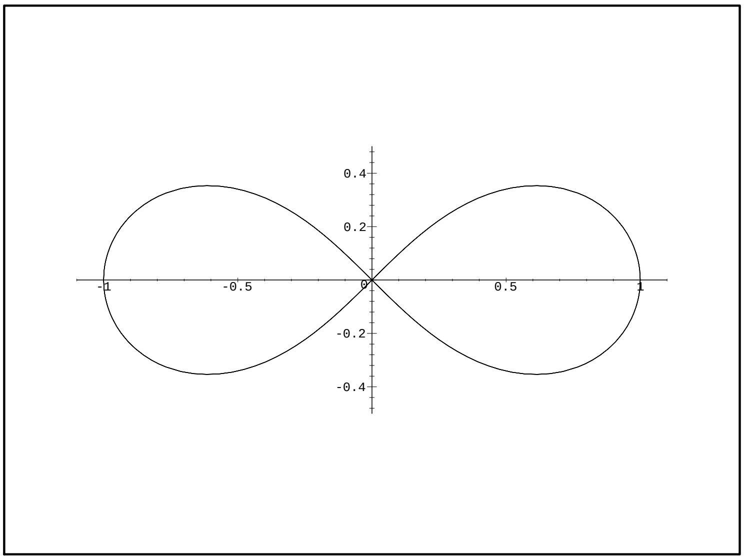
$$z(w) = \int_0^w \frac{dx}{\sqrt{1 - x^2}}\,, \qquad\qquad z(w) = \int_0^w \frac{dx}{\sqrt{1 - x^4}}\,.$$

genauer untersuchen und uns dabei an die Vorlage von Abel halten, der etwas allgemeiner die Umkehrfunktionen von Integralen $\int_0^w \frac{dx}{\sqrt{(1-c^2x^2)(1+e^2x^2)}}$ untersucht hat. Die entsprechenden Arbeiten in seinen gesammelten Werken [Ab] sind – im Gegensatz zu vielen andern Veröffentlichungen aus dieser Zeit – auch heute noch sehr gut lesbar. Ein Blick in Allings Aufbereitung [All] kann aber dennoch nicht schaden.

Die Funktion $z(w)$ ist eine wohldefinierte differenzierbare Funktion auf dem offenen Intervall $(-1, 1)$. Die Transformation $\phi : [0, 1] \longrightarrow [0, 1]$, die durch

$$x \longmapsto \frac{2t}{1 + t^2} \qquad\qquad x \longmapsto \sqrt{\frac{2t^2}{1 + t^4}}$$

Abbildung 1.9. Die Lemniskate $(x^2+y^2)^2 = x^2 - y^2$



gegeben ist, zeigt, daß

$$\int_0^1 \frac{dx}{\sqrt{1-x^2}} = 2\int_0^1 \frac{dt}{1+t^2} \qquad \int_0^1 \frac{dx}{\sqrt{1-x^4}} = \sqrt{2}\int_0^1 \frac{dt}{\sqrt{1+t^4}}$$

konvergiert. Eine numerische Integration ergibt

$$\pi = 2\int_0^1 \frac{dx}{\sqrt{1-x^2}} = 3.141592\ldots \qquad \omega = 2\int_0^1 \frac{dx}{\sqrt{1-x^4}} = 2.62205\ldots$$

Da die Funktion $z(w)$ auf dem Intervall $[-1,1]$ streng monoton steigt, existiert eine Umkehrfunktion $w(z)$, die wir mit

$$w = \sin z \qquad w = \operatorname{sin\,lemn} z =: \operatorname{sl} z$$

bezeichnen. Direkt aus der Definition folgt

$$\sin 0 = 0,\ \sin\tfrac{\pi}{2} = 1 \qquad\qquad \mathrm{sl}\, 0 = 0,\ \mathrm{sl}\,\tfrac{\omega}{2} = 1$$

Diese Funktionen sind differenzierbar, und eine einfache Rechnung liefert

$$\tfrac{d}{dz}\sin z = \sqrt{1-\sin^2 z} =: \cos z \qquad\qquad \tfrac{d}{dz}\mathrm{sl}(z) = \mathrm{f}(z)\mathrm{F}(z),$$

wobei wir $\mathrm{f}(z) = \sqrt{1-\mathrm{sl}^2(z)}$ und $\mathrm{F}(z) = \sqrt{1+\mathrm{sl}^2(z)}$ (mit positivem Vorzeichen um $z = 0$) gesetzt haben.

Übung. 4. Man zeige, daß gilt: $\frac{d}{dz}\mathrm{f}(z) = -\mathrm{sl}(z)\,\mathrm{F}(z)$, $\frac{d}{dz}\mathrm{F}(z) = \mathrm{sl}(z)\,\mathrm{f}(z)$, $\mathrm{F}(0) = 0$, $\mathrm{f}(0) = 1$, $\mathrm{f}(\frac{\omega}{2}) = 0$, $\mathrm{F}(\frac{\omega}{2}) = \sqrt{2}$. Außerdem stelle man fest, welche der Funktionen $\mathrm{sl}(z)$, $\mathrm{f}(z)$ und $\mathrm{F}(z)$ gerade, bzw. ungerade sind.

Die wichtigste Eigenschaft der Funktion $\mathrm{sl}(z)$ ist ihr Additionsgesetz. Um es herzuleiten, setzen wir

$$\varpi = \pi, \qquad\qquad \varpi = \omega,$$

und betrachten $D = \{\alpha,\beta \in \mathbb{R}^2 : -\frac{\varpi}{2} \le \alpha,\beta,\alpha+\beta \le \frac{\varpi}{2}\}$. Die durch $g(\alpha,\beta) =$

$$\sin\alpha\cos\beta + \cos\alpha\sin\beta, \qquad\qquad \frac{\mathrm{sl}(\alpha)\mathrm{f}(\beta)\mathrm{F}(\beta) + \mathrm{sl}(\beta)\mathrm{f}(\alpha)\mathrm{F}(\alpha)}{1+\mathrm{sl}^2(\alpha)\,\mathrm{sl}^2(\beta)}.$$

definierte Funktion $g : D \longrightarrow \mathbb{R}$ ist wohldefiniert (und, wie man zugeben muß, vom Himmel gefallen. Bei Siegel [Sie] kann man nachlesen, was dahinter steckt). Führt man die neuen Variablen $\gamma = \frac{\alpha+\beta}{2}$ und $\delta = \frac{\alpha-\beta}{2}$ ein, so findet man $\partial g/\partial\delta = 0$, mit anderen Worten: $g(\gamma,\delta)$ hängt nur von γ ab. Indem wir g an der Stelle $\delta = \gamma$ auswerten, finden wir

$$g(\gamma) = \sin 2\gamma, \qquad\qquad g(\gamma) = \mathrm{sl}\, 2\gamma.$$

Führen wir jetzt wieder die Variablen α und β ein, so folgen die Additionsformeln:

$$\boxed{\begin{aligned}
\sin(\alpha+\beta) &= \sin\alpha\cos\beta+\cos\alpha\sin\beta,\\
\cos(\alpha+\beta) &= \cos\alpha\cos\beta-\sin\alpha\sin\beta,\\
\mathrm{sl}(\alpha+\beta) &= \frac{\mathrm{sl}(\alpha)\mathrm{f}(\beta)\mathrm{F}(\beta)+\mathrm{sl}(\beta)\mathrm{f}(\alpha)\mathrm{F}(\alpha)}{1+\mathrm{sl}^2(\alpha)\,\mathrm{sl}^2(\beta)},\\
\mathrm{f}(\alpha+\beta) &= \frac{\mathrm{f}(\alpha)\mathrm{f}(\beta)-\mathrm{sl}(\alpha)\mathrm{sl}(\beta)\mathrm{F}(\alpha)\mathrm{F}(\beta)}{1+\mathrm{sl}^2(\alpha)\,\mathrm{sl}^2(\beta)},\\
\mathrm{F}(\alpha+\beta) &= \frac{\mathrm{F}(\alpha)\mathrm{F}(\beta)+\mathrm{sl}(\alpha)\mathrm{sl}(\beta)\mathrm{f}(\alpha)\mathrm{f}(\beta)}{1+\mathrm{sl}^2(\alpha)\,\mathrm{sl}^2(\beta)}.
\end{aligned}}$$

Indem man zu $\alpha=\beta$ spezialisiert, folgen daraus die Verdoppelungsformeln:

$$\boxed{\begin{aligned}
\sin(2\alpha) &= 2\sin\alpha\cos\alpha, & \mathrm{sl}(2\alpha) &= \frac{2\mathrm{sl}(\alpha)\mathrm{f}(\alpha)\mathrm{F}(\alpha)}{1+\mathrm{sl}(\alpha)^4},\\
\cos(2\alpha) &= \cos^2\alpha-\sin^2\alpha, & \mathrm{f}(2\alpha) &= \frac{\mathrm{f}(\alpha)^2-\mathrm{sl}(\alpha)^2\mathrm{F}(\alpha)^2}{1+\mathrm{sl}(\alpha)^4},\\
& & \mathrm{F}(2\alpha) &= \frac{\mathrm{F}(\alpha)^2+\mathrm{sl}(\alpha)^2\mathrm{f}(\alpha)^2}{1+\mathrm{sl}(\alpha)^4}.
\end{aligned}}$$

Diese Verdoppelungsformeln stammen bereits von Fagnano; natürlich hat er sie mit Integralen ausdrücken müssen, weil er die Umkehrfunktionen nicht besaß.

Übung. Zeige, daß die Verdoppelungsformel für $\mathrm{sl}(z)$ zur Aussage

$$2\int_0^w \frac{dx}{\sqrt{1-x^4}} = \int_0^v \frac{dx}{\sqrt{1-x^4}} \quad \text{mit} \quad v = 2w\frac{\sqrt{1-w^4}}{1+w^4}$$

äquivalent ist.

Setzt man in den Additionsformeln dagegen $\beta=\frac{1}{2}\varpi$, so erhält man

$$\sin(\alpha+\tfrac{\pi}{2}) = \cos\alpha, \qquad\qquad \mathrm{sl}(\alpha+\tfrac{\omega}{2}) = \tfrac{\mathrm{f}(\alpha)}{\mathrm{F}(\alpha)} =: \mathrm{cl}\,\alpha.$$

Diese beiden Formeln haben wir streng betrachtet nur für $-\frac{1}{2}\varpi\le\alpha\le 0$ bewiesen. Für $0\le\alpha\le\frac{\omega}{2}$ können wir sie aber andererseits benutzen, um $\sin(z)$ und $\mathrm{sl}(z)$ auf das Intervall $[0,\frac{\varpi}{2}]$ fortzusetzen. Indem wir mit $\cos(z)$, $\mathrm{f}(z)$ und $\mathrm{F}(z)$ ähnlich verfahren und den Prozeß wiederholen, können wir diese Funktionen so auf die ganze reelle Achse ausdehnen.

Jetzt muß man sich davon überzeugen, daß die Additonsformeln auch für die so erweiterten Funktionen gültig sind und nachrechnen, daß die Funktionen (außerhalb etwaiger Polstellen) differenzierbar sind. Weiter findet man, daß $\sin z$ eine 2π-periodische und $\mathrm{sl}\, z$ eine 2ω-periodische Funktion ist. Ableiten zeigt dasselbe für $\cos z$, $\mathrm{f}(z)$ und $\mathrm{F}(z)$. Schließlich gelten folgende Beziehungen:

$$\sin^2 z + \cos^2 z = 1, \qquad \mathrm{sl}^2 z + \mathrm{cl}^2 z + \mathrm{sl}^2 z\, \mathrm{cl}^2 z = 1.$$

Diejenige für $\sin z$ und $\cos z$ folgt direkt aus der Definition, im Fall der lemniskatischen Funktionen folgt die Behauptung so: $\mathrm{sl}^2 z + \mathrm{cl}^2 z + \mathrm{sl}^2 z\, \mathrm{cl}^2 z = (1+\mathrm{sl}^2 z)^{-1}((1+\mathrm{sl}^2 z)\,\mathrm{sl}^2 z + (1-\mathrm{sl}^2 z) + (1-\mathrm{sl}^2 z)\,\mathrm{sl}^2 z) = 1$.

Übung. Man verifiziere folgende Tabelle von Funktionswerten:

z	$\sin z$	$\cos z$	$\mathrm{sl}\, z$	$\mathrm{f}(z)$	$\mathrm{F}(z)$
$m\varpi$	0	$(-1)^m$	0	$(-1)^m$	1
$(m+\frac12)\varpi$	$(-1)^m$	0	$(-1)^m$	0	$\sqrt{2}$

Das nächste Ziel sind entsprechende Halbierungsformeln. Dazu definieren wir

$$x = \sin\tfrac{\alpha}{2}, \quad y = \cos\tfrac{\alpha}{2}, \qquad x = \mathrm{sl}(\tfrac{\alpha}{2}),\ y = \mathrm{f}(\tfrac{\alpha}{2}),\ z = \mathrm{F}(\tfrac{\alpha}{2}).$$

Damit gilt

$$y^2 = 1 - x^2, \qquad y^2 = 1 - x^2,\ z^2 = 1 + x^2$$

Jetzt ersetzen wir α durch $\alpha/2$ in den Verdopplungsformeln; das liefert

$$\cos\alpha = y^2 - x^2, \quad \sin\alpha = 2xy \qquad \mathrm{f}(\alpha) = \tfrac{y^2 - x^2z^2}{1+x^4} = \tfrac{1-2x^2-x^4}{1+x^4},\quad \mathrm{F}(\alpha) = \tfrac{z^2+x^2y^2}{1+x^4} = \tfrac{1+2x^2-x^4}{1+x^4},$$

Auflösen nach x und y gibt die Identitäten

$$x^2 = \frac{1-\cos\alpha}{2}, \quad y^2 = \frac{1+\cos\alpha}{2} \qquad x^2 = \tfrac{1-\mathrm{f}(\alpha)}{1+\mathrm{F}(\alpha)} = \tfrac{\mathrm{F}(\alpha)-1}{1+\mathrm{f}(\alpha)},\quad y^2 = \tfrac{\mathrm{f}(\alpha)+\mathrm{F}(\alpha)}{1+\mathrm{F}(\alpha)},\ z^2 = \tfrac{\mathrm{f}(\alpha)+\mathrm{F}(\alpha)}{1+\mathrm{f}(\alpha)}.$$

Es ist eine einfache Übungsaufgabe zu zeigen, daß die Ausdrücke auf der rechten Seite alle nichtnegativ sind. Damit können wir problemlos die Quadratwurzel ziehen und finden

$$\sin\tfrac{\alpha}{2} = \sqrt{\tfrac{1-\cos\alpha}{2}}\,, \qquad \mathrm{sl}\,(\tfrac{\alpha}{2}) = \sqrt{\tfrac{1-\mathrm{f}(\alpha)}{1+\mathrm{F}(\alpha)}}\,,$$

$$\mathrm{f}(\tfrac{\alpha}{2}) = \sqrt{\tfrac{\mathrm{f}(\alpha)+\mathrm{F}(\alpha)}{1+\mathrm{F}(\alpha)}}\,,$$

$$\cos\tfrac{\alpha}{2} = \sqrt{\tfrac{1+\cos\alpha}{2}}\,, \qquad \mathrm{F}(\tfrac{\alpha}{2}) = \sqrt{\tfrac{\mathrm{f}(\alpha)+\mathrm{F}(\alpha)}{1+\mathrm{f}(\alpha)}}\,.$$

Die Vorzeichen der Quadratwurzeln werden so bestimmt, daß $\mathrm{sl}\,x \geq 0$ ist für $x \in [0,\omega]$ und $\mathrm{sl}\,x \leq 0$ für $x \in [\omega, 2\omega]$ etc. Insbesondere finden wir folgende Werte:

$$\sin(\tfrac{\pi}{4}) = \tfrac{1}{2}\sqrt{2}, \qquad \mathrm{sl}(\tfrac{\omega}{4}) = \sqrt{\sqrt{2}-1},$$

$$\cos(\tfrac{\pi}{4}) = \tfrac{1}{2}\sqrt{2}. \qquad \mathrm{f}(\tfrac{\omega}{4}) = \sqrt{2-\sqrt{2}},\ \mathrm{F}(\tfrac{\omega}{4}) = \sqrt[4]{2}.$$

Man beachte, daß dies alles algebraische Zahlen sind. Weiter erzeugen diese Zahlen abelsche Erweiterungen von $\mathbb{Q}$ bzw. $\mathbb{Q}(i)$, und sie können mit Zirkel und Lineal konstruiert werden. Bekanntlich hat Gauss gezeigt, daß man den Kreis mit Zirkel und Lineal in $p = 2^{2^n}+1$ Teile teilen kann, sofern p nur prim ist; er hat auch angedeutet, daß sich dies für $p = 5$ auf den Lemniskatenfall überträgt. Das Analogon zum Kreisteilungssatz von Gauss wurde aber erst von Abel bewiesen – siehe dazu Rosens [Ros] Darstellung von Abels Beweis, daß sich die Lemniskate mit Zirkel und Lineal fünfteilen läßt (die Arbeit ist nicht ohne gewisse Mängel: einerseits wird so getan, als enthalte die Arbeit eine elementare Herleitung der Abelschen Ergebnisse, andererseits werden wesentliche (und nicht-elementare) Ingredienzien ohne Beweis zitiert).

Abels nächster Schritt besteht darin, die bisher betrachteten Funktionen von der reellen Achse auf die ganze komplexe Ebene auszudehnen. Dazu werden die Funktionen zuerst auf die imaginäre y-Achse ausgedehnt; für beliebiges $x + iy$ erhält man die Funktionswerte dann aus dem Additionstheorem!

Um zu sehen, wie wir unsere Funktionen auf der imaginären Achse zu definieren haben, setzen wir – ohne uns um die Bedeutung zu kümmern – in den Ausgangsintegralen $x = -iu$ und finden

$$iz = \int_0^{iw} \frac{du}{\sqrt{1+u^2}}\,, \qquad iz = \int_0^{iw} \frac{dx}{\sqrt{1-u^4}}\,.$$

Also setzen wir

$$\sin(iz) = i\sinh z, \qquad \mathrm{sl}(iz) = i\mathrm{sl}(z).$$

Hierbei ist $\sinh z$ als Umkehrfunktion des Integrals $z = \int_0^w dx/\sqrt{1+x^2}$ definiert.

Mittels der Additionsformeln können wir $\sin z$ und $\mathrm{sl}\, z$ nun zu Funktionen $\mathbb{C} \longrightarrow \mathbb{C} \cup \{\infty\}$ auf der ganzen komplexen Ebene machen. Da wir den Funktionswert an der Stelle $x + iy$ durch die Additionsformel für x und iy gewonnen haben, muß als nächstes gezeigt werden, daß die Additionsformel allgemein für die Addition von Funktionswerten an $x + iy$ und $u + iv$ gilt. Ebenso muß man sich davon überzeugen, daß die so definierten Funktionen außerhalb ihrer Pole differenzierbar (genauer: holomorph) sind, und daß die Pole diskret in $\mathbb{C}$ liegen, d.h. daß in jeder kompakten Teilmenge von $\mathbb{C}$ nur endlich viele Pole liegen.

Die Funktionen f und F werden jetzt durch $\mathrm{f}(iz) = \mathrm{F}(z)$ und $\mathrm{F}(iz) = \mathrm{f}(z)$ fortgesetzt; alle bisher bewiesenen Relationen bleiben damit weiter gültig, und zwar für alle $z \in \mathbb{C}$. Die folgende Tabelle gibt einige Funktionswerte:

z	$\mathrm{sl}\, z$	$\mathrm{f}(z)$	$\mathrm{F}(z)$
$m\omega + n\omega i$	0	$(-1)^m$	$(-1)^n$
$(m+\frac{1}{2})\omega + n\omega i$	$(-1)^{m+n}$	0	$(-1)^n\sqrt{2}$
$m\omega + (n+\frac{1}{2})\omega i$	$(-1)^{m+n} i$	$(-1)^n\sqrt{2}$	0
$(m+\frac{1}{2})\omega + (n+\frac{1}{2})\omega i$	∞	∞	∞

Ein genaueres Studium der Funktionen zeigt, daß $\mathrm{sl}(z)$ nur die Polstellen hat, die sich aus dieser Tabelle ergeben. Die dazugehörigen Rechnungen sind ganz elementar und können bei Abel nachgelesen werden.

Per definitionem haben die Funktionen sl, cl, f und F die beiden Perioden 2ω und $2\omega i$; wir behaupten, daß $\mathrm{sl}\, z$ und $\mathrm{cl}\, z$ tatsächlich $(1+i)\omega$ und $(1-i)\omega$ als Perioden besitzen. Die Verifikation besteht in ein paar einfachen Rechnungen:

$$\begin{aligned}
\mathrm{sl}(\alpha+\omega) &= -\mathrm{sl}(\alpha), & \mathrm{sl}(\alpha+i\omega) &= -\mathrm{sl}(\alpha),\\
\mathrm{f}(\alpha+\omega) &= -\mathrm{f}(\alpha), & \mathrm{f}(\alpha+i\omega) &= \mathrm{f}(\alpha),\\
\mathrm{F}(\alpha+\omega) &= \mathrm{F}(\alpha), & \mathrm{F}(\alpha+i\omega) &= -\mathrm{F}(\alpha),
\end{aligned}$$

und jetzt liefert die Additionsformel $\mathrm{sl}(\alpha + (1+i)\omega) = \mathrm{sl}(\alpha)$ wie behauptet. Ähnlich zeigt man, daß $\mathrm{cl}(\alpha+(1+i)\omega) = \mathrm{cl}(\alpha)$, und daß f und F bei Addition von $(1+i)\omega$ ihr Vorzeichen wechseln. Wer Spaß am Rechnen hat, mag sehen, ob er folgende Ergebnisse beweisen kann:

$$\begin{array}{rclrcl}
\mathrm{sl}\big(\alpha+\frac{\omega}{2}\big) &=& \frac{\mathrm{f}(\alpha)}{\mathrm{F}(\alpha)} & \mathrm{sl}\big(\alpha+\frac{i\omega}{2}\big) &=& i\frac{\mathrm{f}(\alpha)\mathrm{F}(\alpha)}{1-\mathrm{sl}(\alpha)^2} \\
\mathrm{f}\big(\alpha+\frac{\omega}{2}\big) &=& -\sqrt{2}\,\frac{\mathrm{sl}(\alpha)\mathrm{F}(\alpha)}{1+\mathrm{sl}(\alpha)^2} & \mathrm{f}\big(\alpha+\frac{i\omega}{2}\big) &=& \sqrt{2}\,\frac{\mathrm{f}(\alpha)}{1-\mathrm{sl}(\alpha)^2} \\
\mathrm{F}\big(\alpha+\frac{\omega}{2}\big) &=& \sqrt{2}\,\frac{\mathrm{F}(\alpha)}{1+\mathrm{sl}(\alpha)^2} & \mathrm{F}\big(\alpha+\frac{i\omega}{2}\big) &=& i\sqrt{2}\,\frac{\mathrm{sl}(\alpha)\mathrm{f}(\alpha)}{1-\mathrm{sl}(\alpha)^2}
\end{array}$$

1.5 Die Weierstraßsche $\wp$-Funktion

Die im letzten Abschnitt definierten Funktionen $\mathrm{sl}(z)$ und $\mathrm{cl}(z)$ sind Beispiele für differenzierbare doppeltperiodische komplexwertige Funktionen: solche Funktionen heißen seit Abel und Jacobi *elliptische Funktionen.* Es hat sich herausgestellt, daß Additionstheoreme wie für $\mathrm{sl}(z)$ ganz allgemein für solche elliptischen Funktionen existieren, und im folgenden wollen wir eine knappe Übersicht über deren elementarste Eigenschaften gewinnen.

Es gibt eine Unmenge elliptischer Funktionen und noch mehr Beziehungen zwischen ihnen (man schaue sich nur einmal die bescheidene Formelsammlung in Lawden [Law] an). Die von Weierstraß eingeführte $\wp$-Funktion hat sich als "Grundfunktion" aber inzwischen durchgesetzt. Um sie zu definieren, gehen wir von einem Gitter Λ in $\mathbb{C}$ aus; darunter verstehen wir die Menge der $\mathbb{Z}$-Linearkombination zweier über den reellen Zahlen unabhängiger komplexen Zahlen, also die Menge $\Lambda = \mathbb{Z}\omega_1 \oplus \mathbb{Z}\omega_2 = \{a\omega_1 + b\omega_2 : a, b \in \mathbb{Z}\}$, wobei $\omega_1, \omega_2 \in \mathbb{C}^\times$ und ω_1/ω_2 nicht reell ist. Gitter sind also insbesondere additive Gruppen.

Jedem solchen Gitter ordnet Weierstraß nun die Funktion

$$\wp(z) = \wp(z;\Lambda) = \frac{1}{z^2} + \sum{}'\Big(\frac{1}{(z-\lambda)^2} - \frac{1}{\lambda^2}\Big), \tag{1.5}$$

zu, wobei der Strich am Summenzeichen andeuten soll, daß nur über die $\lambda \in \Lambda \setminus \{0\}$ summiert wird.

Übung. Das eindimensionale Analogon zu $\wp(z)$ ist die durch

$$c(z) = \frac{1}{z} + \sum{}'\Big(\frac{1}{z-\lambda} - \frac{1}{\lambda}\Big)$$

definierte Funktion $c(z)$, wobei hier das "Gitter" $\Lambda = \mathbb{Z}$ in $\mathbb{R}$ zugrunde liegt. Man beweise die absolute Konvergenz von $c(z)$ für alle $z \in \mathbb{R}\setminus\mathbb{Z}$, rechne nach, daß $c'(z)$ $\mathbb{Z}$-periodisch ist, und benutze das, um die $\mathbb{Z}$-Periodizität von $c(z)$ zu beweisen.

Faßt man die Summanden für n und $-n$ zusammen und interpretiert $\frac{2z}{z^2-n^2}$ als die logarithmische Ableitung von $1-(z/n)^2$, so folgt, daß $c(z)$ die logarithmische Ableitung von $s(z) = z\prod'(1-z^2/n^2)$ ist. Diese $\mathbb{Z}$-periodische Funktion hat Nullstellen genau in $\mathbb{Z}$ – wer kann sie identifizieren?

Behauptet wird nun, daß $\wp(z)$ eine elliptische Funktion mit Periodengitter Λ ist, d.h. daß erstens $\wp(z)$ auf kompakten Mengen außerhalb von Λ gleichmäßig konvergiert, und zweitens für alle $\lambda \in \Lambda$ die Gleichung $\wp(z+\lambda) = \wp(z)$ gelten soll. Das erste ist eine Standardabschätzung, die zweite Behauptung benutzt einen Trick: man zeigt, daß die Ableitung von $\wp(z)$ Periodengitter Λ hat; folglich ist die Differenz $\wp(z) - \wp(z+\lambda)$ für $\lambda \in \Lambda$ konstant, und Einsetzen von $z = \lambda/2$ zeigt, daß die Differenz verschwindet.

Übung. Bekanntlich konvergiert $\sum_{n\in\mathbb{N}} n^{-s}$ genau dann, wenn $s > 1$ ist; das zweidimensionale Analogon hiervon ist: $\sum'\lambda^{-s}$ konvergiert genau für $s > 2$. Man beweise das. (Hinweis: beidesmal benutzt man das Integralkriterium.)

Übung. Benutze die geometrische Reihe, um $(z-\lambda)^{-2}$ als Potenzreihe in z zu schreiben; zeige, daß $\wp(z)$ außerhalb von Λ absolut konvergiert, und daß

$$\wp(z) = \frac{1}{z^2} + 3G_4z^2 + 5G_6z^4 + 7G_8z^6 + \ldots$$

gilt mit

$$G_m = G_m(\Lambda) = \sum{}'\lambda^{-m}.$$

Zeige, daß $\wp'(z)$ und $\wp(z)$ Λ-periodische Funktionen sind.

Damit hat man zwei Λ-periodische Funktionen $\wp$ und $\wp'$. Offenbar ist jedes Polynom in $\wp$ und $\wp'$ ebenfalls Λ-periodisch; dasselbe gilt sogar für jede rationale Funktion in $\wp$ und $\wp'$, also für jede Funktion der Form $P(\wp,\wp')/Q(\wp,\wp')$ mit nicht überall verschwindenden Polynomen $P, Q \in \mathbb{C}[X,Y]$. Übrigens gilt davon sogar die Umkehrung: jede Λ-periodische Funktion ist von dieser Gestalt.

Die $\wp$-Funktion genügt nun der Differentialgleichung

$$\wp'(z)^2 = 4\wp(z)^3 - g_2\wp(z) - g_3,$$

wo $g_2 = 60G_4$ und $g_3 = 140G_6$ ist.

Um dies nachzuweisen, müssen wir einen Satz von Liouville benutzen:

Satz 1.4. (Liouville) *Jede elliptische Funktion ohne Pol ist konstant.*

Der Rest ist Rechnung: man zeigt (Übung!)

$$\begin{aligned}\wp'(z) &= -2z^{-3} + 6G_4 z + 20G_6 z^3 + \ldots\,; \\ \wp'(z)^2 &= 4z^{-6} - 24G_4 z^{-2} - 80G_6 + \ldots\,; \\ 4\wp(z)^3 &= 4z^{-6} + 36G_4 z^{-2} + 60G_6 + \ldots\,;\end{aligned}$$

Daraus folgt sofort, daß

$$\wp'(z)^2 - 4\wp(z)^3 + 60G_4\wp(z) - 140G_6 = z * (\text{Potenzreihe})$$

ist. Da beide Seiten außerhalb von Λ auf ganz $\mathbb{C}$ konvergieren, muß die Potenzreihe auf der rechten Seite Konvergenzradius ∞ besitzen. Damit ist die rechte Seite eine beschränkte (warum?) elliptische Funktion, nach dem zitierten Satz von Liouville also konstant; Einsetzen von $z = 0$ liefert jetzt die Behauptung.

Übung. Setze $\wp(z) = (\mathrm{sl}\, z)^{-2}$ und zeige, daß $\wp$ der Differentialgleichung $\wp'(z)^2 = 4\wp(z)^3 - 4\wp(z)$ genügt.

Als nächstes werden wir noch benötigen, daß das Polynom $4x^3 - g_2x - g_3$ für kein Λ mehrfache Nullstellen hat; dies zeigt man durch ein genaueres Studium der Nullstellen von $\wp'$ und ist nicht schwer.

Schließlich betrachten wir noch die Abbildung

$$\phi_\Lambda : z \longmapsto (\wp(z), \wp'(z)).$$

Da die Funktionen $\wp$ und $\wp'$ Pole besitzen, geht diese Abbildung von $\mathbb{C}$ nach $\overline{\mathbb{C}} \times \overline{\mathbb{C}}$ mit $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$. Da $\wp$ und $\wp'$ beide Λ-periodisch sind, hängt das Bild nur von der Restklasse $z + \Lambda$ ab, d.h. ϕ_Λ induziert eine wohldefinierte Abbildung $\phi : \mathbb{C}/\Lambda \longrightarrow \overline{\mathbb{C}} \times \overline{\mathbb{C}}$. Aber auch das Bild können wir einschränken: wegen der Differentialgleichung, der $\wp$ und $\wp'$ genügen, kommen nicht alle Punkte $(x, y) \in \overline{\mathbb{C}} \times \overline{\mathbb{C}}$ als Bild vor, sondern außer den Polen nur diejenigen, die der Gleichung $y^2 = 4x^3 - g_2x - g_3$ genügen; die Menge dieser Punkte (zusammen mit dem Punkt $\phi(0) = (\infty, \infty)$) sei mit $E(\mathbb{C})$ bezeichnet. Jetzt gilt

Satz 1.5. *Die Abbildung*

$$\phi : \mathbb{C}/\Lambda \longrightarrow E(\mathbb{C}) : z \longmapsto (\wp(z), \wp'(z))$$

ist bijektiv.

Damit haben wir eine Bijektion zwischen der Gruppe $\mathbb{C}/\Lambda$ und den Punkten auf $E(\mathbb{C})$; durch Strukturtransport wird daher $E(\mathbb{C})$ ebenfalls eine Gruppe: um $P_1 = (x_1, y_1)$ und $P_2 = (x_2, y_2)$ zu addieren, bestimmt man deren Urbilder unter der Bijektion, addiert diese in der Gruppe $\mathbb{C}/\Lambda$, und bildet die Summe via ϕ wieder nach $E(\mathbb{C})$ ab.

Damit haben wir eine Erklärung dafür, daß wir mit unserer Sekanten-Tangenten-Methode (das englische "chord-tangent method" wird auch mit Sehnen-Tangenten-Methode übersetzt) kein neutrales Element gefunden haben: uns fehlte der unendlich ferne Punkt. In der Tat, das Nullelement von $E(\mathbb{C})$ ist das Bild von $0 + \Lambda$; da $\wp$ und $\wp'$ in $z = 0$ einen Pol haben, ist $\phi(0) = (\infty, \infty)$ der unendlich ferne Punkt.

Ist $P = (x, y)$ auf $E(\mathbb{C})$ gegeben, so können wir nun ohne weiteres angeben, was $-P$ sein soll: dazu wählen wir ein Urbild $z + \Lambda$ von P; dessen Negatives ist $-z + \Lambda$, und unter ϕ landet dies auf $(\wp(-z), \wp'(-z))$. Da $\wp$ gerade, $\wp'$ ungerade ist, haben wir $-P = (x, -y)$. Das Negative eines Punktes erhält man also einfach durch Spiegeln an der x-Achse.

Um in unserer Gruppe $E(\mathbb{C})$ rechnen zu können, müssen wir wissen, wie man $\wp(z + w)$ und $\wp'(z + w)$ berechnet. Dies leisten die Additionstheoreme für die Weierstraßsche $\wp$-Funktion: es gilt nämlich

$$\wp(z + w) = -\wp(z) - \wp(w) + \frac{1}{4}\Big(\frac{\wp'(z) - \wp'(w)}{\wp(z) - \wp(w)}\Big)^2,$$

falls $z, w, z + w \in \mathbb{C} \setminus \Lambda$ ist. Spezialisieren zu $z = w$ (und Anwenden von L'Hôpital) liefert die Verdoppelungsformel

$$\wp(2z) = -2\wp(z) + \frac{1}{4}\Big(\frac{\wp''(z)}{\wp'(z)}\Big)^2.$$

Die Beweisidee ist wie gehabt: man setzt

$$h(z) = \wp(z + w) + \wp(z) + \wp(w) - \frac{1}{4}\Big(\frac{\wp'(z) - \wp'(w)}{\wp(z) - \wp(w)}\Big)^2$$

und zeigt dann, daß h keinen Pol hat und in $z = 0$ verschwindet.

Was hat nun die Addition mittels der $\wp$-Funktion mit der Sekanten-Tangenten-Methode zu tun? Eine ganze Menge. Es gilt nämlich der

Satz 1.6. *Sei $\wp$ die Weierstraßsche $\wp$-Funktion zum Gitter Λ, welche der Differentialgleichung*

$$\wp'(z)^2 = 4\wp(z)^3 - g_2\wp(z) - g_3$$

genügt. Sind $z, w \in \mathbb{C} \setminus \Lambda$, so liegen $\phi(z)$ und $\phi(w)$ auf der Kurve

$$C : y^2 = 4x^3 - g_2 x - g_3.$$

Der dritte Schnittpunkt der Geraden durch $\phi(z)$ und $\phi(w)$ hat dann die Koordinaten $(\wp(z+w), -\wp'(z+w)) = \phi(-z-w)$.

Mit anderen Worten: man addiert die Punkte $\phi(z)$ und $\phi(w)$ auf C, indem man mit der Sekanten-Methode den dritten Schnittpunkt bestimmt und dann an der x-Achse spiegelt.

Beweis. Die Gleichung der Gerade durch $\phi(z)$ und $\phi(w)$ ist

$$y = m(x - \wp(z)) + \wp'(z) \quad \text{mit} \quad m = \frac{\wp'(z) - \wp'(w)}{\wp(z) - \wp(w)}.$$

Schneiden mit C liefert die Gleichung

$$4x^3 - (m(x - \wp(w)) + \wp'(w))^2 - g_2 x - g_3 = 0.$$

Zwei Nullstellen sind durch $x_1 = \wp(z)$ und $x_2 = \wp(w)$ gegeben; Addition aller Nullstellen ergibt nach Vieta das -4-fache des Koeffizienten von x^2, und wir finden

$$x_3 = -x_1 - x_2 + \frac{m^2}{4}. \tag{1.6}$$

Da (x_3, y_3) auf C liegt, gibt es ein $u \in \mathbb{C}/\Lambda$ mit $x_3 = \wp(u)$, und (1.6) zeigt

$$\wp(u) = -\wp(z) - \wp(w) + \frac{1}{4}\Big(\frac{\wp'(z) - \wp'(w)}{\wp(z) - \wp(w)}\Big)^2,$$

d.h. es ist in der Tat $\wp(z+w) = x_3$.

Da (x_3, y_3) auf der Kurve liegt, muß $y_3 = \pm\wp(z+w)$ gelten, und die Frage ist: welches Vorzeichen gilt? Dazu setzen wir $x = x_3$ in der Geradengleichung und erhalten $y_3 - \wp'(w) = m(\wp(z+w) - \wp(w))$. Setzt man $z = 0$ auf der linken Seite, erhält man 0, falls das positive Vorzeichen gilt, und $-2\wp'(w)$ andernfalls. Auf der rechten Seite kann man leider nicht einfach $z = 0$ setzen, aber wir können sagen, was passiert, wenn $z \longrightarrow 0$ geht. Dann wird nämlich für eine geeignete Potenzreihe $P(z)$

$$\begin{aligned} m &= \frac{\wp'(z) - \wp'(w)}{\wp(z) - \wp(w)} = \frac{-2z^{-3} - \wp'(w) + P'(z)}{z^{-2} - \wp(w) + P(z)} \\ &= \frac{-2}{z}\,\frac{1 - z^2\wp'(w) + z^2 P'(z)}{1 - z^2\wp(w) + z^2 P(z)}, \end{aligned}$$

also

$$m(\wp(z+w)-\wp(w)) = -\frac{2-z^3\wp'(w)+z^3P'(z)}{1-z^2\wp(w)+z^2P(z)}\,\frac{\wp(z+w)-\wp(w)}{z}.$$

Läßt man nun $z \longrightarrow 0$ gehen, konvergiert dieser Ausdruck gegen $-2\wp'(w)$. Also gilt das negative Vorzeichen.

Der Fall $z = w$ ist Übungsaufgabe. $\square$

Ein wesentlich eleganterer Beweis ist der folgende (der, wie so oft bei eleganten Beweisen, aus dem Hut gezaubert wird): ersetze in der Additionsformel w durch $-w$ und dann z durch $z+w$. Vergleich mit dem Original liefert

$$\Big(\frac{\wp'(z+w)+\wp'(w)}{\wp(z+w)-\wp(w)}\Big)^2 = \Big(\frac{\wp'(z)-\wp'(w)}{\wp(z)-\wp(w)}\Big)^2.$$

Daher stimmen beide Klammern bis auf das Vorzeichen überein. Weil aber der Quotient beider Klammern eine Λ-periodische elliptische Funktion ist, hängt das Vorzeichen nicht von z oder w ab (denn die Funktion nimmt höchstens zwei Werte an, ist also beschränkt, folglich konstant nach dem Satz von Liouville). Einsetzen von $w = -2z$ zeigt, daß das negative Vorzeichen gilt.

Die so erhaltene Identität läßt sich als Determinantengleichung schreiben, nämlich so:

$$\begin{vmatrix} 1 & \wp(z+w) & -\wp'(z+w) \\ 1 & \wp(z) & \wp'(z) \\ 1 & \wp(w) & \wp'(w) \end{vmatrix} = 1.$$

Diese Gleichung beschreibt aber bekanntlich eine Gerade durch die Punkte $(\wp(z),\wp'(z))$, $(\wp(w),\wp'(w))$ und $(\wp(z+w),-\wp'(z+w))$.

Neben den Invarianten g_2 und g_3 eines Gitters Λ sind noch zwei andere Größen wichtig. Zum einen ist dies die "Diskriminante"

$$\Delta(\Lambda) = g_2(\Lambda)^3 - 27g_3(\Lambda)^2;$$

man kann zeigen, daß $\Delta(\Lambda) \neq 0$ für jedes Gitter Λ in $\mathbb{C}$ gilt. Weiter gibt es zu jedem Paar $g_2, g_3 \in \mathbb{C}$ mit $\Delta \neq 0$ ein Gitter mit eben diesen Invarianten.

Nicht verschwindende Invarianten haben die Tendenz, in irgendwelchen Nennern aufzutauchen; das ist auch hier so: die modulare Invariante j ist definiert durch

$$j(\Lambda) = \frac{g_2(\Lambda)^3}{\Delta(\Lambda)}.$$

Diese Invariante ist uns bereits begegnet: setzt man $\Lambda_\tau = \mathbb{Z} + \tau\mathbb{Z}$ für ein τ der oberen Halbebene, so kann man j als Funktion von τ auffassen. Setzt man dann noch $q = e^{2\pi i\tau}$, so erhält man die j-Invariante aus Abschnitt 1.3; dort war übrigens $\tau = \frac{1}{2}(1 + \sqrt{d})$.

Die j-Invariante dient dazu, im wesentlichen gleiche Gitter als solche zu erkennen. Zwei Gitter Λ und Λ' heißen ähnlich, wenn es ein $\alpha \in \mathbb{C}^\times$ gibt mit $\Lambda' = \alpha\Lambda$. Man kann zeigen, daß zwei Gitter genau dann ähnlich sind, wenn ihre j-Invarianten übereinstimmen.

Ist Λ ein Gitter und $m \in \mathbb{Z}$, dann gilt natürlich $m\Lambda \subseteq \Lambda$. Man sagt, Λ habe komplexe Multiplikation, wenn es ein $\tau \in \mathbb{C}^\times \setminus \mathbb{Z}$ gibt mit $\tau\Lambda \subseteq \Lambda$.

Übung. Sei m eine natürliche Zahl und $\Lambda = \mathbb{Z} \oplus \sqrt{-m}\,\mathbb{Z}$. Man zeige, daß Λ komplexe Multiplikation hat.

Übung. Das Gitter Λ habe komplexe Multiplikation mit $\tau \in \mathbb{C}^\times \setminus \mathbb{Z}$. Zeige, daß τ einer quadratischen Gleichung mit ganzen Koeffizienten und negativer Diskriminante genügt.

Der Einheitskreis

Einiges von dem, was wir im letzten Abschnitt über komplexwertige doppeltperiodische Funktionen erzählt haben, hat ein eindimensionales Analogon: wir können nämlich $\Lambda = \mathbb{Z}$ als Gitter in $\mathbb{R}$ auffassen. Als Λ-periodische Funktion nehmen wir $\sin 2\pi z$ und $\cos 2\pi z$, und wenn wir $x = \cos 2\pi z$ und $y = \sin 2\pi z$ setzen, haben wir eine analytische Parametrisierung $\phi : \mathbb{R}/\mathbb{Z} \longrightarrow C$ des Einheitskreises $C : x^2 + y^2 = 1$ gefunden. Wir behaupten, daß ϕ bijektiv ist. Aus $\phi(z_1 + \mathbb{Z}) = \phi(z_2 + \mathbb{Z})$ folgt aus den bekannten Eigenschaften von sin und cos sofort $z_1 \equiv z_2 \bmod \mathbb{Z}$, also $z_1 + \mathbb{Z} = z_2 + \mathbb{Z}$ und damit die Injektivität. Ist andererseits $x^2 + y^2 = 1$, so folgt $|x| \le 1$, und es gibt ein $z \in \mathbb{R}$ mit $x = \cos 2\pi z$. Wegen $\cos^2 2\pi z + \sin^2 2\pi z = 1$ ist $\sin 2\pi z = \pm y$, und indem man ggf. z durch $-z$ ersetzt, erhält man $(x, y) = \phi(z + \mathbb{Z})$.

Da $\mathbb{R}/\mathbb{Z}$ eine abelsche Gruppe bezüglich der Addition ist, können wir C durch Strukturtransport ebenfalls zu einer abelschen Gruppe machen. Als neutrales Element erhalten wir $\phi(0 + \mathbb{Z}) = (\cos 0, \sin 0) = (1, 0) = \mathcal{O}$; um zu gegebenem (x, y) das Inverse zu finden, gehen wir zum Urbild unter ϕ zurück, nehmen dort das Inverse, und bilden wieder mit ϕ ab: das liefert $-(x, y) = (\cos(-2\pi z), \sin(-2\pi z)) = (x, -y)$, d.h. Inversenbildung ist Spiegeln an der x-Achse. Insbesondere gibt es genau zwei Elemente, die mit ihrem Inversen übereinstimmen, nämlich den Punkt $(-1, 0)$ der Ordnung 2 und das neutrale Element $\mathcal{O} = (1, 0)$.

Um die allgemeinen Additionsformeln herzuleiten, setzen wir $(x_1, y_1) = \phi(z_1+\mathbb{Z})$ und $(x_2, y_2) = \phi(z_2+\mathbb{Z})$. Addition der z-Werte und Abbilden mit ϕ auf den Einheitskreis gibt $(x_1, y_1)+(x_2, y_2) = (\cos 2\pi(z_1+z_2), \sin 2\pi(z_1+z_2))$. Nach dem Additionsgesetz für sin und cos ist aber

$$\cos 2\pi(z_1 + z_2) \;=\; \cos 2\pi z_1 \cos 2\pi z_2 - \sin 2\pi z_1 \sin 2\pi z_2 \;=\; x_1x_2 - y_1y_2$$

Entsprechend folgt $\sin 2\pi(z_1 + z_2) = x_1y_2 + x_2y_1$, und wir haben das Additionsgesetz

$$(x_1, y_1) + (x_2, y_2) = (x_1x_2 - y_1y_2, x_1y_2 + x_2y_1). \tag{1.7}$$

Gemäß dieser Konstruktion werden Punkte auf dem Einheitskreis so addiert, daß sich ihre Winkel addieren. Die Assoziativität der Addition folgt für alle Teilkörper von $\mathbb{R}$ aus der Assoziativität der Addition von $\mathbb{R}/\mathbb{Z}$.

Da die Additionsformel (1.7) auch für endliche Körper $\mathbb{F}_q$ sinnvoll ist (es treten ja keine Nenner auf), können wir damit auch eine Addition von Punkten über $\mathbb{F}_q$ definieren, müssen aber (da die Existenz von Nullelement und Inversem klar ist) noch die Assoziativität nachrechnen. Das ist aber lediglich ein technisches Problem. Die so konstruierte Gruppe wird im folgenden mit $C(\mathbb{F}_q)$ bezeichnet.

Betrachten wir z.B. $C(\mathbb{F}_5)$; da es genau vier Lösungen der Kongruenz $x^2+y^2 \equiv 1 \bmod 5$ gibt, nämlich $(0, \pm1)$ und $(\pm1, 0)$, hat $C(\mathbb{F}_5)$ vier Elemente. Da es nur zwei Gruppen der Ordnung 4 gibt, nämlich die zyklische Gruppe $\mathbb{Z}/4\mathbb{Z}$ und die Kleinsche Vierergruppe $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, muß $C(\mathbb{F}_5)$ zu einer dieser beiden Gruppen isomorph sein. Wegen $(0, 1) + (0, 1) = (-1, 0)$ und $(-1, 0) + (-1, 0) = (1, 0) = \mathcal{O}$ erzeugt aber $(0, 1)$ schon ganz $C(\mathbb{F}_5)$, d.h. wir haben $C(\mathbb{F}_5) \simeq \mathbb{Z}/4\mathbb{Z}$.

Entsprechend findet man über $\mathbb{F}_7$ die acht Punkte $(0, \pm1), (\pm1, 0)$ und $(\pm2, \pm2)$; hier rechnet man nach, daß $2(2, 2) = (0, 1)$, $2(0, 1) = (-1, 0)$ und $2(-1, 0) = \mathcal{O}$ ist: also ist $C(\mathbb{F}_7) \simeq \mathbb{Z}/8\mathbb{Z}$. Im nächsten Kapitel werden wir die Anzahl der Punkte auf $C(\mathbb{F}_p)$ allgemein bestimmen.

Sequenzen Quadratischer Reste

Sei p eine ungerade Primzahl, $n \geq 1$, und seien $\varepsilon_1, \ldots, \varepsilon_n \in \{\pm1\}$. Mit $N(\varepsilon_0, \ldots, \varepsilon_{n-1})$ sei die Anzahl aller n-tupel $a, a + 1, \ldots, a + n - 1$ von Elementen in $(\mathbb{Z}/p\mathbb{Z})^\times$ bezeichnet, für die $(\frac{a+j}{p}) = \varepsilon_j$ für $j = 0, \ldots, n - 1$ gilt. Beispielsweise bezeichnet $N(+1)$ die Anzahl der quadratischen Reste modulo p, es ist also $N(+1) = \frac{p-1}{2}$; entsprechend ist $N(-1) = \frac{p-1}{2}$.

- Man erstelle eine kleine Tabelle für die Anzahl der Sequenzen der Länge 2, d.h. man bestimme $N(+1,+1)$, $N(+1,-1)$, $N(-1,+1)$ und $N(-1,-1)$ für einige Primzahlen. Man vermute eine Formel für diese Anzahlen.

- Man betrachte den Ausdruck $(1+(\frac{a}{p}))(1+(\frac{a+1}{p}))$; er nimmt genau dann den Wert 4 an, wenn a und $a+1$ beides quadratische Reste sind. Man nutze diese Beobachtung aus, um Formeln für die Anzahlen $N(\pm 1, \pm 1)$ hinzuschreiben.

- Unter diesen Formeln bestehen einfache Relationen; man finde und beweise sie und leite so geschlossene Formeln für die $N(\pm 1, \pm 1)$ ab.

- Man stelle ähnliche Formeln für Sequenzen der Länge 3 auf; im Spezialfall $p \equiv 3 \bmod 4$ kann man eine einfache Formel für $N(+1,+1,+1)$ erraten und beweisen. Man untersuche auch den Fall $p \equiv 1 \bmod 4$ und errate eine explizite Formel. Wie stark weichen $\frac{p}{8}$ und $N(+1,+1,+1)$ höchstens voneinander ab?

- Man formuliere eine Vermutung darüber, wie sehr $N(\varepsilon_0, \ldots, \varepsilon_{n-1})$ im allgemeinen von $2^{-n}p$ abweicht.

- Finde und beweise einen Zusammenhang zwischen $N(+1,+1,+1)$ und den $\mathbb{F}_p$-rationalen Punkten auf der elliptischen Kurve $y^2 = x^3 - x$.

Die Kleinsche Kurve $V : xy^3 + yz^3 + zx^3 = 0$

Die Kleinsche Kurve ist keine elliptische Kurve; dennoch zeigen sich Regelmäßigkeiten, wenn man die Punkte von V über endlichen Körpern zählt. Man erstelle eine Tabelle von Punktanzahlen über $\mathbb{F}_p$ und versuche, ein Gesetz zu erraten. Was tut sich, wenn man von $\mathbb{F}_p$ zu $\mathbb{F}_{p^2}$, $\mathbb{F}_{p^3}$ ... etc. übergeht?

Kapitel 2

Das Gruppengesetz

2.1 Projektive Ebenen und singuläre Punkte

Wie wir bei der Diskussion der Bijektion $\mathbb{C}/\Lambda \longrightarrow E(\mathbb{C})$ gesehen haben, brauchen wir einen unendlich fernen Punkt, um $E(\mathbb{C})$ zur Gruppe machen zu können. Einen solchen bekommen wir durch die Einführung projektiver Räume geschenkt.

Projektive Räume

Betrachten wir zuvor die Fermatgleichung $X^p + Y^p + Z^p = 0$. Wenn wir nicht ständig von "nichttrivialen" Lösungen sprechen wollen, sollten wir in einer Menge von Tripeln (x, y, z) arbeiten, die das Element $(0, 0, 0)$ gar nicht enthält. Zweitens ist eine Lösung (tx, ty, tz) für ein $t \in \mathbb{Q}^\times$ genauso interessant wie (x, y, z); anstatt von "wesentlich verschiedenen" Lösungen zu sprechen, machen wir das wie richtige Mathematiker und identifizieren (tx, ty, tz) mit (x, y, z). Etwas genauer: wir führen auf der Menge aller Tripel $(x, y, z) \in \mathbb{Q} \setminus \{(0, 0, 0)\}$ eine Äquivalenzrelation $\sim$ ein, indem wir $(u, v, w) \sim (x, y, z)$ setzen, falls es ein $t \in \mathbb{Q}^\times$ gibt mit $u = tx$, $v = ty$ und $w = tz$.

Übung. Zeige, daß dadurch wirklich eine Äquivalenzrelation definiert ist.

Die Menge aller solchen Äquivalenzklassen nennt man den zweidimensionalen projektiven Raum über $\mathbb{Q}$ und bezeichnet ihn mit $\mathbb{P}^2(\mathbb{Q})$. Die Äquivalenzklasse eines Tripels (x, y, z) wird mit $(x : y : z)$ bezeichnet; es ist also $(x : y : z) = \{(tx, ty, tz) : t \in \mathbb{Q}^\times\}$.

Dieselbe Konstruktion funktioniert natürlich für beliebige Dimension n und beliebige Grundkörper K.

Übung. Führe die Konstruktion von $\mathbb{P}^n(K)$ allgemein durch.

Für uns interessant ist vorläufig nur $\mathbb{P}^2(K)$, dessen Punkte wir $(x : y : z)$ mit $x, y, z \in K$ schreiben. Den Unterraum $\{(x : y : 1) \in \mathbb{P}^2(K)\}$ können wir mit dem affinen Raum $\mathbb{A}^2(K) = \{(x, y) : x, y \in K\}$ identifizieren, d.h. wir können $\mathbb{A}^2(K)$ durch die Abbildung

$$\mathbb{A}^2(K) \longrightarrow \mathbb{P}^2(K) : (x, y) \longmapsto (x : y : 1)$$

in den $\mathbb{P}^2(K)$ einbetten, und umgekehrt entspricht jeder Punkt $(x : y : z)$ mit $z \neq 0$ einem solchen affinen Punkt, da wir ja $(x : y : z) = (x/z : y/z : 1)$ haben. Wir können uns $\mathbb{P}^2(K)$ also vorstellen als einen $\mathbb{A}^2(K)$, zu dem wir alle Punkte der Form $(x : y : 0)$ hinzugenommen haben (die Gesamtheit dieser Punkte heißt die *Gerade in* ∞).

Indem wir unsere Betrachtungen von $\mathbb{A}^2(K)$ nach $\mathbb{P}^2(K)$ verlegen, bekommen wir also unendlich ferne Punkte geschenkt. Anstatt daher die Nullstellen eines Polynoms $f \in K[X, Y]$ in $\mathbb{A}^2(K)$ zu studieren, betrachten wir das "homogenisierte" Polynom $F \in K[X, Y, Z]$ in $\mathbb{P}^2(K)$. Der Vorgang des Homogenisierens ist ganz einfach: sei n der Grad von f (bei $f = \sum a_{i,j} X^i Y^j$ ist n das maximale $i + j$); dann machen wir aus f ein homogenes Polynom in drei Variablen x, y, z, indem wir jeden Term so oft mit z multiplizieren, bis er Grad n hat: $F(X, Y, Z) = \sum a_{i,j} X^i Y^j Z^{n-i-j}$.

Beispiel: Homogenisieren von $Y^2 - X^3 - aX - b$ liefert $F(X, Y, Z) = Y^2 Z - X^3 - aXZ^2 - bZ^3$.

Leider macht es keinen Sinn, homogenen Polynomen $F \in K[X, Y, Z]$ vom Grad n einen Funktionswert an der Stelle $(x : y : z)$ zuzuordnen wegen $(x : y : z) = (\lambda x : \lambda y : \lambda z)$ und $F(\lambda x, \lambda y, \lambda z) = \lambda^n F(x, y, z)$. Allerdings können wir von der Nullstellenmenge $\mathcal{V}(F)$ von F zu sprechen, also von $\mathcal{V}(F) = \{(x : y : z) \in \mathbb{P}^2(K) : F(x, y, z) = 0\}$.

Die Nullstellenmenge von $F(X, Y, Z) = Y^2 Z - X^3 - aXZ^2 - bZ^3$ besteht aus allen $(x : y : z) \in \mathbb{P}^2(K)$ mit $y^2 z = x^3 + axz^2 + bz^3$; ist $z \neq 0$, können wir $z = 1$ annehmen und erhalten genau die "affinen" Punkte unserer elliptischen Kurve, nämlich die $(x : y : 1)$ mit $y^2 = x^3 + ax + b$. Ist dagegen $z = 0$, so folgt $0 = x^3$, also $x = 0$ und $y \in K^\times$ beliebig. Also ist $(0 : 1 : 0)$ der Schnittpunkt von F mit der Gerade im Unendlichen, und $\mathcal{V}(F)$ besteht aus den affinen Punkten der elliptischen Kurve plus einem Punkt $(0 : 1 : 0)$ im Unendlichen.

Übung. Homogenisiere die affine Geradengleichung $ax + by + c = 0$ ($a, b, c \in K$, nicht $a = b = c = 0$). Zeige, daß jede projektive Gerade im $\mathbb{P}^2(K)$ mit $z = 0$ genau einen Schnittpunkt besitzt (d.h. jede Gerade schneidet die Gerade im Unendlichen genau einmal). Zeige allgemein: Zwei Geraden im $\mathbb{P}^2(K)$ besitzen *immer* genau einen Schnittpunkt (m.a.W.: im Projektiven gibt es keine Parallelen!); dieses ist genau dann ein Punkt im Unendlichen, wenn die affinen Geraden parallel sind.

Übung. Schneide die Gerade $y = tx$ mit der elliptischen Kurve $y^2 = x^3 - x$. In $\mathbb{C}$ gibt es dann genau drei Schnittpunkte; für großes t liegen zwei davon nahe bei $(0, 0)$, der dritte bei (t, t^2). Zeige, daß der dritte Schnittpunkt für $t \longrightarrow \infty$ im Affinen gegen (∞, ∞) divergiert, im Projektiven aber gegen $(0 : 1 : 0)$ konvergiert.

Singuläre Punkte

Sei eine Kurve durch ein homogenes Polynom $F(X, Y, Z) \in K[X, Y, Z]$ gegeben; ein Punkt $P \in \mathbb{P}^2(K)$ heißt eine *Singulärität* von F, wenn $F(P) = 0$ ist und die formalen Ableitungen F_X, F_Y und F_Z im Punkt P alle verschwinden.

Folgende Beispiele sollen den Begriff erläutern:

- Sei die affine Kurve durch $y^2 = x^3$ gegeben; dann ist $F = Y^2Z - X^3$, und die partiellen Ableitungen sind $F_X = -3X^2$, $F_Y = 2YZ$ und $F_Z = Y^2$; da diese im Ursprung $(0 : 0 : 1)$ verschwinden, ist dieser Punkt singulär.
- Sei die affine Kurve durch $y^2 = 1 - x^4$ über einem Körper der Charakteristik $\neq 2$ gegeben; dann ist $F = Y^2Z^2 - Z^4 - X^4$, und die partiellen Ableitungen sind $F_X = -4X^3$, $F_Y = 2YZ^2$ und $F_Z = 2Y^2Z - 4Z^3$. Wenn diese alle verschwinden sollen, muß erstens $x = 0$ sein; aus $Y^2Z^2 = Z^4$ folgt dann $y \neq 0$ (sonst wäre $x = y = z = 0$), aus $F_Y = 0$ schließlich $z = 0$, d.h. wenn $P = (x : y : z)$ singulär ist, dann ist $P = (0 : 1 : 0)$ der Punkt im Unendlichen; dieser ist auch tatsächlich singulär.
- Der unendlich ferne Punkt einer Weierstraßgleichung ist nie singulär: mit $F = Y^2Z - X^3 - aXZ^2 - bZ^3$ ist nämlich $F_X = -3X^2 - aZ^2$, $F_Y = 2YZ$ und $F_Z = Y^2 - 2aXZ - 3bZ^2$. In $(0 : 1 : 0)$ ist also $F_Z = 1$, unabhängig von der Charakteristik von K.

Übung. Zeige, daß die Fermatgleichungen $X^n + Y^n + Z^n = 0$ über jedem Körper, dessen Charakteristik kein Teiler von n ist, nicht singulär ist.

Übung. Bestimme alle singulären Punkte auf $y^2 = x^3 - x^2$.

Übung. Ist die durch $x^2 + y^2 + x^2y^2 = 1$ über $\mathbb{Q}$ definierte Kurve singulär?

4.9.1999

Übung. Man zeige, daß die über einem Körper K definierte Kleinsche Kurve $xy^3 + yz^3 + zx^3 = 0$ genau dann singulär ist, wenn $\operatorname{char} K = 7$ ist.

Im ersten Kapitel haben wir die Vorstellung gewonnen, singuläre Punkte P auf kubischen Kurven seien solche, für die jede Gerade durch P mit der Kurve einen doppelten Schnittpunkt besitzt. Wir wollen jetzt zeigen, daß dieser Sachverhalt für die oben definierten Singularitäten in der Tat richtig ist. Dazu sei $P = (x_0, y_0)$ ein singulärer Punkt im Endlichen auf der durch $f(x,y) = 0$ beschrieben Kurve, und der Einfachkeit halber sei der Grundkörper K in $\mathbb{C}$ enthalten. Sei weiter $y = mx + c$ eine Gerade durch P. Um die Schnittpunkte mit der Kurve zu bestimmen, setzt man $y = mx+c$ in $f(x,y) = 0$ ein und findet $F(x) := f(x, mx+c) = 0$. Diese Gleichung soll in $x = x_0$ eine doppelte Nullstelle haben, d.h. es soll gelten $F(x) = (x-x_0)^2 g(x)$ für eine (lineare) Funktion g. Dies ist offenbar genau dann der Fall, wenn $F(x_0) = F'(x_0) = 0$ ist. Nun ist $F(x_0) = 0$, weil P nach Voraussetzung auf der Kurve liegt; wir müssen also nur noch zeigen, daß $F'(x_0) = 0$ ist.

Dazu berechnen wir die Ableitung von $F(x)$; wir finden

$$\begin{aligned} F'(x) &= \lim_{h\to 0} \tfrac{F(x+h)-F(x)}{h} \\ &= \lim_{h\to 0} \tfrac{f(x+h,mx+c+mh)-f(x,mx+c)}{h} \end{aligned}$$

Nach Taylor ist nun aber $f(x+h, mx+c+mh) = f(x, mx+c) + hf_x(x, mx+c) + mhf_y(x, mx+c) + h^2 G(x, mx+c)$, somit $F'(x_0) = 0$, da f_x und f_y in P verschwinden. Damit ist die Behauptung gezeigt.

Singuläre Punkte auf Weierstraßkurven

Wir werden uns im folgenden vor allem für Kurven E in Weierstraßform

$$E : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6 \tag{2.1}$$

interessieren. E heißt definiert über K, wenn alle Koeffizienten in K liegen. E heißt elliptische Kurve, wenn E nichtsingulär ist, d.h. wenn es keinen singulären Punkt auf $E(\overline{K})$ gibt.

Auch bei langer Weierstraßform ist der Punkt im Unendlichen nie singulär: setzt man $Z = 0$ in der homogenisierten Weierstraßform, so folgt $x = 0$, d.h. auch hier ist $(0:1:0)$ der unendlich ferne Punkt, und dieser ist nicht singulär, weil wie oben $F_Z = 1$ in diesem Punkt ist.

Wir wollen noch zeigen, daß man singuläre Punkte, die im Endlichen liegen, schon am affinen Modell ablesen kann. Genauer:

Proposition 2.1. *Sei $F(X,Y,Z) \in K[X,Y,Z]$ eine elliptische Kurve in homogenisierter langer Weierstraßform, und*

$$f(x,y) = F(X,Y,1) = y^2 + a_1xy + a_3y - x^3 - a_2x^2 - a_4x - a_6$$

das dazugehörige dehomogenisierte Polynom. Dann ist der endliche Punkt $P = (x : y : 1)$ auf der von F definierten projektiven Kurve genau dann singulär, wenn $f(p) = f_x(p) = f_y(p) = 0$ ist, wo $p = (x,y)$ der Punkt P in affiner Beschreibung ist.

Beweis. Sei $(x : y : 1)$ singulär; dann ist $0 = F(P) = F(x,y,1) = f(p)$, sowie $F_X(P) = f_x(p) = a_1y - 3x^2 - 2a_2x - a_4$ und $F_Y(p) = f_y(p) = 2y + a_1x + a_3$. Es genügt daher zu zeigen, daß $F_Z(P) = 0$ ist, wenn f, f_x und f_y in p verschwinden. Dazu wiederum genügt es, $F_Z(x,y,1)$ als Linearkombination von f, f_x und f_y zu schreiben, falls dies überhaupt möglich ist. Aber nun ist $F_Z(P) = y^2 + a_1xy + 2a_3y - a_2x^2 - 2a_4x - 3a_6$; den Koeffizienten a_6 können wir nur mit $f(p)$ wegbekommen, da er in den Ableitungen nicht vor kommt. Wir bilden also $F_Z(P) - 3f(p) = -2y^2 - 2a_1xy - a_3y + 3x^3 + 2a_2x^2 + a_4x$. Der Koeffizient a_4 kommt nur in $f_x(p)$ vor, also bilden wir $F_Z(P) - 3f(p) + xf_x(p) = -2y^2 - a_1xy - a_3y = -yf_y(p)$ und erhalten $F_Z(P) = 3f(p) - xf_x(p) - yf_y(p)$ wie gewünscht. □

Ist K ein Körper der Charakteristik $\neq 2$, so können wir $\eta = y + (a_1x + a_3)/2$ setzen (quadratische Ergänzung) und finden

$$\eta^2 = x^3 + \frac{b_2}{4}x^2 + \frac{b_4}{2}x + \frac{b_6}{4} \tag{2.2}$$

mit $b_2 = a_1^2 + 4a_2$, $b_4 = a_1a_3 + 2a_4$, $b_6 = a_3^2 + 4a_6$.

Ist darüberhinaus auch char $K \neq 3$, so setzen wir $\xi = x + b_2/12$ und erhalten die kurze Weierstraßnormalform

$$\eta^2 = \xi^3 - \frac{c_4}{48}\xi - \frac{c_6}{864}, \tag{2.3}$$

wobei

$$c_4 = b_2^2 - 24b_4, \quad c_6 = -b_2^3 + 36b_2b_4 - 216b_6$$

ist. Man setzt noch

$$b_8 = a_1^2a_6 - a_1a_3a_4 + 4a_2a_6 + a_2a_3^2 - a_4^2$$

und definiert die *Diskriminante* Δ durch

$$\Delta = -b_2^2b_8 - 8b_4^3 - 27b_6^2 + 9b_2b_4b_6.$$

Da die Diskriminante in den für uns interessanten Fällen von nichtsingulären Kurven von 0 verschieden ist, können wir auch die Definition der j-Invariante imitieren durch $j = c_4^3/\Delta$.

Die oben gemachten Definitionen sind Tabelle 2.1 noch einmal versammelt. Jetzt gilt

TABLE 2.1.

b_2	$=$	$a_1^2 + 4a_2\,,$
b_4	$=$	$a_1a_3 + 2a_4\,,$
b_6	$=$	$a_3^2 + 4a_6,$
b_8	$=$	$a_1^2a_6 - a_1a_3a_4 + 4a_2a_6 + a_2a_3^2 - a_4^2\,,$
c_4	$=$	$b_2^2 - 24b_4\,,$
c_6	$=$	$-b_2^3 + 36b_2b_4 - 216b_6\,,$
Δ	$=$	$-b_2^2b_8 - 8b_4^3 - 27b_6^2 + 9b_2b_4b_6\,,$
$4b_8$	$=$	$b_2b_6 - b_4^2\,,$
1728Δ	$=$	$c_4^3 - c_6^2\,,$
j	$=$	$c_4^3/\Delta \;=\; 1728 + c_6^2/\Delta\,.$

Satz 2.2. *Die über einem Körper K definierte elliptische Kurve E ist singulär genau dann, wenn $\Delta = 0$ in K gilt. In diesem Fall gibt es genau einen singulären Punkt P, und zwar ist dieser wie folgt bestimmt:*

- *Ist* $\operatorname{char} K = 2$ *und E in langer Weierstraßform (2.1) gegeben, so ist*

$$P = \begin{cases} (\sqrt{a_4}, \sqrt{a_2a_4 + a_6}\,) & \textit{falls } c_4 = 0 \;(\iff a_1 = 0) \\ (a_3/a_1, (a_3^2 + a_1^2a_4)/a_1^3) & \textit{falls } c_4 \neq 0 \;(\iff a_1 \neq 0) \end{cases}$$

- *Ist* $\operatorname{char} K = 3$ *und E in der Form (2.2) gegeben, dann ist*

$$P = \begin{cases} (-\sqrt[3]{b_6}, 0) & \textit{falls } c_4 = 0 \;(\iff b_2 = 0) \\ (-b_4/b_2, 0) & \textit{falls } c_4 \neq 0 \;(\iff b_2 \neq 0) \end{cases}$$

- *Ist schließlich* $\operatorname{char} K \neq 2,3$ *und* E *in kurzer Weierstraßform (2.3) gegeben, so ist*

$$P = \begin{cases} (0,0) & \textit{falls } c_4 = 0 \\ (-c_6/12c_4, 0) & \textit{falls } c_4 \neq 0 \end{cases}$$

  der singuläre Punkt.

Insbesondere ist der einzige singuläre Punkt, wenn K *endlich oder von Charakteristik* 0 *ist, stets* K*-rational.*

Beweis. Wir betrachten zuerst den Fall, wo K Charakteristik $\neq 2,3$ hat. Dann dürfen wir E in kurzer Weierstraßnormalform (2.3) annehmen (denn lineare Transformationen $x \longmapsto x+t$ ändern das Singularitätsverhalten ebensowenig wie die Diskriminante Δ), und dann ist (mit $Z = 1$, denn der unendlich ferne Punkt ist ohnehin nicht singulär) $F_x = -3x^2 + c_4/48$ und $F_y = 2y$. Die erste Gleichung gibt $x = \pm\sqrt{c_4}/12$, die zweite $y = 0$; setzt man dies in (2.3) ein, so folgt noch $c_6 = \mp\sqrt{c_4}^3$ (insbesondere ist $\sqrt{c_4} \in K$). Daraus ergibt sich sofort $\Delta = 0$. Ist also $c_4 = 0$, so folgt $x = 0$ und $y = 0$; ist aber $c_4 \neq 0$, so ist $x = \pm\sqrt{c_4}/12 = -c_6/12c_4$. Wir haben gesehen: gibt es einen singulären Punkt, so ist es der angegebene, und es gilt $\Delta = 0$. Ist umgekehrt $\Delta = 0$, so verschwinden im angegebenen Punkt die Ableitungen F_x und F_y, und $\Delta = 0$ ist damit gleichbedeutend, daß der Punkt auf E liegt.

Sei jetzt char $K = 2$. Dann ist $b_2 = a_1^2$, $b_4 = a_1a_3$, $c_4 = a_1^4$, folglich $c_4 = 0 \iff a_1 = 0$.

Eine endliche Singularität ist, wie wir gesehen haben, eine gemeinsame Nullstelle der Gleichungen

$$\begin{aligned} f &= y^2 + a_1xy + a_3y + x^3 + a_2x^2 + a_4x + a_6, \\ f_x &= a_1y + x^2 + a_4, \\ f_y &= a_1x + a_3. \end{aligned}$$

Wer meint, in der ersten Zeile müsse $f = y^2 + \ldots - x^3 - a_2x^2 \ldots$ stehen, hat wegen $+1 = -1$ durchaus recht. Ist jetzt $a_1 = 0$, dann ist $f_y = 0$ nur dann, wenn auch $a_3 = 0$, und in diesem Fall folgt $\Delta = 0$, sowie $x_0 = \sqrt{a_4}$, $y_0 = \sqrt{a_2a_4 + a_6}$.

Ist $a_1 \neq 0$, dann ist $x = a_3/a_1$ (wegen $f_y = 0$) und $y = (a_3^2 + a_1^2a_4)/a_1^3$ (wegen $f_x = 0$). Die Bedingung $f = 0$ liefert jetzt $\Delta = 0$: zum einen ist nämlich (man beachte $2 = 0$ und $-1 = 1$)

$$\begin{aligned} \Delta &= b_2^2b_8 + b_6^2 + b_2b_4b_6 \\ &= a_1^6a_6 + a_1^5a_3a_4 + a_1^4a_2a_3^2 + a_1^4a_4^2 + a_3^4 + a_1^3a_3^3, \end{aligned}$$

zum andern ergibt sich

$$\begin{aligned} a_1^6 f(x,y) &= a_1^6(y^2 + a_1xy + a_3y + x^3 + a_2x^2 + a_4x + a_6) \\ &= a_3^4 + a_1^4a_4^2 + a_1^3a_3^3 + a_1^3a_3^3 + a_1^5a_3a_4 + a_1^3a_3^3 + a_1^4a_2a_3^2 + a_6a_1^6, \end{aligned}$$

und wegen $3a_1^3a_3^3 = a_1^3a_3^3$ ist $\Delta = 0$ in der Tat äquivalent dazu, daß der Punkt (x,y) auf der Kurve liegt, also $f(x,y) = 0$ genügt.

Der Fall char $K = 3$ wird als Übung diskutiert.

Schließlich noch ein Wort zur K-Rationalität im Falle char $K = 2$: ist K ein endlicher Körper der Charakteristik 2, so ist $x \longmapsto x^2$ ein Automorphismus und folglich jedes Element von K ein Quadrat. $\square$

Ist P der singuläre Punkt einer Weierstraßkurve E, so nennt man $E_{\rm ns} = E(\mathbb{Q}) \setminus \{P\}$ ihren nichtsingulären Teil.

Übung. Verifiziere Satz 2.2 für char $K = 3$.

Übung. Mit $(\sqrt{a_4}, \sqrt{a_2a_4 + a_6})$ ist im Falle char $K = 2$, $c_4 = 0$ auch der Punkt $(\sqrt{a_4}, -\sqrt{a_2a_4 + a_6})$ singulär. Warum widerspricht dies nicht der Behauptung, es gäbe höchstens einen singulären Punkt auf kubischen Weierstraßkurven?

Übung. Sei $F(X,Y,Z) = 0$ die Gleichung einer Kurve C; zeige, daß C genau dann singulär ist, wenn die Kurve $C' : F(X',Y',Z') = 0$ dies ist, wobei $X' = X + a$, $Y' = Y + b$ und $Z' = Z + c$ ist.

Übung. Sei $f(x) \in \mathbb{Q}[x]$ ein Polynom vom Grad n. Zeige, daß der unendlich ferne Punkt genau dann singulär ist, wenn $n \geq 4$ gilt.

2.2 Additionsformeln

Sei E eine über einem Körper K definierte elliptische Kurve in langer Weierstraßform (2.1). Für $P \in E(K)$ definieren wir $-P$ als den dritten Schnittpunkt der Geraden durch P und $\mathcal{O}$ mit E. Mit $P = [x_1 : y_1 : 1]$ und $\mathcal{O} = [0 : 1 : 0]$ ist die Gerade gegeben durch $x = x_1$; Schneiden mit der Kurvengleichung gibt

$$y^2 + (a_1x_1 + a_3)y - (x_1^3 + a_2x_1^2 + a_4x_1 + a_6) = 0,$$

d.h. die Summe der beiden Wurzeln ist $-(a_1x_1 + a_3)$. Da eine Wurzel durch $y = y_1$ gegeben ist, muß $-(a_1x_1 + a_3) - y_1$ die andere sein.

Sind zwei Punkte $P_1 = (x_1, y_1)$ und $P_1 = (x_2, y_2)$ in $E(K)$ mit $x_1 \neq x_2$ gegeben, so sei $-P_1 - P_2$ der dritte Schnittpunkt dieser Geraden mit der

Kurve. Die Sekante durch P_1 und P_2 hat die Gleichung $y = y_1 + m(x - x_1)$ mit $m = (y_2 - y_1)/(x_2 - x_1)$; Einsetzen in (2.1) liefert eine Gleichung in x mit den Wurzeln x_1, x_2 und x_3. Die Summe dieser Wurzeln ist der negative Koeffizient von x^2, und es folgt $x_3 = -x_1 - x_2 + m(a_1 + m)$. Setzt man diesen Wert in die Geradengleichung ein, erhält man die y-Koordinate von $-P_3$, und dies liefert $y_3 = -[y_1 + m(x_3 - x_1) + a_1x_3 + a_3]$.

Im Falle $x_1 = x_2$ ist entweder $y_1 = -y_2$: dann setzen wir $P_1 + P_2 = \mathcal{O}$; oder es ist $y_1 = y_2$, also $P_1 = P_2$. Dann sei $-2P_1$ der dritte Schnittpunkt der Tangente mit der elliptischen Kurve. Eine Rechnung wie oben gibt dann

Satz 2.3. *Sei E eine über dem Körper K definierte elliptische Kurve in langer Weierstraßnormalform (2.1). Das durch die Sekanten-Tangenten-Methode definierte Additionsgesetz wird durch die folgenden Formeln beschrieben:*

$$(x_1, y_1) + (x_2, y_2) = (x_3, y_3),$$

wobei

$$\begin{aligned} x_3 &= -x_1 - x_2 - a_2 + a_1m + m^2 \\ y_3 &= -y_1 - (x_3 - x_1)m - a_1x_3 - a_3 \end{aligned}$$

und

$$m = \begin{cases} \dfrac{y_2 - y_1}{x_2 - x_1} & \textit{falls} \quad x_1 \neq x_2 \\ \dfrac{3x_1^2 + 2a_2x_1 + a_4 - a_1y_1}{2y_1 + a_1x_1 + a_3} & \textit{falls} \quad x_1 = x_2 \end{cases}$$

Insbesondere ist für $P = (x, y)$ die x-Koordinate von $2P$ gegeben durch

$$x_{2P} = \frac{x^4 - b_4x^2 - 2b_6x - b_8}{4x^3 + b_2x^2 + 2b_4x + b_6}. \tag{2.4}$$

Noch spezieller gilt für Kurven in kurzer Weierstraßform

$$x_3 = -x_1 - x_2 + m^2, \ \ y_3 = -y_1 - m(x_3 - x_1),$$

wobei $m = \frac{y_2 - y_1}{x_2 - x_1}$ ist, falls $x_2 \neq x_1$ ist, und $m = \frac{3x^2 + a}{2y}$ für $x_2 = x_1$ und $y \neq 0$. Die Fälle, die dadurch nicht überdeckt werden, sind a) $x_1 = x_2$ und $y_1 = -y_2$: hier ist $P_1 = -P_2$ und folglich $P_1 + P_2 = \mathcal{O}$; und b) $2(x, y)$ mit $y = 0$: hier ist $(x, 0)$ ein Punkt der Ordnung 2, also $2(x, y) = \mathcal{O}$.

Daß die in Satz 2.3 angegebene Addition kommutativ ist, das Nullelement $\mathcal{O}$ besitzt, und daß zu jedem $P \in E(K)$ ein Inverses $-P$ existiert, ist klar. Was nicht klar ist (und aus dem Stand heraus relativ schwer zu beweisen ist),

ist die Assoziativität der Addition. Für Teilkörper $K \subseteq \mathbb{C}$ folgt dies aus dem Additionsgesetz der Weierstraßschen $\wp$-Funktion: wir haben ja bereits gezeigt, daß die Sekanten-Tangenten-Addition der Addition in $\mathbb{C}/\Lambda$ entspricht, und die Addition der Restklassen $z + \Lambda$ ist offensichtlich assoziativ. Das Problem ist, die Assoziativität z.B. in endlichen Körpern nachzurechnen. Natürlich würde es genügen, die Formeln für $P + (Q + R)$ und $(P + Q) + R$ hinzuschreiben und zu vergleichen, aber das ist eine herkulaneische Aufgabe und sicherlich kein Beweis, der große Einsichten vermittelt. In den meisten Quellen wird an dieser Stelle der Satz von Bezout (ohne Beweis) zitiert und dann ein Beweis der Assoziativität zumindest für die Fälle gegeben, in denen keine der Punkte P, Q, R, $P + Q$, etc. zusammenfallen.

Der wohl eleganteste Beweis der Assoziativität benutzt Divisoren; ich halte es aber nicht für ratsam, deren Theorie am Anfang einer Einführung in elliptische Kurven zu entwickeln, weil man dann eine ganze Weile lang glauben muß, daß das alles irgendwann einmal einen Sinn bekommt. Daher verschieben wir beides auf später und glauben vorläufig einmal, daß die oben eingeführte Addition tatsächlich für jeden Grundkörper K assoziativ ist.

Um diese Formelsammlung einmal in Aktion zu sehen, wählen wir die Kurve $E : y^2 + xy = x^3 - 18x + 27$. Hier ist $a_1 = 1$, $a_2 = a_3 = 0$, $a_4 = -18$ und $a_6 = 27$. Wir finden $b_2 = 1$, $b_4 = -36$, $b_6 = 108$, $b_8 = -324$, also $c_4 = 865$, $c_6 = -24625$ und $\Delta = 23625 = 3^3 \cdot 5^3 \cdot 7$. Also ist E eine elliptische Kurve über $\mathbb{F}_p$ für alle $p \neq 3, 5, 7$. Der Punkt $P = (1, 1)$ liegt offensichtlich auf $E(\mathbb{F}_2) : y^2 + xy = x^3 + 1$; wir wollen jetzt einmal die Vielfachen von P ausrechnen.

Nach dem Additionsgesetz ist

$$m = \frac{3x_1^2 + 2a_2x_1 + a_4 - a_1y_1}{2y_1 + a_1x_1 + a_3} = \frac{x_1^2 - y_1}{x_1} = 0,$$

somit $x_{2P} = -2x_P + m + m^2 = 0$ und $y_{2P} = -y_P - (x_{2P} - x_P)m - x_{2P} = 1$, also $2P = (0, 1)$

Den Punkt $3P$ erhalten wir durch Addition von P und $2P$; hier ist $m = (y_2 - y_1)/(x_2 - x_1) = 0$ und daher $x_{3P} = -x_P - x_{2P} = 1$, sowie $y_{3P} = -y_P - x_{3P} = 0$ und $3P = (1, 0)$.

Schließlich ist einerseits $4P = P + 3P$, also m wegen $x_P = x_{3P}$ nicht definiert. Das bedeutet (wegen $P \neq 3P$), daß $4P = \mathcal{O}$ sein muß. In der Tat haben wir bereits gesehen, daß $-(x, y) = (x, -a_1x - a_3 - y)$ gilt, in unserem Fall also $-(x, y) = (x, -x - y)$ und insbesondere $-(1, 1) = (1, 0)$.

Also erzeugt P eine Gruppe der Ordnung 4. Abzählen aller möglichen Punkte von $E(\mathbb{F}_2)$ liefert, daß dies schon alle sind; also ist $E(\mathbb{F}_2) \simeq \mathbb{Z}/4\mathbb{Z}$.

Übung. Beweise die Existenz eines neutralen und eines inversen Elements direkt aus der Definition des Additionsgesetzes für elliptische Kurven. Man mache sich anhand einer Skizze klar, daß die Assoziativität dagegen keine triviale Folge der Definition ist.

Der Einheitskreis

Bevor wir uns der Anzahl der Punkte über $\mathbb{F}_p$ auf singulären Weierstraßkurven zuwenden, beschäftigen wir uns mit der analogen Frage für den Einheitskreis $C : x^2 + y^2 = 1$. Wir stellen zwei verschiedene Methoden vor, diese Punkte zu zählen; die erste arbeitet mit der Parametrisierung des Einheitskreises, die wir in Kapitel 1 gewonnen haben.

Wir erinnern also daran, daß wir die von $(-1, 0)$ verschiedenen rationalen Punkte auf E zu

$$(x, y) = \Big(\frac{1-t^2}{1+t^2}, \frac{2t}{1+t^2}\Big)$$

bestimmt haben.

Ist $p \equiv 3 \bmod 4$, so ist -1 quadratischer Nichtrest, folglich $1 + t^2$ niemals $\equiv 0 \bmod p$, und die Werte $t = 0, 1, \ldots p-1$ liefern genau p Punkte auf $E(\mathbb{F}_p)$, und zwar lauter verschiedene: aus $(1-t^2)/(1+t^2) \equiv (1-s^2)/(1+s^2) \bmod p$ folgt nämlich $2s^2 \equiv 2t^2 \bmod p$, wegen $p \geq 3$ also $s \equiv \pm t \bmod p$. Gleichsetzen der y-Koordinaten liefert dann $s \equiv t \bmod p$. Also gibt es im Falle $p \equiv 3 \bmod 4$ insgesamt $p+1$ Punkte (die obigen p plus den Punkt $(-1, 0)$, der in der Liste nicht auftritt).

Ist dagegen $p \equiv 1 \bmod 4$, so gibt es zwei Punkte weniger, weil für zwei Werte von t (nämlich für die beiden Lösungen der Kongruenz $t^2 \equiv -1 \bmod 4$) die Parametrisierung keinen Punkt über $\mathbb{F}_p$ liefert.

Also hat die Gruppe $E(\mathbb{F}_p)$ mindestens $p-1$ bzw. $p+1$ Punkte, je nachdem $p \equiv 1 \bmod 4$ oder $p \equiv 3 \bmod 4$ ist. Daß dies schon alle sind, sieht man so: bei der Herleitung der Parametrisierung hatten wir zwar $K = \mathbb{Q}$ angenommen, aber wenn wir "rational" durch "K-rational" ersetzen (wobei ein K-rationaler Punkt einer sein soll, dessen Koordinaten in K liegen), geht alles genauso durch; man muß sich lediglich davon überzeugen, daß $t^2 + 1 \neq 0$ ist: das kann man aber aus der Gleichung $1 - x = t^2(1 + x)$ sofort ablesen (beachte char $K \neq 2$).

Schließlich funktioniert genau derselbe Beweis, wenn man $\mathbb{F}_p$ durch einen beliebigen endlichen Körper $\mathbb{F}_q$ ungerader Charakteristik ersetzt, und man erhält

Proposition 2.4. *Sei p ungerade Primzahl und $q = p^f$. Dann liegen auf dem Einheitskreis genau*

$$q + (-1)^{(q+1)/2} = \begin{cases} q-1 & \textit{falls } q \equiv 1 \bmod 4 \\ q+1 & \textit{falls } q \equiv 3 \bmod 4 \end{cases}$$

$\mathbb{F}_q$-rationale Punkte.

Die zweite Methode funktioniert etwas allgemeiner für Kurvengleichungen der Form $y^2 = f(x)$; im Falle des Einheitskreises ist $f(x) = 1 - x^2$. Wir werden daher etwas allgemeiner fragen, wieviele Lösungen in $\mathbb{F}_p$ eine Gleichung $y^2 = f(x)$ besitzt, anders ausgedrückt: wieviele Lösungen die Kongruenz $y^2 \equiv f(x) \bmod p$ für primes $p > 2$ besitzt.

Dazu nehmen wir uns ein festes $x \in \mathbb{F}_p$ her und Fragen, wieviele Lösungen die Gleichung

$$y^2 = f(x) \tag{2.5}$$

besitzt. Die Antwort: es sind genau $1 + \big(\frac{f(x)}{p}\big)$.

In der Tat: ist $p \mid f(x)$, so ist das Legendresymbol gleich 0, und (2.5) hat genau die eine Lösung $(x, 0)$. Ist dagegen $\big(\frac{f(x)}{p}\big) = -1$, also $f(x)$ kein Quadrat in $\mathbb{F}_p$, dann hat (2.5) keine Lösung; ist schließlich $\big(\frac{f(x)}{p}\big) = +1$, so ist $f(x) = z^2$ für ein $z \in \mathbb{F}_p^\times$, und (2.5) hat die beiden Lösungen (x, z) und $(x, -z)$ (hier verwenden wir, daß $p \neq 2$ ist).

Damit können wir ganz leicht eine Formel für die Anzahl aller Lösungen hinschreiben: da (2.5) für festes $x \in \mathbb{F}_p$ genau $1 + \big(\frac{f(x)}{p}\big)$ Lösungen hat, gibt es insgesamt $\sum_{x=0}^{p-1} \big\{1 + \big(\frac{f(x)}{p}\big)\big\} = p + \sum_{x=0}^{p-1} \big(\frac{f(x)}{p}\big)$ Lösungen. Wir haben gezeigt:

Proposition 2.5. *Die Kongruenz $y^2 \equiv f(x) \bmod p$ hat für prime $p > 2$ genau $p + \sum_{x=0}^{p-1} \big(\frac{f(x)}{p}\big)$ Lösungen.*

Jetzt gehen wir daran, diese Formel für $f(x) = 1 - x^2$ auszuwerten. Wegen $1 - x^2 = (1-x)(1+x)$ ist

$$S := \sum_{x=0}^{p-1} \Big(\frac{1-x^2}{p}\Big) = \sum_{x=0}^{p-1} \Big(\frac{1-x}{p}\Big)\Big(\frac{1+x}{p}\Big).$$

Setzen wir $s = x + 1$, so wird daraus

$$S = \sum_{s=0}^{p-1} \Big(\frac{2-s}{p}\Big)\Big(\frac{s}{p}\Big).$$

Da mit s auch $2s$ ganz $\mathbb{F}_p$ durchläuft, können wir noch $s = 2t$ setzen und finden unter Berücksichtigung von $(2/p)^2 = +1$

$$S = \sum_{t=0}^{p-1} \left(\frac{2-2t}{p}\right)\left(\frac{2t}{p}\right) = \sum_{t=0}^{p-1} \left(\frac{t}{p}\right)\left(\frac{1-t}{p}\right).$$

Um diese letzte Summe zu bestimmen, wenden wir einen hübschen Trick an: wir zeigen, daß gilt:
a) $|S| \le p-2$;
b) $S \equiv (-1)^{(p+1)/2} \bmod p$.
Daraus folgt dann, daß $S = (-1)^{(p+1)/2}$ sein muß, weil unter den Zahlen mit Betrag $\le p-2$ keine andern vorkommen, die $\equiv (-1)^{(p+1)/2} \bmod p$ sind.

Die Behauptung a) ist trivial: in der Summe über alle $t \in \mathbb{F}_p$ sind zwei Summanden gleich 0 (nämlich für $t = 0$ und $t = 1$), die andern sind $+1$ oder -1, und a) folgt.

Für die Behauptung b) bemerken wir zuerst, daß nach dem Eulerschen Kriterium

$$\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \bmod p$$

ist (schreibe $a \equiv g^k$ für eine Primitivwurzel $g \bmod p$. Genau dann ist a quadratischer Reste, wenn k gerade ist. Aber $a^{\frac{p-1}{2}} \equiv g^{k\frac{p-1}{2}} \bmod p$, und die rechte Seite ist nach dem kleinen Fermatschen Satz $\equiv +1 \bmod p$, wenn k gerade ist, und $\equiv -1 \bmod p$, wenn k ungerade ist). Damit finden wir

$$\begin{aligned} S &= \sum_{t=0}^{p-1} \left(\tfrac{t}{p}\right)\left(\tfrac{1-t}{p}\right) \\ &\equiv \sum_{t=0}^{p-1} t^{(p-1)/2}(1-t)^{(p-1)/2} \equiv \sum_{t=0}^{p-1} (t-t^2)^{(p-1)/2} \\ &\equiv \sum_{t=0}^{p-1} \left\{t^{(p-1)/2} - \tbinom{(p-1)/2}{1} t^{(p+1)/2} \pm \ldots + (-1)^{(p-1)/2} t^{p-1}\right\}. \end{aligned}$$

Jetzt benutzen wir den folgenden

Hilfssatz 2.6. *Sei p eine ungerade Primzahl. Dann gilt*

$$\sum_{x=0}^{p-1} x^k \equiv \begin{cases} 0 & \textit{falls } (p-1) \nmid k \\ -1 & \textit{falls } (p-1) \mid k \end{cases}$$

Damit sind dann alle Summen über die Polynome t^k kongruent $0 \bmod p$ bis auf die letzte, und wir finden

$$S \equiv \sum_{t=0}^{p-1}(-1)^{(p-1)/2}t^{p-1} \equiv -(-1)^{(p-1)/2} = (-1)^{(p+1)/2} \bmod p.$$

Das war zu zeigen.

Nachzutragen bleibt der Beweis von Hilfssatz 2.6. Ist $(p-1) \mid k$, so ist nach Fermat jeder Summand $x^k \equiv 1 \bmod p$, mit Ausnahme von $x = 0$. Also ist $S \equiv p-1 \equiv -1 \bmod p$. Ist aber $(p-1) \nmid k$, so ist $g^k \not\equiv 1 \bmod p$, wo g eine Primitivwurzel modulo p ist. Andererseits ist $g^k S \equiv \sum_{t=0}^{p-1} t^k \equiv \sum_{t=0}^{p-1}(gt)^k \equiv \sum_{s=0}^{p-1} s^k = S \bmod p$, also $p \mid (g^k-1)S$. Da p aber kein Teiler von $g^k - 1$ ist, folgt $p \mid S$ wie behauptet.

Schließlich kann man auch in diesem Beweis $\mathbb{F}_p$ durch $\mathbb{F}_q$ ersetzen: da $\mathbb{F}_q^\times$ zyklisch ist, können wir das Legendresymbol $(\frac{a}{p})$ ersetzen durch $\chi(a) = a^{(q-1)/2} = \pm 1$ für $a \in \mathbb{F}_q^\times$, und alles geht genauso durch. Damit haben wir einen zweiten Beweis für Proposition 2.4 gefunden.

Als nächstes zeigen wir, daß die Gruppen $C(\mathbb{F}_p)$ immer zyklisch sind. Dazu betrachten wir bei gegebenem Grundkörper K der Charakteristik $\ne 2$ die Abbildung $\psi : C \longrightarrow L^\times : (x,y) \longmapsto x + iy$; hierbei ist i eine Wurzel aus -1, die nicht in K zu liegen braucht, und $L = K(i)$. Wir behaupten zuerst, daß ψ ein Gruppenhomomorphismus ist, d.h. daß $\psi(P_1+P_2) = \psi(P_1)\cdot\psi(P_2)$ gilt. Wegen $\psi(P_1)\cdot\psi(P_2) = (x_1+iy_1)(x_2+iy_2) = x_1x_2 - y_1y_2 + i(x_1y_2+x_2y_1) = \psi(P_1+P_2)$ ist das aber klar.

Die Injektivität von ψ ist ganz einfach zu zeigen: aus $1 = \psi(x,y) = x+iy$ folgt wegen $1 = x^2+y^2 = (x+iy)(x-iy)$ $x+iy = 1$ und $x - iy = 1$; Addition bzw. Subtraktion ergeben dann $x = 1$ und $y = 0$, also $(x,y) = (1,0) = \mathcal{O}$.

Betrachten wir zuerst den Fall $K \ne L$, also $i \notin K$. Das Bild von ψ ist offensichtlich gleich der Untergruppe $G_m[-1] := \{x+iy \in L : x^2+y^2 = 1\}$ von $L^\times$; insbesondere induziert also ψ einen Gruppenisomorphismus $\psi : C(\mathbb{F}_p) \longrightarrow G_m[-1]$. Im Falle, daß $K = \mathbb{F}_p$ mit $p \equiv 3 \bmod 4$ ist, ist $G_m[-1]$ als Untergruppe des endlichen Körpers $L = K(i) = \mathbb{F}_{p^2}$ automatisch zyklisch.

Sei jetzt $K = L$, also $i \in K$. Wir behaupten, daß dann jedes Element $r \in K^\times$ sich in der Form $x+iy$ mit $x, y \in K$ und $i^2 = -1$ schreiben läßt, d.h. daß $\psi : C(K) \longrightarrow K^\times$ surjektiv ist. Dazu setzen wir einfach $x = \frac{1}{2}(r+\frac{1}{r})$ und $y = \frac{1}{2i}(r-\frac{1}{r})$ und rechnen nach, daß $r = x+iy$ und $x^2+y^2 = 1$ ist. Die Gruppe $\mathbb{F}_p^\times$ ist als multiplikative Gruppe eines endlichen Körpers natürlich zyklisch. Damit haben wir gezeigt:

Proposition 2.7. *Die Abbildung* $\psi : (x, y) \longmapsto x + iy$ *von* $C(K)$ *nach* $G_m[-1]$ *bzw.* $K^\times$ *ist ein Gruppenisomorphismus; insbesondere ist* $C(\mathbb{F}_p)$ *zyklisch.*

Was hat es mit diesem Gruppenisomorphismus auf sich? Eine Antwort darauf gibt die analytische Parametrisierung $\phi : z + \mathbb{Z} \longmapsto (\cos 2\pi z, \sin 2\pi z)$, denn wir finden $\psi \circ \phi(z + \mathbb{Z}) = \cos 2\pi z + i \sin 2\pi z = e^{2\pi i z}$, und $\psi \circ \phi$ ist damit in der Tat der ganz gewöhnliche Gruppenisomorphismus zwischen $\mathbb{R}/\mathbb{Z}$ und dem komplexen Einheitskreis. Im Falle endlicher Körper bricht dieses Bild natürlich zusammen, gibt aber trotzdem den entscheidenden Hinweis darauf, wie man die Abbildung von $C(K)$ nach $L^\times$ zu definieren hat.

Singuläre Weierstraßkurven

Wir werden im folgenden zeigen, daß auch auf singulären kubischen Kurven durch das Sekanten-Tangenten-Verfahren ein Additionsgesetz gegeben ist, solange man vom singulären Punkt weg bleibt. Auch die Anzahl aller Punkte auf singulären kubischen Kurven über dem endlichen Körper $\mathbb{F}_p$ werden wir bestimmen.

Da sich singuläre kubische Kurven in etwa wie quadratische Kurven verhalten, darf man annehmen, daß die Berechnung der Anzahl der $\mathbb{F}_p$-rationalen Punkte ebenfalls möglich ist. Tatsächlich ist die Lage hier sogar noch einfacher. Über endlichen Körpern der Charakteristik $p > 3$ läßt sich nämlich jede singuläre Kurve auf die Form $y^2 = x^3 + ax^2$ bringen (das werden wir weiter unten zeigen), und deren $\mathbb{F}_p$-rationale Punkte lassen sich leicht zählen.

Als erstes versuchen wir, mit der Parametrisierung von E_c durchzukommen. Mit $y = tx$ wird $t^2x^2 = x^3 + ax^2$, und der dritte Schnittpunkt ist durch $x = t^2 - a$, $y = tx$ gegeben. Betrachten wir also $\psi : K \longrightarrow E(K) : t \longmapsto (t^2 - a, t^3 - ta)$. Offenbar ist ψ injektiv (Vergleich der x-Koordinaten gibt Gleichheit bis auf's Vorzeichen, Vergleich der y-Koordinaten dann Gleichheit). Außerdem ist ψ "fast" surjektiv, da der singuläre Punkt $(0, 0)$ der einzige ist, der möglicherweise nicht erwischt wird: ist nämlich $P = (x, y)$ gegeben und $x \neq 0$, so folgt $x + a = (y/x)^2$, und mit $t = y/x$ wird $P = \psi(t)$. Der singuläre Punkt $(0, 0)$ wird schließlich genau dann parametrisiert, wenn $a = t^2$ ein Quadrat in K ist. Wir nehmen nun $K = \mathbb{F}_p$ an und unterscheiden drei Fälle:
1. $a = 0$: dann ist $t \longmapsto (t^2, t^3)$ eine Bijektion zwischen $\mathbb{F}_p$ und $E(\mathbb{F}_p) \setminus \mathcal{O}$, und insbesondere finden wir für die Anzahl aller Punkte auf dem nichtsingulären Teil $\#E_{\rm ns}(\mathbb{F}_p) = \#E(F_p) = \#\mathbb{F}_p = p$, da wir den Punkt $(0, 0)$ weglassen und den Punkt $\mathcal{O}$ hinzunehmen müssen.

2. $a = t^2$ ist ein Quadrat in $\mathbb{F}_p^\times$: Dann gibt es zwei Werte von t, die den Punkt $(0,0)$ liefern, und wie oben folgt $\#E_{\rm ns}(\mathbb{F}_p) = \#E(F_p)+1-2 = p-1$.
3. a ist kein Quadrat in $\mathbb{F}_p^\times$. Dann taucht der Punkt $(0,0)$ nicht im Bild der Parametrisierung auf, und wir finden $\#E_{\rm ns}(\mathbb{F}_p) = \#E(F_p) + 1 = p+1$, da wir den Punkt $\mathcal{O}$ noch hinzuzählen müssen.

Da der Wert des Legendresymbols (a/p) in den Fällen 1., 2. und 3. gleich 0, +1 bzw. −1 ist, haben wir damit gezeigt:

Proposition 2.8. *Die Anzahl der $\mathbb{F}_p$-rationalen Punkte auf dem nichtsingulären Teil E_{ns} der Kurve $E: y^2 = x^3 + ax^2$ beträgt $p - \left(\frac{a}{p}\right)$.*

Aber auch mit der im letzten Abschnitt eingeführten Methode lassen sich die Punkte auf $E_{\rm ns}(\mathbb{F}_p)$ zählen: in der Tat ist deren Anzahl N gleich

$$N = p + \sum_{x=0}^{p-1} \Big(\frac{x^3+ax^2}{p}\Big) = p + \sum_{x=1}^{p-1} \Big(\frac{x+a}{p}\Big).$$

Die Summation geht dabei nur über $x \neq 0$, weil $\left(\frac{x}{p}\right)^2 = 1$ genau für $x \neq 0$ ist. Daher ist

$$N = p - \Big(\frac{a}{p}\Big) + \sum_{x=0}^{p-1} \Big(\frac{x+a}{p}\Big) = p - \Big(\frac{a}{p}\Big) + \sum_{t=0}^{p-1} \Big(\frac{t}{p}\Big),$$

wobei wir $t = x + a$ gesetzt haben. Die letzte Summe ist aber $= 0$: sie hat nämlich Betrag $< p$ und ist wegen Hilfssatz 2.6 (plus Eulersches Kriterium) durch p teilbar. Also haben wir $N = p - \left(\frac{a}{p}\right)$; zieht man den singulären Punkt ab und den nichtsingulären unendlich fernen Punkt hinzu, ist N also die Anzahl der $\mathbb{F}_p$-rationalen Punkte auf $E_{\rm ns} = E(\mathbb{Q}) \setminus \{P\}$ für die Kurve $E: y^2 = x^3 + ax^2$.

Damit haben wir einen zweiten Beweis für Prop. 2.8 gefunden.

Da es bis auf Isomorphie nur eine Gruppe von Primzahlordnung p gibt, nämlich die zyklische Gruppe $\mathbb{Z}/p\mathbb{Z}$, muß das Additionsgesetz auf $y^2 = x^3$ zu $\mathbb{Z}/p\mathbb{Z} \simeq (\mathbb{F}_p, +)$ isomorph sein. Tatsächlich läßt sich ein solcher Isomorphismus hinschreiben; allgemeiner kann man sogar die Gruppenstruktur aller singulären Weierstraßkurven bestimmen.

Bevor wir das tun, wollen wir noch zwei Dinge tun: a) ein paar Sachen zum Thema algebraische Gruppen sagen: und b) singuläre Weierstraßkurven klassifizieren.

Ad a). Algebraische Gruppen sind Objekte, die eine ganze Menge Struktur haben. Zum einen sind sie Varietäten, also gewisse Nullstellengebilde mit einer Topologie, die nach Zariski benannt ist; als Varietät besitzen diese Objekte insbesondere eine Dimension. Zum anderen besitzen sie eine Addition, die sich lokal als polynomiale Abbildung schreiben läßt. Man kann zeigen, daß die eindimensionalen algebraischen Gruppen über einem perfekten Körper K genau die folgenden sind:

- elliptische Kurven; dies sind die einzigen projektiven Kurven mit einer Gruppenstruktur.
- die additive Gruppe $\mathbb{G}_a = \mathbb{A}^1$ mit Gruppengesetz $(x, y) \longmapsto x + y$.
- die multiplikative Gruppe $\mathbb{G}_m = \mathbb{A}^1 \setminus \{0\}$ mit Gruppengesetz $(x, y) \longmapsto xy$.
- die verschränkte multiplikative Gruppe $\mathbb{G}_m[a]$ für $a \in K^\times \setminus K^{\times 2}$; diese besteht aus allen $\alpha \in L^\times$ mit $L = K(\sqrt{a})$, für die $N_{L/K}\alpha = 1$ ist, also aus allen $x + y\sqrt{a}$ mit $x^2 - ay^2 = 1$. Das Gruppengesetz ist gegeben durch $(x, y) + (x', y') = (xx' + ayy', xy' + x'y)$. Für $a = -1$ erhält man übrigens die Kreisgruppe auf $E : x^2 + y^2 = 1$.

Übung. Zeige, daß $\mathbb{G}_m[a]$ eine Gruppe ist.

Ad b). Wir wollen zeigen, daß singuläre Weierstraßkurven über einem Körper der Charakteristik $\neq 2, 3$ immer auf die Form $y^2 = x^3 + ax^2$ gebracht werden können. Dazu gehen wir aus von der kurzen Weierstraßform

$$y^2 = x^3 - \frac{1}{48}c_4 x - \frac{1}{864}c_6.$$

Ist $c_4 = 0$, so muß wegen $\Delta = 0$ auch $c_6 = 0$ sein, d.h. die Kurvengleichung $y^2 = x^3$ hat bereits die gewünschte Form. Ist $c_4 \neq 0$, so ist der singuläre Punkt gegeben durch $(-c_6/12c_4, 0)$. Die Transformation $Y = y$, $X = x + c_6/12c_4$ (lediglich eine Verschiebung der y-Achse) führt den singulären Punkt in den Ursprung über, und die Kurvengleichung wird $Y^2 = X^3 + aX^2$ mit $a = -c_6/4c_4$.

Wie wir bei der Berechnung der Anzahl der Punkte auf einer über einem Körper K der Charakteristik $\neq 2, 3$ (diese Einschränkung ist rein technischer Natur) definierten Kurve $E_a : y^2 = x^3 + ax^2$ schon festgestellt haben, sind hier drei Fälle zu unterscheiden, je nachdem $a = 0$, $a \neq 0$ ein Quadrat oder a kein Quadrat ist.

Daß die Frage, ob a ein Quadrat ist, etwas mit der Geometrie der Kurve zu tun hat, sieht man so: sei $a \neq 0$; damit $y = mx$ eine Tangente im Punkt $(0,0)$ ist, muß sie mit der Kurve eine dreifache Nullstelle gemeinsam haben (eine doppelte besitzt sie ohnehin schon). Einsetzen gibt $m^2x^2 = x^3 + ax^2$, also $x^2(a - m^2) = 0$. Die Tangentensteigungen sind also durch $m = \sqrt{a}$ und $m = -\sqrt{a}$ gegeben, und genau dann sind diese Steigungen K-rational, wenn $a \neq 0$ ein Quadrat in K ist.

Behandeln wir zuerst den Fall $a = 0$, also $y^2 = x^3$. Wir behaupten, daß durch

$$\phi : \begin{cases} \mathcal{O} & \longmapsto 0, \\ (x,y) & \longmapsto \frac{x}{y} \end{cases} \tag{2.6}$$

ein Gruppenhomomorphismus $\phi : E_{\mathrm{ns}} \longrightarrow (K,+)$ des nichtsingulären Teils in die additive Gruppe des Grundkörpers definiert ist (da $(x,y) \in E_{\mathrm{ns}}(\mathbb{Q})$ ist, muß $y \neq 0$ sein).

Zu zeigen ist dazu, daß drei Punkte P_1, P_2, P_3 genau dann kollinear sind, wenn $\phi(P_1) + \phi(P_2) + \phi(P_3) = 0$ ist: denn P_1, P_2, P_3 sind ja nach dem Additionsgesetz genau dann kollinear, wenn $P_1 + P_2 + P_3 = 0$ ist, und da offensichtlich $\phi(-P) = -\phi(P)$ ist, ist die Homomorphiebedingung $-\phi(P_3) = \phi(-P_3) = \phi(P_1 + P_2) = \phi(P_1) + \phi(P_2)$. Das läßt sich natürlich nachrechnen, indem man $P_j = (x_j, y_j)$ setzt und das Additionsgesetz benutzt – allerdings artet das in eine wüste Rechnerei aus, und es ist mir nicht gelungen, die Behauptung so zu beweisen. Verwenden wir also den Standardtrick und rechnen erstmal projektiv, also mit der Kurvengleichung $Y^2Z = X^3$. Jede Gerade, die nicht durch den singulären Punkt $(0,0)$ geht, hat die Form $Z = lX + mY$. Einsetzen gibt $X^3 = lXY^2 + mY^3$, und Dividieren durch $Y \neq 0$ gibt $U^3 - lU - m = 0$, wobei $U = X/Y$ gesetzt ist. Die Lösungen dieser Gleichung sind $u_j = x_j/y_j$ mit $j = 1,2,3$, und nach Vieta gilt $0 = u_1 + u_2 + u_3$. Das war aber zu zeigen.

Als nächstes sei $y^2 = x^3 + ax^2$ und $0 \neq a = c^2$ ein Quadrat in K. Wie eben erhalten wir die Schnittgleichung $(Y^2 - c^2X^2)(lX + mY) = X^3$. Mit $U = Y + cX$ und $V = Y - cX$ folgt $8c^3UV(lX + mY) = (U - V)^3$, wegen $2cX = U - V$ und $2Y = U + V$ also $4c^2UV(l(U-V) + mc(U+V)) = (U-V)^3$. Division durch V^3 gibt mit $W = U/V$ die Gleichung $W^3 - 3W^2 + 3W - 1 - 4c^2W(l(W-1) + mc(W-1)) = 0$. Das Produkt der drei Nullstellen $w_j = u_j/v_j = (y_j + cx_j)/(y_j - cx_j)$ ist daher gleich dem Negativen des konstanten Terms, d.h. gleich 1. Also liefert $(x,y) \longmapsto (y_j + cx_j)/(y_j - cx_j)$ einen Gruppenisomorphismus von $E_{\mathrm{ns}}(K) \longrightarrow K^\times$.

Schließlich ist noch $y^2 = x^3 + ax^2$ zu untersuchen, wo $0 \neq a$ kein Quadrat in $K^\times$ ist. Dann führt dieselbe Rechnung wie eben auf den durch $(x, y) \longmapsto (y_j + cx_j)/(y_j - cx_j)$ definierten Gruppenisomorphismus $E(K) \longrightarrow G_m[c]$.

Damit haben wir folgende Erkenntnis gewonnen:

Satz 2.9. *Es gibt drei Typen singulärer Weierstraßkurven:*

1. *Die durch die Kurve $y^2 = x^3$ repräsentierte Klasse mit Spitze; hier ist $E_{ns}(K) \simeq (K, +)$ zur additiven Gruppe von K isomorph.*
2. *Die durch $y^2 = x^3 + ax^2$ ($a = c^2$) repräsentierte Klasse von Kurven mit Knoten und rationalen Tangenten; hier ist $E_{ns}(K) \simeq K^\times$.*
3. *Die durch $y^2 = x^3 + ax^2$ ($a \notin K^2$) repräsentierte Klasse von Kurven mit Knoten und irrationalen Tangenten; hier ist $E_{ns}(K) \simeq G_m[a]$.*

Ist E eine über $\mathbb{Q}$ definierte elliptische Kurve mit ganzen Koeffizienten, so kann man E auch als Kurve über $\mathbb{F}_p$ auffassen, indem man die Koeffizienten modulo p reduziert. Genau dann ist $E/\mathbb{F}_p$ eine elliptische Kurve, wenn $p \nmid \Delta$ gilt. Man sagt nun, $E/\mathbb{Q}$ habe

- *gute Reduktion*, wenn $p \nmid \Delta$ gilt;
- *zerfallende multiplikative Reduktion*, wenn $E/\mathbb{F}_p$ einen Knoten mit $\mathbb{F}_p$-rationalen Tangenten hat;
- *nicht zerfallende multiplikative Reduktion*, wenn $E/\mathbb{F}_p$ einen Knoten mit $\mathbb{F}_p$-irrationalen Tangenten hat.
- *additive Reduktion*, wenn $E/\mathbb{F}_p$ eine Spitze hat.

Liegt gute oder multiplikative Reduktion vor, so spricht man von semistabiler Reduktion. Eine Kurve E heißt semistabil, wenn sie an jeder Primstelle p semistabile Reduktion besitzt. Man beachte auch, daß nicht zerfallende multiplikative Reduktion zu zerfallender multiplikativer Reduktion wird, wenn man von K zur quadratischen Erweiterung $K(\sqrt{a}\,)$ übergeht.

Wir weisen gleich an dieser Stelle darauf hin, daß diese Definitionen auch für elliptische Kurven Sinn machen, die über dem Körper $\mathbb{Q}_p$ der p-adischen Zahlen definiert sind. Außerdem verlangt man gewöhnlich, daß man die Kurve, bevor man modulo p reduziert, durch "zulässige Transformationen" auf eine Form bringt, in der Diskriminante p-minimal wird, d.h. für die die in Δ aufgehende p-Potenz minimal wird.

Auch hier lassen sich die etwas seltsam aussehenden Homomorphismen analytisch leicht verstehen. Beginnen wir mit dem Fall $y^2 = x^3$; diese Kurve wird natürlich nicht von einer Weierstraßschen $\wp$-Funktion parametrisiert wegen $\Delta = 0$. Eine analytische Funktion $w = f(z)$, die der Differentialgleichung $w'^2 = 4w^3$ genügt und wie die $\wp$-Funktion an der Stelle $z = 0$ einen Pol zweiter Ordnung hat, gibt es tatsächlich, nämlich $w = z^{-2}$. Also wird die singuläre Kurve $y^2 = x^3$ von $x = w(z)$ und $y = \frac{1}{2}w'(z)$ parametrisiert, und durch Komposition mit $(x, y) \longmapsto \frac{x}{y}$ erhalten wir nichts anderes als $z \longmapsto (w, \frac{1}{2}w') \longmapsto 2w/w' = -z$, also Multiplikation mit -1 (ein dezenter Hinweis darauf, daß wir das Vorzeichen in der Definition (2.6) anders wählen sollten).

Der Fall, wo die Singularität ein Knoten ist, ist interessanter. Zuerst suchen wir eine analytische Funktion mit doppeltem Pol in $z = 0$, welche der Differentialgleichung $w'^2 = 4w^3 - 4w^2$ genügt. Dazu könnten wir einen Potenzreihenansatz $w(z) = z^{-2} + a_0 + a_1 z + a_2 z^2 + \ldots$ versuchen; dieser führt auch tatsächlich ans Ziel, aber es ist leichter, die Funktion direkt anzugeben: $w(z) = (\sin z)^{-2}$. Damit ist $w'(z) = -2\frac{\cos z}{\sin^3 z}$ und $w'^2(z) = 4w(z)^3 - 4w(z)^2$ wie gewünscht. Also erhalten wir die Parametrisierung $z + 2\pi\mathbb{Z} \longrightarrow (w, \frac{1}{2}w')$, und Komposition mit der Abbildung $(x, y) \longrightarrow \frac{y+xi}{y-xi}$ liefert $z \longmapsto (w, \frac{1}{2}w') \longmapsto \frac{\cos z + i\sin z}{\cos z - i \sin z} = e^{2iz}$.

Bestimmung von $\# E(\mathbb{F}_p)$ für elliptische Kurven

Im allgemeinen ist unsere Zähltechnik für elliptische (also nichtsinguläre) Kurven unbrauchbar, auch wenn es einige wenige Spezialfälle gibt, in denen sie unter großen Anstrengungen zum Ziel führt. Manchmal aber kann man doch zu einer Antwort gelangen: betrachten wir z.B. die elliptische Kurve $E_c : y^2 = x^3 + cx$ über $\mathbb{F}_p$ mit $p = 2m + 1$ und $c \in \mathbb{Z}$. Dann folgt

$$\begin{aligned} S &= \sum_{x=1}^{2m} \Big(\frac{x}{p}\Big)\Big(\frac{x^2+c}{p}\Big) \\ &= \sum_{x=1}^{m} \Big(\frac{x}{p}\Big)\Big(\frac{x^2+c}{p}\Big) + \sum_{x=1}^{m} \Big(\frac{p-x}{p}\Big)\Big(\frac{(p-x)^2+c}{p}\Big) \\ &= \sum_{x=1}^{m} \Big(\frac{x}{p}\Big)\Big(\frac{x^2+c}{p}\Big) + \sum_{x=1}^{m} \Big(\frac{-x}{p}\Big)\Big(\frac{x^2+c}{p}\Big). \end{aligned}$$

Ist nun m ungerade, also $p \equiv 3 \bmod 4$, so ist $(-1/p) = -1$ und folglich $S = 0$; ist dagegen m gerade, also $p \equiv 1 \bmod 4$, so folgt nur, daß S eine gerade Zahl sein muß.

Nach unseren obigen Überlegungen ist also $\# E(\mathbb{F}_p) = p+1$, falls $p \equiv 3 \bmod 4$ ist. Das wirft natürlich die Frage auf, was sich für prime $p \equiv 1 \bmod 4$ tut. Eine Berechnung von $S = S_p$ für $c = -1$ und kleine Werte von p liefert

p	5	13	17	29	37	41	53	61
S_p	2	−6	−2	10	2	−10	−14	10
a	1	3	1	5	1	5	7	6

Schreibt man $p = a^2 + b^2$ als Summe zweier Quadrate, wobei b gerade sein soll, so findet man sofort, daß $S_p = \pm 2a$ ist. Die Bestimmung des Vorzeichens ist auch nicht schwer: setzt man $a \equiv 1 \bmod 4$, falls $p \equiv 1 \bmod 8$ und $a \equiv 3 \bmod 4$, falls $p \equiv 5 \bmod 8$, so findet man die Vermutung $S_p = -2a$. Diese wurde in dieser Form zuerst von Jacobsthal, in leicht anderer Form bereits von Gauß bewiesen. Damit hat man

Proposition 2.10. *Wir betrachten die elliptische Kurve $E : y^2 = x^3 - x$ über $\mathbb{F}_p$; dann ist*

$$\# E(\mathbb{F}_p) = \begin{cases} p+1 & \textit{falls } p \equiv 3 \bmod 4, \\ p+1-2a & \textit{falls } p \equiv 1 \bmod 4, \end{cases}$$

wobei $p = a^2 + b^2$ mit $a \equiv (-1)^{(p-1)/4} \bmod 4$ ist. Insbesondere gilt die Hasse-Schranke $|\# E(\mathbb{F}_p) - (p+1)| < 2\sqrt{p}$ wegen $2a < 2\sqrt{p}$.

Etwas ähnliches funktioniert für andere Kurven mit “komplexer Multiplikation” (macht man in $y^2 = x^3 - x$ die Substitution $x = -x'$, $y = iy'$, so geht E in sich über, und man sagt, E besitze komplexe Multiplikation mit i. Entsprechend gibt es Kurven, die z.B. komplexe Multiplikation mit $\sqrt{-2}$ besitzen).

Im allgemeinen aber muß man, um Formeln für die Anzahl von Punkten auf elliptischen Kurven über $\mathbb{F}_p$ zu bekommen, *sehr* weit ausholen. Berechnen wir einmal die Summen S_p für die elliptische Kurve $E : y^2 = x^3 - 4x^2 + 16$ mit Diskriminante $\Delta = -45056 = -2^{12} \cdot 11$. Wir finden

p	3	5	7	11	13	17	19	23	29	31
S_p	1	−1	2	−1	−4	2	0	1	0	−7

Hier sieht man überhaupt keine Regel; und wäre man auf den Zufall angewiesen gewesen, hätte man auch nicht entdeckt, daß diese Anzahlen in der Funktion

$$f(q) = q \prod_{n=1}^{\infty} (1-q^n)^2 (1-q^{11n})^2$$

kodiert sind. Durch Ausmultiplizieren findet man nämlich

$$\begin{aligned} f(q) = q &- 2q^2 - q^3 + 2q^4 + q^5 - 2q^7 - 2q^9 - 2q^{10} + q^{11} - 2q^{12} + 4q^{13} \\ &+ 4q^{14} - q^{15} - 4q^{16} - 2q^{17} + 4q^{18} + 2q^{20} + \ldots \end{aligned}$$

Setzt man also $f(q) = \sum_{n=1}^{\infty} a_n q^n$, so scheint $S_p = -a_p$ zu gelten. Dies ist tatsächlich richtig:

Satz 2.11. *(Eichler) Ist die elliptische Kurve E gegeben durch $y^2 = x^3 - 4x^2 + 16$, und schreibt man $f(q) = q\prod_{n=1}^{\infty}(1-q^n)^2(1-q^{11n})^2$ in der Form $f(q) = \sum_{n=1}^{\infty} a_n q^n$, so ist $\# E(\mathbb{F}_p) = p + 1 - a_p$ für alle primen $p \geq 3$.*

Übrigens gibt es für die hier betrachtete elliptische Kurve ein "schöneres Modell": die Transformation $y = 8Y + 4$, $x = 4X$ liefert $64Y^2 + 64Y = 64X^3 - 64X$, also nach Kürzen von 64 die "lange" Weierstraßgleichung $E : Y^2 + Y = X^3 - X^2$. Da die Transformation linear mit Koeffizienten aus $\mathbb{Z}$ ist, und die Umkehrtransformation nur den Nenner 2 hat, haben beide Kurven über jedem Körper der Charakteristik $\neq 2$ die gleiche Anzahl von Punkten; Eichlers Satz gilt also auch für letztere Kurve, die übrigens $\Delta = -11$ hat.

Was hat es mit dieser Funktion f auf sich? Um diese Frage zu beantworten, setzen wir $q = e^{2\pi i z}$ für ein z aus der oberen Halbebene (damit ist $|q| < 1$) und betrachten f als Funktion von z. Man kann dan zeigen, daß für alle $a, b, c, d \in \mathbb{Z}$ mit $ad - bc = 1$ und $11 \mid c$ gilt:

$$f\Big(\frac{az+b}{cz+d}\Big) = (cz+d)^2 f(z). \tag{2.7}$$

Insbesondere ist $f(z+1) = f(z)$ (man wähle $a = d = 1$, $b = c = 0$). Die Eigenschaft (2.7) ist die wichtigste Eigenschaft von Modulformen (im allgemeinen Fall muß man die 11 durch ein $N \in \mathbb{N}$ ersetzen, den sogenannten "Führer" der Kurve. Dieser ist im wesentlich das Produkt aller in Δ aufgehenden Primzahlen, wobei nur 2 und 3 in höherer als der ersten Potenz auftreten können); in der Tat ist f eine solche. Die Vermutung von Taniyama-Shimura besagt nun, daß es für jede über $\mathbb{Q}$ definierte elliptische Kurve eine dazugehörige Modulform f gibt, deren Fourierkoeffizienten a_p der Beziehung $\# E(\mathbb{F}_p) = p + 1 - a_p$ genügen. Diese tiefe Vermutung hat Wiles zumindest für die große Klasse semistabiler Kurven bewiesen.

4.9.1999

2.3 Faktorisierung mit elliptischen Kurven

Die $(p-1)$-Methode

Vorbild für das ECM-Verfahren war Pollards $(p-1)$-Methode, hinter der eine fast schon unverschämt einfache Idee steckt. Sei dazu p ein Primfaktor von N und $a \in \mathbb{N}$ nicht durch p teilbar; nach dem kleinen Fermatschen Satz ist $a^{p-1} \equiv 1 \bmod p$ und folglich $(a^{p-1}-1, N)$ ein Teiler von N (möglicherweise der triviale Teiler N, falls z.B. sogar $N \mid (a^{p-1}-1)$ ist). Nun kennen wir ja den Exponenten $p-1$ genausowenig wie p selbst; glücklicherweise brauchen wir aber nur ein Vielfaches von $p-1$ zu kennen: denn wenn $m \equiv 0 \bmod p-1$ ist, gilt ja erst recht $a^m \equiv 1 \bmod p$. Daß in dieser letzten Kongruenz noch p auftritt, macht nichts, denn wir rechnen einfach modulo N, will heißen: ist $(p-1) \mid m$ und $p \mid N$, so ist mit $b :\equiv a^m - 1 \bmod N$ der ggT (b, N) ein Teiler von N.

Beispiel: sei $N = 2807$; mit $a = 2$ und $m = 12$ ist $2^{12} - 1 = 4095 \equiv 1288 \bmod 2807$, und ggT $(4095, 1288) = 7$ (nach dem euklidischen Algorithmus). Mit diesem Wert von m hätten wir jeden Primteiler p von N gefunden, für den $p-1$ ein Teiler von 12 ist, also $p = 3, 5, 7, 13$.

Wenn man a^k für große k berechnen will (und die hier auftretenden k *sind* groß), so genügt eine simple `for - next`-Schleife nicht. Hätte man keinen besseren Algorithmus, wäre die $(p-1)$-Methode auch nicht besser als die Division durch alle ungeraden Zahlen unterhalb von $\sqrt{n}$. Man hat aber. Tatsächlich gibt es zwei einfache Algorithmen, die auf der Binärdarstellung von k beruhen und diese von links bzw. von rechts abarbeiten.

Schreiben wir also $k = k_0 + k_1 \cdot 2^1 + \ldots + k_r \cdot 2^r$ mit $k_j \in \{0, 1\}$. Dann ist $a^k = a^{k_0} \cdot (a^2)^{k_1} \cdots (a^{2^r})^{k_r}$, und natürlich braucht man nur diejenigen Potenzen zu berechnen, für die $k_0 = 1$ ist. Man berechnet also durch wiederholtes Quadrieren $a^2 = a \cdot a$, $a^4 = a^2 \cdot a^2$, ..., und $a^{2^r} = a^{2^{r-1}} \cdot a^{2^{r-1}}$, und dann multipliziert man die benötigten Potenzen auf. Wir müssen also insgesamt $r \leq \log_2 k$ mal quadrieren, und dann höchstens r Produkte bilden; insgesamt kommen wir also mit weniger als $2 \log_2 k$ Multiplikationen aus. Demgegenüber frißt eine `for - next`-Schleife $k-1$ Multiplikationen: der Zeitgewinn ist also beträchtlich!

Das obige Verfahren liefert folgenden Algorithmus zur Berechnung von $y = g^n$, $n \geq 0$:

1. `Setze` $y := 1$, $z := g$; `falls` $n = 0$, `gebe` y `aus; ende`.

2. `Ist` n `ungerade, so setze` $y := y \cdot z$;

3. `Setze` $n := \lfloor n/2 \rfloor$; `falls` $n = 0$, `gebe` y `aus; ende. Sonst setze` $z := z^2$ `und gehe nach 2.`

Wir wollen einmal testen, ob 1003 prim ist. Dazu berechnen wir 2^{1002} mod 1003; wenn das Ergebnis nicht $\equiv 1 \bmod 1003$ ist, kann 1003 nach dem kleinen Satz von Fermat nicht prim sein. Nun ist $1002 = (1111101010)_2$, sowie $a^2 \equiv 4$, $a^{2^2} \equiv 16$, $a^{2^3} \equiv 256$, $a^{2^4} \equiv 341$, $a^{2^5} \equiv 936$, $a^{2^6} \equiv 477$, $a^{2^7} \equiv 851$, $a^{2^8} \equiv 35$, $a^{2^9} \equiv 222$, folglich $a^{1002} \equiv 222 \cdot 35 \cdot 851 \cdot 477 \cdot 936 \cdot 256 \cdot 4 \equiv 990 \bmod 1003$, und insbesondere ist 1003 nicht prim.

Wie hat man nun m bei gegebenem N zu wählen? Das einfachste ist, sich eine Schranke B vorzugeben (z.B. $B = 10^4$, wenn man nicht viel Zeit hat, oder $B = 10^6$ auf einem großen Rechner). Dann bildet man das Produkt aller Primzahlpotenzen p^r mit $p^r < B \le p^{r+1}$, und verfährt dann wie oben. Bei $B = 20$ wäre beispielsweise $m = 2^4 \cdot 3^2 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 19$. Für detaillierte Hinweise zur Implementierung sowie zu Modifikationen des Algorithmus siehe [Co1].

Mit $B = 20$ erhalten wir (in unserem Beispiel $n = 1003$) $k = 2^4 \cdot 3^2 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 19 = 232792560$, und es ergibt sich $2^k \equiv 222 \bmod 1003$, sowie $\operatorname{ggT}(221, 1003) = 17$. Die Faktorisierungen $1003 = 17 \cdot 59$, $16 = 2^4$ und $58 = 2 \cdot 29$ erklären, warum wir den Faktor 17 gefunden haben, den Faktor 59 aber nicht.

ECM

Nachdem sich die $(p-1)$-Methode beim Auffinden kleiner Primfaktoren bewährt hatte, versuchte man, ähnliche Methoden zu finden. Erfahrungsgemäß lassen sich viele Sätze, die für $p-1$ gelten (z.B. daß es unendlich viele Primzahlen p gibt, für die $p-1$ durch ein gegebenes $m \ge 2$ teilbar ist) auch für $p+1$ beweisen, indem man die Gruppe $(\mathbb{Z}/p\mathbb{Z})^\times = \mathbb{F}_p^\times$ durch $\mathbb{F}_{p^2}^\times$ ersetzt, oder genauer durch die Untergruppe der Ordnung $p+1$ in $\mathbb{F}_{p^2}^\times$. Dies funktionierte auch hier, und aus der $(p-1)$-Methode wurde eine $(p+1)$-Methode gewonnen, die solche Primfaktoren p finden konnte, für die $p+1$ ein Produkt kleiner Primzahlen ist.

Lenstra hat sich dann gefragt, ob man die Gruppen der Ordnung $p-1$ bzw. $p+1$ nicht ersetzen kann durch elliptische Kurven über $\mathbb{F}_p$; von diesen weiß man (sh. Kapitel 5), daß deren Ordnung den Ungleichungen $p - 2\sqrt{p} < \#E(\mathbb{F}_p) < p + 2\sqrt{p}$ genügt. Während die $(p-1)$-Methode Primteiler findet, für die $p-1$ Produkt kleiner Potenzen von kleinen Primzahlen ist (wir wollen solche Zahlen künftig *glatt* nennen), findet ECM Primteiler, für die

die Ordnung $\#E(\mathbb{F}_p)$ glatt ist. Der Vorteil liegt auf der Hand: kann man mit der elliptischen Kurve E_1 keinen Teiler finden, wählt man einfach eine zweite Kurve E_2; moderne Implementierungen rechnen gewöhnlich mit Hunderten elliptischer Kurven gleichzeitig.

Wie funktioniert Lenstras ECM nun? Nehmen wir einmal an, es sei $p \mid N$, E eine elliptische Kurve, und $k = \#E(\mathbb{F}_p)$ glatt. In diesem Fall wird kP für jeden festen Punkt P auf E gleich dem Punkt $\mathcal{O}$ sein, d.h. die Nenner von x_{kP} und y_{kP} sind beide durch p teilbar. Wenn wir die Addition auf E aber nicht modulo p, sondern modulo N vornehmen, wird im allgemeinen der Nenner z.B. von x_{kP} *nicht* durch N, sondern nur durch p teilbar sein; wenn dies der Fall ist, haben wir p gefunden. Hinter ECM steckt also die Tatsache, daß Addition modulo N (für nicht prime N) im allgemeinen keine Gruppe liefert.

Der genaue Algorithmus sieht daher so aus:

1. `Ist ggT(N,6) = 1?`
2. `Wähle eine elliptische Kurve E und einen Punkt P auf E`
3. `Wähle eine Schranke B`
4. `Sind N und die Diskriminante von E teilerfremd?`
5. `Berechne kP`

Wenn man Glück hat, geht bei der Berechnung von kP etwas schief: ist nämlich k ein Vielfaches der Gruppenordnung von $E(\mathbb{F}_p)$, dann muß bei der Berechnung von kP eine Steigung vorkommen, deren "Nenner" durch p teilbar ist; außer in dem unwahrscheinlichen Fall, daß der Nenner sogar durch N teilbar ist, hat man damit einen Faktor von N gefunden. Ist die Berechnung von kP möglich, so hat man keinen Teiler gefunden und erhöht entweder B oder wählt eine andere elliptische Kurve. (In Wirklichkeit wählt man zuerst den Punkt $P = (x, y)$, und bestimmt danach z.B. den Koeffizienten $b \in \mathbb{Z}$ von $E : y^2 = x^3 + x + b$ so, daß P auf E liegt.)

Beispiel: Sei $E : y^2 = x^3 + x - 9$ und $P = (2, 1)$. Wegen $\Delta = -2^4 \cdot 7 \cdot 313$ ist dies eine elliptische Kurve über $\mathbb{F}_p$ für alle $p \neq 2, 7, 313$. Wählen wir $B = 8$, so ist $k = 8 \cdot 3 \cdot 5 \cdot 7 = 840$, sowie $840 = (1101001000)_2$. Jetzt berechnen wir $2P = (289, 641)$, $4P = (314, 713)$, $8P = (571, 704)$, $16P = (64, 719)$, $32P = (550, 949)$, $64P = (276, 114)$, $128P = (772, 188)$, $256P = (668, 300)$, $512P = (984, 704)$; damit ist $kP = 512P + 256P + 64P + 8P$, aber bereits die Berechnung von $8P + 64P$ geht schief: die Steigung hat nämlich den Nenner

$x_1 - x_0 = 571 - 276 = 295$, und es ist ggT $(295, 1003) = 59$. Tatsächlich tritt dies schon bei $32P + 4P$ auf, wo $550 - 314$ durch 59 teilbar ist.

Übung. Man entwickle eine Faktorisierungsmethode, welche die Addition auf dem Einheitskreis benutzt. (Bem.: Darüber ist anscheinend nichts veröffentlicht.)

2.4 Birationale Transformationen

Nicht immer sieht man einer Kurve an, daß sie elliptisch ist. In diesem Abschnitt wollen wir an einigen Beispielen zeigen, wie man bestimmte Kurven auf Weierstraßnormalform bringen kann.

Wir beginnen mit quartischen, über einem Körper K der Charakteristik $\neq 2$ definierten Kurven der Form

$$C : v^2 = f(u) = au^4 + bu^3 + cu^2 + du + e. \tag{2.8}$$

Solche Kurven sind im unendlichen Punkt singulär und im allgemeinen keine elliptischen Kurven; etwas anders sieht es aus, wenn C einen K-rationalen Punkt (p, q) besitzt. In diesem Fall können wir (nach einer Transformation $u \longmapsto u + p$) annehmen, daß der rationale Punkt von der Form $(0, q)$ ist; die Kurvengleichung ist dann

$$v^2 = au^4 + bu^3 + cu^2 + du + q^2.$$

Multiplikation mit $q^2 u^{-4}$ liefert, wenn man $t = vq/u^2$ und $s = q/u$ setzt,

$$t^2 = s^4 + \frac{d}{q}s^3 + cs^2 + bqs + aq^2.$$

Durch quadratische Ergänzung findet man jetzt Polynome $g(s) = s^2 + g_1 s + g_0$ und $h(s) = h_1 s + h_0$ mit

$$s^4 + \frac{d}{q}s^3 + cs^2 + dqs + aq^2 = g(s)^2 + h(s);$$

in der Tat ist $g_1 = d/2q$ und $g_0 = c/2 - d^2/8q^2$.

Damit kann man die Kurvengleichung in der Form

$$(t - g(s))(t + g(s)) = h(s)$$

schreiben. Jetzt setzt man $t + g(s) = x$, so ist $t - g(s) = h(s)/x$ und $2g(s) = x - h(s)/x$. Multipliziert man die letzte Gleichung mit x^2, so folgt

$$x^3 - h_1 xs + h_0 x = 2x^2(s^2 + g_1 s + g_0),$$

und mit $xs = y$ schreibt sich das als

$$x^3 - h_1 y + h_0 x = 2y^2 + 2gh_1 xy + 2g_0 x^2.$$

Multiplikation mit 8 und Ersetzen von (x, y) durch $(2x, 4y)$ liefert dann eine lange Weierstraßform, und der Rest geht wie am Anfang dieses Kapitels.

Übung. Man transformiere die “Fermatgleichung” $y^2 = 1 + x^4$ auf Weierstraßform.

Übung. Man stelle fest, wie die expliziten Substitutionen lauten, die (2.8) auf lange Weierstraßform bringen.

Übung. Man bestimme die Umkehrabbildung, die die Weierstraßgleichung auf die Form $v^2 = au^4 + bu^3 + cu^2 + du + q^2$ bringt. Welchem Punkt auf der Weierstraßkurve entspricht $(0, q)$?

FLT für $n = 4$

Der Beweis, daß die Gleichung $x^4 + y^4 = z^4$ in ganzen Zahlen x, y, z nur die trivialen Lösungen mit $xyz = 0$ besitzt, ist der einfachste Fall unter den Gleichungen vom Typ $x^n + y^n = z^n$ für $n \geq 3$. Auf den ersten Blick erstaunlich ist die Tatsache, daß ein direkter Beweis dieser Aussage sehr viel schwieriger ist als ein Beweis der allgemeineren Behauptung, wonach sogar $x^4 + y^4 = z^2$ nur triviale ganzzahlige Lösungen hat; auf den zweiten Blick wird dies aber verständlich: denn $x^4 + y^4 = z^2$ ist, wie wir oben gesehen haben, eine elliptische Kurve, $x^4 + y^4 = z^4$ dagegen nicht.

Wir beginnen mit einem Beweis, der im wesentlichen auf Fermat selbst zurückgeht. Zuvor aber erinnern wir an folgenden

Hilfssatz 2.12. *Sind a, b teilerfremde natürliche Zahlen mit $ab = c^n$, dann ist $a = r^n$ und $b = s^n$ mit $r, s \in \mathbb{N}$.*

Beweis des Hilfssatzes. Man braucht nur a und b in ihre Primfaktoren zu zerlegen: kommt in a genau p^m vor, darf b nicht durch p teilbar sein; da ab eine n-te Potenz ist, muß also auch p^m eine sein, und dies ist genau dann der Fall, wenn m durch n teilbar ist. Also sind a und b bis auf das Vorzeichen selbst n-te Potenzen. $\square$

Die Beweisidee für den Fall $n = 4$ von FLT ist, wie wir noch sehen werden, im Grunde dieselbe wie diejenige, die hinter dem zweiten 2-Abstieg steckt: will man alle rationalen Punkte auf $y^2 = f(x^2)$ bestimmen, wo f ein quadratisches Polynom (selbstverständlich mit rationalen Koeffizienten) ist, so untersucht man zuerst, ob $y^2 = f(x)$ einen rationalen Punkt enthält (dafür gibt es einen Algorithmus: vgl. das Hassesche Lokal-Global-Prinzip). Gibt es keinen solchen Punkt, kann es auch auf $y^2 = f(x^2)$ keinen geben; gibt es einen, so können wir (nach Diophant) *alle* rationalen Punkte finden, und müssen dann "nur noch" nachschauen, ob einer dieser Punkte ein Quadrat als x-Koordinate besitzt.

Dies wenden wir nun auf $x^4 + y^4 = z^2$ an und nehmen an, diese Gleichung besitze eine Lösung mit $xy \neq 0$. Unter diesen wählen wir eine aus, für die $\max\{|x|, |y|\}$ minimal ist, für die also insbesondere x, y und z paarweise teilerfremd sind. Die rationalen Punkte auf $X^2 + Y^2 = Z^2$ haben wir bereits bestimmt: $X = 2T$, $Y = 1 - T^2$ und $Z = 1 + T^2$ mit $T = U/V$ liefert die ganzzahligen Lösungen $X = 2UV$, $Y = V^2 - U^2$ und $Z = V^2 + U^2$. Angewandt auf dieses Problem erhalten wir $x^2 = 2UV$, $y^2 = V^2 - U^2$ und $z = V^2 + U^2$ für $U, V \in \mathbb{Z}$ mit $(U, V) = 1$. Aus $x^2 = 2UV$ folgt aber, daß $U = 2u^2$ und $V = v^2$ oder $U = u^2$ und $V = 2v^2$ ist.

Der zweite Fall gibt $y^2 = 4v^4 - u^4$ und ist modulo 4 unmöglich; also ist $U = 2u^2$ und $V = v^2$, somit $y^2 = v^4 - 4u^4$. Aus $4u^4 = v^4 - y^2 = (v^2 - y)(v^2 + y)$ folgt, daß $\frac{1}{2}(v^2 - y)$ und $\frac{1}{2}(v^2 + y)$ teilerfremd sind; nach Hilfssatz 2.12 ist somit $v^2 + y = 2r^4$ und $v^2 - y = 2s^4$. Addition gibt $v^2 = r^4 + s^4$, also eine weitere Lösung der Ausgangsgleichung, und wegen $x^4 = 16u^4v^4 = 16u^4r^4s^4$ ist sicherlich $\max\{|r|, |s|\} < |x| < \max\{|x|, |y|\}$: dies ist ein Widerspruch.

Wir halten für später fest, daß der Beweis, ausgehend von einem Punkt $(y^2/x^2, z/x^2)$ auf $Y^2 = X^4 + 1$, über einen Punkt auf der Kurve $Y^2 = X^4 - 4$ zu einem "kleineren" Punkt auf der Ausgangskurve führt.

Übung. Man transformiere $Y^2 = X^4 - 4$ auf Weierstraßform.

Schnitt zweier Quadriken

Wir wollen anhand eines Beispiels zeigen, wie man ein System zweier quadratischer Gleichungen unter Umständen als elliptische Kurve erkennt. Sei dazu das System

$$\begin{aligned} Q_1 &: U^2 - V^2 + kX^2 = 0 \\ Q_2 &: W^2 - V^2 - kX^2 = 0 \end{aligned} \tag{2.9}$$

gegeben. Elimination von kX^2 gibt die Gleichung

$$U^2 + W^2 = 2V^2, \tag{2.10}$$

die aber nicht zu dem System (2.9) äquivalent ist: jeder Lösung (u, v, w) von (2.10) entsprechen i.a. zwei Lösungen $(u, v, w, \pm x)$ des Systems.

Gleichung (2.10) enthält den rationalen Punkt $(1, 1, 1)$, ist folglich rational parametrisierbar. Dehomogenisieren mit $V = 1$ gibt $x^2 + w^2 = 2$ mit rationalem Punkt $(1, 1)$; die Gerade $u = t(w - 1) + 1$ gibt den zweiten Schnittpunkt

$$w = \frac{t^2 - 2t - 1}{t^2 + 1}, \quad u = \frac{-t^2 - 2t + 1}{t^2 + 1}.$$

Also ist $(U, V, W) = (t^2 - 2t - 1, t^2 + 1, -t^2 - 2t + 1)$ eine Parametrisierung der homogenen Form, und der Schnitt von Q_1 und Q_2 wird beschrieben durch

$$kX^2 = V^2 - U^2 = W^2 - V^2 = -4t^3 + 4t.$$

Die Transformation $X = 2y/k^2$, $t = -x/k$ liefert uns schließlich die kurze Weierstraßform

$$E : y^2 = x^3 - k^2 x.$$

Übung. Man bringe das System $x^2 + x + 2 = y^2$, $x^2 - x - 2 = z^2$ auf Weierstraßform.

Selmerkurven

Kubische Kurven der Form

$$au^3 + bv^3 = c \tag{2.11}$$

mit $abc \neq 0$ sind von Selmer ausführlich untersucht worden; für solche Kurven hat er erstmals auch eine Gruppe definiert, die heute Selmergruppe heißt. Hier wollen wir nur zeigen, wie man Selmerkurven auf Weierstraßnormalform bringt.

Am einfachsten geht das für Kurven der Form $u^3 + v^3 = c$; deren homogene Form ist $u^3 + v^3 = cw^3$. Setzt man $u = x - y$ und $v = x + y$, so folgt $cw^3 = 2x^3 + 6xy^2$. Division durch x^3 gibt $c(w/x)^3 = 6(y/x)^2 + 2$; multipliziert man das ganze mit 6^3c^2, so folgt $(6cw/x)^3 - 2^4 \cdot 3^3 c^2 = (36cy/x)^2$, also $Y^2 = X^3 - 432c^2$ mit $Y = 36cy/x$ und $X = 6cw/x$. Insbesondere ist die Fermatkubik $u^3 + v^3 = 1$ birational isomorph zu $y^2 = x^3 - 432$.

Etwas allgemeiner gilt

Proposition 2.13. *Eine Selmerkurve der Form (2.11), welche über einem Körper der Charakteristik $\neq 2, 3$ definiert ist, und für die (eventuell nach* *einer Permutation von (a, b, c) die Wurzel $\theta := \sqrt[3]{c/b} \in K$ liegt, ist birational isomorph zur Weierstraßkurve*

$$y^2 = x^3 - 432a^2b^2c^2$$

unter den Transformationen

$$\begin{array}{rclrcl} u & = & -\frac{6b\theta^2 x}{y-36abc}, & v & = & \frac{y+36abc}{y-36abc}, \\ x & = & -\frac{12ab\theta^2 u}{v-\theta}, & y & = & 36abc\frac{v+\theta}{v-\theta} \end{array}$$

Der Beweis ist jetzt eine einfache Übungsaufgabe: man beginnt damit, (2.11) durch a zu dividieren, der Rest geht von selbst.

Das folgende Resultat stammt von Euler:

Proposition 2.14. *Ist $au^3 + bv^3 + cw^3 = 0$, dann gilt $r^3 + s^3 + abct^3 = 0$ mit*

$$\begin{array}{rcl} r & = & -6bc^2w^6 - c^3w^9 - 3b^2cv^6w^3 + b^3v^9, \\ s & = & -3bc^2w^6 + c^3w^9 - 6b^2cv^6w^3 - b^3v^9, \\ t & = & -3uvw(b^2v^6 + bcv^3w^3 + c^2w^6). \end{array}$$

Wie man diese Formeln herleiten kann, werden wir noch zeigen; hier bemerken wir lediglich, daß dieses Ergebnis hervorragend geeignet ist, rationale Punkte auf Kurven der Form $r^3 + s^3 + dt^3 = 0$ zu finden: man faktorisiert $d = abc$ und sucht Punkte auf allen Kurven $au^3 + bv^3 + cw^3 = 0$; diese werden im allgemeinen kleinere Koordinaten haben als die entsprechenden auf $r^3 + s^3 + dt^3 = 0$, d.h. eine Suche auf den assoziierten Kurven ist i.a. erfolgreicher.

Kapitel 3

Torsionspunkte: Satz von Nagell-Lutz

3.1 Überblick

Unser Fernziel ist es, für die Gruppe $E(\mathbb{Q})$ der rationalen Punkte einer über $\mathbb{Q}$ definierten elliptischen Kurve $E : y^2 = x^3 + ax + b$ zu zeigen, daß $E(\mathbb{Q}) = E(\mathbb{Q})_{\text{tors}} \times \mathbb{Z}^r$ gilt, wobei $E(\mathbb{Q})_{\text{tors}}$ die Gruppe aller Punkte endlicher Ordnung (also die Torsionsgruppe von $E(\mathbb{Q})$) bezeichnet, und daß sowohl $E(\mathbb{Q})_{\text{tors}}$, als auch r endlich sind.

Übung. Sei G eine abelsche Gruppe und G_{tors} die Menge aller Elemente von G mit endlicher Ordnung. Zeige, daß G_{tors} eine Gruppe ist.

Hier besteht ein wesentlicher Unterschied zwischen elliptischen Kurven und singulären kubischen Kurven: für $C : y^2 = x^3$ war ja $C(\mathbb{Q}) = (\mathbb{Q}, +)$, und dies ist eine nicht endlich erzeugte Gruppe. Betrachten wir nämlich $G = \langle \alpha_1, \ldots, \alpha_n \rangle$, und ist n der Hauptnenner der α_j, so enthält $G \subset (\mathbb{Q}, +)$ nur solche Elemente, deren Nenner ein Teiler von n ist; folglich ist niemals $G = (\mathbb{Q}, +)$.

Ähnlich sieht es mit $C : y^2 = x^3 + x^2$ aus, wo $C(\mathbb{Q}) \simeq \mathbb{Q}^\times$ war. Hier ist $C(\mathbb{Q})_{\text{tors}} \simeq \{-1, +1\}$ (denn die einzigen Elemente endlicher Ordnung in $\mathbb{Q}^times$ sind ± 1), und es gilt $C(\mathbb{Q}) = C(\mathbb{Q})_{\text{tors}} \times \bigoplus_p \mathbb{Z}$, wo die direkte Summe über alle Primzahlen geht: diese Aussage ist nichts anderes als der Satz von der eindeutigen Primfaktorzerlegung, wonach jedes Element aus $\mathbb{Q}^\times$

sich eindeutig in der Form $(-1)^\alpha \prod_p p^{\alpha(p)}$ schreiben läßt, wobei nur endlich viele $\alpha(p)$ von 0 verschieden sind (deswegen die direkte Summe).

Zurück zu elliptischen Kurven über $\mathbb{Q}$. Als Einstimmung bestimmen wir die Punkte der Ordnung 2 oder 3 explizit. Für die Gruppe (!) aller K-rationalen Punkte, deren Ordnung ein Teiler von $N \in \mathbb{N}$ ist, schreiben wir auch $E(K)[N]$.

Sei zuerst P ein Punkt der Ordnung 2 auf $E(\mathbb{C})$, also $P = (x, y)$ mit $x, y \in \mathbb{C}$ und $2P = \mathcal{O}$. Letzteres ist äquivalent zu $P = -P$, also zu $(x, y) = -(x, y) = (x, -y)$. Damit haben wir gesehen: ein Punkt $P = (x, y) \neq \mathcal{O}$ hat genau dann Ordnung 2, wenn $y = 0$ ist. Die dazugehörigen x-Koordinaten erhält man aus $0 = y^2 = x^3 + ax + b$: in $\mathbb{C}$ hat diese Gleichung genau drei Lösungen x_1, x_2, x_3, und diese sind alle verschieden, weil sonst die Diskriminante von $x^3 + ax + b$ und damit auch Δ verschwinden würde. Es gibt also genau vier Punkte $P \in E(\mathbb{C})$ mit $2P = 0$, nämlich $\mathcal{O}$, $(x_1, 0)$, $(x_2, 0)$ und $(x_3, 0)$. Da jeder dieser Punkte Ordnung höchstens 2 hat, muß $E(\mathbb{C})[2] \simeq \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ sein.

Wegen $\mathbb{Q} \subset \mathbb{C}$ ergeben sich daraus für die rationalen 2-Teilungspunkte folgende Möglichkeiten:

- $f(x)$ ist irreduzibel über $\mathbb{Q}$, hat also keine rationale Nullstelle; dann ist $E(\mathbb{Q})[2] = \{\mathcal{O}\}$.
- $f(x)$ ist über $\mathbb{Q}$ Produkt eines linearen und eines irreduziblen quadratischen Polynoms, hat also genau eine rationale Nullstelle x_1; dann ist $E(\mathbb{Q})[2] = \{\mathcal{O}, (x_1, 0)\} \simeq \mathbb{Z}/2\mathbb{Z}$.
- $f(x)$ hat drei rationale Nullstellen $x_1, x_2, x_3 \in \mathbb{Q}$; dann gilt $E(\mathbb{Q})[2] = \{\mathcal{O}, (x_1, 0), (x_2, 0), (x_3, 0)\} \simeq \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$.

Die Bestimmung von $E(\mathbb{Q})[3]$ ist sehr einfach, wenn man sich der geometrischen Methode bedient. Dazu sei P ein Wendepunkt des Graphen von $E : y^2 = x^3 + ax + b$; eine Tangente an E in P hat dann Vielfachheit 3, folglich ist $P + P = -P$, d.h. $3P = \mathcal{O}$. Umgekehrt folgt aus $3P = \mathcal{O}$ für einen Punkt $P \neq 0$ mit reellen Koordinaten, daß P Wendepunkt ist. Nun hat das Schaubild von E aber entweder keinen oder genau zwei Wendepunkte, folglich ist entweder $E(\mathbb{R})[3] = \{\mathcal{O}\}$ oder $E(\mathbb{R})[3] = \{\mathcal{O}, P, -P\} \simeq \mathbb{Z}/3\mathbb{Z}$. Wegen $\mathbb{Q} \subset \mathbb{R}$ impliziert dies sofort, daß $E(\mathbb{Q})[3]$ entweder trivial oder $\simeq \mathbb{Z}/3\mathbb{Z}$ ist; insbesondere gibt es keine über $\mathbb{Q}$ definierte elliptische Kurve mit $E(\mathbb{Q})[3] \supseteq \mathbb{Z}/3\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$. Dabei haben wir aber die Anschauung benutzt: daß nämlich der Graph von E maximal zwei Wendepunkte besitzt, ist etwas, das man noch zeigen muß.

4.9.1999

Übung. Bestimme $E(\mathbb{R})[2]$ in Abhängigkeit der Anzahl der reellen Nullstellen von f, wenn $E : y^2 = f(x)$ ist.

Übung. Vervollständige obigen Beweis von $\#E(\mathbb{R})[3] \leq 3$.

Übung. Berechne $E(\mathbb{Q})[2]$ für die elliptischen Kurven $E_1 : y^2 = x^3 - 2$, $E_2 : y^2 = x^3 + x$ und $E_3 : y^2 = x^3 - x$.

Ziel dieses Kapitels ist der Beweis des Satzes von Nagell-Lutz:

Satz 3.1. *Sei $y^2 = x^3 + ax + b$ eine elliptische Kurve mit Koeffizienten $a, b \in \mathbb{Z}$. Ist $P = (x, y)$ ein Torsionspunkt, dann gilt*
i) $x, y \in \mathbb{Z}$;
ii) entweder ist $y = 0$ oder $y^2 \mid D = 4a^3 + 27b^2$.

Damit ist die Bestimmung von $E(\mathbb{Q})_{\text{tors}}$ zumindest für Kurven mit kleinem D ein Kinderspiel: nehmen wir beispielsweise $E : y^2 = x^3 + x$, so ist $D = 4$ und damit $y \in \{\pm 1, \pm 2\}$. Keiner dieser Werte führt auf einen rationalen Punkt, also ist $E(\mathbb{Q})_{\text{tors}} = E(\mathbb{Q})[2] = \{\mathcal{O}, (0,0)\}$.

Übung. Bestimme die Torsionsgruppen $E(\mathbb{Q})_{\text{tors}}$ für folgende elliptische Kurven: $y^2 = x^3 - x$, $y^2 = x^3 - 43x + 166$, $y^2 = x^3 - 219x + 1654$. (Lösung: $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, $\mathbb{Z}/7\mathbb{Z}$, $\mathbb{Z}/9\mathbb{Z}$).

Übung. Zeige $E(\mathbb{Q})_{\text{tors}} = \{\mathcal{O}\}$ für $E : y^2 = x^3 - 2$ und folgere aus $(5, 3) \in E(\mathbb{Q})$, daß E unendlich viele rationale Punkte besitzt, also Rang ≥ 1 hat.

Die wesentliche Aussage des Satzes von Nagell-Lutz ist die Ganzzahligkeit von x und y; die Teilbarkeitsrelation ii) ist eine ganz einfache Konsequenz:

Hilfssatz 3.2. *Ist $P = (x_P, y_P)$ ein Punkt auf $E : y^2 = x^3 + ax + b$, und haben P und $2P$ ganzzahlige Koordinaten, dann ist $y_P = 0$ oder $y_P^2 \mid D = 4a^3 + 27b^2$.*

Beweis. Nach der Verdoppelungsformel gilt

$$x_{2P} = \frac{\phi(x_P)}{4\psi(x_P)} \quad \text{mit} \quad \begin{cases} \phi(X) &= X^4 - 2aX^2 - 8bX + a^2 \text{ und} \\ \psi(X) &= X^3 + aX + b \end{cases}$$

Mit $f(X) = 3X^2 + 4a$ und $g(X) = 3X^3 - 5aX - 27b$ gilt, wie man sofort nachrechnet, $f(X)\phi(X) - g(X)\psi(X) = D$ (verifizieren ist natürlich einfach; man kann solche Identitäten aber auch herleiten: sh. Anhang A). Setzen wir $X = x_P$ in dieser Identität und beachten $\phi(x_P) = 4x_{2P}\psi(x_P)$, sowie $\psi(x_P) = y_P$, dann folgt

$$y_P^2[4x_{2P}f(x_P) - g(x_P)] = D.$$

Da nach Voraussetzung x_P, y_P und x_{2P} ganze Zahlen sind, bedeutet das $y_P^2 \mid D$, und das war zu zeigen. $\square$

Damit können wir die Implikation i) $\Longrightarrow$ ii) ganz einfach beweisen: ist $P = (x, y)$ ein Torsionspunkt, so auch $2P$ (denn mit $nP = \mathcal{O}$ ist natürlich erst recht $2nP = \mathcal{O}$); nach Teil i) haben beide Punkte ganze Koordinaten, und der Hilfssatz liefert dann $y^2 \mid D$.

Als sofortige Folgerung aus dem Satz von Nagell-Lutz halten wir fest, daß es auf einer gegebenen elliptischen Kurve über $\mathbb{Q}$ nur endlich viele Torsionspunkte gibt: dies liegt daran, daß es nur endlich viele y mit $y^2 \mid D$ und folglich auch nur endlich viele Punkte (x, y) mit dieser Eigenschaft gibt.

Im nächsten Abschnitt wenden wir uns dem Beweis der Ganzzahligkeit von rationalen Torsionspunkten zu; das wesentliche Hilfsmittel hierzu ist die "Reduktion" elliptischer Kurven.

3.2 Reduktion modulo p

Ist $y^2 = x^3 + ax + b$ eine über $\mathbb{Q}$ definierte elliptische Kurve, so kann man durch einfache Transformationen erreichen, daß a und b ganze Zahlen sind: ist nämlich t der Hauptnenner von a und b, so folgt durch Multiplikation mit t^6 die Gleichung $(t^3y)^2 = (t^2x)^3 + t^5a(tx) + t^6b$, mit $Y = t^3y$ und $X = t^2X$ also $Y^2 = X^3 + a'X + b'$, wobei jetzt $a' = t^5a$ und $b' = t^6b$ ganze Zahlen sind.

Für elliptische Kurven $y^2 = x^3 + ax + b$ macht es aber Sinn, die Gleichung modulo p zu lösen; für alle $p \nmid \Delta$ ist das Ergebnis dann eine elliptische Kurve $\overline{E} : y^2 = x^3 + \overline{a}\,x + \overline{b}$ über $\mathbb{F}_p$ (der Querstrich steht hier für die Restklasse modulo p).

Tatsächlich ist es vorteilhaft, diese Reduktion nicht nur von $\mathbb{Q}$ nach $\mathbb{F}_p$, sondern allgemeiner von $\mathbb{Q}_p$ nach $\mathbb{F}_p$ zu studieren (wegen $\mathbb{Q} \subset \mathbb{Q}_p$ ist das in der Tat allgemeiner). Dazu nehmen wir an, $E : y^2 = x^3 + ax + b$ sei eine über $\mathbb{Q}_p$ definierte elliptische Kurve mit Koeffizienten in $\mathbb{Z}_p$; indem wir jedem Element $a = a_0 + a_1p + a_2p^2 + \ldots \in \mathbb{Z}_p$ das Element $\overline{a} = a_0 \in \mathbb{F}_p$ zuordnen, erhalten wir den Reduktionshomomorphismus $\overline{} : \mathbb{Z}_p \longrightarrow \mathbb{F}_p : a \longmapsto \overline{a}$.

Natürlich wollen wir auch untersuchen, wie sich die Punkte auf E bzw. $\overline{E}$ bei diesen Reduktionen verhalten. Dazu sei $(x_0 : y_0 : z_0) \in \mathbb{P}^2(\mathbb{Q}_p)$ irgend ein Punkt; durch Multiplikation mit einem geeigneten $\lambda \in \mathbb{Q}_p$ (es genügt sogar ein geeignetes $\lambda \in \mathbb{Q}$) können wir erreichen, daß $(x_0 : y_0 : z_0) = (x : y : z)$ mit $x, y, z \in \mathbb{Z}_p$ und $\max\{|x|_p, |y|_p, |z|_p\} = 1$ ist (hier bezeichnet $|x|_p$ den p-adischen Betrag von x: für $p^m \parallel x$ ist $|x|_p = p^{-m}$), m.a.W.: daß die Koordinaten ganz, aber nicht alle durch p teilbar sind. Dann setzen wir $\overline{P} = (\overline{x} : \overline{y} : \overline{z})$.

Auf dieselbe Weise können wir Geraden reduzieren: ist $g : rx+sy+tz = 0$ eine Geradengleichung in $\mathbb{P}^2(\mathbb{Q}_p)$, so dürfen wir oBdA annehmen, daß die Koeffizienten $a, b, c \in \mathbb{Z}_p$ und nicht alle drei durch p teilbar sind. Dann definieren wir wie Reduktion von g durch $\overline{g} : \overline{r}x + \overline{s}y + \overline{t}z = 0$.

Trotz dieser auf den ersten Blick etwas holprig aussehenden Definition erweist sich die Reduktion als eine ziemlich gutartige Abbildung: sie respektiert nämlich das Gruppengesetz:

Proposition 3.3. *Sei $E : y^2 = x^3 + ax + b$ eine über $\mathbb{Q}_p$ definierte Kurve mit Koeffizienten $a, b \in \mathbb{Z}_p$; sind dann $P_1, P_2, P_3 \in E(\mathbb{Q}_p)$ kollinear, so gilt dasselbe für die reduzierten Punkte $\overline{P}_1, \overline{P}_2$ und $\overline{P}_3$, und zwar mit der korrekten Vielfachheit: fallen z.B. P_1 und P_2 bei der Reduktion zusammen (ist also $\overline{P}_1 = \overline{P}_2$), so schneidet die Reduktion der Geraden durch die P_j die reduzierte Kurve $\overline{E}$ im Punkt $\overline{P}_1$ mit Vielfachheit ≥ 2.*

Beweis. Wir beginnen etwas allgemeiner mit dem Studium einer beliebigen kubischen Kurve C, die in projektiver Form durch $F(X, Y, Z) = 0$ beschrieben wird, wobei F ein Polynom in drei Variablen mit Koeffizienten aus $\mathbb{F}_p$ sein soll. Dabei nehmen wir ohne Beschränkung der Allgemeinheit an, daß mindestens einer der Koeffizienten von F eine p-adische Einheit, also nicht durch p teilbar ist.

Weiter sei $g : l_1X + l_2Y + l_3Z = 0$ eine Gerade; wir nehmen an, daß mindestens einer der Koeffizienten eine p-adische Einheit ist, und indem wir notfalls die Koordinaten vertauschen, dürfen wir annehmen, daß dies l_3 ist (dies ist der Grund, warum wir nicht von vornherein Kurven in Weierstraßform betrachten: die Weierstraßform geht bei solchen Vertauschungen verloren). Indem wir die Geradengleichung durch die Einheit $-l_3$ dividieren, dürfen wir sogar annehmen, daß die Gerade durch $g : Z = lX + mY$ gegeben ist. Die Schnittpunkte von g mit C sind durch die Nullstellen von $G(X, Y) := F(X, Y, lX + mY) = 0$ gegeben; Reduktion liefert $\overline{G}(X, Y) = \overline{F}(X, Y, \overline{l}X + \overline{m}Y) = 0$.

Es kann durchaus vorkommen, daß $\overline{G}$ identisch verschwindet: geometrisch bedeutet dies, daß die reduzierte Kurve eine Gerade enthält. Ein Beispiel für ein solches Verhalten ist die kubische Fermatkurve $C : X^3 + Y^3 = Z^3$, deren Reduktion modulo 3 sich in der Form $(X+Y-Z)^3 = 0$ schreiben läßt und insbesondere die Gerade $Z = X + Y$ enthält (dies erklärt auch, warum C über $\mathbb{F}_3$ nur aus singulären Punkten besteht). Glücklicherweise kann dies bei Weierstraßkurven nicht vorkommen: eine Kurve der Form $Y^2Z = X^3 +$

$aXZ + bZ^3$ kann durch Ersetzen von z.B. $Z = lX + mY$ plus Reduktion modulo p nicht identisch zum Verschwinden gebracht werden, weil der Term X^3 überlebt.

Wir dürfen also annehmen, daß die Reduktion $\overline{G}$ nicht identisch verschwindet. Betrachten wir nun die Punkte $P_j = (x_j : y_j : z_j)$ mit $j = 1, 2, 3$; wir nehmen an, daß die Koordinaten in $\mathbb{Z}_p$ liegen und für gegebenes j nicht alle durch p teilbar sind. Wir bemerken, daß $(\overline{x}_j, \overline{y}_j)$ für kein j gleich $(0, 0)$ sein kann: sonst wäre nämlich auch $\overline{z}_j = \overline{l}\,\overline{x}_j + \overline{m}\,\overline{y}_j$ gleich 0, was der Normierung unserer Koordinaten widerspräche.

Da die Punkte P_j auf der Geraden und der Kurve liegen, gibt es ein $\lambda \in \mathbb{Q}_p$ mit $F(X, Y, lX + mY) = \lambda H(X, Y)$ mit $H(X, Y) = (y_1X - x_1Y)(y_3X - x_3Y)(y_3X - x_3Y)$. Wie wir eben gesehen haben, kann die Reduktion von H nicht identisch verschwinden. Falls also $\lambda \in \mathbb{Z}_p$ ist, haben wir $\overline{F}(X, Y, \overline{l}X + \overline{m}Y) = \overline{\lambda H}(X, Y)$, und da auch $\overline{F}$ nicht identisch verschwindet, muß $\overline{\lambda} \neq 0$, also λ eine Einheit in $\mathbb{Z}_p$ sein. Wäre $\lambda \notin \mathbb{Z}_p$, so ist $\lambda^{-1} \in p\mathbb{Z}_p$ und folglich $0 = \overline{\lambda F} = \overline{H}$: Widerspruch.

Also ist $\overline{F} = c\overline{H}$ für ein $c = \overline{\lambda} \in \mathbb{F}_p^\times$. Damit sind die reduzierten Punkte $\overline{P}_j$ kollinear; außerdem haben die reduzierten Punkte in der Tat die korrekten Vielfachheiten. □

Wichtig für die Bestimmung der Struktur von $E(\mathbb{Q}_p)$ (genauer: von deren unten definierten Untergruppe $E^{(0)}$ von endlichem Index) ist nun, daß die Reduktion, sieht man einmal von singulären Punkten ab, surjektiv ist:

Proposition 3.4. *Sei E wie oben; ist Q ein Punkt auf dem nichtsingulären Teil $\overline{E}_{ns}$ der reduzierten Kurve, so existiert ein $P \in E(\mathbb{Q}_p)$ mit $Q = \overline{P}$.*

Zum Beweis benötigen wir ein Standardresultat aus der Theorie der p-adischen Zahlen, nämlich das Henselsche Lemma, und zwar genügt uns eine ganz einfache Version:

Hilfssatz 3.5. *Sei $f \in \mathbb{Z}_p[T]$ ein Polynom in T mit Koeffizienten aus $\mathbb{Z}_p$, und sei $|f(t_0)|_p < 1$ und $|f'(t_0)|_p = 1$ für ein $t_0 \in \mathbb{Z}_p$. Dann existiert ein $t \in \mathbb{Z}_p$ mit $f(t) = 0$ und $|t - t_0|_p \leq |f(t_0)|_p$.*

Der Beweis von Proposition 3.4 ist damit ganz einfach:

Beweis. Sei $F(X, Y, Z) = 0$ die projektive Weierstraßgleichung, die E beschreibt. Da $Q = (x : y : z)$ nicht singulär ist, können wir oBdA annehmen, daß $\partial F/\partial X$ in Q nicht verschwindet. Jetzt wählen wir $a, b, c \in \mathbb{Z}_p$ beliebig mit $\overline{a} = x$, $\overline{b} = y$ und $\overline{c} = z$. Dann genügt $f(T) := F(T, b, c)$ den Bedingungen von Hensels Lemma, und der Punkt $P = (t, b, c)$ tut's. □

Das Henselsche Lemma wiederum ist im wesentlichen nichts anderes als das Newtonverfahren in $\mathbb{Q}_p$: ausgehend von der Näherung t_0 konstruieren wir eine Folge von Approximationen $t_{n+1} = t_n - f(t_n)/f'(t_n)$ und zeigen, daß diese gegen das gewünschte Element konvergiert. Dazu starten wir mit t_0 und bilden $u_0 = -f(t_0)/f'(t_0)$. Wir hätten gern, daß $t_1 = t_0 + u_0$ eine bessere Approximation an (das bis jetzt hypothetische) $t \in \mathbb{Z}_p$ ist als t_0; um $f(t_1)$ zu berechnen, entwickeln wir das Polynom in zwei Variablen $f(T+U)$ in der Form $f(T+U) = f(T) + Uf_1(T) + U^2 f_2(T) + ... + U^d f_d(T)$ mit $d = \deg f$; hier sind $f_1, ..., f_d(T)$ Polynome mit ganzen Koeffizienten. Ein Vergleich mit der Taylorentwicklung zeigt $f_1(T) = f'(T)$, folglich ist $f(t_1) = f(t_0 + u_0) = f(t_0) + u_0 f'(t_0) + u_0^2 h$ für ein $h \in \mathbb{Z}$, nach Wahl von u also $f(t_1) = u_0^2 h$. Nun ist $|u_0|_p = |f(t_0)|_p / |f'(t_0)|_p$; nach Voraussetzung ist $f'(t_0)$ eine p-adische Einheit, folglich $|u_0|_p = |f(t_0)|_p < 1$ und damit wegen $|h|_p \le 1$ (denn h ist ganz) $|f(t_1)|_p = |u_0^2 h|_p \le |u_0^2|_p = |f(t_0)|_p^2$. Mit anderen Worten: wegen $|f(t_0)|_p < 1$ ist $f(t_1)$ p-adisch sicherlich näher an der 0 als $f(t_0)$ (wie das gewöhnliche Newtonverfahren ist die Konvergenz hier quadratisch).

Um diesen ersten Schritt wiederholen zu können, müssen wir sicherstellen, daß die zweite Voraussetzung $p \nmid f'(t_0)$ nicht verlorengegangen ist. Wegen $f'(t_1) = f'(t_0) + uh_1$ für ein p-ganzes h_1 ist aber p ein Teiler der Differenz $f'(t_1) - f'(t_0)$, insbesondere also $f'(t_1)$ nicht durch p teilbar.

Jetzt konstruieren wir wie oben ein $t_2 = t_1 + u_1$ mit $|u_1|_p = |f(t_1)|_p$, $|f(t_2)|_p \le |f(t_1)|_p^2$ und $p \nmid f'(t_2)$. Iterieren liefert eine Folge $t_0, t_1, t_2, \ldots$, die wegen $|t_{m+1} - t_m|_p = |u_m|_p = |f(t_m)|_p \longrightarrow 0$ gegen ein $t \in \mathbb{Z}_p$ konvergiert. Da Polynome stetig bezüglich der von $|\cdot|_p$ induzierten Topologie sind, folgt aus $|f(t_n)|_p \longrightarrow 0$ sofort $f(t) = 0$. Damit ist Hensels Lemma bewiesen.

3.3 Lokale Kriterien

Sei wie oben $E : y^2 = x^3 + ax + b$ eine über $\mathbb{Q}_p$ definierte elliptische Kurve mit $a, b \in \mathbb{Z}_p$ und $\pi : \mathbb{P}^2(\mathbb{Q}_p) \longrightarrow \mathbb{P}^2(\mathbb{F}_p)$ die oben definierte Reduktion.

Die Beweisidee zum Satz von Nagell-Lutz sieht wie folgt aus: sei $\overline{E}^{(0)} = E_{\mathrm{ns}}(\mathbb{F}_p)$ die Gruppe der nichtsingulären Punkte auf der Reduktion von E und $E^{(0)} = \{P \in E(\mathbb{Q}_p) : \overline{P} \in \overline{E}^{(0)}\}$ die Menge der Punkte, die bei Reduktion nicht auf dem singulären Punkt landen. $E^{(0)}$ ist eine Gruppe: sind nämlich $P, Q \in E^{(0)}$, so ist $\pi(P+Q) = \pi(P) + \pi(Q) \in \overline{E}^{(0)}$, da $E_{\mathrm{ns}}(\mathbb{F}_p)$ eine Gruppe ist; nach Definition von $E^{(0)}$ ist damit $P + Q \in E^{(0)}$.

Da insbesondere $\mathcal{O}$ ein nichtsingulärer Punkt ist, ist $E^{(1)} := \ker \pi$ eine Untergruppe von $E^{(0)}$. Deren wichtigste Eigenschaft ist

Satz 3.6. *Die Gruppe $E^{(1)}$ ist torsionsfrei.*

Damit haben wir gewonnen:

Korollar 3.7. *Sei E wie oben. Ist dann $(x, y) \in E(\mathbb{Q}_p) \setminus \{\mathcal{O}\}$ ein Torsionspunkt, so gilt $x, y \in \mathbb{Z}_p$.*

Beweis. Torsionspunkte $\neq \mathcal{O}$ können nicht in $E^{(1)}$ liegen, folglich landen sie bei Reduktionen in $\overline{E}^{(0)}$ oder auf dem singulären Punkt. In projektiver Darstellung ist also $P = (x : y : z)$ mit p-ganzen x, y, z und einer p-Einheit z (andernfalls wäre $p \mid z$ und die Reduktion gleich $(0 : 1 : 0)$, also $P \in E^{(1)}$). Die affinen Koordinaten von P sind daher $(x/z, y/z)$, und z ist eine p-Einheit. ☐

Korollar 3.8. *Sei $E : y^2 = x^3 + ax + b$ eine über $\mathbb{Q}$ definierte elliptische Kurve mit $a, b \in \mathbb{Z}$. Ist $(x, y) \neq \mathcal{O}$ ein Torsionspunkt in $E(\mathbb{Q})$, so gilt $x, y \in \mathbb{Z}$.*

Beweis. Wegen $\mathbb{Q} \subset \mathbb{Q}_p$ für jede Primzahl p ist nach dem vorhergehenden Korollar $x, y \in \mathbb{Z}_p$ für jedes p. Außerdem ist natürlich $x, y \in \mathbb{Q}$ wegen $(x, y) \in E(\mathbb{Q}) \setminus \{\mathcal{O}\}$. Die einzigen rationalen Zahlen, die p-ganz für jede Primzahl p sind, sind aber die ganzen Zahlen. ☐

Gehen wir nun an die Torsionsfreiheit von $E^{(1)}$. Dazu definieren wir eine Funktion $\ell : E^{(1)} \longrightarrow \mathbb{N}$ durch $\ell(x, y) = n$, wo $2n$ der genaue Exponent von p ist, der im Nenner von x aufgeht. Diese Funktion können wir auf $E^{(0)}$ fortsetzen, indem wir für $(x, y) \in E^{(0)}$ einfach $\ell(x, y) = 0$ setzen. Den Wert $\ell(x, y)$ nennen wir auch das Level von (x, y).

Um diese Definition zu rechtfertigen, schreiben wir $x = r/c$, $y = s/c$, und setzen $v_p(r) = \rho$, $v_p(s) = \sigma$, sowie $v_p(c) = \gamma$; hierbei ist $v_p(m)$ durch $p^{v_p(m)} \parallel m$ definiert. Da (x, y) im Kern der Reduktion liegt, muß c durch p teilbar sein, genauer muß $\gamma > \rho$ gelten. Aus der Gleichung $s^2c = r^3 + arc^2 + bc^3$ ersehen wir dann, daß $v_p(arc^2) \geq \rho + 2\gamma > v_p(r^3)$ und $v_p(bc^3) \geq 3\gamma > v_p(r^3)$ ist, d.h. r^3 ist der Summand auf der rechten Seite, der am wenigsten oft durch p teilbar ist. Also ist $2\sigma + \gamma = v_p(s^2c) = v_p(r^3) = 3\rho$, somit $3v_p(x) = 3(\rho - \gamma) = 2(\sigma - \gamma) = 2v_p(y)$. Insbesondere ist $v_p(x)$ gerade.

Für später halten wir fest:

$$\ell(x, y) = n, \; 2n = \gamma - \rho = -v_p(x), \; 3n = \gamma - \sigma = -v_p(y) \tag{3.1}$$

Inbesondere ist $n = \ell(x, y) = v(x/y)$.

Sei nun $E : Y^2Z = X^3 + aXZ^2 + bZ^3$ eine in homogener Form gegebene elliptische Kurve. Für eine natürliche Zahl $N \geq 1$ setzen wir $X_N = p^{2N}X$, $Y_N = p^{3N}Y$ und $Z_N = Z$; aus der Gleichung für E wird dann $E_N : Y_N^2 Z_N = X_N^3 + p^{4N}aX_N Z_N^2 + p^{6N}BZ_N^3$ (die dadurch induzierte Abbildung $E \longrightarrow E_N$ ist natürlich ein Gruppenisomorphismus $E(K) \longrightarrow E_N(K)$: die Abbildung ist ja eine lineare Abbildung der Ebene auf sich und bildet insbesondere Geraden auf Geraden ab). Die modulo p reduzierte Kurve ist $\overline{E}_N : Y_N^2 Z_N = X_N^3$; wir schreiben π_N für die Abildung $E(K) \longrightarrow \overline{E}_N(\mathbb{F}_p)$.

Was passiert mit einem affinen Punkt $P = (x, y) \in E(K)$ unter diesen Abbildungen? In projektiver Form ist $P = (r : s : c)$; sei $\ell(P) = n$, also $v_p(x) = -2n$ und $v_p(y) = -3n$. Dann geht P unter π_N auf die Reduktion von $Q = (p^{2N}r : p^{3N}s : c)$, also

- auf den singulären Punkt $(0, 0)$ von $E_N(\mathbb{F}_p)$, falls $\ell(x, y) < N$ ist;
- auf einen nichtsingulären endlichen Punkt von $E_N(\mathbb{F}_p)$, falls $\ell(x, y) = N$ ist;
- auf den unendlich fernen Punkt von $E_N(\mathbb{F}_p)$, falls $\ell(x, y) > N$ ist.

In der Tat: wegen (3.1) gilt im Falle $n \leq N$, daß $Q = (p^{2N-\gamma}r : p^{3N-\gamma}s : p^{-\gamma}c)$ lauter p-ganze Koordinaten hat, von denen die letzte eine p-adische Einheit ist; die Reduktion gibt also $\overline{\mathbb{Q}} = (0 : 0 : 1)$, falls $2(N - n) = 2N - \gamma + \rho = v_p(p^{2N-\gamma}r) > 0$ ist, und einen Punkt $\overline{\mathbb{Q}} = (\overline{r} : \overline{s} : 1)$ mit $\overline{r} \neq 0$, falls $N = n$ ist. Ist dagegen $n > N$, so enthält die mittlere Koordinate die größte Potenz von p, und die Reduktion gibt den unendlich fernen Punkt $(0 : 1 : 0) = \mathcal{O}$.

Betrachten wir nun π_0 genauer: die Abbildung $E \longrightarrow E_0$ ist die Identität, und $E_0 \longrightarrow \overline{E}_0$ die gewöhnliche Reduktion. Schränken wir π_0 auf $E^{(0)} = E_0^{(0)}$ ein, so erhalten wir einen Gruppenhomomorphismus $\pi_0 : E^{(0)} \longrightarrow \overline{E}_0^{\text{ns}}(\mathbb{F}_p)$, dessen Kern per definitionem $E^{(1)}$ ist, also die Menge aller Punkte vom Level ≥ 1. Nach dem, was wir eben gesehen haben, wird π_1 Punkte aus $E^{(1)}$ auf nichtsinguläre Punkte von $\overline{E}_0^{\text{ns}}(\mathbb{F}_p)$ abbilden, folglich ist die Einschränkung von π_1 auf $E^{(1)}$ ein Gruppenhomomorphismus, dessen Kern wir mit $E^{(2)}$ bezeichnen. Indem wir so fortfahren, erhalten wir Untergruppen

$$E^{(0)} \supset E^{(1)} \supset \ldots \supset E^{(N)} \supset \ldots$$

von $E(\mathbb{Q}_p)$ und dazugehörige Homomorphismen $\pi_N : E^{(N)} \longrightarrow \overline{E}_N^{\text{ns}}(\mathbb{F}_p)$, deren Kern $E^{(N+1)}$ genau aus den Punkten vom Level $\geq N + 2$ besteht.

Tatsächlich gilt viel mehr: der Homomorphismus $\pi_0 : E^{(0)} \longrightarrow \overline{E}_{\rm ns}(\mathbb{F}_p)$ ist surjektiv nach Proposition 3.4, und sein Kern besteht nach Definition aus $E^{(1)}$; nach dem Isomorphiesatz ist also $E^{(0)}/E^{(1)} \simeq \overline{E}_{\rm ns}(\mathbb{F}_p)$. Entsprechend ist $\pi_N : E^{(N)} \longrightarrow \overline{E}_N^{\rm ns}(\mathbb{F}_p)$ ein Epimorphismus mit Kern $E^{(N+1)}$; wegen $\overline{E}_N^{\rm ns}(\mathbb{F}_p) \simeq \mathbb{Z}/p\mathbb{Z}$ ist also $E^{(N)}/E^{(N+1)} \simeq \mathbb{Z}/p\mathbb{Z}$ für alle $N \ge 1$. Wir haben damit gezeigt:

Proposition 3.9. *Die eben definierten Teilmengen $E^{(N)}$ von $E(\mathbb{Q}_p)$ sind abelsche Gruppen; für $N \ge 1$ sind die Faktorgruppen $E^{(N)}/E^{(N+1)}$ zyklisch der Ordnung p; der Quotient $E^{(0)}/E^{(1)}$ ist isomorph zur Gruppe der nichtsingulären Punkte auf $\overline{E}$. Schließlich ist $\bigcap_N E^{(N)} = \{\mathcal{O}\}$.*

Daraus folgt sofort

Korollar 3.10. *Sei $(x,y) \in E(\mathbb{Q}_p)$ ein Punkt endlicher Ordnung n mit $(n,p) = 1$. Dann ist $x, y \in \mathbb{Z}_p$.*

Beweis. Wenn nicht, dann ist $\ell(x,y) \ge 1$. Wegen $\bigcap_N E^{(N)} = \{\mathcal{O}\}$ gibt es ein $N \in \mathbb{N}$ mit $(x,y) \in E^{(N)} \setminus E^{(N+1)}$. Die Abbildung $E \longrightarrow E^{(N)} \longrightarrow E^{(N)}/E^{(N+1)}$ bildet (x,y) dann auf ein Element der Ordnung > 1 (also der Ordnung p) in $E^{(N)}/E^{(N+1)}$; aber Gruppenhomomorphismen können ein Element mit Ordnung prim zu p nicht auf ein Element der Ordnung p abbilden: Widerspruch! $\square$

Der Rest dieses Abschnitts ist dem Problem gewidmet, wie man die Voraussetzung $(n,p) = 1$ umgehen kann. Dazu definieren wir eine Abbildung $u : E^{(1)} \longrightarrow \mathbb{Z}_p$ durch $u(x,y) = x/y$ (das geht: wegen $(x,y) \in E^{(1)}$ haben x und y beide eine p-Potenz im Nenner stehen, sind also sicher nicht p-ganz und insbesondere $\neq 0$) und $u(\mathcal{O}) = 0$. Wegen $|u(x,y)|_p = p^{-n}$ mit $n = \ell(x,y)$ ist $u(x,y)$ in der Tat p-adisch ganz.

Bemerkung. Was macht u mit einem Punkt $P = (x,y) \in E^{(1)}$? Übergang zur projektiven Schreibweise gibt $(r : s : c)$ mit $x = r/c$, $y = s/c$, und wenn man jetzt durch $s \neq 0$ (wegen $P \in E^{(1)}$) teilt, hat man $(x/y : 1 : 1/y)$. Mit anderen Worten: liegt P auf $E^{(1)}$, so entspricht ihm der Punkt $(x/y, 1/y)$ auf $z = x^3 + axz^2 + bz^3$, und u bildet P ab auf die x-Koordinate in der $x-z$-Ebene. Genau diese Kurve wird im Buch von Silverman-Tate [ST] zum Beweis des Satzes von Nagell-Lutz verwendet; im wesentlichen handelt es sich also um denselben Beweis wie in Cassels [Cas].

Wir haben also folgende Situation:

$$\begin{array}{ccccccccccc}
E(\mathbb{Q}_p) & \longleftarrow & E^{(0)} & \longleftarrow & E^{(1)} & \longleftarrow & E^{(2)} & \longleftarrow & E^{(3)} & \longleftarrow & \dots \\
 & & & & \big\downarrow{\scriptstyle u} & & \big\downarrow{\scriptstyle u} & & \big\downarrow{\scriptstyle u} & & \\
 & & & & p\mathbb{Z}_p & \longleftarrow & p^2\mathbb{Z}_p & \longleftarrow & p^3\mathbb{Z}_p & \longleftarrow & \dots
\end{array}$$

Die waagerechten Abbildungen sind hier Inklusionen, die vertikalen werden von u induziert. Nehmen wir nun einen Punkt $P \in E^{(n)} \setminus E^{(n+1)}$; für alle $s \in \mathbb{Z}$ mit $p \nmid s$ ist dann auch $sP \in E^{(n)} \setminus E^{(n+1)}$ (denn die Faktorgruppe $E^{(n)}/E^{(n+1)}$ hat Ordnung p), folglich hat auch sP Level n und es gilt $|u(sP)|_p = |u(P)|_p$. Für $s = p$ andererseits ist $pP \in E^{(n+1)}$, folglich gilt $\ell(pP) \geq n+1$ und damit $|u(pP)|_p \leq |p|_p |u(P)|_p$. Mit Induktion folgt damit $|u(sP)|_p \leq |s|_p |u(P)|_p$ für alle $s \in \mathbb{Z}$.

Wenn wir wüßten, daß hier das Gleichheitszeichen steht (mit anderen Worten: daß im obigen Beispiel $pP \in E^{(n+1)} \setminus E^{(n+2)}$ ist), so würde folgen, daß z.B. $E^{(n)}/E^{(n+2)} \simeq \mathbb{Z}/p^2\mathbb{Z}$ ist, und mit Induktion könnten wir schließen, daß sogar $E^{(n)}/E^{(n+m)} \simeq \mathbb{Z}/p^m\mathbb{Z}$ für alle $m, n \geq 1$ gilt. Insbesondere könnte dann $E^{(1)}$ keinen Punkt mit p-Potenzordnung enthalten, und wir könnten unseren Beweis der Torsionsfreiheit von $E^{(1)}$ beenden. Wir formulieren daher den

Hilfssatz 3.11. *Für alle $P \in E^{(1)}$ und all $s \in \mathbb{Z}$ gilt*

$$|u(sP)|_p = |s|_p |u(P)|_p.$$

Danach ist alles ganz einfach: Ist $P \in E^{(1)}$ ein Punkt endlicher Ordnung, also $sP = \mathcal{O}$ mit $s \in \mathbb{N}$, so folgt aus diesem Hilfssatz, daß $0 = |u(sP)|_p = |s|_p |u(P)|_p$ gilt; wegen $s \geq 1$ ist $|s|_p \neq 0$, folglich ist $P = \mathcal{O}$; mit anderen Worten: $E^{(1)}$ ist torsionsfrei. Damit ist der Beweis von Satz 3.6 beendet.

Dieser Hilfssatz wiederum wäre trivial, wenn u ein Gruppenhomomorphismus wäre: außerdem wäre, da u dann injektiv ist, $E^{(1)}$ einer Untergruppe von $p\mathbb{Z}_p$ isomorph; dann folgt aber die Torsionsfreiheit von $E^{(1)}$ aus derjenigen von $p\mathbb{Z}_p$. Leider ist u aber kein Gruppenhomomorphismus: die Differenz $u(P_1+P_2) - u(P_1) - u(P_2)$ ist nämlich i.a. nicht gleich 0; allerdings ist diese Differenz p-adisch sehr klein; wir werden nämlich zeigen, daß die Ungleichung

$$\begin{aligned} |u(P_1+P_2) - u(P_1) - u(P_2)|_p \leq \max\,\{|u(P_1)|_p^5, |u(P_2)|_p^5\} \\ \text{für } P_1, P_2 \in E^{(1)} \end{aligned} \tag{3.2}$$

gilt, aus der sich dann Hilfssatz 3.11 ergeben wird. Die Ungleichung (3.2) ist sicher richtig, wenn einer der Punkte gleich $\mathcal{O}$ ist (ist z.B. $P_1 + P_2 = \mathcal{O}$, also $P_2 = -P_1$, so folgt $u(P_1+P_2) - u(P_1) - u(P_2) = u(\mathcal{O}) - u(P_1) - u(-P_1) = 0$).

Seien also alle vorkommenden Punkte endlich; da diese nicht über dem singulären Punkt liegen, kann man die Gerade durch die P_j in der Form $Z_N = lX_N + mY_N$ schreiben, wo $|l|_p \le 1$ und $|m|_p \le 1$ ist. Sei weiter oBdA $|u(P_2)|_p \le |u(P_1)|_p = p^{-N}$; Schneiden mit E_N liefert dann

$$\begin{aligned}0 &= X_N^3 + p^{4N} aX_N(lX_N + mY_N)^2 + p^{6N} b(lX_N + mY_N)^3 - Y_N^2(lX_N + mY_N) \\ &= c_3X_N^3 + c_2X_N^2Y_N + c_1X_NY_N^2 + c_0Y_N^3\end{aligned}$$

mit $c_3 = 1 + p^{4N}al^2 + p^{6N}bl^3$ und $c_2 = 2p^{4N}alm + 3p^{6N}bl^2m$. Insbesondere ist $|c_3|_p = 1$ und $|c_2|_p \le p^{-4N}$. Die Wurzeln X_N/Y_N der Gleichung sind $-p^{-N}u(P_1+P_2)$, $p^{-N}u(P_1)$ und $p^{-N}u(P_2)$; da ihre Summe gleich $-c_2/c_3$ ist, folgt die Behauptung.

Beweis von Hilfssatz 3.11. Wir zeigen zuerst

$$|u(sP) - su(P)|_p \le |u(P)|_p^5. \tag{3.3}$$

Da dies unter der Transformation $s \longmapsto -s$ invariant und für $s = 0$ trivial ist, genügt es, die Behauptung für $s \in \mathbb{N}$ zu zeigen, und dazu machen wir Induktion. Mit $n = \ell(P)$ ist zu zeigen, daß $p^{5n} \mid (u(sP) - su(P))$ gilt. Für $s = 1$ ist dies sicher richtig; für den Induktionsschluß nehmen wir an, es sei $p^{5n} \mid (u(sP) - su(P))$. Dann betrachten wir, daß nach (3.2)

$$|u((s+1)P) - u(sP) - u(P)|_p \le \max\,\{|u(sP)|_p^5, |u(P)|_p^5\}$$

gilt. Wegen $|u(sP)|_p \le |u(P)|_p$ ist daher $u((s+1)P) - u(sP) - u(P)$ durch p^{5n} teilbar; nach Induktionsannahme bleibt dies richtig, wenn wir $u(sP)$ durch $su(P)$ ersetzen. Folglich ist $u((s+1)P) - (s+1)u(P)$ durch p^{5n} teilbar, und (3.3) ist bewiesen.

Nun zur eigentlichen Behauptung: sei $n = \ell(P)$ (wegen $P \in E^{(1)}$ ist $n \ge 1$); im Falle $p \nmid s$ ist dann $su(P)$ genau durch p^n teilbar, während $u(sP) - su(P)$ durch p^{5n} teilbar ist. Dies geht nur dann, wenn $p^n \parallel u(sP)$ gilt. Damit ist der Hilfssatz für alle s mit $p \nmid s$ bewiesen (sogar für alle s mit $p^{4n} \nmid s$).

Ist die Aussage aber für s richtig, dann auch für ps: aus der Ungleichung $|u(psP) - pu(sP)|_p \le |u(sP)|_p^5$ (setze $P = sP$ und $s = p$ in (3.3)) folgt ja, daß $u(psP) - pu(sP)$ durch $u(sP)^5$ teilbar ist, was nur geht, wenn $|u(psP)|_p = |pu(sP)|_p$ ist. Die Induktionsvoraussetzung gibt $|u(sP)|_p = |s|_p|u(P)|_p$, somit ist auch $|u(psP)|_p = |pu(sP)|_p = |ps|_p|u(P)|_p$. Die Behauptung folgt. □

Einige Bemerkungen

Die Gruppen $E^{(N)}$ wurden erstmals von Elisabeth Lutz [L1, L2] eingeführt und untersucht. Wir haben bereits bemerkt, daß $E^{(1)}$ zu einer Untergruppe $\neq 0$ von $p\mathbb{Z}_p$ (und damit zu $\mathbb{Z}_p$) isomorph wäre, wenn die Abbildung $u : E^{(1)} \longrightarrow p\mathbb{Z}_p$ ein Gruppenhomomorphismus wäre. Die Aussage $E^{(1)} \simeq \mathbb{Z}_p$ ist aber trotzdem richtig, und der Beweis benutzt nicht viel mehr als Hilfssatz 3.11.

Über die Gruppe $E^{(0)}$ wissen wir inzwischen ganz gut bescheid: es ist ja $E^{(0)}/E^{(1)} \simeq \overline{E}(\mathbb{F}_p)$ und $E^{(1)} \simeq \mathbb{Z}_p$. Um behaupten zu können, daß damit die Struktur von $E(\mathbb{Q}_p)$ im wesentlichen bekannt ist, müßten wir noch zeigen, daß die Faktorgruppe $E(\mathbb{Q}_p)/E^{(0)}$ endlich ist, und ggf. ihre Struktur bestimmen. Dies wurde von Kodaira und Néron gemacht, und der in Silverman [Sil] gegebene Beweis beginnt so:

> Die Endlichkeit von $E(\mathbb{Q}_p)/E^{(0)}$ folgt aus der Existenz des Néron-Modells; dieses ist ein Gruppenschema über Spec($\mathbb{Z}$), dessen generische Faser $E/\mathbb{Q}$ ist.

Ist man nur an der Endlichkeit interessiert, gibt es aber (für Kurven über $\mathbb{Q}_p$) einen Beweis, der mit ganz elementaren topologischen Mitteln auskommt: dazu faßt man $\mathbb{Q}_p$ als topologischen Raum auf (die Topologie wird von der Metrik $|\cdot|_p$ induziert; man beachte, daß die Mengen $\{x \in \mathbb{Z}_p : |x| < p^{-n}\}$ und $\{x \in \mathbb{Z}_p : |x| \leq p^{-n-1}\}$ identisch, also gleichzeitig offen und abgeschlossen sind) und zeigt, daß Addition, Multiplikation und Division bezüglich dieser Topologie stetig sind (man darf $\mathbb{Q}_p$ dann einen topologischen Körper nennen). Man rechnet nach, daß $(\mathbb{Q}_p, +)$ damit eine lokal-kompakte topologische Gruppe ist.

Jetzt überträgt man diese Topologie auf $\mathbb{P}^2(\mathbb{Q}_p)$, und zwar wie folgt: man versieht $\mathbb{Q}_p \times \mathbb{Q}_p \times \mathbb{Q}_p$ mit der Produkttopologie, $\mathbb{Q}_p \times \mathbb{Q}_p \times \mathbb{Q}_p \setminus \{(0,0,0)\}$ mit der Unterraumtopologie, und $\mathbb{P}^2(\mathbb{Q}_p)$ mit der Quotiententopologie bezüglich der Projektion $\mathbb{Q}_p \times \mathbb{Q}_p \times \mathbb{Q}_p \setminus \{(0,0,0)\} \longrightarrow \mathbb{P}^2(\mathbb{Q}_p)$. Dadurch wird $\mathbb{P}^2(\mathbb{Q}_p)$ zum kompakten topologischen Raum, in dem $E(\mathbb{Q}_p)$ als Nullstellenmenge eines Polynoms abgeschlossen ist. Weiter überzeugt man sich davon, daß $E^{(0)}$ offen in $E(\mathbb{Q}_p)$ ist, und damit hat man gewonnen: die Faktorgruppe $E(\mathbb{Q}_p)/E^{(0)}$ ist dann nämlich als Quotient einer kompakten Gruppe wieder kompakt, andererseits aber diskret, weil $E^{(0)}$ offen ist (nach bekannten elementaren Resultaten aus der Theorie topologischer Gruppen). Diskrete kompakte Mengen sind aber endlich.

Komplizierter dagegen ist der Beweis von

Satz 3.12. *Sei E eine über $\mathbb{Q}_p$ definierte elliptische Kurve mit Minimalmodell $y^2 = x^3 + ax + b$. Dann ist die Faktorgruppe $E(\mathbb{Q}_p)/E^{(0)}$ endlich, und zwar ist ihre Ordnung ≤ 4, wenn E additive Reduktion hat (d.h. wenn $\overline{E}(\mathbb{F}_p)$ eine Spitze besitzt), und zyklisch der Ordnung $v_p(\Delta)$ sonst.*

Prinzipiell kann man diesen Satz durch eine mühsame Fallunterscheidung per Hand beweisen; wir geben zwei einfache Spezialfälle, um die Idee klarzumachen.

1. Ist $v_p(a) \ge 1$ und $v_p(b) = 1$, so ist $E(\mathbb{Q}_p) = E^{(0)}$. Dies ist recht einfach: ist $(x, y) \in E(\mathbb{Q}_p)$ ein Punkt mit Reduktion $(0, 0)$, so muß $p \mid x$ und $p \mid y$ gelten. Aus $y^2 = x^3 + ax + b$ folgt dann $p^2 \mid b$: Widerspruch!

2. Ist $v_p(a) = 1$ und $v_p(b) \ge 2$, so ist $E(\mathbb{Q}_p)/E^{(0)} \simeq \mathbb{Z}/2\mathbb{Z}$. Dazu nehmen wir an, es seien $P_i = (x_i, y_i)$, $i = 1, 2$, Punkte mit Reduktion $(0, 0)$. Wir wollen zeigen, daß dann $(x_3, y_3) = P_1 + P_2$ in $E^{(0)}$ liegt. Dazu setzen wir $x_i = p\xi_i$, $y_i = p\eta_i$, $a = p\alpha$, $b = p^2\beta$ und finden $\eta_i^2 = p\xi_i^3 + \xi_i\alpha + \beta$. Differenzbildung liefert $\eta_2^2 - \eta_1^2 = (\xi_2 - \xi_1)\alpha$. Gilt $\eta_1 + \eta_2 = 0$, so folgt wegen $p \nmid \alpha$, daß $\xi_1 = \xi_2$ und folglich $P_1 = -P_2$ ist; insbesondere ist $P_1 + P_2 = \mathcal{O} \in E^{(0)}$.

Ist dagegen $\eta_2 - \eta_1 = 0$, so folgt wie oben $P_1 = P_2$. Ist $y_1 = 0$, so hat P_1 Ordnung 2, folglich ist dann ebenfalls $P_1 + P_2 = \mathcal{O} \in E^{(0)}$; andernfalls ist $x_3 = -2x_1 + m^2$ mit $m = (3x_1^2 + a)/2y_1$. Da p im Zähler genau einmal und im Nenner mindestens einmal aufgeht, ist entweder m p-ganz und nicht durch p teilbar, also die Reduktion von x_3 ungleich 0, oder m hat ein p im Nenner, und P_3 hat Level ≥ 1, liegt also in $E^{(1)} \subset E^{(0)}$.

Sei also $\eta_2^2 \ne \eta_1^2$. Dann folgt $m = (\eta_2 - \eta_1)/(\xi_2 - \xi_1) = \alpha/(\eta_2 + \eta_1)$, und wie eben ist $v_p(m) \le 0$, insbesondere die Reduktion von P_3 nicht gleich $(0, 0)$.

3. Ist $v_p(a) \ge 2$ und $v_p(b) = 2$, so ist $E(\mathbb{Q}_p)/E^{(0)} \simeq \mathbb{Z}/3\mathbb{Z}$. Das wollen wir nicht mehr vormachen – man kann sich vorstellen, daß die Rechnungen recht umständlich werden.

Der (nach dem Satz von Kodaira-Néron endliche) Index $c_p = (E(\mathbb{Q}_p) : E^{(0)})$ heißt p-te *Tamagawa-Zahl* von E. Wegen $c_p = 1$ für alle $p \nmid \Delta$ ist $\prod_p c_p$ endlich; dieser Faktor spielt eine große Rolle für die Feinstruktur der L-Reihe der elliptischen Kurve, namentlich bei der Formulierung der Vermutung von Birch und Swinnerton-Dyer.

3.4 Anwendungen und der Satz von Mazur

Wir ziehen jetzt aus den Resultaten des letzten Abschnitts noch einige Folgerungen.

Satz 3.13. *Sei $E : y^2 = x^3 + ax + b$ eine über $\mathbb{Q}$ definierte elliptische Kurve und p eine ungerade Primzahl mit $p \nmid (4a^3 + 27b^2)$. Dann ist $E(\mathbb{Q})_{\text{tors}}$ einer Untergruppe von $\overline{E}(\mathbb{F}_p)$ isomorph.*

Beweis. Wegen $p \nmid 2(4a^3 + 27b^2)$ ist $\overline{E}(\mathbb{F}_p)$ eine elliptische Kurve (d.h. sie ist nicht singulär); daher ist $E(\mathbb{Q}_p) = E^{(0)}$. Weiter ist $E(\mathbb{Q})_{\text{tors}} \subseteq E(\mathbb{Q}_p)_{\text{tors}}$ wegen $\mathbb{Q} \subset \mathbb{Q}_p$, und Komposition der Abbildungen $E(\mathbb{Q})_{\text{tors}} \longrightarrow E(\mathbb{Q}_p)_{\text{tors}} \longrightarrow E(\mathbb{Q}_p) \longrightarrow E(\mathbb{Q}_p)/E^{(1)}$ ist ein Gruppenhomomorphismus, dessen Kern aus allen $P \in E(\mathbb{Q})_{\text{tors}}$ besteht, die Elemente von $E^{(1)}$ sind. Da $E^{(1)}$ torsionsfrei ist, ist dieser Kern trivial, und wir haben eine Injektion $E(\mathbb{Q})_{\text{tors}} \longrightarrow E^{(0)}/E^{(1)} \simeq \overline{E}(\mathbb{F}_p)$. Das war zu beweisen. □

Der folgende Satz läßt sich nun ohne große Mühe beweisen:

Satz 3.14. *Sei E die elliptische Kurve $y^2 = x^3 + ax$ mit einer ganzen Zahl a, die durch keine vierte Potenz $\neq 1$ teilbar ist. Dann gilt*

$$E(\mathbb{Q})_{\text{tors}} \simeq \begin{cases} \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}, & \textit{falls } -a \textit{ ein Quadrat ist;} \\ \mathbb{Z}/4\mathbb{Z}, & \textit{falls } a = 4; \\ \mathbb{Z}/2\mathbb{Z} & \textit{sonst.} \end{cases}$$

Beweis. Der Fall $E(\mathbb{Q})_{\text{tors}} = \{\mathcal{O}\}$ tritt deswegen nicht auf, weil $P = (0,0)$ ein Punkt der Ordnung 2 auf E ist.

Der wesentliche Punkt im Beweis dieses Satzes ist der Nachweis, daß $E(\mathbb{Q})_{\text{tors}}$ maximal vier Punkte enthält. Dies macht man mit folgendem Standardtrick: nach Dirichlet gibt es unendlich viele Primzahlen $p \equiv 3 \bmod 8$. Insbesondere gibt es ein solches p mit $p \nmid 2(4a^3 + 27b^2) = 8a^3$. Damit ist $\#E(\mathbb{Q})_{\text{tors}} \mid \#\overline{E}(\mathbb{F}_p)$; wegen $p \equiv 3 \bmod 4$ ist aber $\#\overline{E}(\mathbb{F}_p) = p{+}1 \equiv 4 \bmod 8$, und wir sehen, daß $\#E(\mathbb{Q})_{\text{tors}}$ jedenfalls nicht durch 8 teilbar ist.

Ebenso zeigen wir, daß es keine Primzahl $q \geq 3$ mit $q \mid \#E(\mathbb{Q})_{\text{tors}}$ gibt. Dazu wählen wir ein primes p mit $p \equiv 1 \bmod q$ und $p \equiv 3 \bmod 4$ prim (nach dem chinesischen Restsatz und Dirichlet). Wie eben ist $\#E(\mathbb{Q})_{\text{tors}} \mid \#\overline{E}(\mathbb{F}_p) = p{+}1$, aber wegen $p{+}1 \equiv 2 \bmod q$ kann q kein Teiler von $\#E(\mathbb{Q})_{\text{tors}}$ sein.

Also ist $\#E(\mathbb{Q})_{\text{tors}} \mid 4$, und es gibt folgende Möglichkeiten:

1. Es gibt drei $\mathbb{Q}$-rationale Punkte der Ordnung 2: dies ist genau dann der Fall, wenn $x^3 + ax = x(x^2 + a)$ drei rationale Wurzeln besitzt, d.h. genau dann, wenn $-a$ ein Quadrat ist;

2. es gibt genau einen $\mathbb{Q}$-rationalen Punkte der Ordnung 2: dann ist nach oben entweder $\#E(\mathbb{Q})_{\text{tors}} = 2$, oder aber der Punkt $(0,0)$ der Ordnung 2 ist

das Doppelte eines anderen Punktes, also $(0,0) = 2(x,y)$ mit gewissen $x,y \in \mathbb{Z}$. Der Satz von Nagell-Lutz besagt zwar, daß dann $y^2 \mid 4a^3$ gilt, aber das hilft uns nicht weiter. Daher schauen wir uns an, was die Verdoppelungsformel hergibt: der Nenner der x-Koordinate von $2(x,y)$ ist $x^4-2ax^2+a^2 = (x^2-a)^2$; dies wird genau dann gleich 0, wenn $x^2 = a$ ist, d.h. wenn a ein Quadrat ist. Da a durch keine vierte Potenz teilbar ist, muß x quadratfrei sein. Aus $y^2 = x(x^2+a) = 2x^3$ folgt dann, daß $x \mid 2$ sein muß, und dies führt auf $x = 2$, $a = 4$, und $y = \pm 4$.

Damit ist alles gezeigt. □

Ganz entsprechend zeigt man

Satz 3.15. *Sei E die elliptische Kurve $y^2 = x^3 + b$ mit einer ganzen Zahl b, die durch keine sechste Potenz $\neq 1$ teilbar ist. Dann gilt*

$$E(\mathbb{Q})_{\text{tors}} \simeq \begin{cases} \mathbb{Z}/6\mathbb{Z}, & \textit{falls } b = 1; \\ \mathbb{Z}/3\mathbb{Z}, & \textit{falls } b = -432 \textit{ oder } 1 \neq b \textit{ Quadrat ist;} \\ \mathbb{Z}/2\mathbb{Z} & \textit{falls } 1 \neq b \textit{ eine dritte Potenz ist;} \\ 0 & \textit{sonst.} \end{cases}$$

Beweis. Übungsaufgabe. □

Dazu braucht man die Anzahl der Punkte von $y^2 = x^3 + b$ über gewissen endlichen Körpern:

Proposition 3.16. *Ist $E : y^2 = x^3 + b$ eine elliptische Kurve mit $b \in \mathbb{Z}$ und $p \equiv 2 \bmod 3$ eine ungerade Primzahl, so gilt $\#E(\mathbb{F}_p) = p + 1$.*

Beweis. Man betrachte den durch $\pi(x) = x^3$ definierten Gruppenhomomorphismus $\pi : \mathbb{F}_p^\times \longrightarrow \mathbb{F}_p^\times$. Dessen Kern besteht aus allen $x \in \mathbb{F}_p^\times$ mit $x^3 = 1$. Gäbe es ein Element x der genauen Ordnung 3, so würde 3 die Gruppenordnung $p - 1$ teilen, was wegen $p \equiv 2 \bmod 3$ nicht der Fall ist. Also ist π injektiv, und damit, da beide Seiten dieselbe Kardinalität haben, automatisch surjektiv.

Schreiben wir nun die Kurvengleichung in der Form $x^3 = y^2 - b$, so sehen wir, daß es zu jedem $y \in \mathbb{F}_p$ genau ein $x \in \mathbb{F}_p$ gibt, sodaß (x,y) auf der Kurve liegt. Zusammen mit dem unendlich fernen Punkt hat die Kurve also genau $p + 1$ Punkte. □

Daß eine über $\mathbb{Q}$ definierte elliptische Kurve nicht beliebig große Torsionsgruppen haben kann, ist zwar nicht offensichtlich, liegt aber wegen Satz

3.13 und des Satzes von Nagell-Lutz doch nahe. Ist beispielsweise E eine elliptische Kurve, deren Diskriminante nicht durch 5 teilbar ist, so wissen wir nach Satz 3.13, daß $E(\mathbb{Q})_{\rm tors}$ einer Untergruppe von $\overline{E}(\mathbb{F}_5)$ ismorph ist; diese hat maximal $10 = \lfloor p+1+2\sqrt{5}\rfloor$ Elemente, folglich ist $\#E(\mathbb{Q})_{\rm tors} \le 10$ für solche elliptischen Kurven.

Bereits im letzten Jahrhundert hat man elliptische Kurven konstruiert, deren Torsionsgruppe eine der folgenden ist:

$$E_{\rm tors} \simeq \begin{cases} \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2m\mathbb{Z} & \text{für } 1 \le m \le 4; \\ \mathbb{Z}/(2m-1)\mathbb{Z} & \text{für } 1 \le m \le 5; \\ \mathbb{Z}/2m\mathbb{Z} & \text{für } 1 \le m \le 6. \end{cases} \tag{3.4}$$

Insbesondere hat Beppo Levi (besser bekannt aus der Theorie des Lebesgue-Integrals) 1906 zeigen können, daß es zu jeder dieser Fälle für jeweils unendlich viele nicht-isomorphe elliptische Kurven auftritt. Darüberhinaus konnte er beweisen, daß die Gruppen $\mathbb{Z}/m\mathbb{Z}$ für $m = 14, 16, 20$, sowie die Gruppen $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2m\mathbb{Z}$ für $m = 5, 6$ nicht als Torsionsgruppen vorkommen können. Auf dem Internationalen Mathematikerkongress 1908 in Rom hat Beppo Levi dann die Vermutung ausgesprochen, daß daß die Liste in (3.4) (möglicherweise mit der Ausnahme $\mathbb{Z}/24\mathbb{Z}$, die er nicht erwähnt) komplett ist. Vierzig Jahre nach Beppo Levi äußert Mordell dieselbe Vermutung, und schließlich ist sie noch einmal 20 Jahre später von Ogg als "folklore conjecture" bezeichnet (und prompt nach ihm benannt) worden.

Nachdem man über Jahrzehnte hinweg weitere Möglichkeiten ausschließen konnte, gelang Barry Mazur 1976 der Beweis der Vermutung von Beppo-Levi, Mordell und Ogg:

Satz 3.17. *Sei E eine über $\mathbb{Q}$ definierte elliptische Kurve. Dann ist E_{tors} zu einer der* 15 *in* (3.4) *angegebenen Gruppen isomorph.*

Dieser Satz sieht zwar unscheinbar aus, ist aber extrem tief. Die berühmte "Beschränktheitsvermutung", wonach es zu jedem gegebenen Zahlkörper K eine Schranke $c(K)$ gibt, sodaß die Kardinalität der Torsionsgruppe jeder über K definierten elliptischen Kurve durch $c(K)$ beschränkt ist (für $K = \mathbb{Q}$ wäre $c(\mathbb{Q}) = 16$ bestmöglich), ist erst vor kurzem bewiesen worden.

Knapp gibt in seinem Buch [Kna] Beispiele für jede über $\mathbb{Q}$ vorkommende Torsionsgruppe (vgl. die Tabelle 1).

Schließlich erwähnen wir noch, daß Doud [D] einen Algorithmus implementiert hat, der für große $D = 4a^3 + 27b^2$ schneller läuft als der auf Nagell-Lutz basierende. (Man beachte, daß Nagell Lutz die Faktorisierung von D verwendet!) Die Idee ist einfach und steht schon bei Mahler [1]: man hat die

E		$E(\mathbb{Q})_{\rm tors}$	Δ
y^2	$= x^3 + 2$	0	$-2^6 3^3$
y^2	$= x^3 + x$	$\mathbb{Z}/2\mathbb{Z}$	-2^6
y^2	$= x^3 + 4$	$\mathbb{Z}/3\mathbb{Z}$	$-2^8 3^3$
y^2	$= x^3 + 4x$	$\mathbb{Z}/4\mathbb{Z}$	-2^{12}
$y^2 + y$	$= x^3 - x^2$	$\mathbb{Z}/5\mathbb{Z}$	-11
y^2	$= x^3 + 1$	$\mathbb{Z}/6\mathbb{Z}$	$-2^4 3^3$
$y^2 - xy + 2y$	$= x^3 + 2x^2$	$\mathbb{Z}/7\mathbb{Z}$	$-2^7 13$
$y^2 + 7xy - 6y$	$= x^3 - 6x^2$	$\mathbb{Z}/8\mathbb{Z}$	$2^8 3^4 17$
$y^2 + 3xy + 6y$	$= x^3 + 6x^2$	$\mathbb{Z}/9\mathbb{Z}$	$-2^9 3^5$
$y^2 - 7xy - 36y$	$= x^3 - 18x^2$	$\mathbb{Z}/10\mathbb{Z}$	$-2^5 3^{10} 11^2$
$y^2 + 43xy - 210y$	$= x^3 - 210x^2$	$\mathbb{Z}/12\mathbb{Z}$	$2^{12} 3^6 5^3 7^4 13$
y^2	$= x^3 - x$	$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$	2^6
y^2	$= x^3 + 5x^2 + 4x$	$\mathbb{Z}/4\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$	$2^8 3^2$
$y^2 + 5xy - 6y$	$= x^3 - 3x^2$	$\mathbb{Z}/6\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$	$2^2 3^6 5^2$
y^2	$= x^3 + 337x^2 + 20736x$	$\mathbb{Z}/8\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$	$2^{20} 3^8 5^4 7^2$

elliptische Parametrisierung von $E : y^2 = 4x^3 - g_2x - g_3$ durch die Weierstraßsche $\wp$-Funktion, also einen Grupenisomorphismus $\mathbb{C}/\Lambda \longrightarrow E(\mathbb{C})$: $z + \Lambda \longmapsto (\wp(z), \wp'(z))$. Nun kann man die Torsionspunkte $\mathbb{C}/\Lambda\,[m]$ sofort hinschreiben: ist $\Lambda = \omega_1\mathbb{Z} \oplus \omega_2\mathbb{Z}$, so sind die m-Torsionspunkte gegeben durch $\{\frac{1}{m}(a_1\omega_1 + a_2\omega_2) : a_j \in \mathbb{Z}, 0 \le a_1, a_2 < m\}$. Da rationale Torsionspunkte nach dem Satz von Mazur höchstens Ordnung 16 haben, berechnet man für alle möglichen m, ob $\wp(\frac{1}{m}\omega_1)$, $\wp(\frac{1}{m}\omega_2)$, $\wp(\frac{1}{m}(\omega_1 + \omega_2))$ ungefähr ganzzahlige Koordinaten hat; wenn nicht, gibt es keine rationalen Punkte der Ordnung m, wenn ja, testet man, ob der gefundene Punkt P wirklich ein Torsionspunkt ist, indem man mP ausrechnet.

Kapitel 4

Rang: Satz von Mordell-Weil

4.1 2-Isogenien

"Isogenien" stecken, wie wir sehen werden, hinter hinreichend vielen Beweisen in der Theorie elliptischer Kurven, um ihre genauere Untersuchung zu rechtfertigen. Isogenien sind "rationale" Abbildungen zwischen elliptischen Kurven, welche das Gruppengesetz respektieren. Man kann zeigen, daß es zu jeder Isogenie $\alpha : E \longrightarrow E'$ eine dazu duale Isogenie $\widehat{\alpha} : E' \longrightarrow E$ gibt, sodaß $\alpha' \circ \alpha = m$ die Multiplikation mit einer ganzen Zahl $m \in \mathbb{Z}$ ist; dieses $m \neq 0$ heißt der Grad der Isogenie α. Zu einer gegebenen elliptischen Kurve $E/\mathbb{Q}$ und einem $0 \neq m \in \mathbb{Z}$ gibt es nicht immer eine elliptische Kurve $E'/\mathbb{Q}$ und eine rationale Isogenie $\alpha : E \longrightarrow E'$ vom Grad m; wenn aber E einen Torsionspunkt der Ordnung m enthält, dann gibt es ein solches α. Insbesondere existieren 2-Isogenien immer dann, wenn einer der drei Punkte in $E[2]$ rational ist.

Es wird sich auszahlen, einige Aussagen statt über $\mathbb{Q}$ allgemeiner über einem beliebigen Körper K der Charakteristik $\neq 2, 3$ zu beweisen. Sei also E eine über K definierte elliptische Kurve mit einem K-rationalen Punkt der Ordnung 2. Diesen können wir in den Punkt $T = (0, 0)$ verschieben, und dann wird E von der Gleichung $E : y^2 = x(x^2 + ax + b)$ beschrieben. Wegen $\Delta(E) = 16b^2(a^2 - 4b)$ ist E genau dann nicht singulär, wenn $b \neq 0$ und $a^2 - 4b \neq 0$ gilt; das wollen wir ab jetzt voraussetzen.

Die Addition des Torsionspunktes $(0,0)$ induziert eine Abbildung $E \longrightarrow E$; mit $(x_1,y_1) = (x,y) + (0,0)$ findet man $x_1 = b/x$ und $y_1 = -by/x^2$, falls nicht gerade $x = 0$ und damit $(x,y) = (0,0)$ ist. Damit hat man alle Punkte auf $E(\overline{K})$ in Paare $\{(x,y),(x_1,y_1)\}$ eingeteilt. Jedem solchen Paar kann man die Invarianten $\lambda = x + x_1$, bzw. $\mu = y + y_1$ zuordnen. Eine kleine Rechnung zeigt

$$\begin{aligned}\mu^2 &= (y - y(1 - b/x^2))^2 = y^2(x - b/x)^2/x^2 \\ &= (x + a + b/x)[(x + b/x)^2 - 4b] = (\lambda + a)(\lambda^2 - 4b)\end{aligned}$$

Wir haben gesehen: ist $(x_1,y_1) = (x,y) + (0,0)$, so liegt der Punkt $(x + x_1, y + y_1)$ auf der Kurve $E' : v^2 = (u + a)(u^2 - 4b)$. Wir können E' sogar dieselbe Form wie E geben, wenn wir u durch $u + a$ ersetzen; dann wird nämlich $\overline{E} : \overline{y}^2 = \overline{x}(\overline{x}^2 + \overline{a}\,\overline{x} + \overline{b})$. Dies ist eine elliptische Kurve, da $\overline{\Delta} = 16\overline{b}^2(\overline{a}^2 - 4\overline{b}) = 16b^2(a^2 - 4b)^2 \neq 0$ ist. Die Abbildung $E \longrightarrow \overline{E}$ ist dann gegeben durch $\overline{x} = x + x_1 + a = y^2/x^2$ und $\overline{y} = y + y_1 = y(x^2 - b)/x^2$ für alle $(x,y) \neq \mathcal{O}, T$. Nun behaupten wir

Proposition 4.1. *Sei $E : y^2 = x(x^2 + ax + b)$ eine elliptische Kurve über einem Körper der Charakteristik $\neq 2, 3$. Dann ist die durch*

$$\alpha(P) = \begin{cases} \mathcal{O}, & \text{falls } P = \mathcal{O} \text{ oder } P = T, \\ (y^2/x^2, y(x^2 - b)/x^2), & \text{falls } P = (x,y) \neq \mathcal{O}, T. \end{cases} \tag{4.1}$$

definierte Abbildung ein Gruppenhomomorphismus $E(K) \longrightarrow \overline{E}(K)$, wo $\overline{E} : \overline{y}^2 = \overline{x}(\overline{x}^2 + \overline{a}\,\overline{x} + \overline{b})$ eine zu E isogene elliptische Kurve mit $\overline{a} = -2a$ und $\overline{b} = a^2 - 4b$ ist.

Beweis. Zu zeigen ist lediglich die Behauptung, daß α ein Gruppenhomomorphismus ist. Dazu machen wir die übliche Fallunterscheidungen.

a) Die Aussage ist offensichtlich richtig, wenn einer oder mehrere der Punkte P_1, P_2 oder $P_1 + P_2$ gleich dem neutralen Element $\mathcal{O}$ sind.

b) Sei $P_2 = T$; der Fall $P_1 = T$ ist wegen $P_1 + P_2 = \mathcal{O}$ unter a) abgehandelt. Sei also $P_1 = P = (x,y)$ mit $x \neq 0$. Dann ist $P + T = (b/x, -by/x^2)$, somit

$$\begin{aligned}\alpha(P+T) &= \Big(\frac{x_{P+T}^2}{y_{P+T}^2}, \frac{Y_{P+T}(x_{P+T}^2 - b)}{x_{P+T}^2}\Big) = \Big(\frac{b^2y^2/x^4}{b^2/x^2}, \frac{-by(b^2/x^2 - 1)}{b^2}\Big) \\ &= \Big(\frac{y^2}{x^2}, \frac{y(x^2 - b)}{x^2}\Big) = \alpha(P) = \alpha(P) + \alpha(T).\end{aligned}$$

c) P_1 und P_2 sind 2-Teilungspunkte $\neq T$. Wir dürfen annehmen, daß $P_1, P_2 \neq T$ sind; der Fall $P_1 + P_2 = T$ läßt sich wegen $\alpha(-P) = -\alpha(P)$ auf b) zurückführen, d.h. wir dürfen auch $P_1 + P_2 \neq T$ annehmen. Dann muß aber $P_1 = P_2$ und somit $P_1 + P_2 = \mathcal{O}$ sein, und dieser Fall ist in a) abgehandelt worden.

d) Wegen $\alpha(-P) = -\alpha(P)$ genügt es im allgemeinen Fall zu zeigen, daß $\alpha(P_1) + \alpha(P_2) + \alpha(P_3) = \mathcal{O}$ ist, falls $P_1 + P_2 + P_3 = \mathcal{O}$ ist. Die letzte Bedingung ist äquivalent dazu, daß die P_i kollinear sind (und auf E liegen). Sei $y = mx + c$ die entsprechende Geradengleichung (die Gerade kann nicht parallel zur y-Achse sein, da sonst einer der P_i ein 2-Teilungspunkt wäre). Im dem Falle, daß alle P_i paarweise verschieden sind, genügt es zu zeigen, daß $\alpha(P_i)$ für $i = 1, 2, 3$ auf der Geraden $y = \overline{m}x + \overline{c}$ liegt für $\overline{m} = \frac{1}{c}(mc - b)$ und $\overline{c} = \frac{1}{c}(c^2 - amc + bm^2)$.

Schreiben wir nun $\alpha(x_i, y_i) = (\overline{x}_i, \overline{y}_i)$; dann wird

$$\begin{aligned}
\overline{m}\,\overline{x}_1 + \overline{c} &= \frac{mc-b}{c}\Big(\frac{y_1}{x_1}\Big)^2 + \frac{c^2 - amc + bm^2}{c} \\
&= \frac{(mc-b)y_1^2 + (c^2 - amc + bm^2)x_1^2}{cx_1^2} \\
&= \frac{mc(y_1^2 - ax_1^2) - b(y_1 - mx_1)(y_1 + mx_1) + c^2x_1^2}{cx_1^2};
\end{aligned}$$

Jetzt verwenden wir $y_1^2 - ax_1^2 = x_1^3 + bx_1$, sowie $y_1 - mx_1 = c$, und erhalten

$$\begin{aligned}
\overline{m}\,\overline{x}_1 + \overline{c} &= \frac{m(x_1^3 + bx_1) - b(y_1 + mx_1) + cx_1^2}{x_1^2} \\
&= \frac{x_1^2(mx_1 + c) - by_1}{x_1^2} = \frac{(x_1^2 - b)y_1}{x_1^2} = \overline{y}_1.
\end{aligned}$$

Die Rechnungen für P_2 und P_3 gehen analog.

e) Der Fall $P_1 = P_2$ ist eine Übungsaufgabe. □

Hinter der Konstruktion der 2-Isogenie steckt etwas mehr als wir oben angedeutet haben. Dazu gehen wir von der gegebenen elliptischen Kurve $E : y^2 = x(x^2 + ax + b)$ über zum rationalen Funktionenkörper $F = K(x)$ und dessen quadratischer Erweiterung $E = K(x, y)$ mit $y = \sqrt{x(x^2 + ax + b)}$. Der Körper E besteht aus Elementen der Form $f(x) + g(x)y$ und deren Quotienten. Die durch $x \longmapsto x_1$, $y \longmapsto y_1$ definierte Abbildung α ist dann ein Endomorphismus von E und induziert, wie man leicht sehen kann, einen K-Automorphismus der Ordnung 2 auf E: daß $\alpha \circ \alpha$ die Identität ist, ist klar; um zu beweisen, daß α ein Ringhomomorphismus ist, muß man im wesentlichen zeigen, daß $\alpha(y^2) = \alpha(y)^2$ ist. Dies folgt aber so:

$$\begin{aligned} \alpha(y^2) = \alpha(x(x^2+ax+b)) &= \frac{b}{x}\Big(\frac{b^2}{x^2} + \frac{ab}{x} + \frac{b}{x}\Big) \\ &= \frac{b^2}{x^4}(x^3+ax^2+bx) = \alpha(y)^2. \end{aligned}$$

Nach der Galoistheorie besitzt die von diesem Automorphismus der Ordnung 2 erzeugte Untergruppe von $\mathrm{Gal}\,(E/K)$ einen Fixkörper L mit $(E : L) = 2$, und offensichtlich ist $K(\lambda,\mu) \subseteq L$, wo wie oben $\lambda = x + x_1 + a$ und $\mu = y + y_1$ gesetzt sind. Um hier Gleichheit zu zeigen, genügt (wegen Galoistheorie) der Nachweis, daß $(E : K) = 2$ ist. Das wiederum können wir zeigen, indem wir x und y durch λ und μ ausdrücken. Zuerst ist einmal $\lambda^{1/2} = y/x$, weiter $\lambda^{-1/2}\mu + \lambda = x - b/x + x + b/x + a = 2x + a$, somit $x = \frac{1}{2}(\lambda + \lambda^{-1/2}\mu - a)$. Also ist $E = K(\lambda^{1/2})$, insbesondere $(E : L) \le 2$. Damit gilt $(E : L) = 2$ wegen $E \ne L$.

Wir haben jetzt folgendes erreicht: zu jeder elliptischen Kurve $E : y^2 = x(x^2 + ax + b)$ haben wir eine "isogene" elliptische Kurve $\overline{E} : \overline{y}^2 = \overline{x}\,(\overline{x}^2 + \overline{a}\,\overline{x} + \overline{b})$ gefunden, wo $\overline{a} = -2a$ und $\overline{b} = a^2 - 4b$ ist. Dabei ist die 2-Isogenie $\alpha : E \longrightarrow \overline{E}$ gegeben durch (4.1).

Da auch $\overline{E}$ dieselbe Form wie E hat, können wir unsere bisherigen Ergebnisse auf $\overline{E}$ anwenden und finden eine 2-Isogenie $\overline{\alpha} : \overline{E} \longrightarrow \overline{\overline{E}}$ von $\overline{E}$ auf $\overline{\overline{E}} : \overline{\overline{y}}^2 = \overline{\overline{x}}(\overline{\overline{x}}^2 + \overline{\overline{a}}\,\overline{\overline{x}} + \overline{\overline{b}})$ mit $\overline{\overline{a}} = -2\overline{a} = 4a$ und $\overline{\overline{b}} = \overline{a}^2 - 4\overline{b} = 4a^2 - 4(a^2 - 4b) = 16b$. Jetzt stellt man aber fest, daß die beiden elliptischen Kurven E und $\overline{\overline{E}}$ isomorph sind via $\overline{\overline{x}} = 4x$ und $\overline{\overline{y}} = 8y$. Indem wir also unsere 2-Isogenie $\overline{\alpha}$ ersetzen durch

$$\beta(P) = \begin{cases} \mathcal{O}, & \text{falls } P = \mathcal{O} \text{ oder } P = \overline{T}, \\ (\overline{y}^2/4\overline{x}^2, \overline{y}(\overline{x}^2 - \overline{b})/8\overline{x}^2), & \text{falls } P = (\overline{x},\overline{y}) \ne \mathcal{O}, \overline{T}, \end{cases} \tag{4.2}$$

erhalten wir eine 2-Isogenie $\beta : \overline{E} \longrightarrow E$. Die Hintereinanderausführung von α und β liefert damit einen Gruppenendomorphismus $\beta \circ \alpha : E \longrightarrow E$; es stellt sich heraus, daß dies genau die Multiplikation mit 2 auf E ist: das erklärt auch den Namen 2-Isogenie.

Satz 4.2. *In obiger Notation ist $\beta \circ \alpha = [2]_E$ und $\alpha \circ \beta = [2]_{\overline{E}}$.*

Beweis. Auf $E : y^2 = x(x^2+ax+b)$ gilt, wie man leicht nachrechnet, folgende Verdoppelungsformel:

$$2(x,y) = \left(\left(\frac{x^2-b}{2y}\right)^2, \frac{(x^2-b)(x^4+2ax^3+6bx^2+2abx+b^2)}{8y^3}\right). \tag{4.3}$$

Der Rest folgt leicht:

$$\frac{\overline{\overline{x}}}{4} = \frac{\overline{y}^2}{4\overline{x}^2} = \frac{y^2(x^2-b)^2x^4}{4y^4x^4} = \frac{(x^2-b)^2}{4y^2},$$

falls $xy \neq 0$. Ebenso erhält man

$$\begin{aligned}\frac{\overline{\overline{y}}}{8} &= \frac{1}{8}\left(y\frac{x^2-b}{x^2}\left(1-(a^2-4b)\frac{x^4}{y^4}\right)\right) \\ &= \frac{(x^2-b)(y^4-(a^2-4b)x^4)}{8x^2y^3} = \frac{(x^2-b)((x^2+ax+b)^2-(a^2-4b)x^4)}{8y^3} \\ &= \frac{(x^2-b)(x^4+2ax^3+6bx^2+2abx+b^2)}{8y^3}.\end{aligned}$$

Die Spezialfälle $xy = 0$ sind dabei getrennt nachzuprüfen. □

Beim Fermatschen Beweis der Unlösbarkeit von $C_1 : z^2 = x^4+y^4$ in ganzen Zahlen sind wir so vorgegangen: wir haben die Existenz eines Punktes P auf C_1 angenommen, haben dann einen Punkt Q auf der Kurve $C_2 : y^2 = u^4-4v^4$ konstruiert und schließlich einen Punkt R auf C_1 gefunden, der "kleiner" war als P. Bringt man die Kurven C_j auf Weierstraßform, so findet man $C_1 : y^2 = x^3-4x$ und $C_2 : y^2 = x^3+x$. Insbesondere ist C_2 eine zu C_1 2-isogene Kurve; bezeichnet man die 2-Isogenie $C_1 \longrightarrow C_2$ mit α, die dazu duale mit β, so kann man nachrechnen, daß $Q = \alpha(R)$, $P = \beta(Q)$ und damit $P = 2R$ ist. Mit anderen Worten: Fermats Beweis arbeitet mit einem 2-Abstieg durch 2-Isogenien.

4.2 Der schwache Satz von Mordell-Weil

Der schwache Endlichkeitssatz von Mordell-Weil besagt

Satz 4.3. *Sei E eine über $\mathbb{Q}$ definierte elliptische Kurve. Dann ist die Faktorgruppe $E(\mathbb{Q})/2E(\mathbb{Q})$ endlich.*

Hierzu lassen sich einige Bemerkungen machen. Zum einen ist zu sagen, daß sich der Satz auch richtig ist, wenn man die 2 durch irgendeine ganze Zahl $m \ge 2$ ersetzt. Zweitens werden wir den Satz nicht allgemein beweisen können: der eigentliche Beweis lebt nämlich in dem Körper, den man erhält, wenn man zu $\mathbb{Q}$ die 2-Teilungspunkte von E adjungiert, also in $\mathbb{Q}(E[2])$. Indem man 2-Isogenien benutzt, kann man den Beweis schon dann in $\mathbb{Q}$ führen, wenn bereits ein einziger 2-Teilungspunkt rational ist. Tatsächlich haben Birch und Swinnerton-Dyer einen Beweis angegeben, der ohne die Arithmetik von Zahlkörpern auskommt, dafür aber mit der Reduktionstheorie quartischer Formen rechnet; dieser Beweis ist für die Praxis von erheblicher Bedeutung.

Wie bereits erwähnt, arbeitet unser Beweis mit elliptischen Kurven, die einen rationalen 2-Teilungspunkt besitzen. Indem man diesen Punkt nach $(0,0)$ bringt, kann man der elliptischen Kurve die Form $E : y^2 = x(x^2+ax+b)$ geben. Man findet $\Delta = 16b^2(a^2-4b)$, folglich ist E genau dann eine elliptische Kurve, wenn $b \ne 0$ und $a^2 - 4b \ne 0$ ist (und wenn $\operatorname{char} \mathrm{K} \ne 2$ ist, aber wir rechnen ohnehin über $\mathbb{Q}$). Den 2-Teilungspunkt $(0,0)$ bezeichnen wir im folgenden mit T.

Neben E wird auch die dazu isogene Kurve $\overline{E} : \overline{y}^2 = \overline{x}(\overline{x}^2 + \overline{a}\,\overline{x} + \overline{b})$ mit $\overline{a} = -2a$ und $\overline{b} = a^2 - 4b$ eine Rolle spielen; man findet $\overline{\Delta} = 16\overline{b}^2(\overline{a}^2 - 4\overline{b}) = 16b^2(a^2-4b)^2 \ne 0$. Weiter hat $\overline{E}$ den 2-Teilungspunkt $\overline{T}$. Die Isogenie $\phi : E \longrightarrow \overline{E}$ haben wir im vorigen Abschnitt angegeben: es ist $\alpha(0) = \mathcal{O}$, $\alpha(T) = \mathcal{O}$, und $\alpha(x,y) = (\overline{x}, \overline{y})$ mit $\overline{x} = y^2/x^2$ und $\overline{y} = y^2(x^2-b)/x^2$ sonst.

Kern des Beweises ist der Weil-Homomorphismus; dieser läßt sich ganz allgemein für Körper K der Charakteristik $\ne 2, 3$ definieren. In der Tat erhält man durch

$$\phi(P) = \begin{cases} K^{\times 2}, & \text{falls } P = \mathcal{O}; \\ bK^{\times 2}, & \text{falls } P = T; \\ xK^{\times 2}, & \text{falls } P = (x,y) \text{ und } P \ne \mathcal{O}, T. \end{cases}$$

eine Abbildung $\phi : E \longrightarrow K^\times/K^{\times 2}$, welche sich trotz der etwas seltsamen Definition als angenehme Abbildung entpuppt:

Proposition 4.4. *Die Abbildung $\phi : E \longrightarrow K^\times/K^{\times 2}$ ist Gruppenhomomorphismus.*

Beweis. Zu zeigen ist, daß $\phi(P_1 + P_2) = \phi(P_1)\phi(P_2)$ für alle P_j auf $E(K)$ gilt. Dazu unterscheiden wir

- $P_1 = \mathcal{O}$: dann folgt $\phi(P_1 + P_2) = \phi(P_2) = \phi(P_1)\phi(P_2)$.
- $P_1 + P_2 = \mathcal{O}$: dann ist $\phi(P_1 + P_2) = K^{\times 2}$; da aber P_1 und P_2 dieselbe x-Koordinate haben, ist $\phi(P_1) = \phi(P_2)$ und damit $\phi(P_1)\phi(P_2) = K^{\times 2}$.
- $P := P_1 = P_2$, aber $P \notin E(K)[2]$: nach (4.3) ist die x-Koordinate von $2P$ ein Quadrat, also $\phi(2P) = K^{\times 2} = \phi(P)\phi(P)$.
- allgemeiner Fall: sei $P_1 + P_2 + P_3 = \mathcal{O}$ und $P_j = (x_j, y_j)$. Die Gerade durch diese drei Punkte hat die Form $y = mx + c$, und die x_j genügen der Gleichung $x(x^2 + ax + b) - (mx + c)^2 = 0$. Also ist die linke Seite gleich $(x - x_1)(x - x_2)(x - x_3)$, und Vergleich des konstanten Terms liefert $x_1 x_2 x_3 = c^2$. Also ist $\phi(P_1)\phi(P_2)\phi(P_3) = K^{\times 2}$, also $\phi(P_1)\phi(P_2) = \phi(P_3) = \phi(-P_3) = \phi(P_1 + P_2)$.

Das war zu zeigen. □

Ganz entsprechend definiert man die Weil-Abbildung $\overline{\phi}$ für die zu E isogene Kurve $\overline{E}$; erstaunlicherweise hat $\overline{\phi}$ etwas mit der Isogenie $\alpha : E \longrightarrow \overline{E}$ zu tun:

Proposition 4.5. *Es ist* $\ker \alpha = \{\mathcal{O}, T\}$ *und* $\operatorname{im} \alpha = \ker \overline{\phi}$.

Mittels exakter Sequenzen kann man das prägnanter so sagen: die Sequenz

$$0 \longrightarrow \{\mathcal{O}, T\} \longrightarrow E(K) \xrightarrow{\ \alpha\ } \overline{E}(K) \xrightarrow{\ \overline{\phi}\ } K^\times / K^{\times 2}$$

ist exakt.

Beweis. Daß $\mathcal{O}$ und T im Kern von α liegen, ist klar nach Konstruktion. Andererseits kann $\alpha(P) = \mathcal{O}$ für ein $P \neq \mathcal{O}$ nur dann gelten, wenn $x_P = 0$ und damit $y_P = 0$ ist.

Jetzt behaupten wir, daß $\overline{T}$ genau dann im Bild von α liegt, wenn $\overline{b} \in K^{\times 2}$ ist. Denn $\alpha(P) = (0, 0)$ mit $P = (x, y)$ ist wegen $\alpha(P) = (y^2/x^2, y(x^2 - b)/x^2)$ äquivalent zu $x \neq 0$, $y = 0$, also zu $x^2 + ax + b = 0$ in K. Diese Gleichung ist in K aber genau dann lösbar, wenn ihre Diskriminante ein Quadrat ist, wenn also $a^2 - 4b = \overline{b} \in K^{\times 2}$ gilt.

Schließlich fragen wir uns, wann ein $(\overline{x}, \overline{y})$ Bild eines $P = (x, y) \neq \mathcal{O}, T$ ist. Zeigen müssen wir, daß dies genau dann der Fall ist, wenn $\overline{x} \in K^{\times 2}$ ist.

Die Richtung $\operatorname{im} \alpha \subseteq \ker \overline{\phi}$ ist dabei einfach: denn $(\overline{x}, \overline{y}) \in \operatorname{im} \alpha$ bedeutet ja $\overline{x} = (y/x)^2$, und das ist offensichtlich ein Quadrat. Sei daher umgekehrt $(\overline{x}, \overline{y}) \in \ker \overline{\phi}$, also $\overline{x} = w^2$ für ein $w \in K^\times$. Dann setzen wir für $j = 1, 2$

$$x_j = \frac{1}{2}\Big(\overline{x} - a + (-1)^j \frac{\overline{y}}{w}\Big), \quad y_j = (-1)^j w x_j.$$

Wegen $x_1x_2 = b$ (man benutze $\overline{y}^2 = \overline{x}\,(\overline{x}^2 + \overline{a}\,\overline{x} + \overline{b})$) sind die $x_j \neq 0$; um nachzuweisen, daß die Punkte (x_j, y_j) auf $E(K)$ liegen, müssen wir $y_j^2/x_j^2 = x_j + a + b/x_j$ zeigen, also $\overline{x} = x_1 + a + b/x_1 = x_1 + a + x_2$ und $\overline{x} = x_2 + a + b/x_2 = x_2 + a + x_1$; das ist aber klar. Schließlich haben wir noch $\alpha(x_j, y_j) = ((y_j/x_j)^2, y_j(x_j^2 - b)/x_j^2) = (w^2, w(x_2 - x_1)) = (\overline{x}, \overline{y})$. Das war's. $\square$

Damit haben wir die beiden exakten Sequenzen

$$0 \longrightarrow \{\mathcal{O}, T\} \longrightarrow E(K) \xrightarrow{\ \alpha\ } \overline{E}(K) \xrightarrow{\ \overline{\phi}\ } K^\times/K^{\times\,2}$$
$$0 \longrightarrow \{\mathcal{O}, \overline{T}\} \longrightarrow \overline{E}(K) \xrightarrow{\ \beta\ } E(K) \xrightarrow{\ \phi\ } K^\times/K^{\times\,2}.$$

An dieser Stelle kommt die Arithmetik von K ins Spiel, weswegen wir uns auf die Fälle $K = \mathbb{Q}_p$ und $K = \mathbb{Q}$ beschränken werden. Ist $K = \mathbb{Q}_p$, so ist $\mathbb{Q}_p^\times/\mathbb{Q}_p^{\times\,2}$ endlich; beispielsweise sind für $p = 2$ nur die von -1, 2 und 5 erzeugten Nebenklassen nichttrivial (sh. Frey [Fr]).

Im Falle $K = \mathbb{Q}$ dagegen ist die Gruppe $\mathbb{Q}^\times/\mathbb{Q}^{\times\,2}$ sicher *nicht* endlich erzeugt: ein unabhängiges Erzeugendensystem sind beispielsweise die Klassen $p\,\mathbb{Q}^{\times\,2}$, p positive Primzahl, zusammen mit $-1\mathbb{Q}^{\times\,2}$. Insbesondere sind hier ϕ und $\overline{\phi}$ nicht surjektiv: wir wollen ja zeigen, daß $E(\mathbb{Q})$ und $\overline{E}(\mathbb{Q})$ endlich erzeugt sind; wenn sie dies sind, dann sind aber auch deren Bilder unter ϕ bzw. $\overline{\phi}$ endlich erzeugt, und Surjektivität würde dann bedeuten, daß $\mathbb{Q}^\times/\mathbb{Q}^{\times\,2}$ endlich erzeugt ist.

Unsere Aufgabe ist jetzt, die endliche Erzeugtheit von $\operatorname{im}\phi$, bzw. $\operatorname{im}\overline{\phi}$ für den Fall $K = \mathbb{Q}$ nachzuweisen. Dazu betrachten wir die Sequenz

$$E(\mathbb{Q}) \xrightarrow{\ \alpha\ } \overline{E}(\mathbb{Q}) \xrightarrow{\ \beta\ } E(\mathbb{Q}),$$

aus der wir (sh. Anhang) die exakte Kern-Kokern-Sequenz

$$\begin{array}{ccccccc}
0 & \longrightarrow & \ker\alpha & \longrightarrow & \ker(\beta\circ\alpha) & \longrightarrow & \ker\beta \\
 & & & & & & \big\downarrow \\
0 & \longleftarrow & \operatorname{coker}\beta & \longleftarrow & \operatorname{coker}(\beta\circ\alpha) & \longleftarrow & \operatorname{coker}\alpha
\end{array}$$

erhalten. Nun ist $\ker\alpha = \{\mathcal{O}, T\}$, $\ker(\beta\circ\alpha) = E(\mathbb{Q})[2]$, $\ker\beta = \{\mathcal{O}, \overline{T}\}$, $\operatorname{coker}\alpha = \overline{E}/\operatorname{im}\alpha \simeq \overline{E}/\ker\overline{\phi} \simeq \operatorname{im}\overline{\phi}$, $\operatorname{coker}(\beta\circ\alpha) = \overline{E}(\mathbb{Q})/2\overline{E}(\mathbb{Q})$ und $\operatorname{coker}\beta \simeq E/\ker\phi \simeq \operatorname{im}\phi$, und obige exakte Sequenz wird zu

$$\begin{array}{ccccccc}
0 & \longrightarrow & \{\mathcal{O}, T\} & \longrightarrow & E(\mathbb{Q})[2] & \longrightarrow & \{\mathcal{O}, \overline{T}\} \\
 & & & & & & \big\downarrow \\
0 & \longleftarrow & \operatorname{im}\phi & \longleftarrow & E(\mathbb{Q})/2E(\mathbb{Q}) & \longleftarrow & \operatorname{im}\overline{\phi}
\end{array}$$

Darauf können wir die Indexformel aus dem Anhang anwenden; setzt man $\#E(\mathbb{Q})[2] = 2^t$, und nimmt an (was wir nachher zeigen werden), daß $E(\mathbb{Q})$ endlich erzeugt ist, so kann man $E(\mathbb{Q}) = E(\mathbb{Q})_{\text{tors}} \oplus \mathbb{Z}^r$ schreiben, wobei $r \in \mathbb{N}_0$ der $\mathbb{Z}$-Rang von $E(\mathbb{Q})$ heißt. Damit ist $\#E(\mathbb{Q})/2E(\mathbb{Q}) = 2^{r+t}$, und wir finden $2 \cdot 2 \cdot 2^{r+t} = 2^t \cdot \#\operatorname{im}\phi \cdot \#\operatorname{im}\overline{\phi}$, also

$$2^r = \frac{\#\operatorname{im}\phi \cdot \#\operatorname{im}\overline{\phi}}{4}.$$

Durch Vertauschen von E und $\overline{E}$ folgt aus dieser Formel übrigens sofort, daß 2-isogene Kurven denselben Rang haben. Die Torsionsgruppen $E(\mathbb{Q})_{\text{tors}}$ und $\overline{E}(\mathbb{Q})_{\text{tors}}$ sind dagegen im allgemeinen nicht isomorph; Beispiele dafür werden wir unten geben.

Zu zeigen ist noch, daß $\operatorname{im}\phi$ und $\operatorname{im}\overline{\phi}$ endlich erzeugt und (weil es Untergruppen der elementar-abelschen Gruppe $\mathbb{Q}^\times/\mathbb{Q}^{\times\,2}$ sind) damit endlich sind; damit wird dann der schwache Endlichkeitssatz bewiesen sein. Daß $\operatorname{im}\phi$ endlich erzeugt ist, folgt sofort aus der folgenden expliziten Beschreibung des Bilds, die wir für einen beliebigen ZPE-Ring R mit Quotientenkörper K formulieren (wir werden die Aussage aber nur für $R = \mathbb{Z}$ und $R = \mathbb{Z}_p$ verwenden):

Proposition 4.6. *Sei R ein ZPE-Ring mit Quotientenkörper K, und $E : y^2 = x(x^2 + ax + b)$ eine über K definierte elliptische Kurve mit $a, b \in R$; dann besteht* $\operatorname{im}\phi$ *aus $1\,K^{\times\,2}$ und allen Klassen $b_1 K^{\times\,2}$, für welche gilt:*

i) $b = b_1 b_2$ mit $b_1, b_2 \in R$;

ii) $N^2 = b_1 M^4 + aM^2e^2 + b_2 e^4$ hat eine Lösung $N, M, e \in R$ mit $e \neq 0$.

Sind i) und ii) für ein $b_1 \mid b$ erfüllt, so ist $(b_1M^2/e^2, b_1MN/e^3) \in E(K)$.

Da es im Falle $K = \mathbb{Q}$ und $R = \mathbb{Z}$ nur endlich viele b_1 mit der Eigenschaft i) gibt, ist klar, daß $\operatorname{im}\phi$ endlich erzeugt ist. Außerdem genügt es, quadratfreie b_1 zu betrachten wegen $\phi(b_1) = \phi(b_1t^2)$.

Bevor wir zum Beweis kommen, formulieren wir eine nützliche Beobachtung, die wir implizit bereits gemacht haben (nämlich für jede Primzahl p einzeln), noch einmal explizit:

Hilfssatz 4.7. *Seien K und R wie oben und $E : y^2 = x^3 + ax^2 + bx + c$ eine über K definierte elliptische Kurve mit Koeffizienten aus R; dann kann man jeden Punkt $P \neq \mathcal{O}$ auf $E(K)$ in der Form $P = (x, y)$ mit $x = m/e^2$, $y = n/e^3$, $m, n, e \in R$, und $(m, e) = (n, e) = 1$ schreiben.*

Beweis. Sei $x = m/M$ und $y = n/N$ mit $M, N \in R$ und $(m, M) = (n, N) = 1$. Wir möchten zeigen, daß $M^3 \mid N^2$ und $N^2 \mid M^3$ ist, denn dann folgt $M^3 = N^2$, also $M = e^2$ und $N = e^3$ für ein $e \in R$.

Aus der Gleichung $y^2 = x^3 + ax^2 + bx + c$ erhält man

$$M^3 n^2 = N^2 m^3 + aN^2 M m^2 + bN^2 M^2 m + cN^2 m^3.$$

Da die rechte Seite durch N^2 teilbar ist und $(n, N) = 1$ gilt, muß $N^2 \mid M^3$ sein. Umgekehrt ist $M \mid N^2 m^3$, wegen $(m, M) = 1$ also $M \mid N^2$. Damit wiederum folgt $M \mid N$, und dann schließlich $M^3 \mid N^2$. □

Beweis von Proposition 4.6. Wir starten mit einem Punkt $P = (x, y) \in E(K)$ und wollen zeigen, daß $\phi(P)$ die angegebene Form hat. Dazu schreiben wir $x = m/e^2$, $y = n/e^3$ wie in Hilfssatz 4.7, und stellen fest, daß

$$n^2 = m(m^2 + ame^2 + be^4) \tag{4.4}$$

gilt. Wir setzen jetzt $b_1 = u(m, b)$, wobei wir die Einheit $u \in R^\times$ später festlegen. Mit $m = b_1 m_1$ und $b = b_1 b_2$ liefert (4.4) dann $b_1^2 \mid n^2$. Wir können also $n = b_1 n_1$ setzen und finden

$$n_1^2 = m_1(b_1 m_1^2 + am_1 e^2 + b_2 e^4). \tag{4.5}$$

Dabei ist $(m_1, b_2) = (m_1, e) = 1$, folglich $m_1 = vM^2$ für eine Einheit $v \in R^\times$ (Beweis wie in Hilfssatz 2.12). Jetzt wählen wir $u \in R^\times$ so, daß $v = 1$ wird, und wegen $M \mid n_1$ ist dann $n_1 = MN$ für ein $N \in R$. Damit wird (4.5) zu

$$N^2 = b_1 M^4 + aM^2 e^2 + b_2 e^4, \tag{4.6}$$

und es ist $\phi(x) = xK^{\times\,2} = mK^{\times\,2} = b_1\mathbb{Q}^{\times\,2}$ wie behauptet.

Sind umgekehrt i) und ii) erfüllt, so zeigt einfaches Nachrechnen, daß der angegebene rationale Punkt auf $E(K)$ liegt. □

Beispiele

1. Sei $E : y^2 = x^3 - 4x$; das ist die Fermatkurve $X^4 + Y^4 = Z^2$ in kurzer Weierstraßform. Weiter ist $\overline{E} : y^2 = x^3 + 16x$. Wir wollen $\#\operatorname{im}\phi$ und $\#\operatorname{im}\overline{\phi}$ berechnen. Wegen $b = -4$ haben wir $b_1 \in \{-1, \pm 2\}$ zu betrachten (die Klasse $1 \cdot \mathbb{Q}^{\times\,2} = \phi(\mathcal{O})$ liegt ohnehin im Bild). Von vornherein wissen wir,

daß $-1 \cdot \mathbb{Q}^{\times 2} = -4 \cdot \mathbb{Q}^{\times 2} = \phi(T)$ im Bild liegt; wir wollen dies trotzdem nachprüfen, indem wir die dazugehörige Gleichung lösen.

b_1	Gleichung	N	M	e	P
-1	$N^2 = -M^4 + 4e^4$	2	0	1	$(0,0)$
2	$N^2 = 2M^4 - 2e^4$	0	1	1	$(2,0)$
-2	$N^2 = -2M^4 + 2e^4$	0	1	1	$(-2,0)$

Hierbei ist P der aus b_1 und der angegebenen Lösung (N, M, e) berechnete Punkt. Insbesondere haben wir $\operatorname{im}\phi = \{1\cdot\mathbb{Q}^{\times 2}, -1\cdot\mathbb{Q}^{\times 2}, 2\cdot\mathbb{Q}^{\times 2}, -2\cdot\mathbb{Q}^{\times 2}\} \simeq \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, also $\#\operatorname{im}\phi = 4$.

Wegen $\overline{\phi}(\overline{T}) = 16 \cdot \mathbb{Q}^{\times 2} = 1 \cdot \mathbb{Q}^{\times 2}$ liefert der Torsionspunkt kein nichttriviales Bild. Weiter hat man, da negative b_1 auf im Reellen nicht lösbare Gleichungen führen, nur noch $b_1 = 2$ zu betrachten. Die Gleichung $N^2 = 2M^4 + 8e^4$ ist aber in $\mathbb{Q}$ nur trivial lösbar: denn offensichtlich muß N gerade sein, also $N = 2n$, folglich $2n^2 = M^4 + 4e^4$ sein. Daher ist $M = 2m$, $n = 2n_1$ und $2n_1^2 = 4m^4 + e^4$; jetzt folgt $e = 2e_1$ und $n_1^2 = 2m^4 + 8e_1^4$. Dies ist aber die Ausgangsgleichung. Wir schließen, daß jede Lösung (N, M, e) die Eigenschaft hat, daß N, M und e beliebig oft durch 2 teilbar sind: also ist $N = M = e = 0$.

Damit ist $\#\operatorname{im}\overline{\phi} = 1$, somit $r = 0$ und

$$E(\mathbb{Q}) = E(\mathbb{Q})_{\text{tors}} = \langle (0,0), (2,0) \rangle \simeq \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}, \quad \text{sowie}$$
$$\overline{E}(\mathbb{Q}) = \overline{E}(\mathbb{Q})_{\text{tors}} = \langle (0,0) \rangle \simeq \mathbb{Z}/2\mathbb{Z}.$$

Als Korollar halten wir fest, daß die Fermatgleichung $Z^2 = X^4 + 1$ keine nichttrivialen Lösungen in $\mathbb{Q}$ hat: wir haben bereits gesehen, daß die Größen $x = 2(Z + X^2)$ und $y = 2xX$ der Weierstraßgleichung $y^2 = x^3 - 4x$ genügen; die Umkehrabbildung wird durch $X = y/2x$ und $Z = x/2 - X^2 = x/2 - y^2/4x^2$ geliefert (in der Tat ist damit $Z^2 = \frac{1}{4}x^2 - y^2/4x + y^4/16x^4 = \frac{1}{4}x^2 - (x^3 - 4x)/4x + X^4 = X^4 + 1$). Was passiert mit den Torsionspunkten von $E(\mathbb{Q})$ unter dieser Abbildung? Offenbar geht $(0,0)$ auf den unendlich fernen Punkt, $(\pm 2, 0)$ dagegen auf die trivialen Lösungen $(X, Z) = (\pm 1, 0)$.

2. Sei $E : y^2 = x^3 - x$. Wir finden $\overline{E} : y^2 = x^3 + 4x$ und wollen $\operatorname{im}\phi$ und $\operatorname{im}\overline{\phi}$ bestimmen.

Zur Berechnung von $\operatorname{im}\phi$ haben wir (4.6) für $b_1 = -1$ zu lösen; offenbar ist $M = 0$, $e = 1$ eine solche. Diese Lösung entspricht dem Punkt $(1, 0) \in E(\mathbb{Q})$, der offensichtlich ein Torsionspunkt (der Ordnung 2) ist.

Entsprechend haben wir für im $\overline{\phi}$ die Gleichung (4.6) für $b_1 = -1, \pm 2$ zu betrachten. Da negative b_1 auf keine lösbaren Quartiken führen, ist $b_1 = 2$ die einzige Möglichkeit, und die entsprechende Gleichung $N^2 = 2M^4 + 2e^4$ hat in der Tat die Lösung $M = e = 1$. Dies führt auf den Punkt $(2, 4) \in \overline{E}(\mathbb{Q})$.

Wir haben also $\#\ \operatorname{im} \phi = \#\ \operatorname{im} \overline{\phi} = 2$, somit $2^r = 1$ und $r = 0$.

3. Sei $E : y^2 = x^3 - 5x$; hier ist $\overline{E} : y^2 = x^3 + 20x$. Die Bestimmung von im ϕ ist einfach:

b_1	Gleichung	N	M	e	P
1	$N^2 = M^4 - 5e^4$	1	3	2	$(\frac{9}{4}, \frac{3}{8})$
-1	$N^2 = -M^4 + 5e^4$	2	1	1	$(-1, -2)$
5	$N^2 = 5M^4 - e^4$	2	1	1	$(5, 10)$
-5	$N^2 = -5M^4 + e^4$	1	0	1	$(0, 0)$

Man beachte, daß $-5 \cdot \mathbb{Q}^{\times 2} = \phi(T)$ ist; selbstverständlich kommt dieser Punkt am Ende wieder heraus.

Um im $\overline{\phi}$ zu bestimmen, müssen wir $N^2 = b_1 M^4 + b_2 e^4$ für $b_1 \in \{1, 2, 5, 10\}$ betrachten (negative b_1 führen auf Gleichungen, die nicht einmal in $\mathbb{R}$ lösbar sind). Wegen $\overline{\phi}(\overline{T}) = 20\mathbb{Q}^{\times 2} = 5\mathbb{Q}^{\times 2}$ muß $N^2 = 4M^4 + 5e^4$ lösbar sein: in der Tat ist $(N, M, e) = (3, 1, 1)$ eine Lösung; diese führt auf den Punkt $\overline{P} = (5, 15)$ auf $\overline{E}$. Wegen $\overline{E}(\mathbb{Q})_{\text{tors}} = \{\mathcal{O}, \overline{T}\}$ muß $\overline{P}$ ein Punkt mit unendlicher Ordnung sein.

Dagegen ist die Gleichung $N^2 = 2M^4 + 10e^4$ nicht rational lösbar: setzen wir $N = 2n$, so folgt $2n^2 = M^4 + 5e^4$; dabei dürfen wir annehmen, daß $5 \nmid M$ ist, weil sonst $5 \mid e$ und $5^2 \mid n$ folgen würde, d.h. wir können so lange kürzen, bis $5 \nmid M$ ist. Jetzt folgt aber $2n^2 \equiv M^4 \bmod 5$, und dies impliziert wegen $5 \nmid M$, daß 2 quadratischer Rest modulo 5 ist. Dies ist aber nicht der Fall, folglich ist $2n^2 = M^4 + 5e^4$ nicht lösbar, und da im $\overline{\phi}$ eine Gruppe ist, muß $\#\operatorname{im} \overline{\phi} = 2$ sein.

Damit folgt insgesamt $2^r = 2$, also $r = 1$. Auf beiden Kurven haben wir sogar einen Punkt unendlicher Ordnung gefunden, nämlich $(-1, -2)$ auf E und $(5, 15)$ auf $\overline{E}$. Um zu zeigen, daß $E(\mathbb{Q}) = \langle (0, 0) \rangle \times \langle (-1, -2) \rangle$, bzw. $\overline{E}(\mathbb{Q}) = \langle (0, 0) \rangle \times \langle (5, 15) \rangle$ ist, braucht man aber den Begriff der Höhe.

Daß es elliptische Kurven über $\mathbb{Q}$ mit beliebig großem Rang gibt, ist eine noch offene Vermutung. Derzeitige Rekordhalterin ist die Kurve $E : y^2 +$

$xy + y = x^3 + ax + b$ mit

$$\begin{aligned} a &= -19252966408674012828065964616418441723 \\ b &= 32685500727716376257923347071452044295907443056345614006 \end{aligned}$$

und Rang ≥ 23; diese wurde von Roland Martin und William McMillen im Juni 1997 gefunden. Der 'kleinste' Punkt unter den 23 Erzeugenden ist

$$(1127027270330215920, 35239781274071006741103777602).$$

Die Berechnung von im ϕ

Damit ist die Berechnung des Rangs einer elliptischen Kurve mit rationalem 2-Torsionspunkt zurückgeführt auf die Berechnung von im ϕ, d.h. auf die Frage, ob quartische Gleichungen der Form (4.6) eine rationale Lösung besitzen. Unglücklicherweise kennt man kein Verfahren, welches allgemein funktioniert. Natürlich kann man in vielen Fällen die Lösbarkeit ausschließen, weil (4.6) nicht einmal modulo einer geeigneten Primzahl lösbar ist, und oft genügt das sogar, um z.B. zu zeigen, daß eine elliptische Kurve Rang 0 hat.

Etwas allgemeiner kann man feststellen, daß Lösbarkeit in rationalen Zahlen natürlich die Lösbarkeit in $\mathbb{R}$ und allen $\mathbb{Q}_p$ impliziert. Wenn wir die Propositionen 4.4 und 4.6 auf $K = \mathbb{Q}_p$ und $R = \mathbb{Z}_p$ anwenden, dann sehen wir, daß die Klassen $b_1\mathbb{Q}^{\times 2}$, für welche (4.6) in $\mathbb{Q}_p$ lösbar ist, eine Gruppe bilden; insbesondere bilden die $b_1\mathbb{Q}^{\times 2}$, für welche (4.6) in *allen* $\mathbb{Q}_p$ (einschließlich $p = \infty$) lösbar ist, eine Gruppe, welche im ϕ als Untergruppe enthält – man nennt sie die zur 2-Isogenie β gehörige *Selmergruppe* $S(E/\mathbb{Q})[\beta]$ von E. Wegen im ϕ = coker $\beta = \overline{E}(\mathbb{Q})/\beta(E(\mathbb{Q}))$ haben wir also einen Monomorphismus $E(\mathbb{Q})/\beta(\overline{E}(\mathbb{Q})) \longrightarrow S(E/\mathbb{Q})[\beta]$. Den Kokern $S(E/\mathbb{Q})[\beta]/$ im ϕ dieser Abbildung nennt man die Tate-Shafarevic-Gruppe $Ш(E/\mathbb{Q})[\beta]$; diese ist also durch die exakte Sequenz

$$0 \longrightarrow E(\mathbb{Q})/\beta(\overline{E}(\mathbb{Q})) \longrightarrow S(E/\mathbb{Q})[\beta] \longrightarrow Ш(E/\mathbb{Q})[\beta] \longrightarrow 0$$

definiert. Wenn die Tate-Shafarevic-Gruppe nicht trivial ist, liefert das hier beschriebene Verfahren i.a. nicht den genauen Rang – es sei denn, es gelingt irgendwie, die Ordnung der auftretenden Tate-Shafarevic-Gruppen genau zu bestimmen.

Eine gute Approximation an die Selmergruppe erhält man, wenn man nach denjenigen $b_1 \in \mathbb{Z}$ fragt, für die die Gleichung

$$N^2 = b_1x^2 + axy + b_2y^2 \tag{4.7}$$

lösbar ist. Ist nämlich (N, M, e) eine Lösung von (4.6), so ist (N, M^2, e^2) eine solche von (4.7), mit anderen Worten: $S(E/\mathbb{Q})[\beta]$ und erst recht im ϕ ist in der Menge aller $b_1\mathbb{Q}^{\times 2}$ enthalten, für welche (4.7) lösbar ist. Schreiben wir (4.7) in der Form $N^2 = b_1(x + \frac{a}{2b_1}y)^2 + \frac{4b-a^2}{4b_1}y^2$, so sehen wir, daß wir Gleichungen der Form $z^2 = rx^2 - sy^2$ zu untersuchen haben. Nach dem Hasseschen Lokal-Global-Prinzip besitzt eine solche Gleichung genau dann eine nichttriviale rationale Lösung, wenn sie für alle $\mathbb{Q}_p$ (einschließlich $p = \infty$) eine solche besitzt.

Wir definieren nun das Hilbertsche Normenrestsymbol (auch einfach Hilbertsymbol genannt) $(r, s)_p$ für alle $r, s \in \mathbb{Q}^\times$ durch

$$(r, s)_p = \Big(\frac{r, s}{p}\Big) = \begin{cases} +1 & \text{falls } r = x^2 - sy^2 \text{ in } \mathbb{Q}_p \text{ lösbar} \\ -1 & \text{sonst.} \end{cases}$$

Das Hilbertsymbol hat folgende Eigenschaften, die sich relativ leicht nachweisen lassen:
$(r_1r_2, s)_p = (r_1, s)_p(r_2, s)_p$ (Multiplikativität im ersten Argument),
$(r^2, s)_p = 1$ (Quadrate sind Normenreste),
$(r, -r)_p = 1$ (klar, weil $r = x^2 + ry^2$ lösbar),
$(a, b)_p = (b, a)_p$ (Symmetrie).
Schließlich zeigt ein einfaches Abzählargument, daß $(a, b)_p = 1$ für alle endlichen $p \nmid 2ab$ gilt. Damit macht das Produkt $\prod_p (a, b)_p$ Sinn, und man kann zeigen, daß die Produktformel

$$\prod_p (a, b)_p = 1$$

äquivalent zum quadratischen Reziprozitätsgesetz ist. Die Details plus diverse Hinweise zur Berechnung von Hilbertsymbolen nebst einem Beweis des angesprochenen Lokal-Global-Prinzips für (4.7) findet man in Freys Büchlein [Fr]; das wichtigste Hilfsmittel zur Berechnung der Hilbertsymbole ist die einfache Beobachtung, daß $(a, p)_p = (a/p)$ ist, falls $a \in \mathbb{Z}$ nicht durch die ungerade Primzahl p teilbar ist (Hensels Lemma!).

Da die Gleichung $z^2 = rx^2 - sy^2$ genau dann lösbar ist, wenn $rz^2 = X^2 - rsy^2$ es ist, ist die Lösbarkeit von (4.7) in $\mathbb{Q}_p$ äquivalent zur Bedingung $1 = (b_1, -b_1\frac{4b-a^2}{4b_1})_p = (b_1, a^2 - 4b)_p$. Aus der Multiplikativität des Hilbertsymbols ergibt sich dann, daß die Menge aller $b_1 \cdot \mathbb{Q}^{\times 2}$, für welche (4.7) in $\mathbb{Q}_p$ lösbar ist, eine Gruppe bildet. Damit ist auch die Menge aller $b_1 \cdot \mathbb{Q}^{\times 2}$, für welche (4.7) in *allen* $\mathbb{Q}_p$ lösbar ist, eine Gruppe, und diese enthält, wie wir gesehen haben, die Selmergruppe und damit im ϕ.

Als Konsequenz dieser Diskussion halten wir fest:

Proposition 4.8. *Die Selmergruppe $S(E/\mathbb{Q})[\beta]$ von E, also die Menge $\{b_1 \cdot \mathbb{Q}^{\times 2}\}$, für welche (4.6) in allen $\mathbb{Q}_p$ lösbar ist, ist enthalten in der Gruppe*

$$\{b_1 \cdot \mathbb{Q}^{\times 2} : (b_1, a^2 - 4b)_p = 1 \textit{ für alle } p\}.$$

Beispiel: die oben betrachtete Gleichung $N^2 = 2M^4 + 10e^4$ ist in $\mathbb{Q}_5$ nicht lösbar, weil $(2,10)_5 = (2,5)_5(2,2)_5 = (2,5)_5 = -1$ ist.

Etwas allgemeiner wollen wir die elliptischen Kurven $E : y^2 = x^3 + px$ für prime $p \geq 2$ betrachten. Hier ist die zu E 2-isogene Familie gegeben durch $\overline{E} : y^2 = x^3 - 4px$.

Wegen $\phi(T) = p\mathbb{Q}^{\times 2}$, und weil negative b_1 hier nicht in Frage kommen (z.B. wegen $(b_1, -4p)_\infty = (b_1, -1)_\infty = -1$ in diesem Fall – dahinter steckt natürlich nur die Unlösbarkeit von Gleichungen der Form $N^2 = b_1 M^4 + b_2 e^4$ im Reellen, wenn b_1 und b_2 negativ sind), haben wir sicher im $\phi = \{1\mathbb{Q}^{\times 2}, p\mathbb{Q}^{\times 2}\}$.

Zur Bestimmung von im $\overline{\phi}$ haben wir dagegen die Werte $b_1 \in \{\pm 1, \pm 2, \pm p, \pm 2p\}$ zu untersuchen. Wegen $\overline{\phi}(\overline{T}) = -4p\mathbb{Q}^{\times 2} = -p\mathbb{Q}^{\times 2}$ geht es nur um $b_1 = -1, \pm 2, -p, \pm 2p$, und weil im $\overline{\phi}$ eine Gruppe ist, genügt die Untersuchung von $b_1 = -1$ und $b_1 = 2$.

Für $b_1 = -1$ haben wir

$$N^2 = -M^4 + 4pe^4 \tag{4.8}$$

zu untersuchen. Die entsprechende quadratische Gleichung in $\mathbb{Q}_q$ lösbar genau dann, wenn $(-1,p)_q = 1$ ist; dies ist automatisch richtig für $q \nmid 2q\infty$, weiter ist $(-1,p)_\infty = 1$ wegen $p > 0$, $(-1,p)_p = (-1/p) = 1$ genau dann, wenn $p \equiv 1 \bmod 4$, und schließlich $(-1,p)_2 = (-1,p)_p$ wegen der Hilbertschen Produktformel. Sei daher $p \equiv 1 \bmod 4$; dann hat (4.8) in $\mathbb{Z}_p$ die Lösung $N = \sqrt{4p-1}$, $M = e = 1$, denn $\sqrt{4p-1} \in \mathbb{Z}_p$ wegen $(\frac{4p-1}{p}) = (\frac{-1}{p}) = 1$ plus Henselschem Lemma. Was die Lösbarkeit in $\mathbb{Z}_2$ angeht, so stellen wir fest, daß $M = 2m$ gerade sein muß; mit $N = 2n$ folgt dann $n^2 = 4m^4 + pe^4$, und diese Gleichung hat die Lösung $(n,m,e) = (\sqrt{p}, 0, 1)$, falls $p \equiv 1 \bmod 8$, sowie $(n,m,e) = (\sqrt{p+4}, 1, 1)$, falls $p \equiv 5 \bmod 8$ (nach dem Henselschen Lemma ist $\sqrt{a} \in \mathbb{Z}_2$ für alle $a \equiv 1 \bmod 8$). Also ist $-1\mathbb{Q}^{\times 2} \in S(E/\mathbb{Q})[\alpha] \iff p \equiv 1 \bmod 4$.

Mit denselben Mitteln kann man zeigen, daß $2\mathbb{Q}^{\times 2} \in S(E/\mathbb{Q})[\alpha] \iff p \equiv 1, 9, 15 \bmod 16$ und $-2\mathbb{Q}^{\times 2} \in S(E/\mathbb{Q})[\alpha] \iff p \equiv 1, 3, 9 \bmod 16$ gilt.

Weil $S(E/\mathbb{Q})[\alpha]$ eine Gruppe ist, erhält man folgende Tabelle:

$S(E/\mathbb{Q})[\alpha]$	p mod 16	$S(E/\mathbb{Q})[\alpha]$	p mod 16
$\langle -1\mathbb{Q}^{\times 2}, 2\mathbb{Q}^{\times 2}, p\mathbb{Q}^{\times 2}\rangle$	1	$\langle -1\mathbb{Q}^{\times 2}, 2\mathbb{Q}^{\times 2}, p\mathbb{Q}^{\times 2}\rangle$	9
$\langle -2\mathbb{Q}^{\times 2}, -p\mathbb{Q}^{\times 2}\rangle$	3	$\langle -p\mathbb{Q}^{\times 2}\rangle$	11
$\langle -1\mathbb{Q}^{\times 2}, p\mathbb{Q}^{\times 2}\rangle$	5	$\langle -1\mathbb{Q}^{\times 2}, p\mathbb{Q}^{\times 2}\rangle$	13
$\langle -p\mathbb{Q}^{\times 2}\rangle$	7	$\langle 2\mathbb{Q}^{\times 2}, -p\mathbb{Q}^{\times 2}\rangle$	15

Folgerung: Ist $p \equiv 1 \bmod 8$ prim, so sind die zu $b_1 = -1, \pm 2$ gehörigen Gleichungen in allen $\mathbb{Q}_p$ sowie in $\mathbb{R} = \mathbb{Q}_\infty$ lösbar. Für z.B. $p = 17$ kann man darüberhinaus zeigen, daß die entsprechenden Gleichungen trotzdem keine Lösung in $\mathbb{Q}$ besitzen. Diese Elemente erzeugen hier also eine nichttriviale Untergruppe von $Ш(E/\mathbb{Q})[\alpha]$.

Der allgemeine Fall

Was kann man tun, wenn die elliptische Kurve keinen rationalen Punkt der Ordnung 2 besitzt? Beweistechnisch ist das kein großes Problem: man ersetzt dann einfach $\mathbb{Q}$ durch den Zahlkörper $K = \mathbb{Q}(E[2])$, den man erhält, wenn man die Koordinaten der 2-Torsionspunkte von $E(\mathbb{C})$ zu $\mathbb{Q}$ adjungiert. Zum Beweis dafür, daß im ϕ dann ebenfalls endlich erzeugt ist, muß man den Einheitensatz von Dirichlet (oder die schwächere Aussage, daß die Einheitengruppe des Rings ganzer Zahlen in K endlich erzeugt ist), sowie den Satz von der Endlichkeit der Klassenzahl benutzen. Die Frage, wie man den Rang in der Praxis berechnet, ist eine ganz andere. In der Regel benutzt man dazu ein Verfahren von Birch und Swinnerton-Dyer, das Rechnungen in algebraischen Zahlkörpern durch das systematische Auflisten von Polynomen vierten Grades mit gegebenen Invarianten ersetzt. Prinzipiell ließe sich die Verwendung algebraischer Zahlentheorie auch bei Beweisen vermeiden, allerdings bekommt man es dann mit Divisoren und Galoiskohomologie zu tun.

4.3 Höhen und der Satz von Mordell-Weil

Ziel ist der Beweis des folgenden Satzes:

Satz 4.9. *Sei* $E: y^2 = x(x^2+ax+b)$ *eine elliptische Kurve mit rationalem 2-Torsionspunkt* $(0,0)$. *Dann ist* $E(\mathbb{Q}) = E_{\rm tors} \oplus \mathbb{Z}^r$ *für ein* $r \in \mathbb{N}_0$*; insbesondere ist* $E(\mathbb{Q})$ *endlich erzeugt.*

Unglücklicherweise folgt die endliche Erzeugtheit nicht sofort aus der Endlichkeit von $E(\mathbb{Q})/2E(\mathbb{Q})$: so ist z.B. auch $\mathbb{Q}/m\mathbb{Q}$ endlich für alle $m \in \mathbb{N}$ (in der Tat ist sogar $\mathbb{Q}/m\mathbb{Q} = 0$ für $m \neq 0$, weil jedes $p/q \in \mathbb{Q}$ von der Form $m \cdot p/mq$ und auch $p/mq \in \mathbb{Q}$ ist; man nennt $\mathbb{Q}$ deswegen "dividierbar"), aber, wie wir gesehen haben, weit davon entfernt, endlich erzeugt zu sein.

Ganz entsprechend haben wir gesehen, daß $E^{(1)} \simeq \mathbb{Z}_p$ ist; damit ist dann, wie man leicht sieht, $E^{(1)}/mE^{(1)} \simeq \mathbb{Z}/p^a\mathbb{Z}$ (für $p^a \parallel m$); andererseits ist $E^{(1)}$ sicher nicht endlich erzeugt: die Gruppe $\mathbb{Z}_p$ ist ja nicht einmal abzählbar!

Es ist also noch etwas zu tun. Die Idee ist die Konstruktion einer "Höhenfunktion"; dies läßt sich ganz allgemein für abelsche Gruppen formulieren:

Satz 4.10. *Sei* G *eine abelsche Gruppe und* $G/2G$ *endlich. Existiert dann eine Abbildung* $h: G \longrightarrow [0,\infty)$ *mit den Eigenschaften*

i) für alle $\kappa > 0$ *ist* $\{P \in G : h(P) < \kappa\}$ *endlich;*

ii) für jedes $P_0 \in G$ *existiert ein* $\lambda_0 > 0$ *derart, daß für alle* $P \in G$ *gilt:*

$$h(P+P_0) \leq 2h(P) + \lambda_0;$$

iii) es existiert ein $\mu \in \mathbb{R}$ *mit* $h(2P) \geq 4h(P) + \mu$ *für alle* $P \in G$,

so ist G *endlich erzeugt.*

Beweis. Nach Voraussetzung ist $n = \#G/2G$ endlich; wir wählen Elemente $Q_1, \ldots, Q_n$ derart, daß G die disjunkte Vereinigung der $Q_i + 2G$ ist, d.h. wir nehmen je ein Element aus jeder Nebenklasse. Zu jedem dieser Elemente Q_i gibt es ein λ_i, sodaß ii) mit $P = -Q_i$ erfüllt ist; sei λ das Maximum dieser endlich vielen λ_i.

Sei nun $P \in G$ gegeben. Dann gibt es ein Q_{i_1} mit $P - Q_{i_1} \in 2G$, d.h. mit $P = 2P_1 + Q_{i_1}$ für ein $P_1 \in G$. So fährt man fort und findet

$$P_1 = 2P_2 + Q_{i_2}, \quad \ldots, \quad P_{\nu-1} = 2P_\nu + Q_{i_\nu}.$$

Für jedes $1 \leq j \leq \nu$ ist dann

$$4h(P_j) \leq h(2P_j) + \mu = h(P_{j-1} - Q_{i_j}) + \mu \leq 2h(P_{j-1}) + \lambda + \mu.$$

Dies schreiben wir in der Form

$$h(P_j) \;\le\; \frac{1}{2}h(P_{j-1}) + \frac{\lambda+\mu}{4} \;=\; \frac{3}{4}h(P_{j-1}) - \frac{1}{4}\big\{h(P_{j-1}) - (\lambda+\mu)\big\}.$$

Solange also $h(P_{j-1}) \ge \lambda + \mu$ ist, gilt $h(P_j) \le \frac{3}{4}h(P_{j-1})$. Irgendwann finden wir daher einen Index m mit $h(P_m) < \lambda + \mu$; folglich läßt sich jedes $P \in G$ schreiben als

$$P = a_1Q_1 + \ldots + a_nQ_n + 2^mR$$

für $a_i \in \mathbb{N}$ und ein (möglicherweise von P abhängiges) $R \in G$ mit $h(R) < \lambda + \mu$. Also wird G von der Menge

$$\{Q_1, \ldots, Q_n\} \;\cup\; \{R \in G : h(R) < \lambda + \mu\}$$

erzeugt, und diese Menge ist wegen i) endlich. □

Bleibt also, auf der Gruppe $E(\mathbb{Q})$ eine Höhe zu definieren. Da man gerne möchte, daß die Menge der Erzeugenden "kleine" Koordinaten hat, wird man versuchen, eine Höhe zu definieren, die die "Kompliziertheit" der Koordinaten mißt. Die Idee, einer rationalen Zahl t die Höhe $h(t) = |t|$ zuzuordnen ist dazu ungeeignet, da er den Punkten $t = 1$ und $t = \frac{10000}{9999}$ in etwa dieselbe Höhe zuordnet, obwohl der zweite Punkt sicher komplizierter ist als der erste.

Definieren wir also $H(P) = \max\{|p|, |q|\}$ für ein $P = (x, y) \in E(\mathbb{Q})$ mit $x = p/q$ und $(p, q) = 1$, und setzen $h(P) = \log H(P)$, sowie $h(\mathcal{O}) = 0$.

Hilfssatz 4.11. *Für alle $\kappa > 0$ ist $\{P \in E(\mathbb{Q}) : h(P) < \kappa\}$ endlich.*

Beweis. Es ist sicherlich die Menge aller $x \in \mathbb{Q}$ mit $H(x) < e^\kappa$ endlich, folglich auch die Menge aller x-Koordinaten von Punkten $P \in E(\mathbb{Q})$ mit $h(P) < \kappa$. Zu jeder x-Koordinate gibt es aber maximal zwei dazugehörige y-Koordinaten. □

Hilfssatz 4.12. *Sei $E : y^2 = x^3 + ax + b$ eine über $\mathbb{Q}$ definierte elliptische Kurve. Dann existiert ein $\mu \in \mathbb{R}$, sodaß $h(2P) - 4h(P) \ge \mu$ für alle $P \in E(\mathbb{Q})$ gilt.*

Beweis. Wir schreiben $P = (x, y)$ und $P^* = 2P = (x^*, y^*)$; dabei nehmen wir an, daß $P^* \ne \mathcal{O}$ ist. Dann gilt nach der Verdoppelungsformel

$$x^* = \frac{x^4 - 2ax^2 - 8bx + a^2}{4(x^3 + ax + b)}.$$

Jetzt setzen wir $x = \frac{p}{q}$ mit $(p,q) = 1$, sowie $x^* = \frac{p^*}{q^*}$. Damit ist

$$p^* = p^4 - 2ap^2q^2 - 8bpq^3 + a^2q^4, \quad q^* = 4q(p^3 + apq^2 + bq^3),$$

und mit $\delta = (p^*, q^*)$ wird $x^* = \widetilde{p}/\widetilde{q}$ mit $p^* = \delta\widetilde{p}$, $q^* = \delta\widetilde{q}$, sowie $(\widetilde{p}, \widetilde{q}) = 1$, und damit $H(P^*) = \max\{|\widetilde{p}|, |\widetilde{q}|\}$.

Mit $D = 4a^3 + 27b^2$ gilt nun

$$4Dq^7 = 4(3p^2q + 4aq^3)p^* - (3p^3 - 5apq^2 - 27bq^3)q^*,$$

d.h. es folgt $|q|^7 \le C_1 \cdot \max\{|p|, |q|\}^3 \cdot \max\{|p^*|, |q^*|\}$. Entsprechend folgt aus

$$\begin{aligned} 4Dp^7 &= f_1p^* + f_2q^* \qquad \text{mit}^1 \\ f_1 &= 4(4a^3 + 27b^2)p^3 - 4a^2bp^2q + 4(3a^4 + 22ab^2)pq^2 + 12(a^3b + 8b^3)q^3, \\ f_2 &= a^2bp^3 + (5a^4 + 32ab^2)p^2q + (26a^3b + 192b^3)pq^2 - 3(a^5 + 8a^2b^2)q^3 \end{aligned}$$

(diese Formeln wurden mit pari kontrolliert und per cut & paste in den Text eingefügt. Herleiten kann man diese Identitäten mit Resultanten) die Abschätzung $|p|^7 \le C_2 \cdot \max\{|p|, |q|\}^3 \cdot \max\{|p^*|, |q^*|\}$, woraus sich dann

$$\max\{|p|, |q|\}^4 \le C_0 \max\{|p^*|, |q^*|\}$$

ergibt. Aus $\delta \mid 4Dp^7$ und $\delta \mid 4Dq^7$ folgt nun $\delta \mid 4D$, d.h. es ist

$$\max\{|p^*|, |q^*|\} \le 4D \max\{|\widetilde{p}|, |\widetilde{q}|\},$$

und insgesamt liefert das

$$\max\{|p|, |q|\}^4 \le C \max\{|\widetilde{p}|, |\widetilde{q}|\}$$

für eine Konstante $C > 0$. Nimmt man den Logarithmus, so heißt das $4h(P) \le h(2P) + \mu$ mit $\mu = \log C$. $\square$

Hilfssatz 4.13. *Sei $E: y^2 = x^3 + ax + b$ eine über $\mathbb{Q}$ definierte elliptische Kurve und $P_0 \in E(\mathbb{Q})$. Dann existiert ein $\lambda_0 > 0$ derart, daß für alle $P \in G$ die Ungleichung $h(P + P_0) \le 2h(P) + \lambda_0$ gilt.*

Beweis. Für $P_0 = \mathcal{O}$ ist die Behauptung trivial (wähle $\lambda_0 = 0$); sei also $P_0 = (x_0, y_0)$. Es genügt dann, die Behauptung für alle P bis auf endlich viele zu beweisen: man braucht nur λ_0 durch $\lambda_0 + \max\{|h(P + P_0) - 2h(P)|\}$

zu ersetzen. Wir werden den Hilfssatz für alle $P \neq \{\mathcal{O}, \pm P_0\}$ beweisen: das erspart uns die Anwendung der Verdoppelungsformel.

Sei also $P_1 = (x_1, y_1) = P + P_0$; mit $m = \frac{y-y_0}{x-x_0}$ ist dann

$$x_1 = m^2 - x - x_0 = \frac{(y-y_0)^2 - (x-x_0)^2(x+x_0)}{(x-x_0)^2}.$$

Ausmultiplizieren liefert im Nenner einen Summanden $y^2 - x^3$, den wir durch $ax + b$ ersetzen; danach haben wir

$$x_1 = \frac{Ay + Bx^2 + Cx + D}{Ex^2 + Fx + G}$$

mit rationalen Zahlen $A, \ldots, G$, die nur von a, b, x_0 und y_0 abhängen. Außerdem dürfen wir annehmen, daß diese Konstanten alle ganz sind (ansonsten multiplizieren wir mit dem Hauptnenner durch – dieser hängt ebenfalls nur von obigen Größen ab). Setzt man jetzt $x = m/e^2$ und $y = n/e^3$ wie in Hilfssatz 4.7, so folgt

$$x_1 = \frac{Ane + Bm^2 + Cme^2 + De^4}{Em^2 + Fme^2 + Ge^4}.$$

Damit ist sicher

$$H(P_1) \le \max\,\{|Ane + Bm^2 + Cme^2 + De^4|, |Em^2 + Fme^2 + Ge^4|\},$$

denn Kürzen macht Zähler und Nenner höchstens kleiner. Jetzt behaupten wir, daß mit $K = \sqrt{1 + |a| + |b|}$

$$e \le H(P)^{1/2}, \quad |m| \le H(P), \quad \text{und } |n| \le K \cdot H(P)^{3/2} \tag{4.9}$$

gilt. Damit wird dann

$$\begin{aligned} |Ane + Bm^2 + Cme^2 + De^4| &\le |Ane| + |Bm^2| + |Cme^2| + |De^4| \\ &\le (|AK| + |B| + |C| + |D|)\, H(P)^2, \quad \text{sowie} \\ |Em^2 + Fme^2 + Ge^4| &\le |Em^2| + |Fme^2| + |Ge^4| \\ &\le (|E| + |F| + |G|)\, H(P)^2. \end{aligned}$$

Also folgt

$$H(P + P_0) \le \max\,\{(|AK| + |B| + |C| + |D|), (|E| + |F| + |G|)\}\, H(P)^2,$$

und nimmt man jetzt den Logarithmus, so folgt die Behauptung.

Zu zeigen bleibt noch (4.9). Die beiden ersten Ungleichungen folgen direkt aus der Definition von $H(P)$. Zum Beweis der dritten schreiben wir

$$n^2 = m^3 + ae^4m + be^6,$$

nehmen den Betrag, und wenden die Dreiecksungleichung an. Dann wird

$$|n|^2 \leq H(P)^3 + |a|H(P)^3 + |b|H(P)^3 \leq KH(P)^3,$$

und Wurzelziehen liefert das gewünschte Ergebnis. $\square$

Die oben eingeführte Höhenfunktion heißt auch die *naive*, die *logarithmische* oder auch die *Weil-Höhe*. Sie hat den Vorteil, leicht berechenbar zu sein. Daneben gibt es den Begriff der *kanonischen* oder auch *Néron-Tate Höhe*, die wie folgt definiert ist: aus der Ungleichung $|h(2P) - 4h(P)| \leq \mu$ (wir haben oben bereits eine Richtung gezeigt; die andere ist einfacher) folgt die Existenz des Grenzwerts

$$\widehat{h}(P) := \lim_{n\to\infty} \frac{h(2^nP)}{4^n}.$$

Die so definierte Funktion hat die Eigenschaften

i) $\widehat{h}(P) = 0$ genau dann, wenn $P \in E(\mathbb{Q})_{\text{tors}}$ ist;

ii) $|\widehat{h}(P) - h(P)| < \mu$ für ein nur von E abhängiges $\mu \in \mathbb{R}$;

iii) $\widehat{h}(mP) = m^2\widehat{h}(P)$ für alle $m \in \mathbb{Z}$;

iv) es ist $\widehat{h}(P+Q) + \widehat{h}(P-Q) = 2\widehat{h}(P) + 2\widehat{h}(Q)$.

Die letzte Eigenschaft erlaubt es, durch

$$\langle\ ,\ \rangle : E(\mathbb{Q}) \times E(\mathbb{Q}) \longrightarrow \mathbb{R} : \langle P, Q\rangle := \widehat{h}(P+Q) - \widehat{h}(P) - \widehat{h}(Q)$$

eine Bilinearform zu definieren. Dies ist ein wichtiges Hilfsmittel, um zu entscheiden, ob vorgegebene Punkte $P_1, \ldots, P_n \in E(\mathbb{Q})$ linear abhängig sind (solche Punkte heißen linear unabhängig, wenn aus $\sum a_iP_i = \mathcal{O}$ immer $a_i = 0$ für alle i folgt): man bildet einfach die Matrix $R = (\langle P_i, P_j\rangle)_{1\leq i,j\leq n}$ und berechnet die reelle Zahl $\det R$; genau dann existieren $a_i \in \mathbb{Z}$, nicht alle gleich 0, mit $\mathcal{O} = \sum a_iP_i$, wenn $\det R = 0$ ist.

Bilden die Punkte $P_1, \ldots, P_n$ eine Basis des freien Anteils von $E(\mathbb{Q})$ (d.h. ist $E(\mathbb{Q}) = E(\mathbb{Q})_{\text{tors}} \oplus \langle P_1\rangle \oplus \ldots \oplus \langle P_n\rangle$), so heißt $R_{E/\mathbb{Q}} = \det R$ der Regulator von E, und dieser hängt nicht von der Wahl der Basis ab.

4.4 Isomorphismen, Isogenien, und Twists

Es ist wohl an der Zeit, einige Begriffe zu klären, die bisher nur ganz am Rande erwähnt wurden. Da ist zum einen der Begriff der Isomorphie, der es erlaubt, gewisse Weierstraßkurven im wesentlichen zu identifizieren. Wenn dieser Begriff seinem Namen Ehre machen soll, dann sollten isomorphe Kurven auch isomorphe Gruppenstruktur tragen – aber bereits hier wird klar, daß dabei der Grundkörper eine Rolle spielt. Sei also

$$E : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$$

eine über einem Körper K definierte elliptische Kurve in langer Weierstraßform. Eine Variablentransformation

$$x = u^2x' + r, \quad y = u^3y' + su^2x' + t \tag{4.10}$$

mit $r, s, t \in K$ und $u \in K^\times$ heißt K-*zulässig*, und führt E über in

$$E : y'^2 + a_1'x'y' + a_3'y' = x'^3 + a_2'x'^2 + a_4'x' + a_6',$$

wobei die neuen Koeffizienten gegeben sind durch

$$\boxed{\begin{array}{rcl}
ua_1' & = & a_1 + 2s, \\
u^2a_2' & = & a_2 - sa_1 + 3r - s^2, \\
u^3a_3' & = & a_3 + ra_1 + 2t, \\
u^4a_4' & = & a_4 - sa_3 + 2ra_2 - (t + rs)a_1 + 3r^2 - 2st, \\
u^6a_6' & = & a_6 + ra_4 + r^2a_2 + r^3 - ta_3 - t^2 - rta_1, \\
 & & \\
a^2b_2' & = & b_2 + 12r, \\
u^4b_4' & = & b_4 + rb_2 + 6r^2, \\
u^6b_6' & = & b_6 + 2rb_4 + r^2b_2 + 4r^3, \\
u^8b_8' & = & b_8 + 3rb_6 + 3r^2b_4 + r^3b_2 + 3r^4, \\
 & & \\
u^4c_4' & = & c_4, \\
u^6c_6' & = & c_6, \\
u^{12}\Delta' & = & \Delta, \\
j' & = & j.
\end{array}}$$

Wir nennen zwei Kurven E und E' *isomorph über* K, wenn es eine K-zulässige Transformation gibt, die E in E' überführt. Man überzeugt sich leicht davon, daß die eine Äquivalenzrelation auf der Menge der in langer Weierstraßform gegebenen und über K definierten elliptischen Kurven sind.

Man sieht auch leicht, daß Kurven in kurzer Weierstraßform (über einem Körper der Charakteristik $\neq 2, 3$) genau dann wieder in solche übergehen, wenn $r = s = t = 0$ ist.

Da sich die absolute Invariante j unter Isomorphie nicht ändert, können Kurven mit verschiedener j-Invariante nicht isomorph sein. Umgekehrt gilt

Satz 4.14. *Sind E und E' zwei über K definierte elliptische Kurven, so ist $j_E = j_{E'}$ genau dann, wenn E und E' über dem algebraischen Abschluß von K isomorph sind, d.h. wenn es eine zulässige Transformation (4.10) mit Koeffizienten $u, r, s, t \in \overline{K}$ gibt.*

Diesen Satz werden wir weder beweisen, noch benutzen. Zwei Kurven E und E', die beide über K definiert sind, aber erst in einer Körpererweiterung von K isomorph werden, heißen *Twists*. Ist z.B. $d \in \mathbb{Z}$ quadratfrei und $E : y^2 = x^3 + ax + b$ eine elliptische Kurve über $\mathbb{Q}$, so ist $E_d : dy^2 = x^3 + ax + b$ (in Weierstraßform: $y^2 = x^3 + ad^2x + bd^3$) ebenfalls eine elliptische Kurve über $\mathbb{Q}$, und zwar ist sie offensichtlich über $\mathbb{Q}(\sqrt{d})$ zu E isomorph: man braucht nur $y' = \sqrt{d}y$ und $x' = x$ zu substituieren. Wegen $\Delta(E_d) = d^6\Delta(E)$ können sie im Falle $d \neq \pm 1$ über $\mathbb{Q}$ aber nicht isomorph sein.

Da Isomorphismen invertierbare lineare Abbildungen der affinen Ebene auf sich sind, induzieren sie Gruppenisomorphismen $E(K) \longrightarrow E'(K)$; insbesondere haben isomorphe elliptische Kurven dieselbe Torsionsgruppe und denselben Rang. Für Isogenien gilt das nicht: zwei Kurven E/K und E'/K heißen *isogen*, wenn es eine nichtkonstante über K definierte rationale (d.h. gegeben durch einen Quotienten zweier Polynome mit Koeffizienten aus K) Abbildung $E \longrightarrow E'$ gibt, die $\mathcal{O}$ in $\mathcal{O}'$ überführt. Man kann zeigen, daß Isogenien *automatisch* (!) Gruppenhomomorphismen $E(K) \longrightarrow E'(K)$ induzieren (wir haben das bei unseren 2-Isogenien nachrechnen müssen), daß sie aber andererseits auch *immer* einen nichttrivialen Kern haben (2-Isogenien verschlucken notwendig einen Punkt der Ordnung 2).

Weiter kann man zeigen, daß es zu jeder Isogenie $\phi : E \longrightarrow E'$ eine dazu duale Isogenie $\psi : E' \longrightarrow E$ gibt, und daß die Komposition der beiden Isogenien jeweils Multiplikation mit einer Zahl $m \in \mathbb{Z}$ auf E bzw. E' ist. Diese Zahl m nennt man den Grad der Isogenie.

Die Isogenien $E \longrightarrow E'$ bilden offenbar eine additive Gruppe; durch Komposition kann man die Isogenien $E \longrightarrow E$ von E auf sich zu einem Ring machen, dem Endomorphismenring $\mathrm{End}(E)$. Da die Multiplikation mit $m \in \mathbb{Z}$ eine solche Isogenie ist, können wir $\mathbb{Z}$ immer als Unterring von $\mathrm{End}(E)$ auffassen (würde man bei der Definition von Isogenien auf die Bedingung $\mathcal{O} \longmapsto \mathcal{O}$ verzichten, so wäre auch die Addition eines festen Punktes $P \in E(\mathbb{Q})$ eine

Isogenie). Ist E eine elliptische Kurve über $\mathbb{Q}$, so gibt es zwei Möglichkeiten: entweder ist $\mathrm{End}(E) \simeq \mathbb{Z}$ (d.h. es gibt keine anderen Isogenien $E \longrightarrow E$ als die Multiplikation mit einem $m \in \mathbb{Z}$), oder $\mathrm{End}(E)$ ist echt größer. Im letzteren Falle kann man zeigen, daß $\mathrm{End}(E) \simeq \mathbb{Z} \oplus \tau\mathbb{Z}$ für eine komplexe Zahl τ ist (genauer ist sogar $\tau = \frac{1}{2}(r + s\sqrt{-d}\,)$ mit gewissen $a, b, d \in \mathbb{N}_0$), und man sagt, E habe komplexe Multiplikation.

Kapitel 5
Die Hasse-Schranke

Daß elliptische Kurven über dem Körper $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ durchaus interessant sind, haben wir bereits festgestellt. In diesem Kapitel wollen wir Nichttriviales zur Anzahl der Punkte auf elliptischen Kurven über $\mathbb{F}_q$ machen; dabei schreiben wir q statt p, weil alle unsere Ergebnisse auch über beliebigen endlichen Körpern mit $q = p^f$ Elementen gelten. Da jeder endliche Punkt sich in der Form (x, y) mit $x, y \in \mathbb{F}_q$ schreiben läß, gibt es auf einer elliptischen Kurve über $\mathbb{F}_q$ jedenfalls höchstens q^2+1 Punkte. Diese triviale Abschätzung ist natürlich ziemlich schlecht. Nehmen wir einmal an, daß $p \geq 5$ ist, also unser endlicher Körper Charakteristik $\neq 2, 3$ hat. Dann können wir die elliptische Kurve in kurzer Weierstraßform schreiben: $y^2 = x^3 + ax + b$ mit $a, b \in \mathbb{F}_q$. Für jeden der q möglichen Werte von x gibt es, da wir uns in einem Körper befinden, höchstens zwei Werte von y, für die (x, y) auf E liegt; dies liefert die weitaus bessere Abschätzung $\#E(\mathbb{F}_q) \leq 2q + 1$. Nun wird aber nicht jedes $x^3 + ax + b$ ein Quadrat in $\mathbb{F}_q$ sein; wären diese Werte zufällig verteilt, dürften wir annehmen, daß in etwa die Hälfte aller x-Werte zu einem Quadrat $x^3 + ax + b$ führt; dies würde bedeuten, daß es in etwa $q + 1$ Punkte auf $E(\mathbb{F}_q)$ gibt. Der Satz von Hasse besagt nun, daß diese Abschätzung gar nicht schlecht ist:

Satz 5.1. *(Hasse) Sei $E : y^2 = x^3 + ax + b$ eine über dem endlichen Körper $\mathbb{F}_q$ definierte elliptische Kurve. Dann gilt $|\#E(\mathbb{F}_q) - (q + 1)| \leq 2\sqrt{q}$.*

Daß dies ein bedeutendes Resultat ist, sieht man dem Satz nicht an. Um dem abzuhelfen, muß man etwas weiter ausholen und einige Sachen über Zetafunktionen und die Riemannsche Vermutung erzählen.

5.1 Die Riemannsche Zetafunktion

Bereits Euler hat die reelle Funktion

$$\zeta(s) = \sum_{n \in \mathbb{N}} n^{-s} \tag{5.1}$$

untersucht; das Integralkriterium zeigt sofort, daß $\zeta(s)$ für alle $s > 1$ konvergiert, und die Divergenz der harmonischen Reihe zeigt, daß $\zeta(s)$ an der Stelle $s = 1$ einen Pol besitzt. Euler's große Leistung war die Erkenntnis, daß

$$\zeta(s) = \prod_{p \in \mathbb{P}} \frac{1}{1 - p^{-s}} \tag{5.2}$$

sich als Produkt über alle Primzahlen von relativ einfach gebauten Funktionen schreiben läßt; tatsächlich ist die Existenz von (5.2) äquivalent zum Satz von der eindeutigen Primfaktorzerlegung in $\mathbb{Z}$. Zum Beweis betrachtet man das Produkt über alle $p \le N$ und entwickelt dann den Bruch via $\frac{1}{1-x} = 1 + x + x^2 + x^3 + \ldots$ in eine unendliche Reihe. Man bekommt dann

$$\prod_{p \le N} \frac{1}{1 - p^{-s}} \;=\; \prod_{p \le N} \Big(1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \ldots\Big) \;=\; \sum_{n \in S_N} \frac{1}{n^s},$$

wobei S_N aus allen Zahlen besteht, deren Primfaktoren $\le N$ sind. Läßt man also $N \longmapsto \infty$ gehen, geht die linke Seite gegen das Produkt in (5.2), die rechte (Cauchy-Kriterium!) gegen die Summe in (5.1).

Euler benutzte (5.2), um folgenden Beweis von der Unendlichkeit der Primzahlen zu geben: gäbe es nur endlich viele Primzahlen, wäre die rechte Seite von (5.2) für $s = 1$ konvergent, die linke, durch (5.1) definierte, aber nicht: Widerspruch. Genau diesen Ansatz hat Dirichlet später wieder ausgegraben, um seinen berühmten Satz von der Existenz unendlich vieler Primzahlen in jeder arithmetischen Progression $ax + b$ mit $\operatorname{ggT}(a, b) = 1$ zu beweisen.

Daß $\zeta(s)$ heute nach Riemann benannt ist, liegt daran, daß Riemann damit begonnen hat, $\zeta(s)$ als Funktion komplexer Zahlen aufzufassen. Er konnte zeigen, daß sich $\zeta(s)$ auf die ganze komplexe Ebene fortsetzen läßt, und zwar als analytische Funktion außerhalb ihres einzigen Pols $s = 1$. Weiter hat er eine Funktionalgleichung bewiesen, die $\zeta(s)$ und $\zeta(1-s)$ zueinander in Beziehung setzt: zu deren Formulierung setzt man gewöhnlich $\xi(s) = \pi^{-s/2}\Gamma(\frac{s}{2})\zeta(s)$ setzt: dann wird nämlich einfach $\xi(s) = \xi(1-s)$. Schließlich hat Riemann einen Beweis des Primzahlsatzes skizziert, wonach für die Funktion $\pi(x) = \#\{p \le x : p \text{ prim}\}$ die Beziehung $\pi(x) \sim \frac{x}{\ln x}$ gilt ($f \sim g$ soll hier bedeuten, daß $f(x)/g(x)$ für $x \longmapsto \infty$ gegen 1 konvergiert). Dabei hat er aber deutliche Lücken gelassen, die erst 1896 unabhängig von Hadamard und de la Vallée-Poussin geschlossen werden konnten. Außerdem hat er vermutet, daß alle Nullstellen von $\zeta(s)$ im kritischen Streifen $0 \le \operatorname{Re} s \le 1$ sogar auf der kritischen Geraden $\operatorname{Re} s = \frac{1}{2}$ liegen. Falls diese Vermutung richtig ist, könnte man für die Differenz $\pi(x) - \frac{x}{\ln x}$ sehr gute Schranken angeben; dies hätte auch für viele Algorithmen der Zahlentheorie Konsequenzen.

Die Riemannsche Vermutung ist bis heute offen, wenn sie auch für die ersten 10^9 Nullstellen (nach wachsendem Imaginärteil geordnet) numerisch verifiziert worden ist. Daß man seit ein paar Jahren wenigstens weiß, in welcher Richtung man einen Beweis zu suchen hat, liegt vor allem daran, daß es zu diesem ganzen Fragenkomplex einen ganz analogen gibt, der von der Zetafunkion elliptischer Kurven ausgeht.

5.2 Die Zetafunktion elliptischer Kurven

Die Zetafunktion elliptischer Kurven wurde ursprünglich ganz analog zur Riemannschen definiert, allerdings in einem anderen Zusammenhang: statt der elliptischen Kurve $y^2 = x^3 + ax + b$ über dem endlichen Körper $K = \mathbb{F}_p$ betrachtet man den rationalen Funktionenkörper $F = \mathbb{F}_p(x)$ und darüber die quadratische Erweiterung $E = F(\sqrt{x^3 + ax + b})$. Solchen Erweiterungen (genauer: dem Ganzheitsring in E) kann man eine Zetafunktion zuordnen, die ganz genauso aussieht wie $\zeta(s)$. Bei deren Studium hat sich aber herausgestellt, daß sie Periode $\frac{2\pi i}{\log p}$ hat, also nur von $t = p^{-s}$ abhängt, und wenn man diese Substitution macht, landet man bei der folgenden Definition: für jede über $k = \mathbb{F}_p$ definierte elliptische Kurve sei

$$Z_E(t) = \exp\left(\sum_{f=1}^{\infty} \#E(\mathbb{F}_{p^f})\frac{t^f}{f}\right). \tag{5.3}$$

Die so definierte Zetafunktion nennt man auch die Kongruenzzetafunktion, weil sie die Lösungsanzahlen der Kongruenz $y^2 \equiv x^3 + ax + b \bmod p$ kodiert. Artin und später etwas allgemeiner F.K. Schmidt konnten zeigen, daß sich die Zetafunktion einer elliptischen Kurve immer in der Form

$$Z_E(t) = \frac{P(t)}{(1-t)(1-pt)} \tag{5.4}$$

schreiben läßt, wobei $P(t) = 1 - a_p t + pt^2$ ist und $a_p = p + 1 - \#E(\mathbb{F}_p)$ gilt! Mit anderen Worten: die Zetafunktion $Z_E(t)$ kodiert die Anzahl der Punkte auf den Kurven $\overline{E}(\mathbb{F}_p)$, wo $\overline{E}$ die Reduktion modulo p einer über $\mathbb{Q}$ definierten elliptischen Kurve E bezeichnet.

Hier sind zwei Wunder passsiert: zum einen ist die Zetafunktion ein Quotient von Polynomen, zum andern taucht im Ergebnis nur a_p auf, obwohl man *alle* a_{p^f} hineingesteckt hat! Nun kann man aus der Kenntnis von $P(t)$ durch Logarithmieren und Entwickeln in Potenzreihen wieder alle $\#E(\mathbb{F}_{p^f})$ berechnen, mit anderen Worten: die Anzahl der Punkte auf $E(\mathbb{F}_{p^f})$ hängen nur von $E(\mathbb{F}_p)$ ab. Daß die Zetafunktion auch noch einer Funktionalgleichung genügt, kann kaum mehr überraschen. Man rechnet einfach nach, daß $Z_E(\frac{1}{pt}) = Z_E(t)$ ist; setzt man $\zeta_E(s) := Z_E(p^{-s})$, so schreibt sich die Funktionalgleichung in der Form $\zeta_E(1-s) = \zeta_E(s)$. Die Ähnlichkeit mit der Riemannschen ζ-Funktion geht noch weiter: dazu rechnen wir einmal die Nullstellen von $\zeta(s)$ aus. Diese sind genau die Nullstellen von $P(p^{-s})$; schreiben wir also $P(t) = (1-\alpha_1 t)(1-\alpha_2 t)$, dann ist $P_1(p^{-s}) = 0$ genau dann, wenn $p^s = \alpha_j$ für $j = 1$ oder $j = 2$ ist. Nun ist $|p^s| = p^{\operatorname{Re} s}$ (allgemein ist $|e^{a+ib}| = |e^a| \cdot |e^{ib}|$ und $|e^{ib}| = 1$ wegen der Eulerschen Formel $e^{ib} = cos b + i \sin b$ und der Identität $\cos^2 b + \sin^2 b = 1$, also $|e^z| = e^{\operatorname{Re} z}$), also $|\alpha_j| = \sqrt{p}$ genau dann, wenn $\operatorname{Re} s = \frac{1}{2}$ ist, mit anderen Worten: die Riemannsche Vermutung für $\zeta_E(s)$ ist äquivalent zur Aussage $|\alpha_j| = \sqrt{p}$, wo α_j die beiden inversen Wurzeln von $P(t)$ sind. Wegen $P(t) = 1 - a_p t + pt^2$ impliziert dies aber $|a_p| = |\alpha_1 + \alpha_2| \le |\alpha_1| + |\alpha_2| \le 2\sqrt{p}$, also die Hasse-Schranke (die Umklehrung gilt übrigens auch: die Hasse-Schranke impliziert die Riemannsche Vermutung für $\zeta_E(s)$).

Damit die Zahlen a_p wirklich von E und nicht von der gewählten Weierstraßgleichung abhängen, müssen wir festlegen, was wir unter "der" Reduktion modulo p einer elliptischen Kurve $E/\mathbb{Q}$ verstehen: denn eine Kurve $E : y^2 = x^3 + 1$ (mit $\Delta = -2^4 3^3$) kann durch die zulässige Transformation $y = 5^{-3}Y$, $x = 5^{-2}X$ in $E' : Y^2 = X^3 + 5^6$ (mit $\Delta' = -2^4 3^3 5^{12}$) verwandelt werden. Über $\mathbb{Q}$ sind diese Kurven isomorph, über $\mathbb{F}_5$ dagegen nicht mehr:

die Reduktion von E modulo 5 ist die elliptische Kurve $\overline{E} : y^2 = x^3 + 1$, diejenige von E' dagegen die singuläre Kurve $\overline{E}' : Y^2 = X^3$. Insbesondere hängt a_5 in diesem Fall von der Wahl des globalen Modells ab. Zur Klärung dieses Problems benötigen wir den Begriff der minimalen Weierstraßgleichung.

Sei E eine über $\mathbb{Q}$ definierte elliptische Kurve in langer Weierstraßform mit ganzen Koeffizienten. Die Weierstraßgleichung heißt p-minimal, wenn die p-adische Bewertung $|\Delta|_p$ ihrer Diskriminante durch eine zulässige Transformation nicht größer gemacht werden kann (d.h. wenn die in Δ aufgehende p-Potenz nicht verkleinert werden kann). Man kann zeigen, daß die Reduktion modulo p einer p-minimalen Weierstraßgleichung bis auf über $\mathbb{F}_p$ zulässige Transformationen eindeutig bestimmt sind. Tatsächlich gibt es für jede über $\mathbb{Q}$ definierte elliptische Kurve E eine Weierstraßgleichung, die für alle p gleichzeitig p-minimal ist (Néron); dieses Ergebnis sieht stärker aus, als es ist: für fast alle p, nämlich mindestens für die $p \nmid \Delta$, ist die Gleichung von vornherein p-minimal. Da zulässige Transformationen die Diskriminante Δ nur um zwölfte Potenzen abändern können, ist jede Weierstraßgleichung mit ganzen Koeffizienten, deren Diskriminante Δ durch keine zwölfte Potenz teilbar ist, notwendig minimal (die Umkehrung gilt nicht: beim Wegtransformieren einer zwölften Potenz könnten die Koeffizienten der Weierstraßgleichung Nenner bekommen, deren Elimination die zwölfte Potenz wieder hineinbringt). Zwei solche minimale Weierstraßgleichungen einer elliptischen Kurve gehen durch eine zulässige Transformation mit $u = \pm 1$ und $r, s, t \in \mathbb{Z}$ auseinander hervor.

Damit sind die Zahlen $a_p = p + 1 - \#E(\mathbb{F}_p)$ durch die elliptische Kurve festgelegt und hängen nicht von der Auswahl der Weierstraßgleichung ab. In dem relativ uninteressanten Fall, daß die Reduktion einer elliptischen Kurve singulär ist, können wir die dazugehörigen ζ-Funktionen leicht berechnen. In Kapitel 2 haben wir gesehen, daß singuläre Weierstraßkurven über $\mathbb{F}_p$ auf die Form $E : y^2 = x^2(x+a)$ gebracht werden können, und daß E_{ns} genau $p - (\frac{a}{p})$ Punkte besitzt. Der dortige Beweis zeigt allgemein, daß für beliebige $q = p^f$ die Formel $\#E_{ns}(\mathbb{F}_q) = p^f - (\frac{a}{p})^f$ gilt (beachte, daß im Falle $(\frac{a}{p}) = -1$ die Wurzel $\sqrt{a} \in \mathbb{F}_{p^2}$ und damit in allen $\mathbb{F}_q$ mit geradem s liegt). Insbesondere ist also $\#E(\mathbb{F}_q) = q + 1 - (\frac{a}{p})^f$, und wenn wir die Definition der Zetafunktion für nichtsinguläre Kurven ausnahmsweise für singuläre Kurven übernehmen, so finden wir

$$\begin{aligned}\log Z_E(t) &= \sum_{f=1}^{\infty} \#E(\mathbb{F}_{p^f})\frac{t^f}{f} = \sum_{f=1}^{\infty}(p^f + 1 - (\tfrac{a}{p})^f)\frac{t^f}{f} \\ &= \sum_{f=1}^{\infty}\frac{(pt)^f}{f} + \sum_{f=1}^{\infty}\frac{t^f}{f} - \sum_{f=1}^{\infty}\frac{((a/p)t)^f}{f} \\ &= -\log(1-pt) - \log(1-t) + \log(1-(\tfrac{a}{p})t),\end{aligned}$$

d.h. $Z_E(t)$ hat die Form (5.4) mit $P(t) = 1 - a_p t$ und $a_p = p + 1 - \#E(\mathbb{F}_p) = (\frac{a}{p})$.

Zu gegebenem E definieren wir jetzt einen Eulerfaktor $L_p(t)$ für jede Primzahl p durch

$$L_p(t) = \begin{cases} \frac{1}{1-a_p t + pt^2}, & \text{falls } p \nmid \Delta, \\ \frac{1}{1-a_p t}, & \text{falls } p \mid \Delta. \end{cases}$$

Man beachte, daß der Eulerfaktor gleich dem Inversen des Zählers der Zetafunktion für $E(\mathbb{F}_p)$ ist. Unter Benutzung von $|a_p| \le p$ (das folgt aus der trivialen Abschätzung $|\#E(\mathbb{F}_p)| \le 2p + 1$) kann man dann ohne weiteres zeigen, daß das Produkt

$$L(s,E) = \prod_{p\in\mathbb{N}} L_p(p^{-s}) = \prod_{p\mid\Delta}\frac{1}{1-a_p p^{-s}} \cdot \prod_{p\nmid\Delta}\frac{1}{1-a_p p^{-s} + p^{1-2s}} \tag{5.5}$$

für alle $s \in \mathbb{C}$ mit $\operatorname{Re} s > 2$ absolut konvergiert. Diese Funktion nennt man die L-Funktion der elliptischen Kurve (Hasse-Weil). Im nächsten Abschnitt werden wir zeigen, daß sogar $|a_p| \le 2\sqrt{p}$ ist; damit folgt dann, daß die L-Funktion sogar für alle s mit $\operatorname{Re} s > \frac{3}{2}$ absolut konvergiert.

Hasse hat nun in Analogie zu den damals bekannten ζ-Funktionen (Riemann, Dedekind, Artin) die folgende Vermutung geäußert:

Vermutung. Die L-Reihe $L(s,E)$ einer über $\mathbb{Q}$ definierten elliptischen Kurve läßt sich analytisch auf die ganze komplexe Ebene fortsetzen; weiter existiert ein $N \in \mathbb{N}$ derart, daß

$$\Lambda(s,E) = N^{s/2}(2\pi)^{-s}\Gamma(s)L(s,E)$$

der Funktionalgleichung $\Lambda(s-2,E) = \pm\Lambda(s,E)$ genügt.

Die Zahl N nennt man den analytischen Führer: dieser ist ein Produkt von Primzahlen, die auch in Δ aufgehen.

4.9.1999

Ist die elliptische Kurve E gegeben, so kann man die dazugehörige L-Funktion an jedem $z \in \mathbb{C}$ numerisch auswerten. Dies haben zuerst Birch und Swinnerton-Dyer gemacht, und aus ihren Berechnungen haben sie folgende Vermutung herausdestilliert:

> Ist r der $\mathbb{Z}$-Rang der elliptischen Kurve E, so hat die dazugehörige L-Funktion an der Stelle $s = 1$ eine r-fache Nullstelle und umgekehrt.

Das ist eine starke Vermutung: sie besagt nicht mehr und nicht weniger, als daß man aus den lokalen Daten a_p (mehr hat man ja in die Definition von $L(s, E)$ nicht reingesteckt) den Rang der Kurve über $\mathbb{Q}$ bestimmen kann! Weitergehende Rechnungen haben zu der folgenden, viel genaueren Vermutung geführt:

Vermutung. Sei E eine über $\mathbb{Q}$ definierte elliptische Kurve, für welche die Hassesche Vermutung der analytischen Fortsetzbarkeit richtig ist. Ist dann r der $\mathbb{Z}$-Rang von E, so gilt

$$\lim_{s\to 1} \frac{L(s,E)}{(s-1)^r} = \frac{c_\infty \prod_p c_p \# Ш(E/\mathbb{Q}) R_{E/\mathbb{Q}}}{(\# E(\mathbb{Q})_{\mathrm{tors}})^2}.$$

Hierbei ist c_∞ eine reelle Periode von E, c_p die p-te Tamagawa-Zahl, $Ш(E/\mathbb{Q})$ die Tate-Shafarevic-Gruppe, und $R_{E/\mathbb{Q}}$ der Regulator von E.

Die hier erwähnte reelle Periode c_∞ ist ein bestimmtes Integral: wir haben im ersten Kapitel erklärt, wie man elliptische Funktionen aus einem Gitter in $\mathbb{C}$ gewinnt; dieses Verfahren kann man umkehren, d.h. man kann jeder elliptischen Kurve über $\mathbb{Q}$ ein entsprechendes Gitter zuordnen, und zwar so, daß eine der beiden Perioden reell ist. Die Tamagawa-Zahlen c_p waren definiert als der Index $c_p = (E : E^{(0)})$, wobei E als elliptische Kurve über $\mathbb{Q}_p$ aufgefaßt wird; ist $p \nmid \Delta$, so folgt aus der Definition von $E^{(0)}$ sofort $c_p = 1$, d.h. das Produkt über alle Primzahlen p macht Sinn. Von der Tate-Shafarevic-Gruppe $Ш(E/\mathbb{Q})$ haben wir nur einen kleinen Teil definiert, nämlich die zu einer 2-Isogenie $\beta : \overline{E} \longrightarrow E$ gehörige Gruppe $Ш(E/\mathbb{Q})[\beta]$, und zwar als die Faktorgruppe derjenigen $b_1 \cdot \mathbb{Q}^{\times\,2}$, für welche die Gleichung

$$N^2 = b_1 M^4 + aM^2e^2 + b_2e^4$$

über jedem $\mathbb{Q}_p$ lösbar ist, modulo der Gruppe aller $b_1 \cdot \mathbb{Q}^{\times\,2}$, für welche diese Gleichung eine rationale Lösung besitzt.

Lange Zeit kannte man keine einzige elliptische Kurve, von der man zeigen konnte, daß ihre Tate-Shafarevic-Gruppe endlich ist; damals stellte die Vermutung von Birch und Swinnerton-Dyer also eine Beziehung her zwischen dem Verhalten einer Funktion an einer Stelle, wo sie nicht definiert ist, und der Ordnung einer Gruppe, von der man nicht weiss, ob sie endlich ist.

Heute ist man schlauer: die Fortsetzbarkeit der L-Funktion war für modulare Kurven schon lange bekannt, und seit Wiles weiß man, daß die meisten elliptischen Kurven über $\mathbb{Q}$ modular sind. Weiter hat es eine von Kolyvagin eingeführte und von Rubin verfeinerte Technik erlaubt, für einige Kurven die Endlichkeit von $Ш(E/\mathbb{Q})$ zu zeigen. Weiter gilt für modulare elliptische Kurven folgendes:

1. $L(1,E) \neq 0 \Longrightarrow r = 0$;
2. $L(1,E) = 0$ und $L'(1,E) \neq 0 \Longrightarrow r = 1$.

5.3 Manins Beweis

Sei $\mathbb{F}_q$ ein endlicher Körper mit $q = p^f$ Elementen und $p \geq 5$. Manins Beweis beginnt mit elliptischen Kurven über $\mathbb{F}_q[t]$, dem Polynomring in der Variablen t über $\mathbb{F}_q$. Dieser Ring besteht also aus allen Polynomen $a_0 + a_1t + \ldots + a_nt^n$ mit Koeffizienten aus $\mathbb{F}_q$. Da $\mathbb{F}_q[t]$ nullteilerfrei ist, kann man den Quotientenkörper $K = \mathbb{F}_q(t)$ einführen, der aus allen Ausdrücken der Form

$$\frac{a_0 + a_1t + \ldots + a_nt^n}{b_0 + b_1t + \ldots + b_mt^m}$$

mit $a_i, b_j \in \mathbb{F}_q$ und $b_m \neq 0$ besteht.

Der elliptischen Kurve

$$E : Y^2 = X^3 + aX + b \tag{5.6}$$

ordnen wir jetzt den quadratischen Twist

$$E_\lambda : \lambda Y^2 = X^3 + aX + b \tag{5.7}$$

zu, wobei $\lambda = \lambda(t) = t^3 + at + b \in \mathbb{F}_q[t]$ ist. Damit ist E_λ eine über K definierte elliptische Kurve. Die Additionsformeln für E_λ sehen folgendermaßen aus: seien $P = (x,y) \neq \mathcal{O}$, $P_j = (x_j, y_j)$ Punkte auf E_λ; dann ist

$$x(P_1 + P_2) = \lambda\Big(\frac{y_1 - y_2}{x_1 - x_2}\Big)^2 - (x_1 + x_2), \tag{5.8}$$

sowie, falls $y \neq 0$ ist,

$$x(2P) = \frac{(3x^2 + a)^2}{4(x^3 + ax + b)} - 2x. \tag{5.9}$$

Die Punkte $(t, 1)$ und $(t, -1)$ liegen offenbar auf $E_\lambda(K)$. Aber auch der Punkt $P_0 = (t^q, (t^3 + at + b)^{(q-1)/2})$ liegt auf $E_\lambda(K)$: es ist nämlich $(t^q)^3 + at^q + b = (t^3 + at + b)^q$ nach dem üblichen Trick des "Anfängerpotenzierens" in endlichen Körpern.

Wir definieren jetzt eine Folge von Punkten

$$P_n = P_0 + n(t, 1), \quad n \in \mathbb{Z} \tag{5.10}$$

auf $E_\lambda(K)$. Die erste Behauptung ist nun

Hilfssatz 5.2. *Ist $P_n = (x_n, y_n) \neq \mathcal{O}$, so gilt $x_n \neq 0$. Schreibt man $x_n = f_n/g_n$ mit $f_n, g_n \in \mathbb{F}_q[t]$, so ist $\deg f_n > \deg g_n$.*

Da $\mathbb{F}_q[t]$ ein euklidischer Ring ist, dürfen wir annehmen, daß f_n und g_n teilerfremd sind. Damit können wir eine Funktion $d : \mathbb{Z} \longrightarrow \mathbb{N}_0$ definieren durch

$$d_n = \begin{cases} 0 & \text{falls } P_n = \mathcal{O}; \\ \deg f_n & \text{sonst.} \end{cases}$$

Die Funktion d_n genügt nun der *Grundrelation*

$$d_{n-1} + d_{n+1} = 2d_n + 2. \tag{5.11}$$

Was hat nun d_n mit der Anzahl $N_q = \#E(\mathbb{F}_q) - 1$ der $\mathbb{F}_q$-rationalen Punkte $\neq \mathcal{O}$ auf E zu tun? Die Antwort besteht aus der Gleichung

$$\#E(\mathbb{F}_q) = N_q + 1 = d_{-1}. \tag{5.12}$$

Diese Beziehung liefert den

Hilfssatz 5.3. *Die Funktion $d(n) := d_n$ ist ein quadratisches Polynom in n:*

$$d(n) = n^2 - (d_{-1} - d_0 - 1)n + d_0.$$

Der Beweis der Hasse-Schranken ist jetzt ganz einfach: das quadratische Polynom $d(x) = x^2 - (d_{-1} - d_0 - 1)x + d_0$ nimmt nach Hilfssatz 5.3 für alle $n \in \mathbb{Z}$ nur nichtnegative Werte an. Wir behaupten nun, daß $d(x)$ sogar für alle $x \in \mathbb{R}$ nichtnegativ ist. Wäre dies nämlich nicht so, so hätte

$d(x)$ zwei einfache reelle Nullstellen (bei einer doppelten Nullstelle wäre die Behauptung trivialerweise richtig).

Sei also z.B. $\xi_1 < \xi_2$; zwischen diesen beiden reellen Zahlen kann kein $n \in \mathbb{Z}$ liegen, da sonst $d(n) < 0$ wäre. Also gilt $n \le \xi_1 < \xi_2 \le n+1$ für ein $n \in \mathbb{Z}$. Würde auf beiden Seiten das Gleichheitszeichen stehen, so wäre $d_n = d_{n+1} = 0$; dies geht nur, wenn $P_n = P_{n+1} = \mathcal{O}$ ist, und dies wiederum impliziert $(t,1) = P_{n+1} - P_n = \mathcal{O}$: Widerspruch. Also ist mindestens eine der beiden Ungleichungen scharf und damit $0 < \xi_2 - \xi_1 < 1$. Weil aber $(\xi_1 - \xi_2)^2$ gleich der Diskriminante von d ist und diese ganz ist, kann das ebenfalls nicht sein.

Also ist $d(x) \ge 0$ für alle $x \in \mathbb{R}$; folglich muß die Diskriminante von d nichtpositiv sein, und wir finden

$$0 \ge \operatorname{disc} d = (d_{-1} - d_0 - 1)^2 - 4d_0 = (N_q - q)^2 - 4q$$

wegen $d_0 = \deg t^q = q$. Dies liefert die Behauptung, wenn wir berücksichtigen, daß wir mit N_q nur die endlichen Punkte gezählt haben.

Nachzutragen sind nun noch die Beweise der Hilfssätze und der beiden Gleichungen (5.11) und (5.12).

Wir beginnen mit (5.12). Dazu stellen wir fest, daß wegen $P_0 = (x_0, y_0)$ mit $x_0 = t^q$ sicherlich $d_0 = q$ ist. Eine kleine Rechnung zeigt

$$x_{-1} = \frac{(t^3 + at + b)[(t^3 + at + b)^{(q-1)/2} + 1]^2}{(t^q - t)^2} - (t^q + t).$$

Mit $\lambda = t^3 + at + b$ erhält man für den Zähler daher

$$\begin{aligned} f_{-1} &= \lambda(\lambda^{(q-1)/2} + 1)^2 - (t^q + t)(t^q - t)^2 \\ &= \lambda^q + 2\lambda^{(q+1)/2} + \lambda - (t^{2q} - t^2)(t^q - t); \end{aligned} \tag{5.13}$$

da wir in Charakteristik p rechnen, ist $\lambda^q = t^{3q} + at^q + b$ (wegen $a \in \mathbb{F}_q$ ist $a^q = a$), folglich hat der Zähler von x_{-1} die Form $t^{2q+1} + h(t)$ mit einem Polynom $h(t)$ vom Grad höchstens $2q$.

Jetzt geht es darum, etwaige gemeinsame Faktoren von Zähler und Nenner zu kürzen. Dazu beachten wir, daß sich der Nenner $t^q - t$ schreiben läßt als Produkt $\prod(t-\alpha)$ über alle $\alpha \in \mathbb{F}_q$. Jetzt gibt es für gemeinsame Teiler $t - \alpha$ zwei Möglichkeiten:

i) Der Zähler ist durch $(t-\alpha)^2$ teilbar;

ii) Der Zähler ist genau durch $t - \alpha$ teilbar.

Im Falle i) zeigt die Darstellung (5.13), daß $t-\alpha$ ein Teiler des Produktes $\lambda(\lambda^{(q-1)/2}+1)^2$ sein muß; weil $\lambda = t^3+at+b$ aber keine mehrfachen Nullstellen in $\mathbb{F}_q$ besitzt (Nichtsingularität!), muß $(t-\alpha) \mid (\lambda^{(q-1)/2}+1)$ sein. Im Falle ii) dagegen teilt $t-\alpha$ notwendig t^3+at+b. Sei nun m die Anzahl aller $t-\alpha$ in i), und n die entsprechende Anzahl in ii). Offenbar ist dann

$$d_{-1} = \deg f_{-1} = 2q+1-2m-n. \tag{5.14}$$

Jetzt zählen wir $\#E(\mathbb{F}_q)$; dazu nehmen wir an, $t-\alpha$ sei ein Teiler von t^3+at+b, also $\alpha^3+a\alpha+b=0$. Dann hat die Gleichung (5.6) für jedes dieser $n \le 3$ verschiedenen α genau eine Lösung.

Ist weiter $t-\alpha$ ein Teiler von $(t^3+at+b)^{(q-1)/2}+1$, d.h. $(\alpha^3+a\alpha+b)^{(q-1)/2} = -1$, so ist $\alpha^3+a\alpha+b$ nach dem Eulerschen Kriterium ein Nichtquadrat in $\mathbb{F}_q$; für diese m Werte von α hat die Gleichung (5.6) aber keine Lösung.

Wir haben also q Werte für α, von denen m gar keine, n genau eine, und $q-m-n$ genau zwei Lösungen für (5.6) liefern. Damit ist die Anzahl aller Lösungen gleich $N_q = 2(q-m-n)+n = 2q-2m-n$. Setzt man dies in (5.14) ein, so erhält man, wie gewünscht, (5.12).

Hilfssatz 5.3 ergibt sich ganz einfach aus der Grundrelation: für $n=-1$ und $n=0$ ist die Aussage trivial; ist sie für $n-1$ und n richtig, so folgt

$$\begin{aligned} d_{n+1} &= 2d_n - d_{n-1} + 2 \\ &= 2[n^2-(d_{-1}-d_0-1)n+d_0] \\ &\quad -[(n-1)^2-(d_{-1}-d_0-1)(n-1)+d_0]+2 \\ &= (n+1)^2-(d_{-1}-d_0-1)(n+1)+d_0. \end{aligned}$$

Entsprechend zeigt man die Behauptung für alle $n \le -1$.

Als nächstes beweisen wir Hilfssatz 5.2. Für $n=0$ gilt er offensichtlich. Ist $P_{n-1} = \mathcal{O}$, so gilt er für n und $n+1$. Wir nehmen nun an, er sei richtig für irgendein $n \ge 0$ mit $P_{n-1} \ne \mathcal{O}$ und zeigen per Induktion, daß er dann auch für $n+1$ gilt. Sei $P_{n+1} \ne \mathcal{O}$; wir nehmen an, es sei $x_{n+1} = 0$ oder $\deg f_{n+1} \le \deg g_{n+1}$ und konstruieren einen Widerspruch. In beiden Fällen ist $x_{n+1}|_\infty$ endlich, also $tx_{n+1}|_\infty = 0$ (d.h. läßt man $t \longrightarrow \infty$ gehen, so geht tx_{n+1} gegen 0). Aus

$$y_{n+1}^2 = \frac{x_{n+1}^3+ax_{n+1}+b}{t^3+at+b}$$

folgt dann, daß $y_{n+1}|_\infty = 0$ ist. Wegen $(x_{n+1},-y_{n+1})+(x_n,y_n)+(t,1) = \mathcal{O}$ sind die drei Punkte $(x_{n+1},-y_{n+1})$, (x_n,y_n) und $(t,1)$ kollinear. Ein Vergleich der Steigungen liefert jetzt

$$y_{n+1} = \frac{1-y_n}{t-x_n}(t-x_{n+1}) - 1,$$

also

$$0 = y_{n+1}|_\infty = \Big\{\frac{1-y_n}{1-t^{-1}x_n}\,(1-t^{-1}x_{n+1}) - 1\Big\}\Big|_\infty.$$

Wegen $t^{-1}x_{n+1}|_\infty = 0$ muß daher

$$\frac{1-y_n}{1-t^{-1}x_n}\Big|_\infty = 1$$

sein. Nach der Additionsformel gilt aber

$$x_{n+1} = \Big(\frac{1-y_n}{t-x_n}\Big)^2(t^3+at+b) - t - x_n,$$

sodaß wir

$$\frac{x_{n+1}}{t} = \Big(\frac{1-y_n}{t-x_n}\Big)^2(1+at^{-2}+bt^{-3}) - 1 - \frac{x_n}{t}$$

erhalten. Aus der Induktionsannahme folgt dann aber

$$\begin{aligned} 0 = \frac{x_{n+1}}{t}\Big|_\infty &= \Big\{\Big(\frac{1-y_n}{1-t^{-1}x_n}\Big)^2(1+at^{-2}+bt^{-3}) - 1 - \frac{x_n}{t}\Big\}\Big|_\infty \\ = -\frac{x_n}{t}\Big|_\infty &\ne 0, \end{aligned}$$

also der gewünschte Widerspruch. Den Beweis für $n \le 0$ führt man analog.

Damit bleibt noch die Grundrelation zu beweisen. Auch diese ist trivial, wenn P_{n-1}, P_n oder $P_{n+1} = \mathcal{O}$ ist: gilt z.B. $P_n = \mathcal{O}$, so ist $x_{n-1} = x_{n+1} = t$, sowie $d_n = 0$ und $d_{n-1} = d_{n+1} = 1$. Ist $P_{n-1} = \mathcal{O}$, so folgt $(x_n, y_n) = (t, 1)$, und die Additionsformel gibt

$$x_{n+1} = \frac{t^4 - 2at^2 - 8bt + a^2}{4(t^3+at+b)}.$$

Also ist $d_{n-1} = 0$, $d_n = 1$ und, da $(f_{n+1}, g_{n+1}) = 1$ ist, $d_{n+1} = 4$.

Wir nehmen daher an, daß die Punkte P_{n-1}, P_n und P_{n+1} alle von $\mathcal{O}$ verschieden sind. Die Additionsformel gibt dann $P_{n-1} = P_n + (t, -1)$, also

$$\begin{aligned} x_{n-1} &= \frac{-(tg_n+f_n)(tg_n-f_n)^2 + (1+y_n)^2(t^3+at+b)g_n^3}{g_n(tg_n-f_n)^2} \\ &= \frac{(tg_n+f_n)(tf_n+ag_n) + 2bg_n^2 + 2y_n(t^3+at+b)g_n^2}{(tg_n-f_n)^2} \qquad (5.15) \\ &= \frac{R}{(tg_n-f_n)^2}, \end{aligned}$$

wobei wir $\lambda y_n^2 = x_n^3 + ax_n + b$ verwendet haben. Ebenso erhält man

$$\begin{aligned} x_{n+1} &= \frac{-(tg_n + f_n)(tg_n - f_n)^2 + (1 - y_n)^2(t^3 + at + b)g_n^3}{g_n(tg_n - f_n)^2} \\ &= \frac{(tg_n + f_n)(tf_n + ag_n) + 2bg_n^2 - 2y_n(t^3 + at + b)g_n^2}{(tg_n - f_n)^2} \\ &= \frac{S}{(tg_n - f_n)^2}. \end{aligned} \tag{5.16}$$

Dabei sind $R, S \in \mathbb{F}_q[t]$, da sich die Nenner von y_n wegheben: wegen $\lambda y^2 = x^3 + ax + b$ ist $\lambda y_n^2 g_n^3$ ganz, und damit ist erst recht $\lambda y_n g_n^2$ ganz. Multipliziert man die Ausdrücke für x_{n-1} und x_{n+1}, so folgt

$$\frac{f_{n-1}f_{n+1}}{g_{n-1}g_{n+1}} = \frac{RS}{(tg_n - f_n)^4} = \frac{(tf_n - ag_n)^2 - 4bg_n(tg_n + f_n)}{(tg_n - f_n)^2}. \tag{5.17}$$

Wenn wir zeigen können, daß

$$g_{n-1}g_{n+1} = c \cdot (tg_n - f_n)^2 \tag{5.18}$$

für ein $c \in \mathbb{F}_q$ gilt, dann folgt

$$f_{n-1}f_{n+1} = c[(tf_n - ag_n)^2 - 4bg_n(tg_n + f_n)]$$

und, unter Berücksichtigung von Hilfssatz 5.2,

$$d_{n-1} + d_{n+1} = \deg(f_{n-1}f_{n+1}) = \deg(t^2 f_n^2) = 2d_n + 2,$$

also die behauptete Relation.

Nun wissen wir aus (5.17), daß $(tg_n - f_n)^2 \mid RS$ ist. Schreiben wir $(tg_n - f_n)^2 = R_1 S_1$ mit $R_1 \mid R$ und $S_1 \mid S$, so folgt

$$x_{n-1} = \frac{R}{(tg_n - f_n)^2} = \frac{R/R_1}{S_1};$$

nun ist $f_{n-1}/g_{n-1} = x_{n-1}$, also $g_{n-1} \mid S_1$. Ähnlich folgt $g_{n+1} \mid R_1$, also $g_{n-1}g_{n+1} \mid (tg_n - f_n)^2$. Es genügt daher,

$$(tg_n - f_n)^2 \mid g_{n-1}g_{n+1} \tag{5.19}$$

zu zeigen.

Nehmen wir an, dies wäre falsch. Dann gibt es ein irreduzibles $f \in \mathbb{F}_q[t]$ mit $2v_f(tg_n - f_n) > v_f(g_{n-1}g_{n+1})$. Aus (5.17) folgt dann $f \mid T$ für $T = (tf_n - ag_n)^2 - 4bg_n(tg_n + f_n)$. Wenn wir zeigen können, daß f ein Teiler von R und S ist, dann teilt f die Polynome $(1 - y_n)(t^3 + at + b)g_n^2$ und $(1 + y_n)(t^3 + at + b)g_n^2$. Da aber $f \nmid g_n$ (sonst würde es f_n und g_n teilen, aber die sind teilerfremd) und f irreduzibel ist, muß $f \mid (t^3 + at + b)$ sein. Teilt man nun T durch $tg_n - f$, so erhält man

$$T = -(tg_n - f_n)[tf_n^2 + (t^3 - 2at - 4b)g_n] + (t^4 - 2at^2 - 8bt + a^2)g_n^2.$$

Also muß $f \mid (t^4 - 2at^2 - 8bt + a^2)$ sein, und zusammen mit $f \mid (t^3 + at + b)$ und

$$(3t^3 - 5at - 27b)(t^3 + at + b) - (3t^2 + a)(t^4 - 2at^2 - 8bt + a^2) = \Delta$$

(mit $\Delta = -4a^3 - 27b^2$; diese Identität ist uns schon beim Beweis des Satzes von Nagell-Lutz über den Weg gelaufen) impliziert dies, daß f die von 0 verschiedene Konstante Δ teilt: Widerspruch.

Es fehlt also noch der Nachweis, daß $f \mid R$ und $f \mid S$ gilt. Wegen $f \mid T$ und $T \mid RS$ ist sicherlich R oder S durch f teilbar. Nehmen wir also an, es wäre z.B. $f \mid R$ und $f \nmid S$. Dann würde wegen $x_{n+1} = f_{n+1}/g_{n+1}$ aus (5.16) folgen, daß $v_f(g_{n+1}) = v_f(tg_n - f_n)^2 > 0$ ist. Da f_{n+1} und g_{n+1} teilerfremd sind, muß dann folglich

$$v_f(f_{n+1}) = 0 \tag{5.20}$$

sein. Mit (5.17) folgt jetzt $0 < v_f(T) = v_f(f_{n-1}) - v_f(g_{n-1})$, d.h. $v_f(f_{n-1}) > v_f(g_{n-1})$. Da auch f_{n-1} und g_{n-1} teilerfremd sind, folgern wir

$$v_f(g_{n-1}) = 0. \tag{5.21}$$

Aus (5.20) und (5.21) ergibt sich jetzt aber

$$v_f(g_{n-1}g_{n+1}) = v_f(tg_n - f_n)^2,$$

und das ist ein Widerspruch, der den Beweis beendet.

Hasses Beweis

In seiner Arbeit [Ma] macht Manin die Bemerkung, daß sein Beweis im Kern mit demjenigen von Hasse übereinstimmt. Im folgenden wollen wir Hasses Beweis skizzieren, ohne alle aufgeführten Aussagen zu beweisen.

Unter einem Endomorphismus einer über einem Körper K definierten elliptischen Kurve E versteht man einen Homomorphismus $E \longrightarrow E$, der sich als rationale Funktion der Koordinaten (mit Koeffizienten in K, versteht sich) schreiben läßt. Standardbeispiel eines Endomorphismus ist die Multiplikation $[m]$ mit einer ganzen Zahl $m \in \mathbb{Z}$. Da Summe und Komposition zweier Endomorphismen wieder ein solcher ist, bilden die Endomorphismen einen Ring $\mathrm{End}(E)$, und man kann die Multiplikationen $[m]$ als einen zu $\mathbb{Z}$ isomorphen Unterring von $\mathrm{End}(E)$ auffassen.

Über einem endlichen Körper $\mathbb{F}_q$ spielt der Frobeniusendomorphismus $\pi : (x, y) \longrightarrow (x^q, y^q)$ eine tragende Rolle. Ist $P \in E(\overline{F})$ ein Punkt auf E über dem algebraischen Abschluß $\overline{F}$ von $\mathbb{F}_q$, so gilt genau dann $E \in E(\mathbb{F}_q)$, wenn $\pi(P) = P$ ist, mit anderen Worten: die $\mathbb{F}_p$-rationalen Punkte auf $E(\overline{F})$ sind genau die Punkte im Kern des Endomorphismus $1 - \pi$.

Nun ist für jeden Endomorphismus $\phi \in \mathrm{End}(E)$ die Spur $\mathrm{Tr}\phi := \phi + \overline{\phi}$ eine ganze Zahl (im Sinne von $\mathbb{Z} \subseteq \mathrm{End}(E)$), und für separable Endomorphismen gilt die Beziehung $\#E(\mathbb{F}_q) = \deg(1 - \phi)$. Insbesondere erhält man daher $\#E(\mathbb{F}_q) = \deg(1 - \pi) = q + 1 - \mathrm{Tr}\phi$. Wegen der Standardabschätzung $|\mathrm{Tr}\phi| \le 2\sqrt{N\phi}$ und $N\pi = 2\sqrt{q}$ folgt daher die Hasse-Schranke.

Bemerkungen zur Literatur

Eine schöne Darstellung von Dirichlets Beweis findet man in dem Büchlein [Fr] von G. Frey. Ein Vierseitenbeweis des Primzahlsatzes (vor nicht allzu langer Zeit brauchte man dafür ein Vielfaches) steht in H. Kochs neuem Buch [Ko]. Der Maninsche Beweis ist Chahal [Ch2] entnommen und ist eine Vereinfachung des Beweises in [Ch1] oder [Kna].

Ein Brief von Peter Roquette

29.4.1998
Lieber Herr Lemmermeyer,

besten Dank für die Zusendung Ihres Maninschen Beweises. Bei der Lektüre Ihres Aufschriebs habe ich mich daran erinnern können, daß ich den Maninschen Beweis seinerzeit studiert habe; es ist schon lange her. Und ich kann mich auch an den Eindruck erinnern, den ich damals nach der Lektüre der Arbeit hatte, nämlich daß dies in der Tat im wesentlichen derselbe Beweis wie bei Hasse ist, nur eben unter Benutzung der expliziten Formel für das

Additionstheorem der elliptischen Funktionen, was Hasse wegen Charakteristik 2 und 3 vollständig vermeiden wollte (und vermieden hat), und unter Weglassung der strukturellen Deutung der eingeführten Begriffe (was ebenfalls nicht im Sinne von Hasse war).

Allerdings hat natürlich der Maninsche Beweis einen gewissen Wert zum Vortrag in einer Vorlesung für Hörer mit wenigen Vorkenntnissen: das sei ihm gerne zugestanden. (Aufgabe: führe diesen Beweis für Charakteristik 3 und 2 durch!)

Lassen Sie mich vielleicht erklären, wie ich die Sache sehe. Die $\mathbb{F}_q$-rationalen Punkte von E sind definitionsgemäß gekennzeichnet als die Fixpunkte der Frobenius-Isogenie π von E. Das ist der Grund dafür, daß der Hassesche Beweis den Begriff „Isogenie“ benutzt (er sagt: „Meromorphismus“).

Sei $X = (x, y)$ ein allgemeiner Punkt von E (über einem Definitionskörper K, den wir der Einfachheit halber als algebraisch abgeschlossen voraussetzen wollen, was aber nicht notwendig ist). Es ist also $y^2 = x^3 + ax + b$. Es ist $K(X) = K(x, y)$ der Funktionenkörper von E. Jede Isogenie μ wird dann gegeben durch den Punkt $\mu X = (x_\mu, y_\mu)$, der rational ist in $K(X)$. Die „Norm“ von μ wird definiert durch den Körpergrad:

$$N(\mu) = [K(X) : K(\mu X)] \tag{5.22}$$

In der Regel ist $N(\mu)$ gleich der Anzahl der Punkte im Kern von μ, nämlich dann wenn μ separabel ist (d.h. wenn $K(X)$ separabel ist über $K(\mu X)$). Hierbei muß man aber den unendlich fernen Punkt mitzählen, die Kurve E also projektiv auffassen. Insbesondere folgt

$$N(\pi - 1) = N_q + 1 \tag{5.23}$$

denn die $\mathbb{F}_q$-rationalen Punkte bilden den Kern von $\pi - 1$. (Die 1 auf der linken Seite bezeichnet die identische Isogenie; die 1 auf der rechten Seite von (5.23) ist natürliche Zahl; sie zählt den unendlich fernen Punkt; wie bei Ihnen schreibe ich hier also N_q für die Anzahl der $\mathbb{F}_q$-rationalen Punkte im endlichen.

Die obige Formel (5.23) ist die Formel (5.12) bei Ihnen. [...]

Der Hassesche Beweis besteht nun darin, die *Normenadditionsformel* zu beweisen:

$$N(\mu + \nu) + N(\mu - \nu) = 2N(\mu) + 2N(\nu), \tag{5.24}$$

welche zeigt, daß die Norm eine quadratische, *positiv definite* Form definiert auf der additiven Gruppe der Isogenien (wozu auch die uneigentliche Isogenie 0 gezählt wird).

Natürlich genügt es im Hinblick auf (5.23), diejenige Untergruppe zu betrachten, die aufgespannt wird von der Eins-Isogenie 1 und der Frobenius-Isogenie $\pi = \pi_q$ zu $\mathbb{F}_q$. Und weiter genügt es, für die Folge $\mu_n = 1 + n\pi$ die Regel

$$N(\mu_{n+1}) + N(\mu_{n-1}) = 2N(\mu_n) + 2 \tag{5.25}$$

zu zeigen (was ein Spezialfall von (5.24) ist). Man sieht den Zusammenhang mit der von Ihnen so genannten „*Grundrelation*" (5.11).

Den einzigen neuen Gedanken von Manin sehe ich darin, die Isogenien μ von E darzustellen als $K(x)$-rationale Punkte der getwisteten Kurve

$$E_\lambda : \lambda z^2 = u^3 + au + b \qquad \text{wobei} \qquad \lambda = x^3 + ax + b\,. \tag{5.26}$$

Zu jeder Isogenie μ von E gehört ein $K(x)$-rationaler Punkt (u, z) von E_λ, nämlich $u = x_\mu$, $z = y_\mu / y$, und zwar ist dabei $v_\infty(u) < 0$, wobei v_∞ die Bewertung der unendlichen Stelle von $K(x)$ ist, also der negative Grad einer rationalen Funktion. Und umgekehrt: jedem $K(x)$-rationalen Punkt (u, z) von E_λ entspricht auf diese Weise eine Isogenie μ, derart daß $x_\mu = u$ und $y_\mu = yz$. (Der unendlich ferne Punkt von E_λ gehört zur uneigentlichen Isogenie $\mu = 0$.)

Dabei entspricht der Addition von Isogenien die Addition von Punkten der getwisteten Kurve. Und die Norm einer Isogenie ist

$$N(\mu) = [K(X) : K(\mu X)] = [K(x) : K(x_\mu)] = [K(x) : K(u)]\,. \tag{5.27}$$

Schreibt man $u = f/g$ mit teilerfremden Polynomen f, g, so ist $v_\infty(u) = -\mathrm{Grad}(f) + \mathrm{Grad}(g) < 0$ und daher

$$[K(x) : K(u)] = \mathrm{Grad}(f)\,. \tag{5.28}$$

Somit sehen wir, daß die auf Seite 3 Ihres Manuskripts eingeführte Zahl d_n nichts anderes ist als die Norm der zugehörigen Isogenie.

Dieser Zusammenhang erlaubt es Manin, dem Leser den Begriff der Isogenie vorzuenthalten und mit rationalen Punkten der getwisteten Kurve zu rechnen. In Wahrheit ist es aber, wie gesagt, der Hassesche Beweis.

Kapitel 6
Geschichte

Griechische Mathematik

Die Frage, warum in der Zeit zwischen Diophant (zwischen 150 v.C. und 250 n.C. – falls er überhaupt gelebt hat) und Newton (1642–1727) keine Fortschritte gemacht wurden, wird von der Geschichte beantwortet. Die größte Bibliothek des Altertums befand sich in Alexandria, in einem Gebäude namens Brucheion im Hafengebiet der Stadt. Als diese Bibliothek aus allen Nähten platzte, legte man eine zweite Sammlung im Tempel des Serapis an. Bei der Eroberung Alexandrias durch Caesars Truppen im Jahre 47 v.C. fiel die Hauptbibliothek im Brucheion, die aus etwa 400.000 Bänden bestand, einem Brand zum Opfer. In den darauffolgenden Jahrhunderten wuchs die Sammlung im Serapistempel wieder an, aber 392. n.C. befahl Kaiser Theodosius die Vernichtung aller heidnischen Tempel, und seine christlichen Horden machten dabei vor der Bibliothek in Alexandria nicht halt. Von da an gab es keine Universalbibliothek mehr.

Wenig später, im Jahre 415 n.C., lynchte ein anderer Mob, vermutlich von Bischof Cyrillus angestiftet, die Heidin Hypatia, Tochter des Theon, der am Museum in Alexandria lehrte und sowohl Euklids Elemente wie den Almagest herausgab. Hypatia, wird berichtet, sei eine höchst gelehrte Frau gewesen; sie ist die erste uns namentlich bekannte Mathematikerin.

Im Jahre 529 ließ Kaiser Justitian, ein christlicher Theologe, die Akademie Platons in Athen schließen, um der Verbreitung von "Irrlehren" entgegenzuwirken, und verbot allen Philosophieunterricht in Athen. Die letzten griechischen Philosophen gingen ins Exil an den Hof des Perserkönigs Chosrau Anoscharwan.

Im Jahre 640 wurde Alexandria von den Arabern erobert, und man hört oft die Geschichte, wonach Omar, Mohammeds zweiter Nachfolger, die letzte alexandrinische Bibliothek vernichten lassen hat mit den Worten: "Entweder enthalten die Bücher das, was im Koran steht, dann brauchen wir sie nicht zu lesen, oder sie enthalten das Gegenteil dessen, was im Koran steht, dann dürfen wir sie nicht lesen." Allerdings ist die Existenz einer solchen Bibliothek recht unwahrscheinlich, da mit der Vernichtung des Brucheion und des Serapistempels wohl nicht mehr viel übrig war; vielmehr vermutet man, daß diese Geschichte von einem syrischen Christen des 13. Jahrhunderts namens Abulpharagius erfunden wurde, um die moslemischen Araber in ein schlechtes Licht zu setzen.

Ende des 15. Jahrhunderts kamen die griechischen Klassiker wieder zu Ehren: Euklids Elemente wurden 1482 aufgelegt, Schriften des Archimedes folgten 1503, und das Werk von Appolonius (262–190 v.C.) über Kegelschnitte erschien 1537.

Für die Naturwissenschaften hatte die Wiederauflage dieser Bücher Folgen: das eben genannte Werk von Appolonius dürfte beispielsweise Kepler (1571–1630) sehr geholfen haben, sich über die damaligen in Stein gemeisselten Gesetze von Aristoteles hinwegzusetzen, wonach für Planeten allein aus Perfektionsgründen keine andere Bahn denkbar sei als die Kreisbahn; Kopernikus (1473–1543) war dieser Schritt noch nicht gelungen. Allerdings hatte auch Kopernikus die Idee eines heliozentrischen Weltbilds von den Griechen geborgt: die Kugelgestalt der Erde wurde bereits von Thales von Milet (624–546 v.C.) gelehrt (und stammt vermutlich aus ägyptischen und babylonischen Quellen), das Verhältnis des Abstands Erde–Sonne und Erde–Mond wurde von Aristarch von Samos (310–250 v.C.) gemessen (!), Erathostenes bestimmte mit seinem bekannten Experiment den Erdumfang zu etwa 40.000 km, und Hipparch (180–125 v.C.) schließlich erklärte nicht nur die Gezeiten, sondern beobachtete auch die Exzentrizität der Erdbahn.

Mit der Entdeckung der sechs Bücher Diophants durch Rafael Bombelli 1570 in der Bibliothek des Vatican beginnt auch der Wiederaufstieg der Mathematik im christlichen Abendland. Rafael Bombelli veröffentlicht 1572 in seinem Algebrabuch 143 Aufgaben aus Diophants Arithmetica, und Wilhelm Holzmann gab 1575 unter dem Pseudonym Xylander (= grch. für Holzmann – das Vergriechen seines Namens war damals en vogue; der "Neandertalmensch" ist von einem Herrn Neumann gefunden worden) die sechs Bücher Diophants in lateinischer Übersetzung heraus; diese wurden von niemand anders als Viéta (1540–1603) studiert, der dann aus Diophants algebraischer Notation das "Buchstabenrechnen" machte, ohne das die heutige Algebra vollkommen undenkbar wäre. Schließlich gab Bachet (1581–1638) im Jahre 1621 Diophants Bücher heraus und versah sie mit Kommentaren; in eines dieser Bücher schrieb ein Justizbeamter in Toulouse namens Fermat dann die Worte

> Cubum autem in duos cubos, aut quadratoquadratum in duos quadratoquadratos, et generaliter nullam in infinitum ultra quadratum potestatem in duos ejusdem nominis fas est dividere: cujus rei demonstrationem mirabilem sane detexi. Hanc marginis exiguitas non caparet.

Die in Kapitel 1 auftauchende Kissoide des Diocles hängt mit einem der drei klassischen Probleme der griechischen Geometrie zusammen, nämlich der *Würfelverdopplung* (die beiden andern sind bekannter: die Quadratur des Kreises und die Dreiteilung des Winkels). Die dahinterstehende Legende ist folgende:

In einem Brief des griechischen Mathematikers Eratosthenes an den ägyptischen König Ptolemäus Euergetes heißt es:

> Dem Könige Ptolemäus wünscht Eratosthenes Glück und Wohlsein. Von den alten Tragödiendichtern, sagt man, habe einer den Minos, wie er dem Glaukos ein Grabmal errichten ließ, und hörte, daß es auf allen Seiten 100 Fuß haben werde, sagen lassen:
>
> > Zu klein entwarfst Du mir die königliche Gruft,
> > Verdopple sie; des Würfels doch verfehle nicht.
>
> Man untersuchte aber auch von seiten der Geometer, auf welche Weise man einen gegebenen Körper, ohne daß er seine Gestalt veränderte, verdoppeln könnte, und nannte die Aufgabe der Art des Würfels Verdoppelung; denn einen Würfel zugrunde legend suchte man diesen zu verdoppeln ... Nach der Zeit, erzählt man, wären die Delier, weil sie von einer Krankheit befallen waren, einem Orakel zufolge geheißen worden einen ihrer Altäre zu verdoppeln und in dieselbe Verlegenheit geraten.

In einer Abhandlung über Brenngläser hat Diocles seine "Kissoide" gefunden; mit dieser Kurve läßt sich das Problem der Würfelverdopplung leicht lösen: man betrachtet nämlich die Gerade durch $(-a, 0)$ und $(0, a/2)$; diese schneidet die Kissoide in einem Punkt (x, y), und man rechnet sofort nach (Übung!), daß $2 = \left(\frac{a-x}{y}\right)^3$ gilt. Insbesondere kann man $\sqrt[3]{2}$ mit Hilfe von Zirkel, Lineal und Kissoide konstruieren. Daß es mit Zirkel und Lineal allein nicht geht, hat man erst viel später zeigen können und wird heutzutage in Algebravorlesungen am Rande mitbewiesen.

Wer sich näher für die Geschichte der Mathematik interessiert, hat mit den drei Bänden von Moritz Cantor [Can] eine ganze Menge Lesestoff. Einige der obigen Bemerkungen findet man in dem etwas kürzeren Artikel [Hil]. Es sei noch bemerkt, daß 1971 vier der dreizehn Bücher des Diophant, von denen bis dahin nur sechs bekannt waren, von Roshdi Rashed in einer arabischen Übersetzung gefunden wurden.

Die bekannteste Aufgabe in Diophant's Büchern verlangt, drei rechtwinklige Dreiecke mit rationalen Seitenlängen und gleicher Fläche zu finden. Elementare Algebra zeigt, daß dies zum Auffinden von Zahlen $v, n \in \mathbb{Q}$ äquivalent ist, für die $v - n$, v und $v + n$ Quadrate rationaler Zahlen sind. Die natürlichen Zahlen n, für welche es solche v gibt, heißen *kongruente Zahlen.* Erst in unserer Zeit hat man festgestellt, daß ein $n \in \mathbb{N}$ genau dann kongruent ist, wenn die elliptische Kurve $E_n : y^2 = x^3 - n^2x$ Rang ≥ 1 hat. Da die Kurve E_2 nichts anderes als die Fermatkurve $X^4 + Y^4 = Z^2$ ist, kann man FLT für $n = 4$ auch so aussprechen: die Zahl 2 ist nicht kongruent.

Die Lemniskate, das AGM, und π

Wir schreiben das Jahr 1691. Jacob Bernoulli beschäftigt sich mit der Kurvengleichung eines gebogenen elastischen Stabs und kommt auf die Funktion

$$y(x) = \int_0^x \frac{z^2\,dz}{\sqrt{1-z^4}}\,, 0 \leq x \leq 1.$$

Die Bogenlänge vom Ursprung $(0, 0)$ zu (x, y) berechnet er zu

$$\int_0^x \frac{dz}{\sqrt{1-z^4}},$$

und folglich hat der Stab die Gesamtlänge $\frac{1}{2}\omega = \int_0^1 \frac{dz}{\sqrt{1-z^4}}$. Drei Jahre später entdeckt er eine Kurve, deren Bogenlänge durch dieselbe Funktion gegeben ist wie die der elastischen Kurve: die Lemniskate. Seine Ergebnisse erscheinen in der Septemberausgabe der Acta Eruditorum. Im Oktoberheft findet sich dagegen eine Arbeit seines jüngeren Bruders Johann, der im Zusammenhang mit der Differentialgleichung $(x\,dx + y\,dy)\sqrt{y} = x\,dy - y\,dx$ auf die Lemniskate und die elastische Kurve gekommen ist. Der darauf folgende Prioritätsstreit hat dann zum endgültigen Zerwürfnis der beiden ohnehin streitlustigen Brüder geführt.

Der nächste Mathematiker, der sich erfolgreich mit der Lemniskate befaßte, war C.G. Fagnano: er zeigte, wie man den Lemniskatenbogen mit Zirkel und Lineal in 2^m, $3 \cdot 2^m$ oder $5 \cdot 2^m$ gleiche Teile teilen kann. Als im Jahre 1750 seine gesammelten Werke erscheinen, schickt er ein Exemplar an die Berliner Akademie. Diese gab sie am 23. 12. 1751 Euler mit der Bitte um ein Gutachten. Euler war von Fagnano's Ergebnissen förmlich elektrisiert und fand schon kurze Zeit später das allgemeine Additionstheorem für lemniskatische Integrale. Zur Geschichte der Lemniskate im Zusammenhang mit elliptischen Integralen ist Ayoubs Arbeit [Ay] empfehlenswert; ebenfalls lesenswert ist Siegels Darstellung [Sie] der Beiträge Fagnanos und Eulers zum Additionsgesetz der lemniskatischen elliptischen Funktionen.

Ein weiteres schönes Resultat Eulers, das allerdings erst nach seinem Tode publiziert wurde, ist die Beziehung

$$\int_0^1 \frac{dz}{\sqrt{1-z^4}} \cdot \int_0^1 \frac{z^2\,dz}{\sqrt{1-z^4}} = \frac{\pi}{2}.$$

Eine numerische Berechnung von π und ω durch Stirling zeigte

$$\int_0^1 \frac{z^2\,dz}{\sqrt{1-z^4}} = \frac{\pi}{\omega} = 1.98140\,23473\,5\ldots$$

Die Theorie der elliptischen Integrale wird in der Folgezeit von Lagrange und Legendre ausgebaut, denen aber wie Euler entgeht, daß die Einführung der Umkehrfunktionen dieser Integrale alles vereinfachen würde.

Kaum hatte Legendre ein dreibändiges Buch über die Theorie elliptischer Integrale veröffentlicht, fand Abel den Zugang via der Umkehrfunktionen, und zusammen mit Jacobi stellte er in den darauffolgenden Jahren die ganze Theorie auf den Kopf und erweiterte sie beträchtlich. Abel legte auch den Grundstein für die Theorie der komplexen Multiplikation: diese Theorie wurde nach seinem Tod von Eisenstein, Kronecker und Weber ausgebaut und durch Takagis Klassenkörpertheorie in gewisser Weise vollendet – andererseits zeigen die Namen Fueter, Hasse, Deuring, Eichler und Shimura, daß dieses Gebiet auch danach noch äußerst fruchtbar war.

Viele, wenn nicht die meisten der Abelschen Ergebnisse hat bereits Gauß gefunden, jedoch nicht publiziert. Schon in seiner Jugend (d.h. ab etwa 14) hat er sich mit dem arithmetisch-geometrischen Mittel (AGM) beschäftigt.

Dieses ist wie folgt definiert: man setzt $a_0 = a$, $b_0 = b$, wo a, b nichtnegative reelle Zahlen sind, und definiert dann rekursiv $a_{n+1} = \frac{1}{2}(a_n + b_n)$, $b_{n+1} = \sqrt{a_n b_n}$. Es ist eine elementare Übungsaufgabe zu zeigen, daß beide Folgen gegen denselben Grenzwert konvergieren, und man setzt $M(a,b) := \lim a_n = \lim b_n$. Am 30. Mai 1799 beschließt Gauß, das AGM von $\sqrt{2}$ und 1 zu berechnen, und er findet

$$M(\sqrt{2}, 1) = 1.98140\,23473\,5\ldots = \frac{\pi}{\omega} \tag{6.1}$$

im Rahmen der Rechengenauigkeit von 11 Dezimalstellen. Sofort vermutet er, daß hinter dieser numerischen Gleichheit tiefe Wahrheiten verborgen sind, und kein halbes Jahr später kann er folgende Gleichung beweisen:

$$M(a,b) \cdot \int_0^{\pi/2} \frac{d\phi}{\sqrt{a^2 \cos^2 \phi + b^2 \sin^2 \phi}} = \frac{\pi}{2}.$$

Die Substitution $z = \cos \phi$ zeigt außerdem

$$\int_0^1 \frac{dz}{\sqrt{1 - z^4}} = \int_0^{\pi/2} \frac{d\phi}{\sqrt{2\cos^2 \phi + \sin^2 \phi}},$$

woraus (6.1) sofort folgt. Publiziert hat er dieses Ergebnis aber erst 1818, als er einen Zusammenhang mit Störungsrechnungen in der Astronomie gefunden hat (siehe [NS]; erstaunlicherweise findet man in diesem Buch auch projektive Geometrie (im Zusammenhang mit Kegelschnitten) und Weierstraßsche $\wp$-Funktionen).

Eine kleine Rechnung mit pari liefert übrigens

$$\begin{aligned}
\pi &= 3.14159265358979323846264338\ldots, \\
\omega &= 2.62205755429211981046483958\ldots, \\
\tfrac{\pi}{\omega} &= 1.19814023473559220743992249\ldots, \\
M(\sqrt{2}, 1) &= 1.19814023473559220743992249\ldots,
\end{aligned}$$

wobei noch zu bemerken ist, daß die Funktion $M(\,\cdot\,,\,\cdot\,)$ auch in pari installiert ist.

Im Laufe der weiteren Untersuchungen hat Gauß festegestellt, daß sich nicht nur π/ω, sondern sogar π allein mit Hilfe des AGM ausdrücken läßt, und zwar so: führt man die Größen $c_{k+1} = \frac{1}{2}(a_k - b_k)$ ein, so gilt

$$\pi = \frac{2M(1, \frac{1}{\sqrt{2}})}{\frac{1}{2} - \sum_{j=1}^{\infty} 2^j c_j^2}. \tag{6.2}$$

Diese Formel liefert sofort den folgenden Algorithmus zur Berechung von π: man setzt $a_0 = 1$, $b_0 = 1/\sqrt{2}$ und $s_0 = \frac{1}{2}$. Dann berechnet man rekursiv $a_{k+1} = \frac{1}{2}(a_k + b_k)$, $b_{k+1} = \sqrt{a_k b_k}$, $c_{k+1}^2 = (a_{k+1} - b_{k+1})^2$ und $s_{k+1} = s_k - 2^{k+1}c_{k+1}^2$, und erhält als $(k+1)$-te Näherung $\pi_k = \frac{(a_k+b_k)^2}{2s_k}$. Mit jedem Schritt dieses Algorithmus verdoppelt sich in etwa die Anzahl der richtigen Dezimalstellen.

Dieser Algorithmus wurde 1976 unabhängig von E. Salamin und R. Brent entdeckt, und erst daraufhin hat man die entsprechenden Formeln in den Arbeiten von Gauß gefunden.

Geometrische Interpretation

Daß die von Diophant benutzte "Sekantenmethode" eine geometrische Interpretation besitzt, hat wohl als erster Newton bemerkt, wenn auch nicht publiziert. Während Euler wie Newton sieht, daß Kegelschnitte mit einem rationalen Punkt deren unendlich viele haben, und daß diese sich parametrisieren lassen, entgeht im wohl der geometrische Hintergrund. Die Tangentenmethode stammt allem Anschein nach von Lagrange (1777); allerdings verzichtet er darauf, diese Methode geometrisch zu interpretieren. Sylvester (1858) hat als erster die Frage gestellt, ob es Punkte gibt, aus denen man alle rationalen Punkte mit der Sekanten-Tangenten-Methode erhalten kann; ähnliche Fragen wurden danach von Lucas (1878) und Desboves (1879) behandelt. Obwohl sich Sylvester (1879) noch einmal ausgiebig mit der Sekanten-Tangenten-Methode auf kubischen Kurven auseinandersetzt, dringt er nicht bis zur Gruppenstruktur elliptischer Kurven durch. Es ist sogar fraglich, ob er das Konzept einer abstrakten Gruppe besessen hat, das damals erst im Entstehen begriffen war. Auch anderen Mathematikern wie Beppo Levi oder Hurwitz, die sich mit der Bestimmung von "Ausnahmepunkten" (Torsionspunkte im heutigen Sprachgebrauch) beschäftigten, bleibt die Gruppenstruktur ebenso verborgen wie Poincaré oder Mordell. Letzterer hat 1922 die endliche Erzeugtheit von $E(\mathbb{Q})$ bewiesen, ohne die Gruppenstruktur zu kennen! Die erste Formulierung des Gruppengesetzes auf elliptischen Kurven in der uns geläufigen Form findet sich in einer Arbeit von Juel [Ju] aus dem Jahre 1896. Aber erst nachdem Weil und Nagell das Gruppengesetz im Jahre 1928 publiziert hatten, scheint es allgemein bekannt gewesen zu sein, daß die rationalen Punkte auf elliptischen Kurven eine abelsche Gruppe bilden.

Der Satz von Mordell-Weil markiert im übrigen einen wichtigen Wendepunkt in der Geschichte diophantischer Gleichungen: waren die Ergebnisse davor lediglich eine Sammlung von Resultaten, die sich auf ganz spezielle Kurven bezogen (Fermatgleichungen zum Beispiel), so ist der Satz von Mordell-Weil ein strukturelles Ergebnis, welches nicht nur für alle elliptischen Kurven über $\mathbb{Q}$, sondern allgemeiner für abelsche Varietäten über beliebigen algebraischen Zahlkörpern gilt.

Analytische Interpretation

Bereits Euler hat sich gefragt, wie man auf Kurven der Form $y^2 = f(x)$ mit Polynomen $f \in \mathbb{Z}[x]$ vom Grad 4 rationale Punkte finden kann. Jacobi hat 1834 darauf hingewiesen, daß Euler seine Additionstheoreme für elliptische Integrale der Form $\int \frac{dx}{\sqrt{f(x)}}$ benutzen hätte können, um aus gegebenen rationalen Punkte neue zu finden, und er fand es erstaunlich, daß Euler dies nicht gesehen hat. Scriba und Weil haben in den 80er Jahren darauf hingewiesen, daß offenbar kein einziger Mathematiker diese Bemerkung Jacobis zur Kenntnis genommen hat – beide haben allerdings eine Arbeit von Kummer über rationale Vierecke übersehen, in der genau diese Idee Jacobis benutzt wird.

Anhang Kapitel A
Resultanten

Sei R ein kommutativer Ring mit 1, und seien $f = a_m X^m + \ldots + a_1 X + a_0$ und $g = b_n X^n + \ldots + b_0$ Polynome aus $R[X]$ mit $a_m b_n \neq 0$. Dann definiert man die Resultante von f und g durch

$$\operatorname{Res}(f, g) := \det \left(\begin{array}{ccccccccc}
a_0 & a_1 & a_2 & \ldots & a_m & & & & \\
 & a_0 & a_1 & a_2 & \ldots & a_m & & & \\
 & & a_0 & a_1 & a_2 & \ldots & a_m & & \\
 & & & \ddots & \ddots & \ddots & \ldots & \ddots & \\
 & & & & a_0 & a_1 & a_2 & \ldots & a_m \\
b_0 & b_1 & b_2 & \ldots & b_n & & & & \\
 & b_0 & b_1 & b_2 & \ldots & b_n & & & \\
 & & b_0 & b_1 & b_2 & \ldots & b_n & & \\
 & & & \ddots & \ddots & \ddots & \ldots & \ddots & \\
 & & & & b_0 & b_1 & b_2 & \ldots & b_n
\end{array}\right)
\begin{array}{l}
\left.\begin{array}{c} \\ \\ \\ \\ \\ \end{array}\right\} n \\
\left.\begin{array}{c} \\ \\ \\ \\ \\ \end{array}\right\} m
\end{array} \tag{A.1}$$

Die Zeilen sind jeweils mit Nullen aufgefüllt.

Im folgenden sei R immer ein Ring mit eindeutiger Primfaktorzerlegung (wir brauchen das Ergebnis ohnehin nur für $R = \mathbb{Z}$). Dann gilt:

Satz A.1. *Seien $f, g \in R[X]$ Polynome vom Grad $m = \deg f$ und $n = \deg g$; dann sind die folgenden Aussagen äquivalent:*

i) f und g haben einen gemeinsamen Teiler vom Grad ≥ 1;
ii) es gibt $a, b \in R[X]$ mit $\deg a < n$ und $\deg b < m$;
iii) es ist $\mathrm{Res}\,(f,g) = 0$.

Beweis. $\boxed{\text{i)} \Longrightarrow \text{ii)}}$ Ist $u \in R[X]$ ein Polynom vom Grad ≥ 1 mit $u \mid f$ und $u \mid g$, so setze man $f = bu$ und $g = -au$.
$\boxed{\text{ii)} \Longrightarrow \text{i)}}$ Zerlege f und bg in Primpolynome; falls i) falsch ist, muß $f \mid b$ sein, was aus Gradgründen aber nicht geht.
$\boxed{\text{ii)} \iff \text{iii)}}$ Sei K der Quotientenkörper von R. Für Polynome $a, b \in K[X]$ mit

$$\begin{aligned} a &= \alpha_0 + \alpha_1 X + \ldots + \alpha_{n-1} X^{n-1}, \\ b &= \beta_0 + \beta_1 X + \ldots + \beta_{m-1} X^{m-1} \end{aligned} \tag{A.2}$$

gilt dann

$$(\alpha_0 \ \alpha_1 \ \ldots \ \alpha_{n-1} \ \beta_0 \ \ldots \ \beta_{m-1})\, R = (c_0 \ c_1 \ \ldots \ c_{m+n-1}) \tag{A.3}$$

für gewisse $c_j \in K$, wobei R die Matrix bezeichnet, deren Determinante auf der rechten Seite von (A.1) steht. Dabei gilt

$$\begin{aligned} c_{m+n-1} &= \alpha_{n-1} a_m + \beta_{m-1} b_n \\ c_{m+n-2} &= \alpha_{n-1} a_{m-1} + \alpha_{n-2} a_m + \beta_{m-1} b_{n-1} + \beta_{m-2} b_n \\ &\ldots \end{aligned}$$

Falls also a, b existieren derart, daß $c = \sum_{i=0}^{m+n-1} c_i X^i \equiv 0$ wird, so ist $\ker R \neq 0$ und folglich $\mathrm{Res}\,(f,g) = \det R = 0$. Ist umgekehrt $\mathrm{Res}\,(f,g) = 0$, so ist $\ker R \neq 0$ und es gibt $a, b \in K[X]$, nicht beide $\equiv 0$, sodaß $c \equiv 0$ wird. Durchmultiplizieren mit dem Hauptnenner der Koeffizienten gibt dann $a, b \in R[X]$.

Sei schließlich $\mathrm{Res}\,(f,g) \neq 0$. Nach der Cramerschen Regel ist dann $R^{-1} = \mathrm{Res}\,(f,g)^{-1} S$ für eine Matrix S mit Einträgen aus $R[X]$ (nämlich gewisse Unterdeterminanten von R). Aus (A.3) folgt dann

$$(\mathrm{Res}\,(f,g) \ 0 \ldots \ 0) R^{-1} = (\alpha_0 \ \ldots \ \alpha_{n-1} \ \beta_0 \ \ldots \ \beta_{m-1}),$$

und (A.2) definiert die gewünschten Polynome. □

Ist $f = a_m X^m + \ldots + a_0$, so heißt $f' = m a_m X^{m-1} + \ldots + a_1$ die formale Ableitung von f. Die Diskriminante $\mathrm{disc}\, f$ von f ist dann definiert als disc $f = (-1)^{n(n-1)/2}\,\mathrm{Res}\,(f, f')$. Satz A.1 sagt aus, daß genau dann disc $f = 0$ ist, wenn f und f' einen gemeinsamen Teiler vom Grad ≥ 1 besitzen, also genau dann, wenn f eine mehrfache Nullstelle hat.

Für Polynome kleinen Grades kann man die Diskriminante explizit hinschreiben:

n	f	disc f
1	$x + b$	1
2	$x^2 + bx + c$	$(b^2 - 4c)$
3	$x^3 + bx^2 + cx + d$	$-(27d^2 - 18bcd + 4c^3 + 4b^3d - b^2c^2)$

Anhang Kapitel B

Exakte Sequenzen

Ein Diagramm

$$A \xrightarrow{f} B \xrightarrow{g} C$$

von abelschen Gruppen A, B, C und Gruppenhomomorphismen $f : A \longrightarrow B$ und $g : B \longrightarrow C$ heißt exakt an der Stelle B, wenn $\operatorname{im} f = \ker g$ gilt. Eine längere Sequenz abelscher Gruppen heißt exakt, wenn sie an jeder Stelle, die nicht am Rand liegt, exakt ist.

Beispiele: die Sequenzen

$$\begin{array}{l} 0 \longrightarrow A \xrightarrow{f} B, \\ \qquad\quad B \xrightarrow{g} C \longrightarrow 0, \\ 0 \longrightarrow A \xrightarrow{h} A \longrightarrow 0 \end{array}$$

sind genau dann exakt, wenn f injektiv, g surjektiv, bzw. h ein Isomorphismus ist. Insbesondere ist

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

eine exakte Sequenz genau dann, wenn f injektiv, $\operatorname{im} f = \ker g$, und g surjektiv ist. Exakte Sequenzen dieser Form nennt man auch kurze exakte Sequenzen. Ist N eine Untergruppe von M, dann ist

$$0 \longrightarrow N \xrightarrow{\imath} M \xrightarrow{\pi} M/N \longrightarrow 0$$

eine kurze exakte Sequenz. Hierbei ist $\imath : N \longrightarrow M$ die Einbettung von N in M und $\pi : M \longrightarrow M/N$ die kanonische Projektion $m \longmapsto m + N$.

Beispiele:

- die Projektion $\pi : a + 4\mathbb{Z} \longmapsto a + 2\mathbb{Z}$ induziert eine exakte Sequenz
$$0 \longrightarrow \mathbb{Z}/2\mathbb{Z} \longrightarrow \mathbb{Z}/4\mathbb{Z} \longrightarrow \mathbb{Z}/2\mathbb{Z} \longrightarrow 0;$$
- bezeichnet $\pi : \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} \longrightarrow \mathbb{Z}/2\mathbb{Z}$ die Projektion auf die erste Koordinate, so ist
$$0 \longrightarrow \mathbb{Z}/2\mathbb{Z} \stackrel{\imath}{\longrightarrow} \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} \stackrel{\pi}{\longrightarrow} \mathbb{Z}/2\mathbb{Z} \longrightarrow 0$$
eine exakte Sequenz, falls die Injektion $\imath$ durch $\imath(a + 2\mathbb{Z}) = (0, a + 2\mathbb{Z})$ gegeben ist.

Proposition B.1. *Ist* $0 \longrightarrow A_1 \longrightarrow \ldots \longrightarrow A_n \longrightarrow 0$ *eine exakte Sequenz endlicher Gruppen, so gilt* $\#A_1 \#A_3 \cdots = \#A_2 \#A_4 \cdots$*; man sagt auch, das alternierende Produkt der Ordnungen sei gleich* 1*.*

Beweis. Die Aussage ist klar für $n = 2$; ist sie für ein $n \in \mathbb{N}$ bewiesen und ist
$$0 \longrightarrow A_1 \longrightarrow A_2 \longrightarrow \ldots \longrightarrow A_n \longrightarrow A_{n+1} \longrightarrow 0$$
eine exakte Sequenz endlicher Gruppen, so gilt dies auch für
$$0 \longrightarrow A_2/A_1 \longrightarrow A_3 \longrightarrow \ldots \longrightarrow A_{n+1} \longrightarrow 0$$
Für die letzte Sequenz gilt nach Induktionsvoraussetzung $\#A_2/A_1\#A_4 \cdots = \#A_3\#A_5 \cdots$, und die Behauptung ist für $n + 1$ bewiesen. □

Proposition B.2. Sandwich-Lemma. *Ist* $A \longrightarrow B \longrightarrow C$ *exakt, und sind* A *und* C *endlich, dann auch* B*.*

Beweis. Seien $\alpha : A \longrightarrow B$ und $\beta : B \longrightarrow C$ die Homomorphismen der exakten Sequenz; dann sind
$$0 \longrightarrow \ker\alpha \longrightarrow A \longrightarrow \operatorname{im}\alpha \longrightarrow 0$$
$$0 \longrightarrow \ker\beta \longrightarrow B \longrightarrow \operatorname{im}\beta \longrightarrow 0$$
ebenfalls exakt. Nun sind $\operatorname{im}\alpha$ als Quotient von A und $\operatorname{im}\beta$ als Untergruppe von C endlich, somit wegen $\operatorname{im}\alpha = \ker\beta$ auch $\#B = \#\ker\beta \cdot \#\operatorname{im}\beta$. □

Die wichtigste Aussage über exakte Sequenzen auf diesem elementaren Niveau ist das Schlangenlemma. Zu dessen Formulierung und Beweis brauchen wir einige Aussagen über "induzierte" Abbildungen:

Lemma B.3. *Sei ein kommutatives Quadrat*

$$\begin{array}{ccc} A & \xrightarrow[\alpha]{} & B \\ \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle g} \\ A' & \xrightarrow[\alpha']{} & B' \end{array}$$

von R-Moduln und R-Homomorphismen gegeben. Dann induzieren α und α' R-Homomorphismen $\overline{\alpha} : \ker f \longrightarrow \ker g$ und $\overline{\alpha}' : \operatorname{coker} f \longrightarrow \operatorname{coker} g$.

Beweis. Wir definieren $\overline{\alpha} : \ker f \longrightarrow \ker g$, indem wir $a \in \ker f$ auf $\alpha(a)$ abbilden; zu zeigen ist, daß $\alpha(a) \in \ker g$ ist. Wegen $g \circ \alpha(a) = \alpha' \circ f(a) = \alpha'(0) = 0$ ist das aber eine Folge aus der Kommutativität des Diagramms.

Entsprechend setzen wir $\overline{\alpha}'(a' + f(A)) = \alpha'(a') + g(B)$. Zu zeigen ist, daß $\alpha' \circ f(A) \subseteq g(B)$ ist: dies ist aber wegen $\alpha' \circ f(A) = g \circ \alpha(A) \subseteq g(B)$ wieder klar.

Daß die induzierten Abbildung $\overline{\alpha}$ und $\overline{\alpha}'$ wieder R-Homomorphismen sind, ist klar, da sie im wesentlichen Einschränkungen von R-Homomorphismen sind. $\square$

Satz B.4. *(Das Schlangenlemma) Sei ein kommutatives Diagramm*

$$\begin{array}{ccccc} A & \xrightarrow[\alpha]{} & B & \xrightarrow[\beta]{} & C \\ \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle g} & & \downarrow{\scriptstyle h} \\ A' & \xrightarrow[\alpha']{} & B' & \xrightarrow[\beta']{} & C' \end{array}$$

mit exakten Reihen gegeben.

a) Die R-Homomorphismen $\alpha : A \longrightarrow B$ etc. induzieren R-Homomorphismen $\overline{\alpha} : \ker f \longrightarrow \ker g$ usw. derart, daß $\ker f \longrightarrow \ker g \longrightarrow \ker h$ und $\operatorname{coker} f \longrightarrow \operatorname{coker} g \longrightarrow \operatorname{coker} h$ 0-Sequenzen von R-Moduln sind;

b) Ist α' injektiv, so ist $0 \longrightarrow \ker \alpha \longrightarrow \ker f \longrightarrow \ker g \longrightarrow \ker h$ exakt.

c) Ist β surjektiv, so ist $\operatorname{coker} f \longrightarrow \operatorname{coker} g \longrightarrow \operatorname{coker} h \longrightarrow \operatorname{coker} \beta' \longrightarrow 0$ exakt.

d) *Sind b) und c) erfüllt, so existiert ein R-Homomorphismus $\delta : \ker h \longrightarrow \operatorname{coker} f$ derart, daß die Sequenz*

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \ker\alpha & \longrightarrow & \ker f & \longrightarrow & \ker g & \longrightarrow & \ker h \\ & & & & & & & & \Big\downarrow{\scriptstyle\delta} \\ 0 & \longleftarrow & \operatorname{coker}\beta' & \longleftarrow & \operatorname{coker} h & \longleftarrow & \operatorname{coker} g & \longleftarrow & \operatorname{coker} f \end{array}$$

exakt wird.

Der R-Homomorphismus δ aus Prop. B.4 wird *Verbindungshomomorphismus* genannt.

Beweis. Die Existenz der induzierten R-Homomorphismen $\overline{\alpha}$, $\overline{\beta}$ etc. folgt aus Lemma B.3.

Damit ist für a) nur noch zu zeigen, daß $\ker f \longrightarrow \ker g \longrightarrow \ker h$ eine 0-Sequenz ist. Sei dazu $b \in \operatorname{im}\alpha$, also $b = f(a)$ für ein $a \in \ker f$. Wegen $\beta(b) = \beta(\alpha(a)) = 0$ (dies folgt aus der ursprünglichen exakten Sequenz $A \longrightarrow B \longrightarrow C$) ist aber in der Tat $\operatorname{im}\alpha \subseteq \ker\beta$. Der Beweis, daß auch $\operatorname{coker} f \longrightarrow \operatorname{coker} g \longrightarrow \operatorname{coker} h$ eine 0-Sequenz ist, läuft genauso.

Um auch $\ker\beta \subseteq \operatorname{im}\alpha$ zu zeigen, dürfen (und müssen) wir voraussetzen, daß $\alpha' : A' \longrightarrow B'$ injektiv ist. Sei nämlich $b \in \ker g$ und $\beta(b) = 0$; wegen der Exaktheit der ursprünglichen Sequenz existiert ein $a \in A$ mit $b = \alpha(a)$. Wir müssen zeigen, daß sogar $a \in \ker f$ gilt. Dazu stellen wir fest, daß $\alpha'(f(a)) = g(\alpha(a)) = g(b) = 0$ ist wegen $b \in \ker g$. Die Injektivität von α' zeigt dann, daß schon $f(a) = 0$ sein muß, d.h. es gilt tatsächlich $a \in \ker f$.

Als Nächstes bestimmen wir den Kern von $\overline{\alpha} : \ker f \longrightarrow \ker g$. Offenbar ist $\ker\overline{\alpha} = \{a \in \ker f : \alpha(a) = 0\} = \ker f \cap \ker\alpha$; wegen $\ker\alpha \subseteq \ker f$ (denn $\alpha(a) = 0$ impliziert $\alpha' \circ f(a) = g \circ \alpha(a) = 0$; da α' injektiv ist, muß sogar $f(a) = 0$, also $a \in \ker f$ sein) folgt aber $\ker\overline{\alpha} = \ker\alpha$. Damit ist b) vollständig bewiesen.

Das Nachrechnen der Exaktheit von $\operatorname{coker} f \longrightarrow \operatorname{coker} g \longrightarrow \operatorname{coker} h$ wird wieder dem ungeübten Leser überlassen. Damit bleibt uns noch, $\phi : \operatorname{coker} h \longrightarrow \operatorname{coker}\beta'$ zu konstruieren und die Surjektivität nachzuweisen. Die Konstruktion ist klar: wir setzen $\phi(c' + \operatorname{im} h) = c' + \operatorname{im}\beta'$. Wegen $\operatorname{im} h = \operatorname{im} h \circ \beta = \operatorname{im}\beta' \circ g$ ist $\operatorname{im} h \subseteq \operatorname{im}\beta'$, die Abbildung also wohldefiniert. Da sie offensichtlich surjektiv ist, haben wir c) bewiesen.

Damit sind wir beim Beweis von d) angelangt; wir beginnen mit der Definition des Verbindungshomomorphismus $\delta : \ker h \longrightarrow \operatorname{coker} f$. Wir sind in folgender Situation:

$$\begin{tikzcd}
 & & \ker g \arrow[r] \arrow[d] & \ker h \arrow[d] & \\
 & A \arrow[r, "\alpha"] \arrow[d, "f"] & B \arrow[r, "\beta"] \arrow[d, "g"] & C \arrow[r] \arrow[d, "h"] & 0 \\
0 \arrow[r] & A' \arrow[r, "\alpha'"] \arrow[d] & B' \arrow[r, "\beta'"] \arrow[d] & C' & \\
 & \operatorname{coker} f \arrow[r] & \operatorname{coker} g & &
\end{tikzcd}$$

Zu konstruieren ist ein R-Homomorphismus $\delta : \ker h \longrightarrow \operatorname{coker} f$; sei also $c \in \ker h$. Die Injektion $\ker h \longrightarrow C$ bekommen wir geschenkt. Auf dem Weg nach $\operatorname{coker} f$ dürfen wir jetzt aber nicht über C' laufen, denn wenn wir unser c mit h nach C' abbilden, bekommen wir die 0. Die einzige Möglichkeit weiterzukommen ist daher, mittels β von C nach B zu gehen. Dies ist in der Tat möglich: da $\beta : B \longrightarrow C$ nach Voraussetzung surjektiv ist, existiert ein $b \in B$ mit $\beta(b) = c$.

Wir dürfen jetzt aber nicht erwarten, mit α nach A zu kommen: wegen der Exaktheit der Sequenz ginge das nur, falls $c = 0$ wäre. Also müssen wir mit g nach B' gehen, d.h. wir bilden $g(b)$. Um jetzt mit α' nach A' zu kommen, müssen wir zeigen, daß $g(b)$ im Bild von α' liegt. Dies ist gleichbedeutend damit, daß $\beta'(g(b)) = 0$ ist. Aber jetzt gilt $\beta'(g(b)) = h(\beta(b)) = h(c) = 0$: also ist in der Tat $g(b) \in \operatorname{im} \alpha'$, d.h. $g(b) = \alpha'(a')$ für ein $a' \in A'$. Damit haben wir einen Kandidaten für $\delta(c)$ gefunden: wir würden gerne $\delta(c) = a' + \operatorname{im} f \in \operatorname{coker} f$ setzen. Dazu ist zu zeigen, daß die Abbildung $c \longmapsto a' + \operatorname{im} f$ nicht von der Auswahl von b abhängt. Nehmen wir also an, es wäre $\beta(b_1) = c$; wegen $\beta(b_1 - b) = 0$ liegt $b_1 - b$ in $\ker \beta = \operatorname{im} \alpha$, d.h. es ist $b_1 = b + \alpha(a)$ für ein $a \in A$. Damit ist $g(b + \alpha(a)) = g(b) + g \circ \alpha(a) = \alpha'(a') + \alpha' \circ f(a) = \alpha'(a' + f(a))$, d.h. a' und $a' + f(a)$ liegen in derselben Restklasse modulo $\operatorname{im} f$, und die Abbildung $c \longmapsto a' + \operatorname{im} f$ ist in der Tat wohldefiniert. Damit ist alles gezeigt. $\square$

Korollar B.5. *Sei eine Sequenz*

$$A \xrightarrow{\alpha} B \xrightarrow{\beta} C$$

gegeben; dann existiert eine exakte Sequenz

$$\begin{array}{ccccccc} 0 \longrightarrow & \ker\alpha & \longrightarrow & \ker(\beta\circ\alpha) & \longrightarrow & \ker\beta \\ & & & & & \downarrow \\ 0 \longleftarrow & \text{coker }\beta & \longleftarrow & \text{coker }(\beta\circ\alpha) & \longleftarrow & \text{coker }\alpha \end{array}$$

Beweis. Wende das Schlangenlemma auf das folgende kommutative und exakte Diagramm an:

$$\begin{array}{ccccccc} & & A & \xrightarrow{\alpha} & B & \longrightarrow & \text{coker }\alpha & \longrightarrow & 0 \\ & & \Big\downarrow{\scriptstyle\beta\circ\alpha} & & \Big\downarrow{\scriptstyle\beta} & & \Big\downarrow & & \\ 0 & \longrightarrow & C & \xrightarrow{id} & C & \longrightarrow & 0 & & \end{array}$$

□

Anhang Kapitel C

Endliche Körper

Die einfachsten endlichen Körper sind sicherlich die Restklassenkörper $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ für prime p. Die Ringe $\mathbb{Z}/p^2\mathbb{Z}$ sind dagegen keine Körper: bezeichnet $\overline{a}$ die Restklasse von $a \bmod p^2$, so ist ja $\overline{p} \cdot \overline{p} = \overline{0}$, d.h. $\overline{p}$ ist Nullteiler und insbesondere nicht invertierbar (natürlich ist $\overline{p} \neq \overline{0}$ auch nicht das Nullelement).

Man muß sich also etwas mehr anstrengen, um endliche Körper mit p^2 Elementen zu funden. Sei dazu p eine ungerade Primzahl und a ein quadratischer Nichtrest modulo p. Dann behaupten wir, daß

$$F = \mathbb{F}_p(\sqrt{a}\,) := \{x + y\sqrt{a} : x, y \in \mathbb{F}_p\}$$

ein endlicher Körper mit p Elementen ist. Addition und Subtraktion sind klar, Nullelement ist $0 = 0 + 0\sqrt{a}$, weiter ist $(x_1 + y_1\sqrt{a}\,)(x_2 + y_2\sqrt{a}\,) = x_1x_2 + ay_1y_2 + (x_1y_2 + x_2y_1)\sqrt{a}$. Schließlich hat $x + y\sqrt{a} \neq 0$ ein Inverses: es ist nämlich

$$\frac{1}{x + y\sqrt{a}} = \frac{x}{x^2 - ay^2} - \frac{y}{x^2 - ay^2}\sqrt{a},$$

und dabei kann nie $x^2 - ay^2 = 0$ sein, weil wegen $y \neq 0$ sonst $(xy^{-1})^2 = a$ wäre im Widerspruch zu $(a/p) = -1$. Da F offenbar p^2 Elemente besitzt, haben wir für alle ungeraden p einen endlichen Körper mit p^2 Elementen gefunden.

Für $p = 2$ scheint dieses Verfahren nicht zu funktionieren, weil $\mathbb{F}_2$ nur Quadrate enthält. Eine kleine Modifikation führt aber auch hier zum Erfolg.

Dazu schauen wir uns noch einmal an, was wir oben gemacht haben: zum endlichen Körper mit p Elementen haben wir eine Nullstelle des Polynoms $x^2 - a$ adjungiert; etwas algebraischer ausgedrückt ist unser obiges F nichts anderes als $\mathbb{F}_p[x]/(x^2 - a)$, wobei $\sqrt{a}$ mit $x \bmod (x^2 - a)$ identifiziert wird (man sehe: das Quadrat von $[x \bmod (x^2 - a)]^2 \equiv x^2 \bmod (x^2 - a) \equiv a \bmod (x^2 - a)$ wie gewünscht). Derselbe Trick funktioniert über $\mathbb{F}_2$, wenn man statt Polynomen $x^2 - a$ das Polynom $x^2 + x + 1$ wählt. Mit $\alpha = x \bmod (x^2 + x + 1)$ ist dann $\alpha^2 = [x^2 \bmod (x^2 + x + 1)] = [x + 1 \bmod (x^2 + x + 1)] = 1 + \alpha$, und wir haben $\mathbb{F}_4 = \{0, 1, \alpha, 1 + \alpha\}$.

Das einzige Problem bei dieser Konstruktion ist das Finden eines irreduziblen Polynoms (mit reduziblen Polynomen geht alles schief: ist $f = gh$ und setzt man $\alpha = g(x) \bmod f(x)$ und $\beta = h(x) \bmod f(x)$, so ist $\alpha\beta = f(x) \bmod f(x) = 0$, d.h. α und β sind Nullteiler). Mit demselben Trick kann man daher Körper mit p^n Elementen konstruieren, sobald es gelingt, ein über $\mathbb{F}_p$ irreduzibles Polynom f vom Grad n zu finden. Daß das aber immer möglich ist, kann man (im wesentlichen durch Abzählen) beweisen: ist z.B. $n = 2$, so gibt es p^2 quadratische Polynome $x^2 + ax + b$; die reduziblen haben die Form $(x - r)(x - s)$ mit $r, s \in \mathbb{F}_p$, und davon gibt es $\frac{1}{2}p(p + 1)$ viele. Also existieren genau $p^2 - \frac{1}{2}p(p + 1) = \binom{p}{2}$ über $\mathbb{F}_p$ irreduzible quadratische Polynome: insbesondere gibt es mindestens eines.

Wir zeigen nun zuerst den "kleinen Fermat":

Proposition C.1. *Für $x \in \mathbb{F}_q$ ist $x^q = x$.*

Beweis. Das ist einfach einzusehen: für $x = 0$ ist es ohnehin klar, für $x \neq 0$ ist x Element der multiplikativen Gruppe $\mathbb{F}_q^\times$; diese hat Ordnung $q - 1$, folglich ist $x^{q-1} = 1$, und Multiplikation mit x liefert die Behauptung. $\square$

Eine wichtige Eigenschaft endlicher Körper ist die folgende: sei $\mathbb{F}_q$ Körper mit $q = p^n$, so ist die Abbildung $\phi : \mathbb{F}_q \longrightarrow \mathbb{F}_q : x \longmapsto x^p$ ein Ringhomomorphismus, d.h. es gilt $\phi(x + y) = \phi(x + y)$ und $\phi(xy) = \phi(x)\phi(y)$. Die letzte Eigenschaft ist klar, die erste beruht auf der Tatsache, daß die Binomialkoeffizienten $\binom{p}{a}$ für $1 \leq a \leq p - 1$ durch p teilbar sind. Damit folgt nämlich

$$\begin{aligned}\phi(x + y) = (x + y)^p &= x^p + \binom{p}{1}x^{p-1}y + \ldots + \binom{p}{p-1}xy^{p-1} + y^p \\ &= x^p + y^p = \phi(x) + \phi(y).\end{aligned}$$

Der Homomorphismus ϕ ist injektiv: aus $0 = \phi(x) = x^p$ folgt nämlich, da $\mathbb{F}_q$ ein Körper ist und somit keine Nullteiler besitzt, daß $x = 0$ ist. Nun sind injektive Abbildungen zwischen endlichen Mengen automatisch surjektiv: folglich ist ϕ ein Isomorphismus! Insbesondere ist in $\mathbb{F}_q$ jedes Element eine p-te Potenz.

Dieser Isomorphismus ist das wesentliche Hilfsmittel zum Studium endlicher Körper. Beispielsweise gilt

Proposition C.2. *Ein $x \in \mathbb{F}_q$ liegt genau dann in $\mathbb{F}_p$, wenn $\phi(x) = x$ gilt.*

In der Tat: falls $x \in \mathbb{F}_p$ ist, so gilt sicher $x^p = x$ nach dem kleinen Satz von Fermat. Damit hat das Polynom $f(x) = x^p - x$ schon mindestens p verschiedene Nullstellen in $\mathbb{F}_q$, nämlich alle Elemente von $\mathbb{F}_p$. Da $\mathbb{F}_q$ ein Körper ist (in einem Körper hat jedes Polynom höchstens so viele Nullstellen wie der Grad angibt), kann es keine weiteren geben, d.h. aus $x = x^p$ und $x \in \mathbb{F}_q$ folgt automatisch $x \in F_p$.

Diese Eigenschaft kann man ausnützen, um zwei wichtige Abbildungen $\mathbb{F}_q \longrightarrow \mathbb{F}_p$ zu definieren, nämlich die Spur und die Norm. Ist $q = p^f$, so setzen wir

$$\mathrm{Tr}\,(x) := x + x^p + x^{p^2} + \ldots + x^{p^{f-1}}$$

und behaupten, daß Tr jedes Element aus $\mathbb{F}_q$ nach $\mathbb{F}_p$ abbildet. Dazu brauchen wir nach Proposition C.2 nur zu zeigen, daß $\mathrm{Tr}\,(x)^p = \mathrm{Tr}\,(x)$ ist. Das ist aber ganz einfach: wegen $x^{p^f} = x^q = x$ ist nämlich

$$\begin{aligned} \mathrm{Tr}\,(x)^p &= (x + x^p + x^{p^2} + \ldots + x^{p^{f-1}})^p \\ &= x^p + x^{p^2} + \ldots + x^{p^f} \;=\; \mathrm{Tr}\,(x) \end{aligned}$$

Die Eigenschaft $\mathrm{Tr}\,(x+y) = \mathrm{Tr}\,(x) + \mathrm{Tr}\,(y)$ ist inzwischen wohl offensichtlich.

Entsprechend definiert man die Norm durch

$$\mathrm{N}(x) := x \cdot x^p \cdot x^{p^2} \cdot \ldots \cdot x^{p^{f-1}}.$$

Aus demselben Grund wie oben ist $\mathrm{N}(x) \in \mathbb{F}_p$ für $x \in \mathbb{F}_q$, und hier ist $\mathrm{N}(xy) = \mathrm{N}(x)\mathrm{N}(y)$.

Satz C.3. *Die Normabbildung $\mathrm{N} : \mathbb{F}_q^\times \longrightarrow \mathbb{F}_p^\times$ ist surjektiv.*

Beweis. Nach Definition der Norm ist

$$\mathrm{N}(x) = x^m \quad \text{mit } m = 1 + p + p^2 + \ldots + p^{f-1}.$$

Damit ist $m = (p^f - 1)/(p-1)$, und der Kern der Normabbildung besteht aus allen Lösungen der Gleichung $x^m = 1$. Da F_q ein Körper ist, hat $x^m = 1$ maximal m Lösungen. Nach dem Isomorphiesatz im $\mathrm{N} \simeq \mathbb{F}_q^\times / \ker \mathrm{N}$ hat dann das Bild mindestens Ordnung $(q-1)/m = p - 1$; da das Bild in $\mathbb{F}_p^\times$ liegt, muß Gleichheit gelten, und wir sind fertig. □

Ganz entsprechend ist auch die Spur $\mathbb{F}_q \longrightarrow \mathbb{F}_p$ immer surjektiv. Zum Abschluß noch ein Resultat, das im Falle eines endlichen Körpers mit p Elementen auf die Existenz einer Primitivwurzel hinausläuft:

Satz C.4. *Die multiplikative Gruppe $\mathbb{F}_q^\times$ eines endlichen Körpers ist zyklisch.*

Beweis. Wir zeigen die beiden folgenden Aussagen:

- Für alle $n \in \mathbb{N}$ gibt es höchstens n Elemente $x \in \mathbb{F}_q^\times$ mit $x^n = 1$;
- Ist G eine endliche Gruppe, und gibt es für jedes $n \in \mathbb{N}$ höchstens n Elemente $g \in G$ mit $g^n = 1$, so ist G zyklisch.

Daraus folgt offenbar der Satz.

Der Beweis der ersten Behauptung ist einfach: jedes $x \in \mathbb{F}_q^\times$ mit $x^n = 1$ ist Nullstelle des Polynoms $X^n - 1$. In einem Körper hat aber jedes Polynom höchstens so viele Nullstellen, wie sein Grad angibt. [Bem.: In Ringen wird das i.a. falsch: z.B. hat das Polynom $X^2 + 1$ in dem Quaternionenring über $\mathbb{R}$ mindestens die Nullstellen i, j und k).

Der Beweis der zweiten Behauptung besteht in einem Vergleich von G und $Z = \mathbb{Z}/N\mathbb{Z}$, wo $N = \#G$ die Ordnung der Gruppe G ist. Zu jedem Element $g \in G$ betrachten wir die von g erzeugte Untergruppe $H = \langle g \rangle$; bezeichnet d deren Ordnung, so ist d ein Teiler der Gruppenordnung N. Ist z ein erzeugendes Element von Z, so ist $W_d = \langle z^{N/d} \rangle$ eine Untergruppe der Ordnung d von Z.

Nun gilt $h^d = 1$ für alle $h \in H$; da H genau d Elemente hat und es höchstens d Elemente $g \in G$ mit $g^d = 1$ gibt, enthält H *alle* Elemente der Ordnung d von G. Wegen $H \simeq W_d$ gilt also

$$\begin{aligned} \#\{\text{Elemente der Ordnung } d \text{ in } G\} &= \#\{\text{Elemente der Ordnung } d \text{ in } W_d\} \\ &\le \#\{\text{Elemente der Ordnung } d \text{ in } Z\} \end{aligned}$$

für jedes $d \in \mathbb{N}$. Nun ist G die disjunkte Vereinigung aller Mengen

$$\{\text{Elemente der Ordnung } d \text{ in } G\}$$

über alle $d \in \mathbb{N}$; wäre auch nur eine der obigen Ungleichungen streng, würde G weniger Elemente haben als Z: Widerspruch. Also enthalten G und Z jeweils gleich viel Elemente der Ordnung d für jedes $d \mid N$; insbesondere gibt es in G ein Element der Ordnung N: also ist G zyklisch. $\square$

References

[Ab] N.H. Abel, *Recherches sur les fonctions elliptiques*, J. Reine Angew. Math. **2** (1827), 101–181; ibid. **3** (1828), 160–190; Œuvres I, 263–388

[All] N.L. Alling, *Real Elliptic Curves*, Mathematics Studies **54**, North Holland, 1981

[Ay] R. Ayoub, *The lemniscate and Fagnano's contributions to elliptic integrals*, Archive for the History of Exact Sciences **29** (1984), 131–149

[Ba1] I.G. Bashmakova, *Diophantus and diophantine equations*, Updated by Joseph Silverman, The Dolciani Mathematical Expositions **20**, MAA (1997), 90 p. $ 21.95;

[Ba2] I.G. Bashmakova, *Diophantine equations and the evolution of algebra*, Transl. Amer. Math. Soc. **147** (1990), 85–100

[Ba3] I.G. Bashmakova, *Arithmetic of algebraic curves from Diophantus to Poincaré*, Hist. Math. **8** (1981), 393–416

[Ba4] I.G. Bashmakova, *Diophantos und Fermat (zur Geschichte der Methode der Tangenten und Extrema)* (Russ.), Istor.-Mat. Issled. **17** (1966), 185–204; frz. Übersetzung Revue d'Histoire des Sciences **19** (1968), 283–306

[BS] .G. Bashmakova, E.I. Slavutin, *Glimpses of algebraic geometry*, Am. Math. Mon. **104** (1997), 62–67

[Can] M. Cantor, *Vorlesungen über die Geschichte der Mathematik*, 3 Bände, Teubner; Johnson reprint 1965

[Cas] J.W.S. Cassels, *Lectures on elliptic curves*, Cambridge University Press, 1991, 143pp, \$ 27

[Ch1] J.S. Chahal, *Topics in number theory*, Plenum Press, 191 pp. \$ 75.00 (1988).

[Ch2] J.S. Chahal, *Manin's proof of the Hasse inequality revisited*, Nieuw Arch. Wiskd., IV. Ser. **13** (1995), 219–232

[Co1] H. Cohen, *A course in computational algebraic number theory*, Graduate Texts in Mathematics. 138. Berlin: Springer-Verlag, xxi, 534 p. DM 88.00 (1993)

[Co2] H. Cohen, *Elliptic Curves*, Waldschmidt, Michel (ed.) et al., From number theory to physics. Lectures of a meeting on number theory and physics held at the Centre de Physique, Les Houches (France), March 7-16, 1989. Berlin: Springer-Verlag, 212-237 (1992)

[Coh] H. Cohn, *Introductory remarks on complex multiplication*, Int. J. Math. Math. Sci. **5** (1982), 675–690

[CSS] G. Cornell, J.H. Silverman, G. Stevens (eds.), *Modular forms and Fermat's last theorem*, Papers from a conference, Boston, MA, USA, August 9–18, 1995; Springer 1997, 582pp, DM 89.–

[Co1] D.A. Cox, *Primes of the form $x^2 + ny^2$. Fermat, class field theory, and complex multiplication*, Wiley, 351 pp \$ 49.95 (1989)

[Co2] D.A. Cox, *Introduction to Fermat's Last Theorem*, Amer. Math. Mon. **101** (1994), 3–14

[Cre] J.E. Cremona, *Algorithms for Modular Elliptic Curves*, 2. Auflage Cambridge Univ. Press 1997, 377 pp, \$ 70

[D] D. Doud, *A procedure to calculate torsion of elliptic curves over $\mathbb{Q}$*, Manuscripta Math. **95** (1998), 463–469

[Enn] A. Enneper, *Elliptische Funktionen. Theorie und Geschichte*, Halle 1890

[Fr] G. Frey, *Elementare Zahlentheorie*, Vieweg & Sohn, 119 S. DM 19.80 (1984)

[Gou] F.Q. Gouvea, *A marvelous proof*, Amer. Math. Mon. **101** (1994), 203–222

[Hil] S. Hildebrandt, *Von mathematischer Kultur*, DMV Mitteilungen **4** (1994), 12–20

[Hur] A. Hurwitz, *Über ternäre Diophantische Gleichungen dritten Grades*, Vierteljahrschrift d. Naturf. Ges. Zürich **62** (1917), 207–229

[Hus] D. Husemöller, *Elliptic curves*, Graduate Texts in Mathematics, Springer-Verlag, 1987

[Ju] C. Juel, *Ueber die Parameterbestimmung von Punkten auf Curven zweiter und dritter Ordnung. Eine geometrische Einleitung in die Theorie der logarithmischen und elliptischen Funktionen*, Math. Ann. **47** (1896), 72–104

[Kna] A. W. Knapp, *Elliptic curves*, Mathematical Notes, Princeton University Press, 1992

[Kob] N. Koblitz, *Introduction to elliptic curves and modular forms*, Graduate Texts in Mathematics, Springer-Verlag, 1984

[Kob] N. Koblitz, *Algebraic aspects of cryptography. With an appendix on Hyperelleptic curves by Alfred J. Menezes, Yi-Hong Wu, and Robert J. Zuccherato*, Springer, 206 pp, DM 98.00; (1998)

[Ko] H. Koch, *Zahlentheorie: algebraische Zahlen und Funktionen*, Vieweg 1997

[KK] M. Koecher, A. Krieg, *Elliptische Funktionen und Modulformen*, Springer 1998

[Kra] H. Kraft, *Algebraische Kurven und diophantische Gleichungen*, in “Lebendige Zahlen. Fünf Exkursionen” (Borho et al.), Birkhäuser Verlag (1981) 93-114

[Law] D.F. Lawden, *Elliptic functions and applications*, Appl. Math. Sci. **80**, Springer-Verlag 1989

[L1] E. Lutz, *Les solutions de l'équation $y^2 = x^3 - Ax - B$ dans les corps p-adiques*, C. R. Acad. Sci., Paris **202** (1936), 20–22

[L2] E. Lutz, *Sur l'équation $y^2 = x^3 - Ax - B$ dans les corps p-adiques*, J. Reine Angew. Math. **177** (1937), 238–247.

[1] K. Mahler, *On the division values of Weierstrass' $\wp$-function*, Quart. J. Math. **6** (1935), 74–77

[Ma] Yu. I. Manin, *On cubic congruences to a prime modulus*, Transl. Amer. Math. Soc. **13** (1960), 1–7

[MM] H. McKean, V. Moll, *Elliptic Curves. Function Theory, Geometry, Arithmetic*, Cambridge Univ. Press 1997, 294pp, \$ 60

[Men] A.J. Menezes, *Elliptic curve public key cryptosystems*, Kluwer Academic Publishers, \$ 79.95 (1993)

[MP] Moonshine Page,
`http://www-cicma.concordia.ca/faculty/cummins/moonshine.html`

[NS] W. Neutsch, K. Scherer, *Celestial Mechanics*, BI 1982

[Pin] R. Pinch, *Computational Number Theory*, Cambridge Univ. Press 1997, 250 pp, \$ 23 (paperback)

[Ros] M. Rosen, *Abel's theorem on the lemniscate*, Amer. Math. Monthly **88** (1981), 387–395

[Sch] N. Schappacher, *Développement de la loi de groupe sur une cubique*, [CA] Semin. Théor. Nombres, Paris/Fr. 1988-89, Prog. Math. **91** (1990), 159–184

[SS] N. Schappacher, R, Schoof, *Beppo Levi and the arithmetic of elliptic curves*, Math. Intell. 18 (1996), 57–69

[Scr] C.J. Scriba, *Zur Geschichte der Bestimmung rationaler Punkte auf elliptischen Kurven: Das Problem von Beha-Eddin Àmuli*, Ber. Sitz. Joachim Jungius-Ges. Wiss., Hamburg **1** (1982/83), No.6, 52 S. (1984).

[Ser] J.-P. Serre, *A course in arithmetic*, Springer-Verlag 1978

[Sie] C.L. Siegel, *Vorlesungen über ausgewählte Kapitel der Funktionentheorie. Teil I*, Mathem. Inst. Univ. Göttingen 1965

[Sil] J. Silverman, *The arithmetic of elliptic curves*, Graduate Texts in Mathematics, Springer-Verlag, 1986

[ST] J. Silverman, J. Tate, *Rational points on elliptic curves*, Springer-Verlag, 1992

[Sk] N. Skoruppa, *Heights*, Vorlesungsskript 1999, `http://www.math.u-bordeaux.fr/~skoruppa/`

[To] G. Toth, *Glimpses of Algebra and Geometry*, Springer 1998

[Wei] A. Weil, *Number theory. An approach through history. From Hammurapi to Legendre*, 1984

Index