

Elliptische Kurven

Sommersemester 2020

Universität Bayreuth

MICHAEL STOLL

INHALTSVERZEICHNIS

1. Einführung	3
2. Affine ebene Kurven	7
3. Projektive ebene Kurven	10
4. Schnitte von Kurven mit Geraden	13
5. Glattheit	15
6. Rationale Abbildungen und Morphismen	17
7. Elliptische Kurven: Definition	19
8. Isomorphismen elliptischer Kurven	23
9. Gruppenstruktur	26
10. Isogenien und Endomorphismen	31
11. Torsion und Weil-Paarung	39
12. Elliptische Kurven über endlichen Körpern	43
13. Faktorisierung und Primzahltest: Grundlagen	48
14. Faktorisierung und Primzahltest mit elliptischen Kurven	56
15. Kryptographie: Grundlagen	62
16. Kryptographie: Elliptische Kurven	69
17. Elliptische Funktionen	73
18. Gitter und elliptische Kurven	80
19. Modulformen und Modulfunktionen	86
20. Die rationale Torsionsuntergruppe	97
21. Gute und schlechte Reduktion	103
22. Der Satz von Mordell	108
23. Der schwache Satz von Mordell	114
24. Eine bessere Schranke für den Rang	120
25. Die Vermutung von Birch und Swinnerton-Dyer	132

1. EINFÜHRUNG

In diesem Einführungskapitel möchte ich — gewissermaßen als Appetithappen — in groben Zügen erklären, wie man elliptische Kurven zur Faktorisierung großer Zahlen verwenden kann. Die Einzelheiten werden im Verlauf der Vorlesung ausführlich erläutert werden.

Für die Zwecke dieser Einführung sei eine elliptische Kurve E einfach eine Gleichung

$$(1.1) \quad E: y^2 = x^3 + ax + b$$

in den Variablen x und y mit Koeffizienten a und b aus einem Körper K (der Charakteristik $\neq 2$), wobei wir noch verlangen, dass $4a^3 + 27b^2 \neq 0$ ist, sonst ist die Kurve nicht „glatt“. Dann können wir die Menge der K -rationalen Punkte von E , geschrieben $E(K)$, definieren als die Menge der Lösungen $(\xi, \eta) \in K \times K$ der Gleichung (1.1). Es gibt gute Gründe (die bald erklärt werden), zu dieser Menge noch einen Punkt O „im Unendlichen“ hinzuzunehmen. Wir setzen also

$$E(K) = \{(\xi, \eta) \in K \times K \mid \eta^2 = \xi^3 + a\xi + b\} \cup \{O\}.$$

Was hat man davon? Einmal davon abgesehen, dass algebraische Kurven wie E an sich ein interessantes Studienobjekt darstellen, ist das Besondere an elliptischen Kurven, dass ihre (rationalen) Punkte in natürlicher Weise eine *abelsche Gruppe* bilden. Diese Gruppenstruktur lässt sich geometrisch kurz und prägnant definieren: O ist das Nullelement, und die Summe dreier Punkte, die auf einer Geraden liegen, ist O . Man muss dabei nur darauf achten, dass man die Schnittpunkte von Gerade und Kurve mit der richtigen Vielfachheit zählt (Tangente in einem Punkt ergibt Vielfachheit 2, eine Wendetangente sogar 3) und dass man im Falle einer senkrechten Geraden O als dritten Schnittpunkt interpretieren muss. Dies ergibt sich ganz natürlich, wenn man E als *projektive Kurve* betrachtet. Aus der geometrischen Interpretation bekommt man schnell folgende Formeln.

$$\begin{aligned} -(\xi, \eta) &= (\xi, -\eta) \\ (\xi, \eta) + (\xi, -\eta) &= O \\ (\xi_1, \eta_1) + (\xi_2, \eta_2) &= (\xi_3, \eta_3) \quad \text{mit} \quad \xi_3 = \lambda^2 - \xi_1 - \xi_2, \quad \eta_3 = -\lambda\xi_3 - \mu, \end{aligned}$$

wobei

$$\lambda = \begin{cases} \frac{\eta_2 - \eta_1}{\xi_2 - \xi_1}, & \text{falls } \xi_1 \neq \xi_2 \\ \frac{3\xi_1^2 + a}{2\eta_1}, & \text{falls } \xi_1 = \xi_2 \text{ und } \eta_1 \neq -\eta_2, \end{cases}$$

und $\mu = \eta_1 - \lambda\xi_1$ ist; dabei ist $y = \lambda x + \mu$ die Gleichung der Geraden durch die beiden Punkte, bzw. der Tangente.

Diese Formeln sehen auf den ersten Blick kompliziert aus, zeigen aber ganz klar, dass man in dieser Gruppe problemlos rechnen kann. (Die Assoziativität der Addition mit diesen Formeln nachzurechnen ist übrigens eine undankbare Aufgabe. Es gibt bessere Möglichkeiten.)

1.1. Beispiel. Als Beispiel betrachten wir die Kurve

$$E: y^2 = x^3 - 43x + 166.$$

Sie hat den rationalen Punkt $P = (3, 8) \in E(\mathbb{Q})$. Wir berechnen

$$2 \cdot P = (-5, -16), \quad 3 \cdot P = P + 2 \cdot P = (11, -32), \quad 4 \cdot P = (11, 32) = -3 \cdot P.$$

BSP
Addition
von Punkten

Also ist $7 \cdot P = O$. (Tatsächlich ist hier $E(\mathbb{Q}) \cong \mathbb{Z}/7\mathbb{Z}$, erzeugt von P . Im Allgemeinen braucht $E(\mathbb{Q})$ nicht endlich zu sein, ist aber immer endlich erzeugt (Satz von Mordell). Später in der Vorlesung werden wir elliptische Kurven über \mathbb{Q} ausführlicher behandeln.) ♣

Wie kann man diese Eigenschaft nun für die Faktorisierung nutzbar machen? Dazu müssen wir zunächst den Fall betrachten, dass der Grundkörper K ein endlicher Körper \mathbb{F}_p ist. In diesem Fall ist die Gruppe $E(K)$ natürlich ebenfalls endlich. Man weiß sogar ziemlich genau, wie groß sie ist: Es gilt $\#E(\mathbb{F}_p) = p + 1 - t$ mit $|t| \leq 2\sqrt{p}$. (Für jedes $\xi \in \mathbb{F}_p$ gibt es durchschnittlich ein $\eta \in \mathbb{F}_p$, das die Gleichung löst. Zusammen mit O ergibt das den Term $p + 1$. Die Aussage gibt also eine genaue Schranke für die Abweichung von diesem durchschnittlichen Verhalten.)

1.2. Beispiel. Wir haben folgende Tabelle für die Größen $\#E_a^\pm(\mathbb{F}_{23})$, wo wir für $a \in \mathbb{F}_{23}$ die Kurven $E_a^\pm: y^2 = x^3 \pm x + a$ betrachten. (Ein Strich steht für eine singuläre Kurve.) **BSP** $\#E(\mathbb{F}_p)$

a	0	1	2	3	4	5	6	7	8	9	10	11
$\#E_a^+$	24	28	24	27	29	22	21	18	28	20	32	33
$\#E_a^-$	24	—	30	30	31	18	22	28	21	32	23	25

a	12	13	14	15	16	17	18	19	20	21	22	
$\#E_a^+$	15	16	28	20	30	27	26	19	21	24	20	
$\#E_a^-$	23	25	16	27	20	26	30	17	18	18	—	

Dazu kommen noch die beiden Kurven $y^2 = x^3 \pm 1$ mit jeweils 24 Punkten. Man kann zeigen, dass in dieser Liste jede elliptische Kurve über \mathbb{F}_{23} genau einmal „bis auf Isomorphie“ vorkommt.

Hier ist $|t| \leq \lfloor 2\sqrt{23} \rfloor = 9$, und wir haben folgende Verteilung:

t	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9
	1	2	1	4	1	4	3	2	2	6	2	2	3	4	1	4	1	2	1

Man sieht, dass alle Möglichkeiten vorkommen, und dass die Verteilung einigermaßen gleichmäßig ist. ♣

Es ist kein Zufall, dass die Tabelle der Häufigkeiten der t -Werte im Beispiel oben symmetrisch zu null ist. Wir nehmen an, dass p ungerade ist, Sei $d \in \mathbb{F}_p^\times$ kein Quadrat. Dann induziert die Abbildung

$$E: y^2 = x^3 + ax + b \quad \mapsto \quad E': y^2 = x^3 + d^2ax + d^3b$$

eine Involution auf der Menge der Isomorphieklassen von elliptischen Kurven über \mathbb{F}_p , und es gilt

$$\#E(\mathbb{F}_p) = p + 1 - t \quad \implies \quad \#E'(\mathbb{F}_p) = p + 1 + t.$$

Es folgt, dass die Anzahl der Isomorphieklassen mit $t = c$ dieselbe ist wie die Anzahl der Isomorphieklassen mit $t = -c$.

Wie kann man nun elliptische Kurven zum Faktorisieren benutzen? Dazu betrachten wir erst einmal eine andere Methode, die den Ansatz mit elliptischen Kurven inspiriert hat. Das ist die „ $(p - 1)$ -Methode von Pollard“.

Sei N eine (große) zusammengesetzte Zahl, die keine Primzahlpotenz ist (beide Eigenschaften lassen sich recht leicht nachweisen, ohne dass man dafür eine Faktorisierung von N finden muss). Wir wollen einen echten Teiler $d \neq 1$ von N finden. Dazu wählen wir zufällig eine Zahl $a \in \{2, \dots, N - 1\}$. Falls

$d = \text{ggT}(a, N) > 1$ ist, dann ist d ein echter Teiler von N , und wir sind schon fertig. Anderenfalls ist a modulo N invertierbar. Wir wählen noch eine Zahl L und setzen $B = \text{kgV}(1, 2, \dots, L)$. Dann berechnen wir $d = \text{ggT}(a^B - 1, N)$. Dazu berechnet man am besten $b = a^B \bmod N$ durch sukzessives Quadrieren und dann $d = \text{ggT}(b - 1, N)$. Der Aufwand dafür ist etwa $(\log B)(\log N)^2$ (oder sogar nur $(\log B)(\log N)(\log \log N)(\log \log \log N)$), und es gilt $\log B \approx L$. Wenn $1 < d < N$ ist, dann haben wir den gesuchten Faktor gefunden.

Wann können wir damit rechnen, einen Faktor zu finden? Das wird wahrscheinlich dann passieren, wenn N Primteiler p und q hat, sodass $p - 1$ ein Teiler von B ist (das bedeutet, dass jede Primzahlpotenz, die $p - 1$ teilt, $\leq L$ sein muss), aber $q - 1$ nicht. Dann ist $a^B - 1$ durch p teilbar, denn $a^{p-1} \equiv 1 \pmod p$. Auf der anderen Seite ist $a^B - 1$ sehr wahrscheinlich nicht durch q teilbar, denn sonst müsste a eine k te Potenz mod q sein mit $k = (q - 1) / \text{ggT}(B, q - 1) \geq 2$. Die Wahrscheinlichkeit dafür ist etwa $1/k$. Wenn wir also mehrere Werte von a versuchen, werden wir ziemlich sicher einen erwischen, für den das Verfahren funktioniert. Wir können also erwarten, dass $d = \text{ggT}(a^B - 1, N)$ durch p , aber nicht durch q teilbar ist; damit ist d ein nichttrivialer Teiler von N .

In der Praxis wird man eine Folge von Werten von B verwenden, die man durch sukzessive Multiplikation

$$2 \cdot 3 \cdot 2 \cdot 5 \cdot 7 \cdot 2 \cdot 3 \cdot 11 \cdot 13 \cdot 2 \cdot 17 \cdot 19 \cdot 5 \cdot 3 \cdot 29 \cdot 31 \cdot \dots$$

erhält; die Folge der Faktoren kommt dabei aus der Folge der Primzahlpotenzen

$$2, 3, 2^2, 5, 7, 2^3, 3^2, 11, 13, 2^4, 17, 19, 5^2, 3^3, 29, 31, \dots$$

Das Problem bei dieser Methode ist, dass sie nur funktioniert, wenn N Primteiler mit den richtigen Eigenschaften hat.

Hier kommen nun elliptische Kurven ins Spiel. Hendrik Lenstra hatte die Idee, die multiplikative Gruppe, die wir eben verwendet haben, durch die Gruppe der Punkte auf einer elliptischen Kurve über $\mathbb{Z}/N\mathbb{Z}$ zu ersetzen. Man hat dann eine recht große Auswahl an Gruppen zur Verfügung und kann hoffen, bald eine zu erwischen, für die die Ordnung über \mathbb{F}_p im obigen Sinne „ L -glatt“ ist, aber die über \mathbb{F}_q nicht. Wir wählen also zufällig eine elliptische Kurve E mit Koeffizienten $a, b \in \mathbb{Z}/N\mathbb{Z}$ zusammen mit einem Punkt $P = (\xi, \eta)$ auf E (mit $\xi, \eta \in \mathbb{Z}/N\mathbb{Z}$). Man kann zum Beispiel a zufällig wählen und

$$E: y^2 = x^3 + ax - a, \quad P = (1, 1)$$

setzen. Wir können E und P auch mit Koeffizienten in \mathbb{F}_p betrachten; dann schreiben wir \tilde{E} und \tilde{P} . Es gilt dann $(p + 1 - t) \cdot \tilde{P} = \tilde{O}$, wenn $\#\tilde{E}(\mathbb{F}_p) = p + 1 - t$ ist. Analog zu eben multiplizieren wir P mit $B = \text{kgV}(1, 2, \dots, L)$. Wenn $p + 1 - t$ ein Teiler von B ist, dann gilt $B \cdot \tilde{P} = \tilde{O}$. Normalerweise wird aber nicht gelten, dass $B \cdot P = O$ ist. Das führt dann dazu, dass während der Rechnung eine Division in $\mathbb{Z}/N\mathbb{Z}$ auszuführen ist durch ein Element, das nicht 0, aber auch nicht invertierbar ist. Die dabei stattfindende ggT-Berechnung liefert uns einen nichttrivialen Teiler von N (üblicherweise ist das p).

Damit das Verfahren in der Praxis funktioniert, muss man eine gute Chance haben, B nicht zu groß zu wählen, sodass $m = \#\tilde{E}(\mathbb{F}_p)$ ein Teiler von B ist. Tatsächlich kann man zeigen, dass man bei optimaler Wahl von L und damit B einen Algorithmus erhält, dessen (erwartete) Laufzeit etwa durch

$$C e^{(\sqrt{2} + o(p)) \sqrt{(\log p)(\log \log p)}}$$



H.W. Lenstra
(*1949)

Foto © MFO

beschränkt ist — der Algorithmus ist *subexponentiell*. Dabei ist p der kleinste Primteiler von N , und $o(p)$ steht für eine Funktion von p , die für $p \rightarrow \infty$ gegen null geht.

1.3. Beispiel. Als Baby-Beispiel wollen wir die Zahl $N = 851$ faktorisieren. Wir nehmen als Kurve $E: y^2 = x^3 + 9x - 9$ über $\mathbb{Z}/851\mathbb{Z}$ mit dem Punkt $P = (1, 1)$. Um $B \cdot P$ zu berechnen, berechnen wir der Reihe nach $P_0 = P$, $P_1 = 2 \cdot P_0$, $P_2 = 3 \cdot P_1$, $P_3 = 2 \cdot P_2$, $P_4 = 5 \cdot P_3$ und so weiter. Auf diese Weise sammelt man gerade die kleinsten gemeinsamen Vielfachen der ersten natürlichen Zahlen an. Nun zur eigentlichen Rechnung.

BSP
Faktorisierung

- (1) $P_1 = 2 \cdot P_0$:
Wir haben $\lambda = 6, \mu = 846$, also $P_1 = (34, 652)$.
- (2) $P_2 = 3 \cdot P_1$:
Zunächst $Q = 2 \cdot P_1$. Wir haben $\lambda = 374, \mu = 701$, also $Q = (244, 802)$. Jetzt $P_2 = P_1 + Q$. Wir haben $\lambda = 487, \mu = 263$ und damit $P_2 = (313, 486)$.
- (3) $P_3 = 2 \cdot P_2$:
 $\lambda = 502, \mu = 795$, also $P_3 = (333, 537)$.
- (4) $P_4 = 5 \cdot P_3$:
Zunächst $Q_1 = 2 \cdot P_3$: $\lambda = 305, \mu = 241$ und $Q_1 = (451, 66)$.
Dann $Q_2 = 2 \cdot Q_1$: $\lambda = 832, \mu = 125$ und $Q_2 = (310, 659)$.
Schließlich $P_4 = P_3 + Q_2$. Der Nenner des Ausdrucks für λ ergibt sich zu 23, was nicht invertierbar ist. Also ist $23 = \text{ggT}(851, 23)$ ein nicht-trivialer Teiler, und wir haben die Faktorisierung $851 = 23 \cdot 37$ gefunden.

Der Hintergrund ist, dass in $E(\mathbb{F}_{23})$ der Punkt P die Ordnung 10 hat, also ist dort $P_4 = O$. Demgegenüber hat P in $E(\mathbb{F}_{37})$ die Ordnung 29, und damit ist P_4 dort nicht der Punkt O . ♣

2. AFFINE EBENE KURVEN

Elliptische Kurven sind spezielle ebene algebraische Kurven. Deswegen müssen wir uns erst einmal ein wenig mit diesen vertraut machen, auch wenn damit zunächst eine Häufung von neuen Begriffen verbunden ist. Allerdings können wir aus Zeitgründen nicht wirklich substantiell in die *Algebraische Geometrie* einsteigen, die für die allgemeine Behandlung derartiger Objekte zuständig ist.

Naiv gesprochen, beschreibt eine *affine ebene Kurve* die Menge der Punkte der Ebene, deren Koordinaten eine Polynomgleichung in zwei Variablen lösen. Um diese Vorstellung zu formalisieren, müssen wir erst einmal die Ebene, in der sich alles abspielt, beschreiben.

Hier und im Folgenden sei K ein (beliebiger) Körper; wir fixieren einen algebraischen Abschluss \bar{K} . Dieser Körper K ist unser *Grundkörper*; aus ihm kommen die Koeffizienten der Gleichungen und (meistens) die Koordinaten der Punkte, die wir betrachten.

2.1. Definition. Die *affine Ebene* \mathbb{A}_K^2 über K hat folgende Eigenschaften.

DEF
affine Ebene

- (1) Für jeden Erweiterungskörper $L \supset K$ ist die Menge der *L-rationalen Punkte* von \mathbb{A}_K^2 gegeben durch

$$\mathbb{A}_K^2(L) = \{(\xi, \eta) \mid \xi, \eta \in L\} = L \times L.$$

- (2) Eine *reguläre Funktion* auf \mathbb{A}_K^2 ist gegeben durch ein Polynom $f \in K[x, y]$. Für jeden Erweiterungskörper $L \supset K$ definiert f (durch Einsetzen der Koordinaten) eine Funktion

$$f_L: \mathbb{A}_K^2(L) \longrightarrow L, \quad (\xi, \eta) \longmapsto f(\xi, \eta).$$

Die Funktion $f_{\bar{K}}$ bestimmt dabei f eindeutig.

Der Ring der regulären Funktionen $K[x, y]$ auf \mathbb{A}_K^2 heißt auch der *affine Koordinatenring* von \mathbb{A}_K^2 und wird mit $K[\mathbb{A}_K^2]$ bezeichnet.

- (3) Eine *rationale Funktion* auf \mathbb{A}_K^2 ist gegeben durch ein Element $f = g/h \in K(x, y)$. Dabei ist $K(x, y)$ der Quotientenkörper von $K[x, y]$.

f heißt *regulär* im Punkt $P = (\xi, \eta) \in \mathbb{A}_K^2(L)$, wenn $h(\xi, \eta) \neq 0$ ist. f definiert dann für jedes $L \supset K$ eine Funktion

$$f_L: \{P \in \mathbb{A}_K^2(L) \mid f \text{ regulär in } P\} \longrightarrow L.$$

Es gilt wieder, dass f durch $f_{\bar{K}}$ eindeutig bestimmt ist.

Die regulären Funktionen sind dann gerade die rationalen Funktionen, die überall (d.h. auf $\mathbb{A}_K^2(L)$ für alle L) regulär sind.

Der Körper $K(x, y)$ der rationalen Funktionen auf \mathbb{A}_K^2 wird auch der *Funktionskörper* von \mathbb{A}_K^2 genannt und mit $K(\mathbb{A}_K^2)$ bezeichnet. \diamond

Diese Definition ist *operational*, d.h. sie sagt nicht so sehr, was \mathbb{A}_K^2 „ist“, sondern eher, was man damit macht. Wer sich damit nicht so wohl fühlt, kann sich in erster Näherung vorstellen, dass die affine Ebene die Zuordnung $L \mapsto L \times L$ „ist“, die einem Erweiterungskörper L von K die Menge der *L-rationalen Punkte* zuordnet. Allerdings gehören die regulären und rationalen Funktionen wesentlich mit zum Bild (wie die differenzierbaren, holomorphen oder meromorphen Funktionen in der Analysis). Wenn man es ganz richtig macht (in der modernen Algebraischen Geometrie), dann definiert man die Objekte wie \mathbb{A}_K^2 als „geringste Räume“, die beide Strukturen beinhalten. (In der klassischen Algebraischen Geometrie ist

der Grundkörper K algebraisch abgeschlossen (oder sogar \mathbb{C}); dann kommt man einigermaßen zurecht, wenn man ein Objekt wie die affine Ebene mit der Menge seiner (K -rationalen) Punkte identifiziert. Über einem beliebigen K ist das nicht mehr sinnvoll.)

2.2. Bemerkung. Völlig analog definiert man \mathbb{A}_K^n , den n -dimensionalen affinen Raum über K .

BEM
♠ n -dim.
affiner Raum

2.3. Definition. Eine affine ebene Kurve C über K ist gegeben durch ein nicht konstantes Polynom $f \in K[x, y]$. Wir schreiben $C: f(x, y) = 0$.

DEF
affine
ebene Kurve

(1) Für jeden Erweiterungskörper $L \supset K$ ist die Menge der L -rationalen Punkte von C gegeben durch

$$C(L) = \{P \in \mathbb{A}_K^2(L) \mid f_L(P) = 0\} = \{(\xi, \eta) \in L \times L \mid f(\xi, \eta) = 0\}.$$

(2) Eine reguläre Funktion auf C ist eine Äquivalenzklasse von Polynomen aus $K[x, y]$, wobei zwei Polynome äquivalent heißen, wenn ihre Differenz durch f teilbar ist. Ist g ein Repräsentant einer solchen Äquivalenzklasse, dann haben wir Funktionen

$$g_L: C(L) \ni (\xi, \eta) \mapsto g(\xi, \eta) \in L,$$

die nur von der Klasse abhängen (denn $f_L = 0$ auf C).

Die regulären Funktionen auf C bilden einen Ring, den affinen Koordinatenring $K[C]$. Er ist isomorph zu $K[x, y]/K[x, y] \cdot f$.

(3) Eine rationale Funktion auf C ist eine Äquivalenzklasse von rationalen Funktionen $g/h \in K(x, y)$, sodass f und h keinen nicht-konstanten gemeinsamen Teiler haben. Dabei sind g_1/h_1 und g_2/h_2 äquivalent, wenn f ein Teiler von $g_1h_2 - g_2h_1$ ist.

Eine rationale Funktion ϕ heißt regulär in $P \in C(L)$, wenn es einen Repräsentanten g/h gibt mit $h_L(P) \neq 0$. Wir haben dann für jedes L eine Funktion

$$\phi_L: \{P \in C(L) \mid g/h \text{ regulär in } P\} \longrightarrow L, \quad P \longmapsto \frac{g_L(P)}{h_L(P)}.$$

(4) C heißt irreduzibel, wenn f irreduzibel ist. C heißt geometrisch irreduzibel, wenn f absolut irreduzibel (d.h. irreduzibel in $\bar{K}[x, y]$) ist.

Wenn C irreduzibel ist, dann ist $K[x, y] \cdot f$ ein Primideal, also ist der Koordinatenring $K[C]$ ein Integritätsbereich. Die rationalen Funktionen auf C bilden dann gerade den Quotientenkörper von $K[C]$, den Funktionenkörper $K(C)$ von C . ◇

Die Bedingung mit dem gemeinsamen Teiler in der Definition der rationalen Funktionen auf C sichert, dass so eine Funktion in allen Punkten von C mit Ausnahme von endlich vielen regulär ist.

2.4. Beispiele.

BSP
affine
ebene Kurven

(1) Als ein triviales Beispiel betrachten wir die „ x -Achse“ $C: y = 0$. Es ist also $f = y$, und die rationalen Punkte sind $C(L) = L \times \{0\}$. Für den Koordinatenring haben wir $K[C] = K[x, y]/K[x, y] \cdot y \cong K[x]$, und der Funktionenkörper ist $K(C) \cong K(x)$.

- (2) Ein weniger triviales Beispiel ist der „Einheitskreis“ $C: x^2 + y^2 = 1$ (also $f = x^2 + y^2 - 1$). Wir setzen voraus, dass die Charakteristik von K nicht 2 ist. Für jedes L haben wir die vier rationalen Punkte $(0, \pm 1)$ und $(\pm 1, 0)$, aber normalerweise natürlich noch mehr. Man kann zeigen, dass

$$C(L) = \left\{ \left(\frac{2t}{1+t^2}, \frac{1-t^2}{1+t^2} \right) \mid t \in L, t^2 \neq -1 \right\} \cup \{(0, -1)\}$$

ist; siehe Übungen.

Als Beispiel einer rationalen Funktion betrachten wir $g = \frac{y-1}{x}$. Wo ist g regulär? Zunächst sicher da, wo die x -Koordinate nicht verschwindet, also in allen Punkten außer $(0, \pm 1)$. Wie verhält es sich in diesen beiden Punkten? In $(0, -1)$ verschwindet der Nenner, aber der Zähler hat den Wert -2 , woraus man schließen kann, dass die Funktion dort nicht regulär ist (sonst müsste $y - 1 = x \frac{y-1}{x}$ dort den Wert 0 haben). In $(0, 1)$ andererseits verschwinden Zähler und Nenner. Hier kann man umformen:

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

und der andere Repräsentant ist in $(0, 1)$ definiert (und hat den Wert 0). Also ist $(0, -1)$ der einzige Punkt, in dem g nicht regulär ist.

- (3) Jede Kurve $C: y^2 = x^3 + ax + b$ ist geometrisch irreduzibel. Denn jede Faktorisierung von $f = y^2 - x^3 - ax - b$ müsste die Form $(y - h_1(x))(y - h_2(x))$ haben, woraus sich $h_2 = -h_1$ und $x^3 + ax + b = h_1(x)^2$ ergibt. Letzteres ist unmöglich, da der Grad der linken Seite 3, der rechten Seite aber gerade ist. ♣

3. PROJEKTIVE EBENE KURVEN

Die affine Ebene und affine ebene Kurven sind zwar relativ anschaulich (jedenfalls wenn $K = \mathbb{R}$ oder in \mathbb{R} enthalten ist), haben aber gewisse Nachteile. Wenn wir $K = \mathbb{C}$ nehmen (in diesem Fall gibt es starke Parallelen zur komplexen Analysis), dann sehen wir an Beispielen, dass die beschriebenen Punktfolgen \mathbb{C}^2 oder $C(\mathbb{C})$ nicht kompakt sind. Das bedeutet, dass sie in einem gewissen Sinn „offen“ sind, dass ihnen „etwas fehlt“. Man kann das in vielen Fällen auch schon am reellen Bild sehen, zum Beispiel bei einer Geraden, einer Parabel oder einer Hyperbel (bei einer Ellipse macht es sich erst über \mathbb{C} bemerkbar).

Eine Auswirkung dieser Unvollkommenheit sind die Ausnahmen und Sonderfälle, die man beachten muss. Beispielsweise schneiden sich zwei verschiedene Geraden stets in genau einem Punkt — außer sie sind parallel. Um diese lästige Ausnahme zu beseitigen, fügt man der affinen Ebene Punkte hinzu. Und zwar gehört zu jeder Schar paralleler Geraden (also jeder „Richtung“) ein neuer Punkt, der auf allen diesen Geraden liegt. Alle diese neuen Punkte gemeinsam bilden ihrerseits eine Gerade, die sogenannte unendlich ferne Gerade. Dann gilt ohne jede Ausnahme, dass sich je zwei verschiedene Geraden in genau einem Punkt treffen und dass durch je zwei verschiedene Punkte genau eine Gerade geht.

Wir werden jetzt diese projektive Ebene formal als Objekt der algebraischen Geometrie definieren, wobei die Definition symmetrischer ist als das eben angedeutete Vorgehen. In der Tat ist die Auszeichnung einer Geraden als „die“ unendlich ferne völlig willkürlich.

3.1. Definition. Die *projektive Ebene* \mathbb{P}_K^2 über K hat folgende Eigenschaften.

DEF
projektive
Ebene

- (1) Zu jedem Erweiterungskörper $L \supset K$ ist die Menge der *L-rationalen Punkte* von \mathbb{P}_K^2 gegeben durch

$$\mathbb{P}_K^2(L) = \{(\xi, \eta, \zeta) \in L^3 \mid (\xi, \eta, \zeta) \neq (0, 0, 0)\} / \sim_L,$$

wobei die Äquivalenzrelation \sim_L gegeben ist durch

$$(\xi, \eta, \zeta) \sim_L (\xi', \eta', \zeta') \iff \exists \lambda \in L^\times : \xi' = \lambda\xi, \eta' = \lambda\eta, \zeta' = \lambda\zeta.$$

Die Koordinaten sind also nur bis auf Skalierung bestimmt.

Der durch (ξ, η, ζ) repräsentierte Punkt wird auch $(\xi : \eta : \zeta)$ geschrieben.

Nach dieser Definition kann man die Punkte der projektiven Ebene auch als die Ursprungsgeraden im dreidimensionalen affinen Raum auffassen. Die affine Ebene findet man wieder, wenn man sie mit der Ebene $z = 1$ identifiziert — die Ursprungsgeraden, die nicht in der xy -Ebene liegen, durchstoßen diese Ebene in einem eindeutig bestimmten Punkt, wodurch wir die Einbettung von \mathbb{A}_K^2 in \mathbb{P}_K^2 bekommen. Die übrigen Geraden entsprechen den unendlich fernen Punkten, entsprechend ihrer Richtung in der xy -Ebene. In Formeln haben wir für die Einbettung:

$$\mathbb{A}_K^2(L) \ni (\xi, \eta) \mapsto (\xi : \eta : 1) \in \mathbb{P}_K^2(L);$$

die Umkehrung ist definiert für die Punkte, deren Z -Koordinate nicht verschwindet (das hängt nicht von der Skalierung ab), und ist gegeben durch $(\xi : \eta : \zeta) \mapsto (\xi/\zeta, \eta/\zeta)$. Die übrigen Punkte sind gerade die L -rationalen Punkte der „unendlich fernen“ Geraden $Z = 0$ (siehe unten).

- (2) Zur Erinnerung: Ein Polynom $f \in K[X, Y, Z]$ heißt *homogen* vom Grad d , wenn es die Form

$$f = \sum_{r+s+t=d} a_{rst} X^r Y^s Z^t$$

hat.

Eine *rationale Funktion* auf \mathbb{P}_K^2 ist gegeben durch ein Element $f/g \in K(X, Y, Z)$, wo f und g homogene Polynome vom selben Grad sind.

f/g heißt *regulär* in $P = (\xi : \eta : \zeta) \in \mathbb{P}_K^2(L)$, wenn $g(\xi, \eta, \zeta) \neq 0$ ist (da g homogen ist, hängt diese Bedingung nicht von der Skalierung ab!). Wir erhalten Funktionen

$$(f/g)_L: \{P \in \mathbb{P}_K^2(L) \mid f/g \text{ regulär in } P\} \ni (\xi : \eta : \zeta) \mapsto \frac{f(\xi, \eta, \zeta)}{g(\xi, \eta, \zeta)} \in L.$$

Beachte: dies ist wohldefiniert, weil f und g beide homogen vom selben Grad sind. \diamond

Man beachte, dass es keine (nicht-konstanten) regulären Funktionen auf der projektiven Ebene gibt — ein Polynom liefert keine wohldefinierte Funktion (außer es ist konstant), und ein Quotient f/g hat immer Punkte in $\mathbb{P}_k^2(\bar{K})$, in denen g verschwindet.

3.2. Bemerkung. Man kann wieder auf analoge Weise den *n-dimensionalen projektiven Raum* \mathbb{P}_K^n über K definieren. \mathbb{P}_K^1 heißt auch die *projektive Gerade* über K . **BEM**
n-dim. proj.
Raum \spadesuit

Projektive ebene Kurven werden im Wesentlichen analog zu den affinen ebenen Kurven definiert. Wir müssen nur aufpassen, dass unsere Polynomgleichung eine wohldefinierte Bedingung liefert. Dies wird dadurch erreicht, dass wir homogene Polynome verwenden.

3.3. Definition. Eine *projektive ebene Kurve* C vom Grad d über K ist gegeben durch ein homogenes Polynom $0 \neq f \in K[X, Y, Z]$ vom Grad $d \geq 1$. (Wir schreiben $C: f(X, Y, Z) = 0$.) **DEF**
projektive
ebene Kurve

- (1) Für einen Erweiterungskörper $L \supset K$ ist die Menge der *L-rationalen Punkte* von C gegeben durch

$$C(L) = \{(\xi : \eta : \zeta) \in \mathbb{P}_K^2(L) \mid f(\xi, \eta, \zeta) = 0\}.$$

- (2) Eine *rationale Funktion* auf C ist eine Äquivalenzklasse rationaler Funktionen auf \mathbb{P}_K^2 , deren Nenner mit f keinen nicht-konstanten gemeinsamen Teiler hat. Dabei heißen g_1/h_1 und g_2/h_2 äquivalent, wenn $f \mid g_1 h_2 - g_2 h_1$.

Eine rationale Funktion ϕ ist *regulär* in $P \in C(L)$, wenn sie einen Repräsentanten g/h hat, sodass h in P nicht verschwindet. Wir haben dann wieder Funktionen

$$\phi_L: \{P \in C(L) \mid g/h \text{ regulär in } P\} \longrightarrow L.$$

- (3) C heißt *irreduzibel*, wenn f irreduzibel (in $K[X, Y, Z]$) ist. C heißt *geometrisch irreduzibel*, wenn f absolut irreduzibel ist.

Ist C irreduzibel, dann bilden die rationalen Funktionen auf C wiederum einen Körper, den *Funktionskörper* $K(C)$ von C . \diamond

Es ist nun ganz einfach, zwischen „affin“ und „projektiv“ hin- und herzuwechseln.

Sei also zunächst $C: f(x, y) = 0$ eine affine Kurve und d der Gesamtgrad des Polynoms f . Dann ist $F(X, Y, Z) = Z^d f(X/Z, Y/Z)$ ein homogenes Polynom vom Grad d (das aus f entsteht, indem wir x durch X und y durch Y ersetzen und dann zu jedem Monom eine Potenz von Z hinzumultiplizieren, sodass der Gesamtgrad gerade d wird). Die projektive Kurve $\bar{C}: F(X, Y, Z) = 0$ heißt dann der *projektive Abschluss* von C ; die „neu hinzugekommenen“ Punkte in $\bar{C}(L) \setminus C(L)$ (das sind die mit Z -Koordinate null) heißen *Punkte im Unendlichen* von C oder \bar{C} .

Ist umgekehrt $C: F(X, Y, Z) = 0$ eine projektive Kurve vom Grad d , dann ist $f(x, y) = F(x, y, 1)$ ein Polynom vom Grad höchstens d , und die affine Kurve $C': f(x, y) = 0$ ist ein *affiner Teil* von C (andere affine Teile bekommt man, indem man X oder Y gleich 1 setzt). Falls $F = aZ^d$ ist, ist allerdings $f = a$ konstant und definiert keine affine Kurve. In diesem Fall hat C nur Punkte auf der unendlich fernen Geraden.

Diese Operationen sind im Wesentlichen invers zueinander: Der affine Teil des projektiven Abschlusses der affinen Kurve C ist wieder C . Umgekehrt gilt, dass der projektive Abschluss des affinen Teils einer projektiven Kurve C wieder C ist, falls das definierende Polynom F nicht durch Z teilbar ist.

3.4. Beispiele.

- (1) Der projektive Abschluss einer affinen Geraden $ax + by = c$ ist die projektive Gerade $aX + bY - cZ = 0$. Sie hat genau einen Punkt $(-b : a : 0)$ im Unendlichen. Alle projektiven Geraden erhält man auf diese Weise, mit Ausnahme der „unendlich fernen“ Geraden $Z = 0$ (die nur aus Punkten im Unendlichen besteht).
- (2) Der projektive Abschluss des Einheitskreises $x^2 + y^2 = 1$ ist $X^2 + Y^2 - Z^2 = 0$. Er hat die beiden L -rationalen Punkte im Unendlichen $(1 : \pm i : 0)$, falls $-1 = i^2$ in L ein Quadrat ist (und $\text{char}(L) \neq 2$, sonst ist es der eine Punkt $(1 : 1 : 0)$). Allgemeiner gilt, dass alle Kreise $(x - a)^2 + (y - b)^2 = r^2$ dieselben zwei Punkte im Unendlichen haben.
- (3) Der projektive Abschluss der affinen Kurve $y^2 = x^3 + ax + b$ ist

$$Y^2Z - X^3 - aXZ^2 - bZ^3 = 0.$$

Er hat genau den einen (stets rationalen) Punkt $(0 : 1 : 0)$ im Unendlichen. ♣

BSP
projektiver
Abschluss

4. SCHNITTE VON KURVEN MIT GERADEN

Wir wollen in diesem Abschnitt beweisen, dass sich eine projektive Gerade und eine projektive Kurve vom Grad d stets in genau d Punkten schneiden. Wir brauchen dieses Resultat für die Definition der Gruppenstruktur auf einer elliptischen Kurve. Damit die Aussage stimmt, müssen die Schnittpunkte aber mit der richtigen Vielfachheit gezählt werden. Deswegen müssen wir erst einmal diese Vielfachheit definieren.

Im Folgenden ist K stets unser Grundkörper und L eine Körpererweiterung von K .

4.1. Definition. Sei $P = (\xi : \eta : \zeta) \in \mathbb{P}_K^2(L)$ ein Punkt, $G: aX + bY + cZ = 0$ eine projektive Gerade über K und $C: F(X, Y, Z) = 0$ eine projektive Kurve über K . Wir setzen voraus, dass $aX + bY + cZ$ kein Teiler von F ist (anderenfalls ist G in C enthalten). Wir definieren $i(G, C; P)$, die *Vielfachheit des Schnittpunkts P von G und C* wie folgt.

DEF
Schnitt-
vielfachheit

Im Fall $P \notin C(L) \cap G(L)$ (wenn also P kein Schnittpunkt von C und G ist) setzen wir $i(G, C; P) = 0$. Anderenfalls lösen wir die Gleichung von G nach einer der Variablen auf, z.B. $Z = -\frac{a}{c}X - \frac{b}{c}Y$ (falls $c \neq 0$ ist), und setzen diesen Ausdruck in F ein. Wir erhalten ein homogenes Polynom H in zwei Variablen, das durch $(\xi Y - \eta X)$ teilbar ist (wenn wir Z eliminiert haben, sonst $(\xi Z - \zeta X)$ bzw. $(\eta Z - \zeta Y)$). Die Vielfachheit dieses Faktors in H ist dann $i(G, C; P)$. \diamond

Die Definition hängt natürlich nicht davon ab, welche Variable wir eliminieren (Übung).

4.2. Beispiel. Wir betrachten die Kurve $C: Y^2Z - X^3 + XZ^2 = 0$. Für die Gerade $Y = 0$ ergibt sich $H = -X^3 + XZ^2 = X(X + Z)(-X + Z)$; wir haben also jeweils Vielfachheit 1 in den Schnittpunkten $(0 : 0 : 1)$, $(-1 : 0 : 1)$ und $(1 : 0 : 1)$. Bei der Geraden $X - Z = 0$ haben wir folgendes Bild. Wir eliminieren Z und bekommen $H = XY^2$, also hat der Schnittpunkt $(1 : 0 : 1)$ die Vielfachheit 2. (Tatsächlich ist die Gerade in diesem Punkt die Tangente an die Kurve.) Der andere Schnittpunkt $(0 : 1 : 0)$ hat dagegen Vielfachheit 1.

BSP
Vielfachheit
von
Schnittpunkten

Schließlich betrachten wir noch die Gerade $Z = 0$. In diesem Fall haben wir $H = -X^3$, also sogar einen Schnittpunkt der Vielfachheit 3 bei $(0 : 1 : 0)$. (Hier ist die Gerade die Wendetangente.) \clubsuit

Aus dem Beispiel lässt sich schon ablesen, dass und warum der folgende Satz richtig ist.

4.3. Satz. Sei $C: F(X, Y, Z) = 0$ eine projektive Kurve vom Grad d über K , und sei $G: aX + bY + cZ = 0$ eine projektive Gerade über K , die nicht in C enthalten ist. Dann gilt

SATZ
Satz von
Bézout
(Spezialfall)

$$\sum_{P \in C(\bar{K}) \cap G(\bar{K})} i(G, C; P) = d.$$

Gilt für einen Erweiterungskörper $L \supset K$

$$\sum_{P \in C(L) \cap G(L)} i(G, C; P) \geq d - 1,$$

so gilt bereits

$$\sum_{P \in C(L) \cap G(L)} i(G, C; P) = d.$$

Die letzte Aussage bedeutet, dass der letzte Schnittpunkt auch L -rational ist, wenn das für alle übrigen gilt.

Beweis. Sei o.B.d.A. $c \neq 0$. Wir setzen $a' = -a/c$, $b' = -b/c$; dann ist die Geradengleichung $Z = a'X + b'Y$. Wir setzen in F ein und bekommen $H(X, Y) = F(X, Y, a'X + b'Y)$; das ist ein homogenes Polynom vom Grad d in $K[X, Y]$. Als solches zerfällt es in $\bar{K}[X, Y]$ in Linearfaktoren:

$$H(X, Y) = \alpha(\eta_1 X - \xi_1 Y)^{d_1} \dots (\eta_k X - \xi_k Y)^{d_k}.$$

Ein Punkt $P = (\xi : \eta : \zeta) \in \mathbb{P}_K^2(\bar{K})$ ist genau dann ein Schnittpunkt von C und G , wenn $H(\xi, \eta) = 0$ und $\zeta = a'\xi + b'\eta$ gilt. Die Schnittpunkte sind also gerade

$$(\xi_1 : \eta_1 : a'\xi_1 + b'\eta_1), \dots, (\xi_k : \eta_k : a'\xi_k + b'\eta_k),$$

und ihre Vielfachheiten sind nach Definition d_1, \dots, d_k mit $d_1 + \dots + d_k = d$. Das beweist den ersten Teil des Satzes.

Für den zweiten Teil beachten wir, dass wir H schreiben können als ein Produkt von d Linearfaktoren, von denen $d - 1$ Koeffizienten in L haben. Dann muss der verbleibende Faktor auch Koeffizienten in L haben. \square

Dieser Satz ist ein Spezialfall des Satzes von Bézout, der sagt, dass sich zwei projektive Kurven der Grade d_1 und d_2 , die keine gemeinsame Komponente haben, stets in genau $d_1 d_2$ Punkten (mit Vielfachheit gerechnet; „Punkt“ heißt hier „ \bar{K} -rationaler Punkt“) schneiden. Um den Satz in dieser Allgemeinheit formulieren zu können, muss man erst die Vielfachheit eines Schnittpunktes von zwei beliebigen Kurven definieren. Dafür muss man aber tiefer in die Algebraische Geometrie einsteigen, als uns das hier möglich ist.

5. GLATTHEIT

In der Analysis legt man üblicherweise Wert darauf, dass die Objekte, die man betrachtet, keine Ecken und Kanten haben, also „glatt“ sind (wie zum Beispiel Mannigfaltigkeiten). Dazu verwendet man Differenzierbarkeitseigenschaften. Dies wird nun auf algebraische Kurven übertragen. Zwar kann man nicht mehr Funktionen ableiten im Sinne eines Grenzwerts von Differenzenquotienten (es gibt ja keine geeignete Topologie), aber man kann in jedem Fall Polynome einfach formal ableiten, indem man den üblichen Rechenregeln folgt. So sind dann auch die folgenden Definitionen zu verstehen.

5.1. Definition.

(1) Eine affine ebene Kurve $C: f(x, y) = 0$ heißt *glatt* in einem Punkt $P = (\xi, \eta) \in C(L)$, wenn nicht beide partielle Ableitungen von f im Punkt P , $\frac{\partial f}{\partial x}(\xi, \eta)$ und $\frac{\partial f}{\partial y}(\xi, \eta)$, verschwinden.

(2) Eine projektive ebene Kurve $C: F(X, Y, Z) = 0$ heißt *glatt* in einem Punkt $P = (\xi : \eta : \zeta) \in C(L)$, wenn

$$\left(\frac{\partial F}{\partial X}(\xi, \eta, \zeta), \frac{\partial F}{\partial Y}(\xi, \eta, \zeta), \frac{\partial F}{\partial Z}(\xi, \eta, \zeta) \right) \neq (0, 0, 0)$$

ist.

(3) Ein Punkt P , in dem C nicht glatt ist, heißt *singulärer Punkt* oder *Singularität* von C . (Dabei kann C affin oder projektiv sein.)

(4) Eine (affine oder projektive) Kurve C heißt *glatt*, wenn sie in allen Punkten $P \in C(\bar{K})$ glatt ist. Anderenfalls heißt C *singulär*. \diamond

Man beachte in Teil (4) der Definition, dass „alle Punkte“ wieder „alle \bar{K} -rationalen Punkte“ bedeutet. Eine Kurve kann also singulär sein, obwohl sie in allen K -rationalen Punkten glatt ist.

Ein Punkt auf einer affinen Kurve ist genau dann glatt, wenn er auf dem projektiven Abschluss glatt ist (Übung).

5.2. Beispiele.

(1) Ist die Kurve $Y^2Z - X^3 - Z^3 = 0$ glatt? Die Punkte $(\xi : \eta : \zeta)$, in denen sie nicht glatt ist, müssten folgende Bedingungen erfüllen:

$$-3\xi^2 = 2\eta\zeta = \eta^2 - 3\zeta^2 = 0.$$

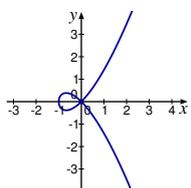
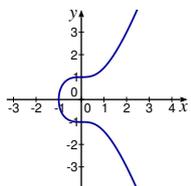
Wenn wir einmal voraussetzen, dass $\text{char}(K) \neq 2, 3$ ist, dann folgt daraus $\xi = \eta = \zeta = 0$. Also kann es einen solchen Punkt nicht geben (es dürfen ja nicht alle projektiven Koordinaten verschwinden), und die Kurve ist glatt.

In Charakteristik 2 sind die Bedingungen äquivalent zu $\xi = 0, \eta = \zeta$; die Kurve ist also in $(0 : 1 : 1)$ singulär. In Charakteristik 3 bleibt nur $\eta = 0$; damit der Punkt auf der Kurve liegt, muss noch $\xi + \zeta = 0$ gelten. Damit ist die Kurve genau in $(-1 : 0 : 1)$ singulär.

(2) Im Gegensatz dazu ist die Kurve $y^2 = x^3 + x^2$ im Punkt $P = (0, 0)$ nicht glatt, denn beide partielle Ableitungen $3x^2 + 2x$ und $2y$ verschwinden dort. Im anschaulichen Bild „kreuzen sich dort zwei Äste“; es liegt ein sogenannter *einfacher Doppelpunkt* vor.

DEF
glatter Punkt
glatte Kurve
Singularität

BSP
glatt/
singulär



Die Skizzen im rechten Rand zeigen die reellen Punkte des affinen Teils der Kurven.



5.3. Bemerkung. Sei $C: F(X; Y, Z) = 0$ eine projektive Kurve und sei weiter $P = (\xi : \eta : \zeta) \in C(K)$. Es ist nicht allzu schwer, Folgendes zu zeigen (Übung):

BEM
Vielfachheit
von P auf C

(1) C ist genau dann glatt in P , wenn

$$i(C; P) := \min\{i(G, C; P) \mid G \text{ eine Gerade durch } P\} = 1$$

gilt. Sonst ist $i(C; P) \geq 2$. Die Zahl $i(C; P)$ heißt auch die *Vielfachheit* von P auf C . Es gilt $i(C; P) \leq d$, wenn d der Grad der Kurve ist.

DEF
Vielfachheit
eines Punktes

(2) Wenn C in P glatt ist, dann gibt es genau eine Gerade G durch P , sodass $i(G, C; P) \geq 2$ ist. Diese Gerade ist die *Tangente* an C in P und hat die Gleichung

DEF
Tangente

$$\frac{\partial F}{\partial X}(\xi, \eta, \zeta) X + \frac{\partial F}{\partial Y}(\xi, \eta, \zeta) Y + \frac{\partial F}{\partial Z}(\xi, \eta, \zeta) Z = 0.$$

Ist $i(G, C; P) = 3$, so heißt P ein *Wendepunkt* von C ; ist $i(G, C; P) \geq 4$, so heißt P ein *Flachpunkt* von C .

DEF
Wendepunkt
DEF
Flachpunkt

6. RATIONALE ABBILDUNGEN UND MORPHISMEN

Wie stets in der Mathematik interessiert man sich auch in der Algebraischen Geometrie nicht nur für die Objekte (wie zum Beispiel algebraische Kurven), sondern auch für die passenden Abbildungen dazwischen. Diese wollen wir jetzt definieren.

6.1. Definition. Seien $C: F(X, Y, Z) = 0$ und $D: G(X, Y, Z) = 0$ zwei irreduzible projektive ebene Kurven über K .

DEF
rationale
Abbildung
Morphismus

- (1) Eine *rationale Abbildung* von C nach D ist eine Äquivalenzklasse von Tripeln (R_1, R_2, R_3) , wo die $R_j \in K[X, Y, Z]$ homogen vom gleichen Grad und nicht alle durch F teilbar sind und außerdem $G(R_1, R_2, R_3)$ durch F teilbar ist. Dabei heißen (R_1, R_2, R_3) und (S_1, S_2, S_3) äquivalent, wenn $F \mid R_i S_j - R_j S_i$ für alle i, j gilt.
- (2) Sei ϕ eine rationale Abbildung von C nach D und $P = (\xi : \eta : \zeta) \in C(L)$. ϕ heißt *regulär* oder *definiert* in P , wenn ϕ einen Repräsentanten (R_1, R_2, R_3) hat, sodass nicht alle $R_j(\xi, \eta, \zeta)$ verschwinden. In diesem Fall ist

$$\phi(P) = (R_1(\xi, \eta, \zeta) : R_2(\xi, \eta, \zeta) : R_3(\xi, \eta, \zeta)) \in D(L)$$

wohldefiniert, und wir erhalten Abbildungen

$$\phi_L: \{P \in C(L) \mid \phi \text{ definiert in } P\} \longrightarrow D(L).$$

- (3) Ein *Morphismus* von C nach D ist eine rationale Abbildung von C nach D , die überall auf C (d.h. auf $C(\bar{K})$) definiert ist.
- (4) Man kann rationale Abbildungen bzw. Morphismen in offensichtlicher Weise miteinander verknüpfen. Dabei spielt die Äquivalenzklasse von (X, Y, Z) die Rolle eines neutralen Elements. Der zugehörige Morphismus ist der Identitätsmorphismus $\text{id}_C: C \rightarrow C$.
- (5) C und D heißen *birational äquivalent*, wenn es rationale Abbildungen $\phi: C \rightarrow D$ und $\psi: D \rightarrow C$ gibt, sodass $\phi \circ \psi = \text{id}_D$ und $\psi \circ \phi = \text{id}_C$ gilt. Dann ist ϕ eine *birationale* Abbildung. Sind ϕ und ψ sogar Morphismen, dann heißen C und D *isomorph*, und ϕ ist ein *Isomorphismus*. \diamond

Es gilt übrigens, dass eine rationale Abbildung von einer *glatten* Kurve in eine andere Kurve automatisch ein Morphismus ist. Genauer gilt, dass jede rationale Abbildung von einer Kurve C in eine Kurve D in jedem glatten Punkt von C definiert ist.

6.2. Beispiele.

BSP
Morphismen

- (1) Je zwei projektive Geraden sind isomorph. Ein Isomorphismus von $Z = 0$ auf $Z = aX + bY$ ist zum Beispiel gegeben durch

$$(X : Y : 0) \mapsto (X : Y : aX + bY).$$

- (2) Es ist möglich, dass ein Morphismus durch konstante Polynome repräsentiert wird. So ein konstanter Morphismus bildet alles auf einen festen (K -rationalen) Punkt ab. Man kann zeigen, dass jeder nicht-konstante Morphismus zwischen irreduziblen projektiven Kurven surjektiv ist, d.h. $\phi_{\bar{K}}$ ist surjektiv. (ϕ_L muss nicht unbedingt surjektiv sein!)

- (3) Hier ist ein nicht-triviales Beispiel für einen Morphismus. Sei C der „Einheitskreis“ $X^2 + Y^2 = Z^2$ über einem Körper K mit $\text{char}(K) \neq 2$. Dann definiert $(X^2 - Y^2, 2XY, Z^2)$ einen Morphismus $\phi: C \rightarrow C$: Es gilt

$$(X^2 - Y^2)^2 + (2XY)^2 - (Z^2)^2 = (X^2 + Y^2 - Z^2)(X^2 + Y^2 + Z^2),$$

also ist die wesentliche Bedingung erfüllt. Die Abbildung ist überall definiert, da alle drei Komponenten nur für $X = Y = Z = 0$ verschwinden, was aber keinem Punkt in \mathbb{P}^2 entspricht. (Auf dem klassischen Einheitskreis $C(\mathbb{R})$ entspricht dieser Morphismus der Verdopplung des Winkels zur positiven x -Achse.)

- (4) Sei C wieder der Einheitskreis und $G: X = 0$ die y -Achse. Dann definieren die Tripel $(0, Y, Z + X)$ und $(0, Z - X, Y)$ denselben Morphismus $\phi: C \rightarrow G$, und $(Z^2 - Y^2, 2YZ, Z^2 + Y^2)$ definiert einen Morphismus $\phi': G \rightarrow C$, der zu ϕ invers ist (falls $\text{char}(K) \neq 2$ gilt): Der Einheitskreis (und damit jeder Kreis mit K -rationalen Punkten) ist zur y -Achse (und damit jeder Geraden) isomorph! (Tatsächlich gilt das sogar für jeden irreduziblen *Kegelschnitt*, also eine Kurve vom Grad 2, mit K -rationalen Punkten.) ♣

7. ELLIPTISCHE KURVEN: DEFINITION

In diesem Abschnitt werden wir elliptische Kurven über einem beliebigen Grundkörper einführen.

Was ist eine elliptische Kurve? Die unten angegebene Definition wirkt etwas ad hoc, ist aber für die Zwecke dieser Vorlesung durchaus angemessen, da uns für das Verständnis „besserer“ Definitionen die nötigen Grundlagen aus der Algebraischen Geometrie fehlen.

Die „richtige“ Definition lautet etwa: Eine elliptische Kurve über K ist eine irreduzible glatte projektive Kurve über K vom Geschlecht 1, auf der ein K -rationaler Punkt fixiert ist. Man kann dann zeigen (mithilfe des Satzes von Riemann-Roch) zeigen, dass eine elliptische Kurve in diesem Sinne stets isomorph ist zu einer elliptischen Kurve im Sinne der unten folgenden Definition, wobei der Isomorphismus den fixierten Punkt auf den Punkt O abbildet.

Warum heißen elliptische Kurven „elliptische Kurven“? Es gibt einen etwas indirekten Zusammenhang mit Ellipsen. Wenn man die Bogenlänge eines Ellipsenstücks berechnen will, dann kommt man auf ein Integral, dessen Integrand die allgemeine Form $R(x, \sqrt{P(x)})$ hat, wobei R eine rationale Funktion in zwei Variablen und P ein Polynom vom Grad 3 oder 4 ist. Solche Integrale heißen wegen dieses Zusammenhangs auch elliptische Integrale. Wenn wir $\sqrt{P(x)}$ oben mit y bezeichnen, dann haben wir die Relation $y^2 = P(x)$, die eine (affine) Kurve C beschreibt, die über einem Körper K , sodass C K -rationale Punkte hat, zu einer elliptischen Kurve birational ist. Den Integranden kann man als eine 1-Form auf dieser elliptischen Kurve auffassen. Man kann also sagen, elliptische Kurven sind die Kurven, auf denen elliptische Integrale „leben“.

Elliptische Kurven sind (als glatte irreduzible Kurven vom Grad 3) natürlich selbst keine Ellipsen (die glatte irreduzible Kurven vom Grad 2 sind).

7.1. Definition. Eine *elliptische Kurve* über dem Körper K ist eine glatte projektive ebene Kurve E vom Grad 3 über K , die durch eine Gleichung der Form

$$Y^2Z + a_1XYZ + a_3YZ^2 = X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3$$

mit Koeffizienten $a_1, a_2, a_3, a_4, a_6 \in K$ gegeben ist.

Der Einfachheit halber benutzen wir meistens die Gleichung des affinen Teils:

$$(7.1) \quad E: y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6.$$

So eine Gleichung heißt (lange) *Weierstraß-Gleichung*. ◇

Die etwas komische Nummerierung der Koeffizienten wird später verständlich werden. Wir hatten uns schon überlegt, dass so eine Kurve jedenfalls im Fall $a_1 = a_2 = a_3 = 0$ geometrisch irreduzibel ist; der Beweis überträgt sich leicht auf beliebige Kurven der Form (7.1).

7.2. Lemma. Sei E eine (nicht notwendig glatte) Kurve, die durch eine Gleichung der Form (7.1) gegeben ist. Dann hat E genau einen Punkt im Unendlichen, nämlich $O = (0 : 1 : 0)$. Dieser Punkt O ist K -rational, E ist in O glatt, und die Tangente an E in O ist die unendlich ferne Gerade $Z = 0$; sie schneidet E in O mit Vielfachheit 3 (d.h. O ist ein Wendepunkt von E).

Beweis. Um die Punkte im Unendlichen zu finden, müssen wir in der (projektiven) Kurvengleichung $Z = 0$ setzen. Es bleibt $X^3 = 0$, also ist der angegebene Punkt $O = (0 : 1 : 0)$ der einzige Punkt, und er hat als Schnittpunkt von E mit der unendlich fernen Geraden die Vielfachheit 3. Da die Vielfachheit ≥ 2 und E

DEF
elliptische
Kurve



Weierstraß

K. Weierstraß
1815–1897

LEMMA
Punkt im
Unendlichen

in O glatt ist (s.u.), ist die unendlich ferne Gerade auch die Tangente. Da die Koordinaten von O in K liegen, ist $O \in E(K)$.

Es bleibt zu zeigen, dass E in O glatt ist. Dazu müssen wir die partiellen Ableitungen bestimmen und in O auswerten. Die Ableitung nach Z ist Y^2 plus Terme, die X oder Z enthalten, also verschwindet sie in O nicht. Damit ist E in O glatt. \square

In vielen Fällen lässt sich die Gleichung einer elliptischen Kurve noch vereinfachen.

7.3. Lemma. Sei E eine elliptische Kurve über K . Wenn die Charakteristik von $K \neq 2$ ist, dann ist E isomorph (als elliptische Kurve, siehe §8) zu einer elliptischen Kurve der Form

$$E': y^2 = x^3 + a'_2 x^2 + a'_4 x + a'_6.$$

Wenn zusätzlich $\text{char}(K) \neq 3$ ist, dann kann man auch noch $a'_2 = 0$ erreichen. Die entstehende Gleichung

$$y^2 = x^3 + ax + b$$

heißt kurze Weierstraß-Gleichung.

Beweis. Der Isomorphismus von E auf E' ist (in projektiven Koordinaten) gegeben durch („quadratische Ergänzung“)

$$(X : Y : Z) \mapsto (X : Y + \frac{a_1}{2} X + \frac{a_3}{2} Z : Z).$$

(„Isomorphismus von elliptischen Kurven“ bedeutet, dass der Punkt $(0 : 1 : 0)$ von E auf den Punkt $(0 : 1 : 0)$ von E' abgebildet wird.) Für die Koeffizienten gilt dann

$$a'_2 = a_2 + \frac{1}{4} a_1^2, \quad a'_4 = a_4 + \frac{1}{2} a_1 a_3, \quad a'_6 = a_6 + \frac{1}{4} a_3^2.$$

Wenn $\text{char}(K) \neq 3$ ist, dann kann man durch eine weitere Transformation der Form $(x, y) \mapsto (x + \frac{1}{3} a'_2, y)$ den Koeffizienten a'_2 ebenfalls zum Verschwinden bringen („kubische Ergänzung“). \square

Nun erhebt sich natürlich die Frage, wann eine (lange oder kurze) Weierstraß-Gleichung tatsächlich eine elliptische Kurve definiert. Anders gesagt, wie erkennt man, ob die definierte Kurve glatt ist oder nicht?

Dazu führen wir einige weitere Größen ein, die von den Koeffizienten abhängen. Die Bezeichnungen sind allgemein gebräuchlich.

7.4. Definition. Sei $E: y^2 + a_1 xy + a_3 y = x^3 + a_2 x^2 + a_4 x + a_6$ eine lange Weierstraß-Gleichung. Wir setzen

$$\begin{aligned} b_2 &= a_1^2 + 4a_2, & b_4 &= a_1 a_3 + 2a_4, & b_6 &= a_3^2 + 4a_6, \\ b_8 &= a_1^2 a_6 - a_1 a_3 a_4 + 4a_2 a_6 + a_2 a_3^2 - a_4^2 \\ c_4 &= b_2^2 - 24b_4, & c_6 &= -b_2^3 + 36b_2 b_4 - 216b_6 \\ \Delta &= -b_2^2 b_8 - 8b_4^3 - 27b_6^2 + 9b_2 b_4 b_6, & j &= c_4^3 / \Delta \end{aligned}$$

Dabei gilt

$$4b_8 = b_2 b_6 - b_4^2 \quad \text{und} \quad 1728 \Delta = c_4^3 - c_6^2.$$

Die Größen c_4 und c_6 werden oft die *Invarianten* der Kurve genannt; Δ ist die *Diskriminante* und j die *j-Invariante* der Kurve. \diamond

LEMMA
kurze
W.-Gleichung

DEF
 $b_2, b_4, b_6, b_8,$
 c_4, c_6, Δ, j

Man beachte, dass sich die vereinfachten Gleichungen (für $\text{char}(K) \neq 2$ bzw. $\text{char}(K) \neq 2, 3$) nach einer zusätzlichen Skalierung der Variablen $((x, y) \mapsto (4x, 8y)$ bzw. $(x, y) \mapsto (36x, 216y)$) auch schreiben lassen als

$$y^2 = x^3 + b_2 x^2 + 8b_4 x + 16b_6$$

bzw.

$$y^2 = x^3 - 27c_4 x - 54c_6.$$

7.5. Lemma. *Eine Weierstraß-Gleichung der Form (7.1) definiert genau dann eine elliptische (d.h. eine glatte) Kurve, wenn die Diskriminante Δ nicht verschwindet.*

LEMMA
Diskriminante
und Glattheit

Beweis. Der Einfachheit halber beschränken wir uns hier auf den Fall, dass die Charakteristik des Grundkörpers weder 2 noch 3 ist. Die anderen Fälle kann man ähnlich behandeln.

In dem betrachteten Fall können wir die ursprüngliche Gleichung in eine kurze Weierstraß-Gleichung $E: y^2 = x^3 + ax + b$ transformieren; man rechnet nach, dass Δ dabei höchstens mit der zwölften Potenz eines invertierbaren Elements multipliziert wird (vergleiche §8). Da es sich um einen Isomorphismus handelt, ändert sich auch nichts daran, ob die Kurve glatt ist oder nicht. Es ist dann

$$\Delta = -16(4a^3 + 27b^2).$$

Wir haben bereits gesehen, dass E im Punkt im Unendlichen glatt ist. Wir können uns also auf den affinen Teil beschränken. Ein Punkt (ξ, η) ist genau dann ein singulärer Punkt auf E , wenn folgende drei Gleichungen erfüllt sind.

$$3\xi^2 + a = 0, \quad 2\eta = 0, \quad \eta^2 = \xi^3 + a\xi + b.$$

Wegen der Annahme über die Charakteristik von K bedeutet das

$$\eta = 0, \quad \xi^2 = -\frac{1}{3}a, \quad \xi^3 + a\xi + b = 0.$$

Einsetzen der zweiten in die dritte Gleichung liefert (falls $a \neq 0$)

$$\xi = -\frac{3b}{2a}.$$

Das System hat also genau dann eine Lösung, wenn

$$\left(\frac{3b}{2a}\right)^2 = -\frac{a}{3},$$

also genau dann, wenn $\Delta = 0$ ist.

Im Fall $a = 0$ vereinfacht sich die Bedingung zu $b = 0$, was dann ebenfalls zu $\Delta = 0$ äquivalent ist. \square

Ist E eine elliptische Kurve über K , dann ist also ihre j -Invariante $j(E) = c_4^3/\Delta$ ein wohldefiniertes Element von K .

7.6. Beispiele.**BSP**
Diskriminante

- (1) Die Kurve $y^2 = x^3$ hat $\Delta = 0$, ist also keine elliptische Kurve. Tatsächlich ist $(0, 0)$ ein singulärer Punkt.
- (2) Die Kurve $y^2 = x^3 + x^2$ hat ebenfalls $\Delta = 0$ und eine Singularität bei $(0, 0)$.
- (3) Die Kurve $y^2 = x^3 + x$ hat $\Delta = -2^6$, ist also eine elliptische Kurve, falls $\text{char}(K) \neq 2$ ist. Ihre j -Invariante ist $12^3 = 1728$.
- (4) Die Kurve $y^2 = x^3 + 1$ hat $\Delta = -2^4 \cdot 3^3$, ist also eine elliptische Kurve, falls $\text{char}(K) \neq 2, 3$ ist. Ihre j -Invariante ist 0. ♣

8. ISOMORPHISMEN ELLIPTISCHER KURVEN

Wir haben im letzten Abschnitt schon einige Male auf den Begriff des Isomorphismus von elliptischen Kurven Bezug genommen. Wir definieren ihn jetzt:

8.1. Definition. Seien E und E' zwei elliptische Kurven über K . Ein Morphismus $\phi: E \rightarrow E'$ ist ein *Isomorphismus elliptischer Kurven*, wenn ϕ die Form

$$(X : Y : Z) \mapsto (u^2X + rZ : u^3Y + su^2X + tZ : Z)$$

mit $r, s, t \in K, u \in K^\times$ hat. ◇

DEF
Isomorphismus
ell. Kurven

Man sieht leicht, dass ein Morphismus dieser Form tatsächlich ein Isomorphismus von Kurven ist; siehe unten.

8.2. Lemma. Wenn E (bzw. E') in der Situation der Definition oben durch eine Weierstraß-Gleichung mit Koeffizienten a_i (bzw. a'_i) gegeben ist, dann gilt

$$\begin{aligned} u a_1 &= a'_1 + 2s \\ u^2 a_2 &= a'_2 - s a'_1 + 3r - s^2 \\ u^3 a_3 &= a'_3 + r a'_1 + 2t \\ u^4 a_4 &= a'_4 - s a'_3 + 2r a'_2 - (t + rs) a'_1 + 3r^2 - 2st \\ u^6 a_6 &= a'_6 + r a'_4 - t a'_3 + r^2 a'_2 - rt a'_1 + r^3 - t^2. \end{aligned}$$

(Das erklärt übrigens die Indizierung der Koeffizienten!) Weiterhin gilt

$$u^4 c_4 = c'_4, \quad u^6 c_6 = c'_6, \quad u^{12} \Delta = \Delta' \quad \text{und} \quad j = j'.$$

Beweis. Das rechnet man nach. □

LEMMA
Transformation
der
Koeffizienten

8.3. Lemma. Seien E und E' elliptische Kurven über K . Dann ist jeder Isomorphismus elliptischer Kurven $\phi: E \rightarrow E'$ auch ein Isomorphismus von ebenen projektiven Kurven, und $\phi(O) = O$.

Ein Isomorphismus von projektiven ebenen Kurven $\phi: E \rightarrow E'$, der durch lineare Polynome gegeben ist und O auf O abbildet, ist bereits ein Isomorphismus elliptischer Kurven.

Beweis. Man prüft nach, dass durch

$$\psi: (X : Y : Z) \mapsto (u^{-2}(X - rZ) : u^{-3}(Y - sX + (sr - t)Z) : Z)$$

der inverse Morphismus gegeben ist. Außerdem ist $\phi(O) = (0 : u^3 : 0) = (0 : 1 : 0)$.

Für den zweiten Teil nehmen wir an, dass ϕ folgende Form hat:

$$(X : Y : Z) \mapsto (\alpha_1 X + \alpha_2 Y + \alpha_3 Z : \beta_1 X + \beta_2 Y + \beta_3 Z : \gamma_1 X + \gamma_2 Y + \gamma_3 Z).$$

Da die unendlich ferne Gerade $Z = 0$ die einzige Gerade ist, die E bzw. E' in O mit Vielfachheit 3 schneidet, und da $\phi(O) = O$ ist, muss ϕ diese Gerade auf sich abbilden. Das bedeutet $\gamma_1 = \gamma_2 = 0$. Dass O fest bleibt, bedeutet $\alpha_2 = 0$. Damit können wir ohne Einschränkung $\gamma_3 = 1$ setzen, und wir sehen, dass der Isomorphismus die angegebene Form hat, jedenfalls bis auf die Relation zwischen den Koeffizienten α_1 und β_2 . Diese ergibt sich aber durch Koeffizientenvergleich nach Einsetzen in die Weierstraß-Gleichung, was die Beziehung $\alpha_1^3 = \beta_2^2$ liefert. Schließlich kann u nicht verschwinden, weil der Morphismus sonst konstant wäre. □

LEMMA
Charakterisierung
von
Isomorphismen

8.4. Bemerkung. Der tiefere algebraisch-geometrische Grund für die Form der Isomorphismen liegt darin, dass die rationale Funktion x (bzw. X/Z) in O einen Pol der Ordnung 2 hat und sonst regulär ist, und alle solche Funktionen die Form $ux + r$ haben mit $u \neq 0$. Ebenso gilt, dass die rationale Funktion y (bzw. Y/Z) in O einen Pol der Ordnung 3 hat und sonst regulär ist, und alle solche Funktionen die Form $uy + sx + t$ haben mit $u \neq 0$. Da der Punkt O fest bleiben soll, bleiben die Polordnungen erhalten, woraus sich die Form des Isomorphismus ergibt. ♠

BEM
Hintergrund

Wir sehen, dass die j -Invariante $j(E)$ unter Isomorphismen invariant ist (daher auch der Name). Damit erhebt sich die Frage, ob davon auch die Umkehrung gilt: Sind zwei elliptische Kurven mit derselben j -Invariante isomorph? Der folgende Satz zeigt, dass die Antwort im Wesentlichen Ja lautet.

8.5. Satz. *Seien E und E' zwei elliptische Kurven über K .*

SATZ
Isomorphie
und
 j -Invariante

- (1) *Sei $\text{char}(K) \neq 2, 3$. E und E' sind genau dann über K isomorph, wenn es ein $u \in K^\times$ gibt mit $c_4(E') = u^4 c_4(E)$ und $c_6(E') = u^6 c_6(E)$.*
- (2) *Wenn $j(E) = j(E')$ ist, dann sind E und E' über \bar{K} isomorph.*
- (3) *Zu jedem $j \in K$ gibt es eine elliptische Kurve E über K mit $j(E) = j$.*

Beweis. Der Einfachheit halber setzen wir für alle Teile $\text{char}(K) \neq 2, 3$ voraus.

- (1) Die gegebenen Kurven sind nach Lemma 7.3 und der Bemerkung vor Lemma 7.5 isomorph zu den Kurven

$$\tilde{E}: y^2 = x^3 - 27c_4(E)x - 54c_6(E) \quad \text{und} \quad \tilde{E}': y^2 = x^3 - 27c_4(E')x - 54c_6(E').$$

Es folgt, dass E und E' genau dann isomorph sind, wenn \tilde{E} und \tilde{E}' isomorph sind.

„ \Rightarrow “: Gibt es einen Isomorphismus $\phi: \tilde{E} \rightarrow \tilde{E}'$ mit Koeffizienten (r, s, t, u) wie in Definition 8.1, dann folgt aus den Relationen in Lemma 8.2, dass $s = r = t = 0$ sein muss und $c_4(E') = u^4 c_4(E)$, $c_6(E') = u^6 c_6(E)$ gilt.

„ \Leftarrow “: Gilt $c_4(E') = u^4 c_4(E)$ und $c_6(E') = u^6 c_6(E)$, dann folgt aus Lemma 8.2, dass diese beiden Kurven durch $(x, y) \mapsto (u^2 x, u^3 y)$ isomorph sind.

- (2) Aus $j(E) = j(E') = j$ folgt entweder $c_4(E) = c_4(E') = 0 = j$ oder $c_6(E) = c_6(E') = 0$, $j = 1728$, oder $j \neq 0, 1728$ und $c_6(E)^2/c_4(E)^3 = c_6(E')^2/c_4(E')^3 \neq 0$. In allen drei Fällen gibt es ein $u \in \bar{K}^\times$, sodass $c_4(E') = u^4 c_4(E)$ und $c_6(E') = u^6 c_6(E)$ ist. Nach Teil (1) sind die Kurven also über \bar{K} isomorph.
- (3) Man prüft nach, dass die Fälle $j = 0$ und $j = 12^3 = 1728$ durch die beiden Kurven

$$y^2 = x^3 + 1 \quad \text{und} \quad y^2 = x^3 + x$$

abgedeckt werden. In den übrigen Fällen tut es die Kurve

$$y^2 = x^3 - \frac{27}{4} \frac{j}{j - 1728} x - \frac{27}{4} \frac{j}{j - 1728}.$$

(Man erhält diese Kurve, indem man in der kurzen Weierstraß-Gleichung $y^2 = x^3 + ax + b$ den Ansatz $a = b$ macht.) □

Wenn K algebraisch abgeschlossen ist, werden die elliptischen Kurven über K also gerade durch die j -Invariante bis auf Isomorphie klassifiziert. Wenn K nicht algebraisch abgeschlossen ist, dann kann es mehrere nicht-isomorphe elliptische Kurven mit derselben j -Invariante geben.

8.6. **Satz.** Seien $\text{char}(K) \neq 2, 3$, $j \in K$ und $E: y^2 = x^3 + ax + b$ eine elliptische Kurve über K mit $j(E) = j$.

SATZ
ell. Kurven
mit $j(E) = j$

- (1) Im Fall $j \neq 0, 1728$ sind die K -Isomorphieklassen elliptischer Kurven E' mit $j(E') = j$ klassifiziert durch $K^\times / (K^\times)^2$. Wenn $d \in K^\times$ eine solche Klasse repräsentiert, dann ist die zugehörige elliptische Kurve gegeben durch

$$y^2 = x^3 + d^2 ax + d^3 b.$$

Diese Kurve heißt der quadratische Twist mit d von E .

DEF
quadratischer
Twist

- (2) Im Fall $j = 0$ ist $a = 0$. Die K -Isomorphieklassen mit $j = 0$ werden klassifiziert durch $K^\times / (K^\times)^6$; die zu $d \in K^\times$ gehörige Kurve ist

$$y^2 = x^3 + db.$$

- (3) Im Fall $j = 1728$ ist $b = 0$. Die K -Isomorphieklassen mit $j = 1728$ werden klassifiziert durch $K^\times / (K^\times)^4$; die zu $d \in K^\times$ gehörige Kurve ist

$$y^2 = x^3 + d ax.$$

Beweis. Es gilt jedenfalls $j = 0 \iff c_4 = 0 \iff a = 0$ und $j = 1728 \iff c_6 = 0 \iff b = 0$.

- (1) Die j -Invariante ist bei einer kurzen Weierstraß-Gleichung eine gebrochen-lineare Funktion von a^3/b^2 . Hier sind $a, b \neq 0$. Daher hat $E': y^2 = x^3 + a'x + b'$ genau dann dieselbe j -Invariante wie E , wenn $a'^3/b'^2 = a^3/b^2$ ist, also $a' = d^2 a$ und $b' = d^3 b$ für ein $d \in K^\times$ gilt. Nach Satz 8.5 sind die beiden Kurven genau dann bereits über K isomorph, wenn d ein Quadrat ist.
- (2) und (3) werden analog bewiesen. □

9. GRUPPENSTRUKTUR

Nun wollen wir beweisen, dass eine elliptische Kurve eine (geometrisch definierte) Gruppenstruktur trägt.

9.1. **Satz.** *Seien E eine elliptische Kurve über K und $L \supset K$ ein Erweiterungskörper. Durch folgende Festlegungen wird $E(L)$ zu einer abelschen Gruppe.*

SATZ
Gruppen-
struktur

- (1) Der Punkt $O \in E(L)$ ist das Nullelement.
- (2) Wenn G eine Gerade ist, die E in den Punkten P, Q, R schneidet (ein Punkt kommt dabei gemäß seiner Vielfachheit als Schnittpunkt evtl. mehrfach vor), dann gilt $P + Q + R = O$.

Etwas konkreter heißt das:

- (1) Der Punkt $-P$ ist der dritte Schnittpunkt der Geraden durch O und P mit E .
- (2) Der Punkt $P + Q$ ist der dritte Schnittpunkt der Geraden durch O und R mit E , wobei R der dritte Schnittpunkt der Geraden durch P und Q mit E ist.

Dabei sind natürlich alle Punkte mit der richtigen Vielfachheit zu zählen. Im Fall, dass P und Q zusammenfallen, muss man zum Beispiel die Tangente an E in $P = Q$ betrachten (anstelle der Geraden durch P und Q), da sie die einzige Gerade ist, die E in diesem Punkt mit Vielfachheit mindestens 2 schneidet.

Formeln für die Addition.

Um es noch konkreter zu machen, sei E durch die Gleichung

$$E: y^2 + a_1 xy + a_3 y = x^3 + a_2 x^2 + a_4 x + a_6$$

gegeben, und P und Q seien die affinen Punkte (ξ, η) und (ξ', η') . Die Gerade durch P und O ist gegeben durch die Gleichung

$$x = \xi$$

und der dritte Schnittpunkt ist

$$-P = (\xi, -\eta - a_1 \xi - a_3).$$

Im Fall $\xi \neq \xi'$ ist die Gerade durch P und Q gegeben durch die Gleichung

$$y = \lambda x + \mu$$

mit

$$\lambda = \frac{\eta' - \eta}{\xi' - \xi} \quad \text{und} \quad \mu = \eta - \lambda \xi = \frac{\xi' \eta - \xi \eta'}{\xi' - \xi}.$$

Wenn $\xi = \xi'$ und $\eta + \eta' \neq -a_1 \xi - a_3$ gilt (dann ist $Q \neq -P$), dann haben wir $\eta = \eta'$ und

$$\lambda = \frac{3\xi^2 + 2a_2 \xi + a_4 - a_1 \eta}{2\eta + a_1 \xi + a_3} \quad \text{und} \quad \mu = \eta - \lambda \xi = \frac{-\xi^3 + a_4 \xi + 2a_6 - a_3 \eta}{2\eta + a_1 \xi + a_3}.$$

Um das zu sehen, kann man entweder die Gleichung der Tangente an E in P bestimmen (siehe Bemerkung 5.3), oder man überlegt sich, dass man den Differenzenquotienten im Fall $\xi \neq \xi'$ mithilfe der Gleichung von E umschreiben kann:

$$\begin{aligned} \frac{\eta' - \eta}{\xi' - \xi} &= \frac{(\eta'^2 + a_1 \xi' \eta' + a_3 \eta') - (\eta^2 + a_1 \xi \eta + a_3 \eta) - a_1(\xi' - \xi)\eta'}{(\xi' - \xi)(\eta' + \eta + a_1 \xi + a_3)} \\ &= \frac{(\xi' - \xi)(\xi'^2 + \xi' \xi + \xi^2 + a_2(\xi' + \xi) + a_4 - a_1 \eta')}{(\xi' - \xi)(\eta' + \eta + a_1 \xi + a_3)} \\ &= \frac{\xi'^2 + \xi' \xi + \xi^2 + a_2(\xi' + \xi) + a_4 - a_1 \eta'}{\eta' + \eta + a_1 \xi + a_3} \end{aligned}$$

und ersetzt dann ξ' und η' durch ξ bzw. η .

Für den dritten Schnittpunkt $R = (\xi'', \eta'')$ dieser Geraden mit E gilt dann

$$\xi + \xi' + \xi'' = \lambda^2 + a_1 \lambda - a_2, \quad \text{also} \quad \xi'' = \lambda^2 + a_1 \lambda - a_2 - \xi - \xi'.$$

Das sieht man, wenn man $y = \lambda x + \mu$ in die Gleichung von E einsetzt:

$$x^3 - (\lambda^2 + a_1 \lambda - a_2)x^2 - (2\lambda\mu + a_1\mu + a_3\lambda - a_4)x - (\mu^2 + a_3\mu - a_6) = 0$$

ξ, ξ', ξ'' sind die drei Lösungen dieser Gleichung, also ist ihre Summe gleich minus dem Koeffizienten von x^2 .

Schließlich haben wir (mit $\eta'' = \lambda \xi'' + \mu$)

$$P + Q = -R = (\xi'', -(\lambda + a_1)\xi'' - \mu - a_3).$$

In vereinfachter Form (nämlich für kurze Weierstraß-Gleichungen) haben wir diese Formeln schon im Einführungskapitel gesehen.

Bevor wir den Satz beweisen, formulieren wir ein Lemma, das wir brauchen, um die Assoziativität nachzuweisen.

9.2. Lemma. *Seien G_i und G'_j (für $i, j \in \{1, 2, 3\}$) paarweise verschiedene Geraden in der projektiven Ebene, sodass die neun Schnittpunkte P_{ij} von G_i und G'_j paarweise verschieden sind. Sei weiter C eine ebene projektive Kurve vom Grad 3, die die acht Punkte P_{ij} mit $(i, j) \neq (3, 3)$ enthält. Dann enthält C auch den neunten Punkt P_{33} .*

LEMMA
Geraden und
Kurven vom
Grad 3

Beweis. Seien G_i und G'_j gegeben durch $L_i(X, Y, Z) = 0$ bzw. $L'_j(X, Y, Z) = 0$ mit linearen Polynomen L_i, L'_j .

Es gibt 10 Monome vom Grad 3 in drei Variablen. Die Bedingung $P_{ij} \in C$ liefert eine homogene lineare Gleichung für die zehn Koeffizienten von C . Der Raum der homogenen Polynome vom Grad 3, die in den acht gegebenen Punkten verschwinden, ist also mindestens zweidimensional. In jedem Fall liegen die Polynome $L = L_1 L_2 L_3$ und $L' = L'_1 L'_2 L'_3$ in diesem Raum und sind linear unabhängig. Wir zeigen, dass die Dimension tatsächlich genau 2 ist, d.h., der Raum wird von L und L' aufgespannt.

Dazu nehmen wir an, die Dimension sei mindestens 3. Dann können wir noch zwei beliebige Punkte vorschreiben, die auf C liegen sollen. Dazu wählen wir einen Punkt P auf G_1 , der von den Schnittpunkten mit den anderen Geraden verschieden ist, und einen Punkt Q , der auf keiner der Geraden liegt. (Dazu muss der Grundkörper K groß genug sein, damit $\mathbb{P}^2(K)$ genügend viele Punkte enthält. Der Satz gilt aber allgemein, da man für den Beweis den Körper vergrößern kann.) Sei $C: F(X, Y, Z) = 0$ eine Kurve vom Grad 3, die die acht gegebenen Punkte

und P und Q enthält. Da G_1 diese Kurve in den vier Punkten P_{1j} ($j = 1, 2, 3$) und P schneidet, muss nach dem Satz 4.3 von Bézout L_1 ein Teiler von F sein: $F = L_1 F'$ mit einem homogenen Polynom F' vom Grad 2. Die durch F' definierte Kurve vom Grad 2 schneidet die Gerade G_2 in den drei Punkten P_{2j} ($j = 1, 2, 3$), also muss L_2 ein Teiler von F' sein: $F' = L_2 F''$. Schließlich hat die durch F'' definierte Gerade mit G_3 die beiden Punkte P_{31} und P_{32} gemeinsam; die beiden Geraden stimmen also überein. Es folgt $F = cL$ mit einer Konstanten c . Das ist aber ein Widerspruch zu $Q \in C$, denn Q liegt auf keiner der Geraden G_i . Also ist die Dimension tatsächlich nur 2.

Sei nun $C: F = 0$ eine Kurve vom Grad 3 durch die acht Punkte. Wir haben gerade gezeigt, dass dann $F = cL + c'L'$ sein muss mit Konstanten c und c' . Da die rechte Seite im Punkt P_{33} verschwindet, gilt dies auch für die linke Seite, also ist $P_{33} \in C$. \square

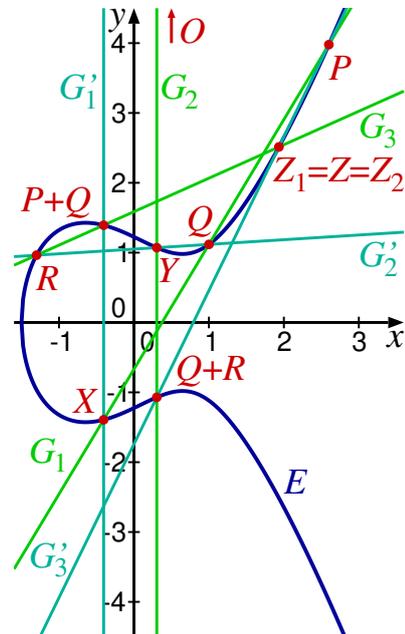
Mit dieser Vorbereitung können wir den Beweis des Satzes angehen.

Beweis von Satz 9.1. Nach dem Zusatz in Satz 4.3 ist die angegebene Verknüpfung auf $E(L)$ wohldefiniert, denn die Gerade durch zwei L -rationale Punkte (oder die Tangente an E in einem L -rationalen Punkt) ist über L definiert, also liegt der dritte Schnittpunkt auch in $E(L)$. Der Punkt O ist nach Definition das Nullelement, und wir haben schon gesehen, dass zu jedem Punkt P das Inverse $-P$ existiert. Die Kommutativität ist auch klar, da die Konstruktion der Summe $P+Q$ in P und Q symmetrisch ist. Es bleibt also noch das Assoziativgesetz

$$(P + Q) + R = P + (Q + R)$$

zu zeigen. Wir betrachten folgende Objekte.

- G_1 sei die Gerade durch P und Q ;
- X sei ihr dritter Schnittpunkt mit E .
- G'_1 sei die Gerade durch O und X ;
- ihr dritter Schnittpunkt mit E ist $P + Q$.
- G'_2 sei die Gerade durch Q und R ;
- Y sei ihr dritter Schnittpunkt mit E .
- G_2 sei die Gerade durch O und Y ;
- ihr dritter Schnittpunkt mit E ist $Q + R$.
- G_3 sei die Gerade durch $P + Q$ und R ;
- Z_1 sei ihr dritter Schnittpunkt mit E .
- G'_3 sei die Gerade durch $Q + R$ und P ;
- Z_2 sei ihr dritter Schnittpunkt mit E .
- Z sei der Schnittpunkt von G_3 und G'_3 .



Wir nehmen erst einmal an, dass die neun Punkte $O, P, Q, R, X, Y, P+Q, Q+R$ und Z paarweise verschieden sind. Wegen

$$Z_1 = -((P + Q) + R) \quad \text{und} \quad Z_2 = -(P + (Q + R))$$

genügt es zu zeigen, dass $Z_1 = Z = Z_2$ ist.

Wir wollen nun Lemma 9.2 anwenden auf unsere Geraden G_i und G'_j . Diese Geraden sind alle verschieden, denn sonst hätten wir mindestens vier Punkte im Schnitt von E mit einer Geraden (fünf oder sechs der neun Punkte liegen auf der

Vereinigung von zwei der Geraden; einer davon könnte der Punkt Z sein, von dem wir noch nicht wissen, dass er auf E liegt), was nach Satz 4.3 bedeuten würde, dass die Gerade in E enthalten ist. E ist aber irreduzibel und kann also keine Gerade als Komponente enthalten. Das Lemma ist also anwendbar. Wir haben folgende Identifikationen.

$$\begin{array}{lll} P_{11} = X, & P_{12} = Q, & P_{13} = P, \\ P_{21} = O, & P_{22} = Y, & P_{23} = Q + R, \\ P_{31} = P + Q, & P_{32} = R, & P_{33} = Z. \end{array}$$

Außerdem ist E eine Kurve vom Grad 3 durch die ersten acht Punkte, also folgt nach dem Lemma $Z \in E$. Damit ist Z der dritte Schnittpunkt sowohl von G_3 als auch von G'_3 mit E , also ist $Z_1 = Z = Z_2$.

Damit ist das Assoziativgesetz im „generischen“ Fall bewiesen. Die Fälle, wo Punkte zusammenfallen, kann man entweder einzeln behandeln, oder man verwendet eine Art „Stetigkeitsargument“ — die beiden Morphismen

$$E \times E \times E \ni (P, Q, R) \mapsto (P + Q) + R \in E$$

und

$$E \times E \times E \ni (P, Q, R) \mapsto P + (Q + R) \in E$$

stimmen auf einer „offenen, dichten“ Teilmenge überein und sind deswegen gleich. Natürlich haben wir hier weder das Produkt auf der linken Seite definiert, noch was in diesem Zusammenhang ein Morphismus ist, noch was die dabei ins Spiel kommende sogenannte *Zariski-Topologie* ist. Darum gehen wir hier etwas anders vor.

Wir stellen erst einmal fest, dass die Assoziativität trivialerweise gilt, wenn $P = O$ oder $R = O$ oder $P = R$ ist (im letzten Fall verwenden wir die Kommutativität). Wenn $P, R \neq O$ und $P \neq R$ gilt, dann gibt es nur endlich viele Punkte Q , so dass die neun Schnittpunkte im Argument oben nicht paarweise verschieden sind (Übung; hierzu ist es nützlich zu wissen, dass für einen nicht konstanten Morphismus $\psi: E \rightarrow E$ die Gleichung $\psi(S) = T$ für einen gegebenen Punkt $T \in E$ nur endlich viele Lösungen hat). Wir betrachten den Morphismus

$$\phi_{P,R}: E \rightarrow E, \quad Q \mapsto ((P + Q) + R) + (-(P + (Q + R))).$$

Die Äquivalenz $P + (-Q) = O \iff P = Q$ ist leicht zu sehen. Es folgt, dass $\phi_{P,R}(Q) = O$ ist für alle bis auf endlich viele $Q \in E(\bar{K})$; dann muss $\phi_{P,R}$ aber konstant $= O$ sein. Das bedeutet, dass die Gleichung $(P + Q) + R = P + (Q + R)$ für *alle* Q gilt. \square

Man kann die Assoziativität natürlich auch mithilfe der expliziten Additionsformeln beweisen, die wir oben hergeleitet haben. Man muss dazu zeigen, dass die Ausdrücke für die Koordinaten von $(P + Q) + R$ und von $P + (Q + R)$ modulo der Gleichung von E ausgewertet in P, Q und R gleich sind. Wenn man das von Hand macht, ist es sehr mühsam; mit einem Computeralgebra-System ist es jedoch ohne große Probleme machbar.

9.3. Bemerkung. Man kann die Gruppenstruktur auch wie folgt „intrinsisch“ charakterisieren. Seien P, Q, R Punkte von E . Es gilt genau dann $P + Q = R$, wenn es eine rationale Funktion ϕ auf E gibt, die in P und Q einfache Nullstellen, in R und O einfache Polstellen und sonst keine Null- oder Polstellen hat. (Falls von den vier Punkten O, P, Q, R welche zusammenfallen, muss man die Null- und Polstellenordnungen entsprechend verrechnen.)

Die eine Implikation ist leicht zu sehen. Sei $L_1(X, Y, Z) = 0$ die Gleichung der Geraden durch P und Q und $L_2(X, Y, Z) = 0$ die Gleichung der Geraden durch R

BEM
Gruppen-
struktur
intrinsisch

und O . Dann ist $\phi = L_1/L_2$ eine passende Funktion: der Zähler verschwindet in P , Q und $-R$, und der Nenner verschwindet in R , O und $-R$, sodass die verlangten Null- und Polstellen auftreten (die Nullstellen von Zähler und Nenner bei $-R$ „kürzen sich weg“).

Diese Charakterisierung impliziert, dass jeder Isomorphismus von Kurven $E \rightarrow E'$, der O auf O abbildet, auch mit der Gruppenstruktur verträglich ist. Für unsere explizite Definition von Isomorphismen elliptischer Kurven folgt diese Aussage daraus, dass so ein Isomorphismus linear ist und daher Geraden auf Geraden abbildet. Tripel von Schnittpunkten der Kurve mit einer Geraden werden also auf ebensolche Tripel abgebildet, und damit bleibt auch die Gruppenstruktur erhalten.

Eine Voraussetzung für diese Charakterisierung ist allerdings, dass man die Ordnung einer Null- bzw. Polstelle einer rationalen Funktion definieren muss. Siehe z.B. das Skript „Arithmetic of hyperelliptic curves“, Abschnitt 2. ♠

9.4. Beispiel. Wir benutzen die Definition der Gruppenstruktur, um uns zu überlegen, welche Punkte der Ordnung 2 bzw. 3 eine elliptische Kurve hat. Wir nehmen wieder an, dass die Charakteristik des Grundkörpers K nicht 2 oder 3 ist, und arbeiten mit einer kurzen Weierstraß-Gleichung

$$E: y^2 = x^3 + ax + b.$$

Wir nehmen außerdem an, dass K algebraisch abgeschlossen ist (bzw. sagen „Punkt“ für „ \bar{K} -rationaler Punkt“).

Ein Punkt $P \in E(\bar{K})$ hat Ordnung 2, wenn $P \neq O$ ist und $2P = P + P = O$ gilt. Das ist äquivalent zu $P + P + O = O$; nach Definition heißt das gerade, dass P, P, O die drei Schnittpunkte von E mit einer Geraden sind. Da diese Gerade O enthält, ist sie senkrecht. Eine solche Gerade hat die Gleichung $x = \xi$ für ein festes ξ ; sie schneidet den affinen Teil von E in den (i.A.) zwei Punkten mit x -Koordinate ξ . Diese Punkte haben die Form (ξ, η) und $(\xi, -\eta)$. In unserem Fall müssen sie übereinstimmen; das bedeutet $\eta = 0$ (tatsächlich ist in einem Punkt mit verschwindender y -Koordinate die Tangente an E senkrecht). Es gibt also genau drei Punkte der Ordnung 2, nämlich die Punkte $(\xi, 0)$ mit $\xi^3 + a\xi + b = 0$.

Ein Punkt $P \in E(\bar{K})$ hat Ordnung 3, wenn $P \neq O$ und $3P = P + P + P = O$ ist. Nach Definition heißt das, dass es eine Gerade geben muss, die E in P mit Vielfachheit 3 schneidet. Das bedeutet, dass P ein Wendepunkt von E ist (und die Gerade die Wendetangente). Man kann zeigen, dass eine glatte kubische Kurve stets genau neun Wendepunkte hat (falls $\text{char}(K) \neq 3$). Einer davon ist O , also gibt es acht Punkte der Ordnung 3. ♣

In Charakteristik 2 gibt es keinen oder einen Punkt der Ordnung 2, je nachdem, ob $a_1 = 0$ oder $a_1 \neq 0$ ist; im letzteren Fall ist es der Punkt $(-a_3/a_1, \sqrt{x^3 + a_2x^2 + a_4x + a_6})$ (man beachte, dass in Charakteristik 2 die Quadratwurzel eindeutig bestimmt ist).

In Charakteristik 3 ist es analog: Es gibt entweder keinen oder zwei Punkte der Ordnung 3. Die Formeln sind etwas komplizierter; die Bedingung für die Existenz von Punkten der Ordnung 3 ist hier, dass $a_1^2 + a_2 \neq 0$ ist.

BSP
Punkte der
Ordnung 2, 3

10. ISOGENIEN UND ENDOMORPHISMEN

Die relevanten Abbildungen zwischen elliptischen Kurven sind Morphismen, welche die Gruppenstruktur respektieren. Bevor wir sie einführen, brauchen wir noch einige Aussagen über den Zusammenhang zwischen rationalen Abbildungen und Funktionenkörpern.

10.1. Satz. *Seien C und D zwei irreduzible (projektive) Kurven über K . Dann gibt es eine Bijektion zwischen der Menge der nicht-konstanten rationalen Abbildungen $\phi: C \rightarrow D$ über K und der Menge der K -linearen Homomorphismen $\phi^*: K(D) \rightarrow K(C)$ der Funktionenkörper. Dabei ist $K(C)$ eine endliche Körpererweiterung von $\phi^*(K(D))$.*

SATZ
rationale
Abbildungen
und
Funktionenkörper

Ist C' eine weitere Kurve und $\psi: D \rightarrow C'$ eine nicht-konstante rationale Abbildung, dann gilt $(\psi \circ \phi)^ = \phi^* \circ \psi^*$.*

Beweis. (Skizze) Es ist etwas einfacher, die Beziehung in affinen Koordinaten zu formulieren. Dazu nehmen wir an, dass weder C noch D die unendlich ferne Gerade $Z = 0$ ist (ansonsten muss man die Überlegung leicht modifizieren). Seien C' und D' die affinen Teile von C und D ; dann gilt $K(C') = K(C)$, $K(D') = K(D)$. Wir bezeichnen die affinen Koordinaten(funktionen) auf C mit x und y und auf D mit u und v . Es gilt dann $K(C) = K(x, y)$ und $K(D) = K(u, v)$.

Eine rationale Abbildung $\phi: C \rightarrow D$ ist in affinen Koordinaten gegeben durch zwei rationale Funktionen $r(x, y), s(x, y) \in K(x, y) = K(C)$, sodass $(r(x, y), s(x, y))$ ein $K(C)$ -rationaler Punkt von D' ist (denn r und s erfüllen die affine Gleichung von D'):

$$\phi: (x, y) \mapsto (r(x, y), s(x, y)).$$

ϕ ist nicht konstant, wenn r und s nicht beide konstant (d.h. in $\bar{K} \cap K(C)$) sind. Der zugehörige Homomorphismus der Funktionenkörper ist dann gegeben durch

$$\phi^*: K(D) \ni f \mapsto f \circ \phi \in K(C).$$

In Koordinaten ausgedrückt, haben wir

$$\phi^*(u) = r(x, y), \quad \phi^*(v) = s(x, y).$$

Ein Homomorphismus von Körpern ist stets injektiv (da das einzige echte Ideal das Nullideal ist). Die Körpererweiterung $K(C)/\phi^*(K(D))$ ist endlich, weil x und y über $\phi^*(K(D)) = K(r(x, y), s(x, y))$ algebraisch sind.

Ist umgekehrt ein K -linearer Homomorphismus $\psi: K(D) \rightarrow K(C)$ gegeben, dann setzen wir $r(x, y) = \psi(u)$, $s(x, y) = \psi(v)$. Sei $g(u, v) = 0$ die Gleichung von D' , dann haben wir

$$g(r(x, y), s(x, y)) = g(\psi(u), \psi(v)) = \psi(g(u, v)) = \psi(0) = 0,$$

also ist

$$\phi: C \longrightarrow D, \quad (x, y) \longmapsto (r(x, y), s(x, y))$$

eine (nicht konstante) rationale Abbildung, und $\psi = \phi^*$.

Außerdem ist für $f \in K(C')$

$$(\psi \circ \phi)^*(f) = f \circ (\psi \circ \phi) = (f \circ \psi) \circ \phi = \phi^*(\psi^*(f)) = (\phi^* \circ \psi^*)(f). \quad \square$$

10.2. Definition. In der Situation von Satz 10.1 heißt der Grad der Körpererweiterung $K(C)/\phi^*(K(D))$ dann auch der *Grad* von ϕ , $\deg \phi$.

DEF
Grad von ϕ
separabel

Sei $\phi^*(K(D)) \subset L \subset K(C)$ der maximale Zwischenkörper, der über $\phi^*(K(D))$ separabel ist. Dann heißt der Grad $[L : \phi^*(K(D))]$ der *separable Grad* von ϕ , $\deg_s \phi$, und der Grad $[K(C) : L]$ der *inseparable Grad* von ϕ , $\deg_i \phi$. Es gilt offensichtlich $\deg \phi = (\deg_s \phi)(\deg_i \phi)$.

ϕ heißt *separabel*, wenn $L = K(C)$ ist (das ist insbesondere stets dann der Fall, wenn $\text{char}(K) = 0$ ist), sonst *inseparabel*. ϕ heißt *rein inseparabel*, wenn $L = \phi^*(K(D))$ ist. \diamond

10.3. Folgerung. Zwei irreduzible Kurven C und D über K sind genau dann birational äquivalent über K , wenn ihre Funktionenkörper $K(C)$ und $K(D)$ isomorphe Körpererweiterungen von K sind.

FOLG
birationale
Äquivalenz

Beweis. „ \Rightarrow “: Sei $\phi: C \rightarrow D$ eine birationale Abbildung mit Inverser $\psi = \phi^{-1}$. Nach Satz 10.1 ist dann $\psi^* \circ \phi^* = (\phi \circ \psi)^* = \text{id}_{K(D)}$ und $\phi^* \circ \psi^* = (\psi \circ \phi)^* = \text{id}_{K(C)}$, also ist ϕ^* ein K -linearer Isomorphismus von $K(D)$ und $K(C)$.

„ \Leftarrow “: Sei $\alpha: K(D) \rightarrow K(C)$ ein K -linearer Isomorphismus mit inversem Isomorphismus $\beta = \alpha^{-1}$. Dann gibt es nach Satz 10.1 eine rationale Abbildung $\phi: C \rightarrow D$ mit $\alpha = \phi^*$ und eine rationale Abbildung $\psi: D \rightarrow C$ mit $\beta = \psi^*$. Es folgt

$$\begin{aligned} (\psi \circ \phi)^* &= \phi^* \circ \psi^* = \alpha \circ \beta = \text{id}_{K(C)} && \text{und} \\ (\phi \circ \psi)^* &= \psi^* \circ \phi^* = \beta \circ \alpha = \text{id}_{K(D)}, \end{aligned}$$

also sind ϕ und ψ zueinander inverse rationale Abbildungen; damit sind C und D birational äquivalent. \square

10.4. Folgerung. Seien C_1, C_2, C_3 irreduzible Kurven über K und $\phi_1: C_1 \rightarrow C_2$, $\phi_2: C_2 \rightarrow C_3$ nicht-konstante rationale Abbildungen. Dann gilt

FOLG
Multiplika-
tät
des Grades

$$\begin{aligned} \deg(\phi_2 \circ \phi_1) &= (\deg \phi_2)(\deg \phi_1), \\ \deg_s(\phi_2 \circ \phi_1) &= (\deg_s \phi_2)(\deg_s \phi_1), \\ \deg_i(\phi_2 \circ \phi_1) &= (\deg_i \phi_2)(\deg_i \phi_1). \end{aligned}$$

Beweis. Die erste Gleichung folgt aus der Multiplikativität der Grade in der Körpererweiterung $K(C_3) \hookrightarrow K(C_2) \hookrightarrow K(C_1)$.

Für die zweite Gleichung sei L der maximale separable Zwischenkörper der Erweiterung $K(C_3) \hookrightarrow K(C_1)$, L' der maximale separable Zwischenkörper der Erweiterung $K(C_3) \hookrightarrow K(C_2)$ und L'' der maximale separable Zwischenkörper von $K(C_2) \hookrightarrow K(C_1)$. Dann gilt $L'' = K(C_2) \cdot L$ und $[L : L'] = [L'' : K(C_2)]$. Damit erhalten wir

$$\begin{aligned} \deg_s(\phi_2 \circ \phi_1) &= [L : K(C_3)] = [L : L'] \cdot [L' : K(C_3)] \\ &= [L'' : K(C_2)] \cdot [L' : K(C_3)] = (\deg_s \phi_1)(\deg_s \phi_2). \end{aligned}$$

Die dritte Gleichung folgt aus den ersten beiden. \square

Jetzt können wir die relevanten Abbildungen einführen. Es zeigt sich, dass es (wie bei Isomorphismen) genügt, die Minimalvoraussetzung $\phi(O) = O$ zu fordern.

10.5. Definition. Seien E und E' elliptische Kurven über K . Eine *Isogenie* von E nach E' ist ein Morphismus $\phi: E \rightarrow E'$, sodass $\phi(O) = O$ ist. Die Kurven E und E' heißen *isogen*, wenn es eine nicht konstante Isogenie $E \rightarrow E'$ gibt. \diamond

DEF
Isogenie

So eine Isogenie ist also entweder konstant: $\phi(P) = O$ für alle $P \in E$, oder surjektiv (als Abbildung $\phi_{\bar{K}}: E(\bar{K}) \rightarrow E'(\bar{K})$).

In der Literatur wird die konstante Abbildung $P \mapsto O$ nicht immer als Isogenie angesehen.



Die wichtigste Eigenschaft von Isogenien ist, dass sie automatisch die Gruppenstrukturen von E und E' respektieren.

10.6. Satz. Sei $\phi: E \rightarrow E'$ eine Isogenie. Dann gilt $\phi_L(P+Q) = \phi_L(P) + \phi_L(Q)$ für alle $P, Q \in E(L)$, d.h., ϕ ist ein Gruppenhomomorphismus.

SATZ
Isogenie
ist Homo-
morphismus

Beweis. Siehe zum Beispiel [Si1, Thm. III.4.8]. Wir können hier nur eine Andeutung bringen: Wir hatten bemerkt, dass die Summe $P+Q$ dadurch charakterisiert ist, dass es eine rationale Funktion f auf E gibt, die in P und Q einfache Nullstellen und in O und $P+Q$ einfache Polstellen hat. Wenn $\phi: E \rightarrow E'$ ein nicht-konstanter Morphismus ist mit $\phi(O) = O$, dann können wir die Norm von f , $N(f) \in K(E')$ bezüglich der durch ϕ^* gegebenen Körpererweiterung betrachten. Diese Funktion $N(f)$ hat dann einfache Nullstellen in $\phi(P)$ und $\phi(Q)$ und einfache Pole in $\phi(O) = O$ und $\phi(P+Q)$. Es folgt, dass $\phi(P+Q) = \phi(P) + \phi(Q)$ ist. \square

Die wichtigsten Beispiele von Isogenien sind die *Multiplikationsabbildungen*. Sei $m \in \mathbb{Z}$ und sei E eine elliptische Kurve. Dann definiert

DEF
Multiplika-
tionsabbildung

$$[m]: E \ni P \mapsto [m](P) = m \cdot P \in E$$

eine Isogenie ($m \cdot P$ ist dabei das m -fache von P als Element einer abelschen Gruppe (= \mathbb{Z} -Modul)). $[m]$ ist für $m \neq 0$ nicht konstant (vgl. [Si1, Prop. III.4.2.(a)]). Man zeigt das direkt für $m = 2$. Es genügt dann, den Fall m ungerade zu behandeln. Im Fall $\text{char}(K) \neq 2$ findet man einen Punkt $P \in E(\bar{K})$ mit $P \neq O$ und $2P = O$; dann ist $mP = P \neq O$, also ist $[m]$ nicht konstant. Im Fall $\text{char}(K) = 2$ zeigt man die Aussage direkt für $m = 3$ und verfährt dann ähnlich mit einem Punkt der Ordnung 3.

Wollen wir andeuten, auf welcher elliptischen Kurve wir die Multiplikationsabbildung betrachten, dann schreiben wir $[m]_E$.

10.7. Definition. Die Isogenien $E \rightarrow E$ (wie zum Beispiel die Multiplikationsabbildungen) heißen dann auch *Endomorphismen*; sie bilden einen Ring $\text{End}_K(E)$ (der ein Unterring des Endomorphismenrings der abelschen Gruppe $E(\bar{K})$ ist) — die Summe ist punktweise definiert: $(\phi + \psi)(P) = \phi(P) + \psi(P)$, das Produkt als Hintereinanderschaltung: $\phi \cdot \psi = \phi \circ \psi$. \diamond

DEF
Endo-
morphismus

Wie für jede nicht-konstante rationale Abbildung zwischen Kurven haben wir für jede nicht-konstante Isogenie $\phi: E \rightarrow E'$ den Grad $\text{deg } \phi$, den separablen Grad $\text{deg}_s \phi$ und den inseparablen Grad $\text{deg}_i \phi$. Der Vollständigkeit halber setzt man noch $\text{deg } 0 = \text{deg}_s 0 = \text{deg}_i 0 = 0$ (wo links 0 die konstante Isogenie $P \mapsto O$ bezeichnet). Es gilt dann

$$\text{deg}(\psi \circ \phi) = (\text{deg } \psi)(\text{deg } \phi), \quad \text{deg } \phi \geq 0 \quad \text{und} \quad \text{deg } \phi = 0 \iff \phi = 0.$$

10.8. Satz. *Der Endomorphismenring $\text{End}_K(E)$ ist ein nullteilerfreier Ring der Charakteristik 0.*

SATZ
Endomorphismenring

Beweis. Seien $\phi, \psi \in \text{End}_K(E)$ mit $\phi \cdot \psi = 0$. Dann folgt $0 = \deg(\phi\psi) = \deg(\phi)\deg(\psi)$, also gilt $\deg(\phi) = 0$ oder $\deg(\psi) = 0$ und damit $\phi = 0$ oder $\psi = 0$. Damit ist gezeigt, dass $\text{End}_K(E)$ nullteilerfrei ist.

Außerdem ist $[m] = 0$ (d.h. konstant) nur dann, wenn $m = 0$ ist; der Homomorphismus $\mathbb{Z} \ni m \mapsto [m] \in \text{End}_K(E)$ ist also injektiv. Das bedeutet, dass der Endomorphismenring Charakteristik null hat. \square

Insbesondere haben wir immer die Einbettung $\mathbb{Z} \ni m \mapsto [m] \in \text{End}_K(E)$.

Man kann die möglichen Endomorphismenringe ziemlich genau klassifizieren. In Charakteristik 0 ist $\text{End}_K(E) = \mathbb{Z}$ der Normalfall. Über endlichen Körpern ist der Endomorphismenring aber stets größer, da man zusätzlich den *Frobenius-Endomorphismus* $(x, y) \mapsto (x^q, y^q)$ hat (wobei q die Größe des Grundkörpers ist). Darauf kommen wir später noch ausführlich zu sprechen.

10.9. Bemerkung. Folgende Aussagen aus dem Beweis von Satz 10.8 gelten allgemeiner für Isogenien zwischen möglicherweise verschiedenen elliptischen Kurven:

BEM
Eigenschaften von Isogenien

$$\begin{aligned} \forall \phi_1, \phi_2: E \rightarrow E', \psi: E' \rightarrow E'': \quad & \psi \circ (\phi_1 + \phi_2) = \psi \circ \phi_1 + \psi \circ \phi_2 \\ \forall \phi: E \rightarrow E', \psi_1, \psi_2: E' \rightarrow E'': \quad & (\psi_1 + \psi_2) \circ \phi = \psi_1 \circ \phi + \psi_2 \circ \phi \\ \forall \phi: E \rightarrow E', \psi: E' \rightarrow E'': \quad & \psi \circ \phi = 0 \iff (\psi = 0 \text{ oder } \phi = 0) \end{aligned}$$

Dabei ist die Summe zweier Isogenien $E \rightarrow E'$ wie in Definition 10.7 punktweise definiert. \spadesuit

Eine ganz wichtige Eigenschaft ist auch die folgende.

10.10. Satz. *Sei $\phi: E \rightarrow E'$ eine nicht-konstante Isogenie vom Grad m . Dann gibt es genau eine Isogenie $\hat{\phi}: E' \rightarrow E$, die zu ϕ duale Isogenie, sodass $\hat{\phi} \circ \phi = [m]_E$ ist. Es gilt dann auch $\phi \circ \hat{\phi} = [m]_{E'}$. Weitere Eigenschaften sind:*

SATZ
duale Isogenie

- (1) Ist $\psi: E' \rightarrow E''$ eine weitere Isogenie, dann gilt $\widehat{\psi \circ \phi} = \hat{\psi} \circ \hat{\phi}$.
- (2) Ist $\psi: E \rightarrow E'$ eine weitere Isogenie, dann gilt $\widehat{\phi + \psi} = \hat{\phi} + \hat{\psi}$.
- (3) $\hat{\hat{\phi}} = \phi$.
- (4) $\deg(\hat{\phi}) = \deg(\phi)$.
- (5) Für $m \in \mathbb{Z}$ gilt $\widehat{[m]_E} = [m]_{E'}$ und $\deg([m]_{E'}) = m^2$.

Man setzt noch $\hat{0} = 0$; dann gelten (1)–(5) für beliebige Isogenien.

Beweis. Siehe zum Beispiel [Si1, Thms III.6.1 und 6.2]. Wir zeigen hier die Existenz von $\hat{\phi}$ nicht. Die Eindeutigkeit ergibt sich so: Seien $\psi, \psi': E' \rightarrow E$ Isogenien mit $\psi \circ \phi = \psi' \circ \phi = [m]$. Dann folgt $(\psi - \psi') \circ \phi = \psi \circ \phi - \psi' \circ \phi = 0$; weil $\phi \neq 0$ ist, folgt daraus $\psi - \psi' = 0$, also $\psi = \psi'$.

Aus $\hat{\phi} \circ \phi = [m]_E$ folgt auch

$$(\phi \circ \hat{\phi}) \circ \phi = \phi \circ (\hat{\phi} \circ \phi) = \phi \circ [m]_E = [m]_{E'} \circ \phi$$

und dann mit einem ähnlichen Argument wie eben $\phi \circ \hat{\phi} = [m]_{E'}$.

Wir zeigen jetzt noch einige der Eigenschaften.

- (1) Sei $m = \deg \phi$, $n = \deg \psi$; dann ist $\deg(\psi \circ \phi) = nm$. Die Isogenie $\hat{\phi} \circ \hat{\psi}$ erfüllt $(\hat{\phi} \circ \hat{\psi}) \circ (\psi \circ \phi) = \hat{\phi} \circ (\hat{\psi} \circ \psi) \circ \phi = \hat{\phi} \circ [n] \circ \phi = [n] \circ (\hat{\phi} \circ \phi) = [n] \circ [m] = [nm]$,
muss also wegen der Eindeutigkeit gleich $\widehat{\psi \circ \phi}$ sein.
- (5) Dass $\deg[m] = m^2$ ist, kann man durch eine explizite Rekursion für die Funktionen $r_m(x)$, die die x -Koordinate von mP für $P = (x, y)$ liefern, beweisen (siehe Lemma 10.11 unten). Wegen $[m] \circ [m] = [m^2] = [\deg[m]]$ folgt daraus $\widehat{[m]} = [m]$.
- (4) Es gilt $m^2 = \deg[m] = (\deg \hat{\phi})(\deg \phi) = (\deg \hat{\phi})m$; es folgt $\deg \hat{\phi} = m = \deg \phi$.
- (3) Es gilt $\phi \circ \hat{\phi} = [m] = [\deg \hat{\phi}]$, also folgt $\phi = \hat{\phi}$.
- (2) Das beweisen wir hier nicht. □

10.11. Lemma. *Ist $\phi: E \rightarrow E'$ eine nicht-konstante Isogenie zwischen elliptischen Kurven, die durch Weierstraß-Gleichungen der Form $y^2 = f(x)$ gegeben sind, dann hat ϕ die Form $(x, y) \mapsto (r(x), s(x)y)$ mit Quotienten von Polynomen $r(x) = p(x)/q(x)$ und $s(x)$. Ist $r(x)$ in gekürzter Form gegeben, dann ist $\deg \phi = \max\{\deg p, \deg q\}$.*

LEMMA
Form von
Isogenien

Beweis. Da y eine quadratische Gleichung über $K(x)$ erfüllt, kann man jede rationale Abbildung $E \rightarrow E'$ eindeutig in der Form

$$(x, y) \mapsto (r_1(x) + r_2(x)y, s_1(x) + s_2(x)y)$$

schreiben, mit rationalen Ausdrücken r_1, r_2, s_1, s_2 . Nun gilt $\phi(x, -y) = -\phi(x, y)$, also

$$(r_1(x) - r_2(x)y, s_1(x) - s_2(x)y) = (r_1(x) + r_2(x)y, -s_1(x) - s_2(x)y).$$

Daher müssen r_2 und s_1 auf E verschwinden.

Außerdem gilt (wenn x', y' die affinen Koordinatenfunktionen auf E' bezeichnen) $[K(E') : K(x')] = [K(x')(y') : K(x')] = 2$ und $[K(E) : K(x)] = 2$, sowie

$$[K(x) : K(x')] = [K(x) : K(r(x))] = \max\{\deg p, \deg q\}.$$

Aus der Multiplikativität der Grade in Körpererweiterungen folgt dann

$$\begin{aligned} \deg \phi &= [K(E) : K(E')] \\ &= \frac{[K(E) : K(x)][K(x) : K(x')]}{[K(E') : K(x')]} \\ &= [K(x) : K(x')] = \max\{\deg p, \deg q\}. \end{aligned} \quad \square$$

Die Aussage, dass die x -Koordinate von $\phi(x, y)$ die Form $r(x)$ hat, gilt allgemein, auch für lange Weierstraß-Gleichungen. Dasselbe gilt für die Formel für den Grad von ϕ .

Nach so vielen neuen Begriffen ist ein Beispiel angebracht.

10.12. **Beispiel.** Sei K ein Körper mit $\text{char}(K) \neq 2$, und sei

$$E: y^2 = x^3 + ax^2 + bx$$

eine elliptische Kurve über K . (Das bedeutet $b \neq 0$ und $a^2 - 4b \neq 0$.) Man beachte, dass der Punkt $(0, 0) \in E(K)$ die Ordnung 2 hat. Dann ist auch

$$E': y^2 = x^3 - 2ax^2 + (a^2 - 4b)x$$

eine elliptische Kurve über K , und wir haben die beiden dualen Isogenien

$$\begin{aligned} \phi: E &\longrightarrow E', & (x, y) &\longmapsto \left(\frac{y^2}{x^2}, \frac{b - x^2}{x^2} y \right) = \left(\frac{x^2 + ax + b}{x}, \frac{b - x^2}{x^2} y \right) \\ \hat{\phi}: E' &\longrightarrow E, & (x, y) &\longmapsto \left(\frac{y^2}{4x^2}, \frac{a^2 - 4b - x^2}{8x^2} y \right) \end{aligned}$$

Man sieht, dass beide Grad 2 haben, und man rechnet nach, dass $\hat{\phi} \circ \phi = [2]_E$ und $\phi \circ \hat{\phi} = [2]_{E'}$ ist, wie es sein muss (Übung).

Der Kern von ϕ besteht offenbar aus den zwei Elementen $O, (0, 0) \in E(K)$; analog besteht der Kern von $\hat{\phi}$ aus den beiden Elementen $O, (0, 0) \in E'(K)$. Dass die Größe des Kerns gerade dem Grad entspricht, ist kein Zufall. Allerdings kann es vorkommen, dass die Punkte im Kern nicht alle K -rational sind. ♣

Wenn wir den Satz über die duale Isogenie auf den Endomorphismenring (also Isogenien $E \rightarrow E$) anwenden, dann bekommen wir folgendes Resultat.

10.13. **Satz.** Die Abbildung $\text{End}_K(E) \rightarrow \text{End}_K(E), \phi \mapsto \hat{\phi}$, ist eine Anti-Involution von $\text{End}_K(E)$ (d.h. ein zu sich selbst inverser Anti-Automorphismus, wobei das „Anti“ sich darauf bezieht, dass die Reihenfolge der Faktoren in einem Produkt vertauscht wird). Wenn wir \mathbb{Z} mit seinem Bild in $\text{End}_K(E)$ identifizieren, dann gilt

$$\phi + \hat{\phi} \in \mathbb{Z} \quad \text{und} \quad \phi\hat{\phi} = \text{deg}(\phi).$$

Außerdem definiert deg eine positiv definite quadratische Form auf $\text{End}_K(E)$.

Beweis. Dass das Dualisieren eine Anti-Involution ist, folgt aus Satz 10.10, (1) bis (3). Der erste Teil dieses Satzes zeigt auch $\phi\hat{\phi} = \text{deg}(\phi)$. Um zu sehen, dass $\phi + \hat{\phi} \in \mathbb{Z}$ ist, betrachten wir

$$\mathbb{Z} \ni \text{deg}(1 + \phi) = (1 + \phi)(\widehat{1 + \phi}) = (1 + \phi)(1 + \hat{\phi}) = 1 + \phi + \hat{\phi} + \text{deg}(\phi).$$

Dass deg eine quadratische Form ist, bedeutet, dass $\text{deg}(n\phi) = n^2 \text{deg} \phi$ ist für $n \in \mathbb{Z}$ und dass

$$(\phi, \psi) \longmapsto \text{deg}(\phi + \psi) - \text{deg}(\phi) - \text{deg}(\psi)$$

\mathbb{Z} -bilinear in ϕ und ψ ist. Die erste Aussage folgt leicht aus Satz 10.10, denn $\text{deg}(n\phi) = (\text{deg } n)(\text{deg} \phi) = n^2 \text{deg} \phi$. Die zweite Aussage folgt daraus, dass sich die rechte Seite umformen lässt zu $\phi\hat{\psi} + \psi\hat{\phi}$ und dass $\phi \mapsto \hat{\phi}$ natürlich \mathbb{Z} -linear ist. deg ist positiv definit, weil $\text{deg}(\phi) \geq 0$ ist für alle ϕ , und $\text{deg}(\phi) = 0$ ist nur für $\phi = 0$. □

Es ist das Vorhandensein dieser Anti-Involution, die eine positiv definite quadratische Form induziert, die die Klassifikation der möglichen Endomorphismenringe ermöglicht (zusammen mit einem weiteren Resultat, das den Rang von $\text{End}_K(E)$ als \mathbb{Z} -Modul durch 4 beschränkt).

BSP
Isogenie

SATZ
Dualisieren
in $\text{End}_K(E)$

Wir brauchen noch ein Resultat über den Zusammenhang zwischen der Größe des Kerns einer Isogenie und ihrem (separablen) Grad.

10.14. **Satz.** *Sei $\phi: E \rightarrow E'$ eine nicht-konstante Isogenie über K . Dann gilt für alle $P \in E'(\bar{K})$*

$$\#\phi_{\bar{K}}^{-1}(P) = \deg_s(\phi).$$

Insbesondere hat der Kern von $\phi_{\bar{K}}$ die Ordnung $\deg_s(\phi)$.

SATZ
Kern
und Grad

Beweis. Siehe [Si1, Thm. III.4.10]. Grob gesagt, erhalten wir eine algebraische Gleichung vom Grad $\deg \phi$ für die x -Koordinaten der Urbilder von $P = (\xi, \eta)$, die sich schreiben lässt als $f(x^{p^k}) = f_1(x^{p^k}) - \xi f_2(x^{p^k}) = 0$, wo p^k der inseparable Grad von ϕ und $\deg_s \phi = \deg f = \max\{\deg f_1, \deg f_2\}$ ist. (Dabei ist p die Charakteristik von K . Im Fall $\text{char}(K) = 0$ ist $p^k = \deg_i \phi = 1$.) Dann ist $f(x)$ für fast alle ξ ein Polynom ohne mehrfache Nullstellen, also gibt es für fast alle ξ genau $\deg_s \phi$ Lösungen in \bar{K} , die zu genau $\deg_s \phi$ Punkten Q führen mit $\phi_{\bar{K}}(Q) = P$. (Genau einer der Punkte mit der jeweiligen x -Koordinate wird auf P abgebildet, der andere (wenn es ihn gibt) auf $-P$.) Da die Mengen $\phi_{\bar{K}}^{-1}(P)$ für verschiedene P durch Translation (Addition eines geeigneten Punktes in $E(\bar{K})$) auseinander hervorgehen, müssen dann alle diese Mengen genau $\deg_s \phi$ Elemente haben. \square

Für den Fall, dass wir in Charakteristik p sind, brauchen wir noch Informationen darüber, wann eine Isogenie (d.h. die von ihr induzierte Körpererweiterung der Funktionkörper) nicht separabel ist.

Sei dazu E eine elliptische Kurve über einem Körper K der Charakteristik p , und sei $q = p^e$ eine Potenz von p . Wenn wir in der Weierstraß-Gleichung von E jeden Koeffizienten a_j durch seine q -te Potenz a_j^q ersetzen, bekommen wir eine Gleichung, die eine elliptische Kurve $E^{(q)}$ über K definiert (die Diskriminante der neuen Gleichung ist die q -te Potenz der Diskriminante der alten Gleichung, also von null verschieden). Außerdem definiert dann

$$\phi: E \longrightarrow E^{(q)}, \quad (x, y) \longmapsto (x^q, y^q)$$

eine Isogenie. Wenn K endlich und q eine Potenz von $\#K$ ist, dann ist $E^{(q)} = E$; ϕ heißt im Fall $q = \#K$ der *Frobenius-Endomorphismus* von E .

DEF
Frobenius-
Endo-
morphismus
LEMMA
inseparable
Endo-
morphismen

10.15. **Lemma.** *Sei $K = \mathbb{F}_q$ mit $q = p^e$ und E eine elliptische Kurve über K .*

- (1) *Sei $\phi: E \rightarrow E^{(p)}$, $(x, y) \mapsto (x^p, y^p)$. Dann ist ϕ rein inseparabel:
 $\deg \phi = \deg_i \phi = p$.*
- (2) *Sei $\pi \in \text{End}_K(E)$ der Frobenius-Endomorphismus und seien $m, n \in \mathbb{Z}$. Der Endomorphismus $m + n\pi$ ist genau dann separabel, wenn m nicht durch p teilbar ist.*

Beweis. Siehe [Si1, Cor. III.5.5 und Prop. II.2.11]. Teil (1) ist klar, denn wir adjungieren eine p -te Wurzel. Die Aussage über den Grad folgt aus Lemma 10.11.

Es gilt $\deg \phi = p$, also $[p] = \hat{\phi}\phi$, und damit $\deg_i [p] \geq \deg_i \phi = p > 1$. Wir können $\pi = \psi\phi$ zerlegen (mit $\psi: E^{(p)} \rightarrow E$, $(x, y) \mapsto (x^{p^{e-1}}, y^{p^{e-1}})$). Ist $m = pm'$ durch p teilbar, dann haben wir $m + n\pi = (m'\hat{\phi} + n\psi) \circ \phi$; dieser Endomorphismus ist also inseparabel. Dass die Multiplikation mit m im Fall $p \nmid m$ separabel ist, folgt aus $p \nmid m^2 = \deg[m]$ und der Tatsache, dass der inseparable Grad stets eine Potenz von p ist. Für die andere Richtung von (2) (die für uns wichtiger ist), braucht man

die Aussage, dass die Summe einer separablen und einer inseparablen Isogenie separabel ist. Das kann man (wie in Silvermans Buch) mit Hilfe des *invarianten Differentials* beweisen. \square

10.16. **Definition.** Ist $\phi: E \rightarrow E'$ eine Isogenie, dann schreiben wir $E[\phi]$ für ihren Kern, d.h. **DEF**
 $E[\phi]$

$$E[\phi] = \ker \phi_{\bar{K}} = \{P \in E(\bar{K}) \mid \phi_{\bar{K}}(P) = O\}.$$

Für die K -rationalen Punkte im Kern schreiben wir $E(K)[\phi]$. Ist $\phi = [m]$ eine Multiplikationsabbildung, dann schreiben wir einfach $E[m]$ für den Kern (also die Gruppe der Punkte, deren Ordnung ein Teiler von m ist). \diamond

11. TORSION UND WEIL-PAARUNG

In diesem Abschnitt wollen wir die Struktur der m -Torsionspunkte einer elliptischen Kurve genauer untersuchen. Das sind die Punkte P mit $m \cdot P = O$.

DEF
 m -Torsionspunkt

11.1. Satz. Sei E eine elliptische Kurve über einem algebraisch abgeschlossenen Körper K und sei $m \in \mathbb{Z}_{>0}$.

SATZ
Torsion

(1) Wenn $\text{char}(K)$ kein Teiler von m ist (z.B. $\text{char}(K) = 0$), dann ist

$$E[m] \cong \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z}.$$

(2) Wenn $\text{char}(K) = p \neq 0$ ist, dann gilt entweder

$$\begin{aligned} E[p^e] &= \{O\} && \text{für } e = 1, 2, 3, \dots, && \text{oder} \\ E[p^e] &\cong \mathbb{Z}/p^e\mathbb{Z} && \text{für } e = 1, 2, 3, \dots \end{aligned}$$

Im ersten Fall heißt E supersingulär, im zweiten Fall gewöhnlich (engl. ordinary).

DEF
supersingulär
gewöhnlich

Beweis.

(1) In diesem Fall ist $[m]$ separabel, also gilt $\#E[m] = \text{deg}([m]) = m^2$. Entsprechend gilt für alle Teiler d von m , dass $\#E[d] = d^2$ ist. Daraus und aus dem Struktursatz für endliche abelsche Gruppen folgt die Behauptung.

(2) Sei $\phi: E \rightarrow E^{(p)}$, $(x, y) \mapsto (x^p, y^p)$, und sei $\hat{\phi}: E^{(p)} \rightarrow E$ die duale Isogenie. Dann gilt (unter Beachtung von $\text{deg}_s \phi = 1$; vgl. Lemma 10.15, (1))

$$\#E[p^e] = \text{deg}_s[p^e] = (\text{deg}_s[p])^e = (\text{deg}_s \hat{\phi})^e = (\text{deg}_s \phi)^e;$$

Außerdem ist $\text{deg}_s \hat{\phi}$ ein Teiler von $\text{deg} \hat{\phi} = \text{deg} \phi = p$. Die beiden möglichen Fälle entsprechen den Möglichkeiten $\text{deg}_s \hat{\phi} = 1$ und $\text{deg}_s \hat{\phi} = p$. \square

Wenn wir eine elliptische Kurve E über einem endlichen Körper K haben, dann ist $E(K)$ endlich, sagen wir der Ordnung $\#E(K) = n$ und damit enthalten in $E[n]$. Nach dem Struktursatz über endliche abelsche Gruppen und unserem Resultat über die n -Torsionspunkte folgt dann $E(K) \cong \mathbb{Z}/d_1\mathbb{Z} \times \mathbb{Z}/d_2\mathbb{Z}$ mit $d_1 \mid d_2$ und $d_1 d_2 = n$. Außerdem muss $p \nmid d_1$ gelten, wenn p die Charakteristik von K ist. Im Folgenden wollen wir eine Zusatzstruktur auf $E[n]$ beschreiben, die die Möglichkeiten für d_1 noch weiter einschränkt.

11.2. Satz. Sei E eine elliptische Kurve über K . Dann gibt es für jede natürliche Zahl m , die kein Vielfaches der Charakteristik von K ist, eine Abbildung

SATZ
Weil-Paarung

$$e_m: E[m] \times E[m] \rightarrow \mu_m$$

(wobei μ_m die Gruppe der m -ten Einheitswurzeln in \bar{K} ist) mit folgenden Eigenschaften.

(1) e_m ist bilinear:

$$e_m(S_1 + S_2, T) = e_m(S_1, T)e_m(S_2, T), \quad e_m(S, T_1 + T_2) = e_m(S, T_1)e_m(S, T_2).$$

(2) e_m ist alternierend: $e_m(T, T) = 1$.

(3) e_m ist nicht-ausgeartet: Wenn $e_m(S, T) = 1$ gilt für alle $T \in E[m]$, dann ist $S = O$. Insbesondere ist e_m surjektiv.

- (4) e_m ist verträglich mit der Operation der Galois-Gruppe von \bar{K} über K , d.h. für $\sigma \in \text{Gal}(\bar{K}/K)$ gilt

$$e_m(\sigma(S), \sigma(T)) = \sigma(e_m(S, T)).$$

- (5) e_m und $e_{mm'}$ sind miteinander kompatibel: Für $S \in E[mm']$ und $T \in E[m]$ gilt

$$e_{mm'}(S, T) = e_m(m'S, T).$$

- (6) Ist $\phi: E \rightarrow E'$ eine Isogenie, dann sind ϕ und $\hat{\phi}$ bezüglich e_m adjungiert, d.h. für $S \in E[m]$ und $T \in E'[m]$ gilt

$$e_m(S, \hat{\phi}(T)) = e_m(\phi(S), T)$$

(wobei das linke e_m zu E und das rechte zu E' gehört).

Diese Abbildung e_m heißt *(m-)Weil-Paarung*.

Beweis. Siehe [Si1, § III.8]. Wir können hier nur einen Teil des Beweises andeuten. Sei $S \in E[m]$. Wir wissen (nach Bemerkung 9.3), dass es rationale Funktionen l_j in $\bar{K}(E)$ gibt, sodass l_j einfache Nullstellen in S und jS und einfache Pole in O und $(j+1)S$ hat (wenn von diesen vier Punkte welche zusammenfallen, müssen die Nullstellen und Pole miteinander verrechnet werden). Für das Produkt $f_S = l_1 l_2 \cdots l_{m-1}$ gilt dann, dass f_S in S eine m -fache Nullstelle und in O einen m -fachen Pol hat. Auf analoge Weise kann man zeigen, dass es eine rationale Funktion auf E gibt mit Nullstellen in P_1, \dots, P_n und Polstellen in Q_1, \dots, Q_n (mit Vielfachheit gerechnet), wenn $P_1 + \cdots + P_n = Q_1 + \cdots + Q_n$ gilt. Sei $Q \in E(\bar{K})$ mit $mQ = S$. Dann gibt es demnach eine Funktion g_S mit einfachen Nullstellen in allen $Q + T$ und einfachen Polstellen in allen T , wobei T die Gruppe $E[m]$ durchläuft (die m^2 Elemente hat). Die Funktion g_S^m hat dann dieselben Null- und Polstellen wie $f_S \circ [m]$, woraus folgt, dass der Quotient dieser beiden Funktionen konstant ist. Sei nun $T \in E[m]$. Wir betrachten die Funktion

$$E \ni P \mapsto \frac{g_S(P+T)}{g_S(P)}.$$

Wenn $g_S(P)$ und $g_S(P+T)$ beide definiert und $\neq 0$ sind, dann folgt

$$\left(\frac{g_S(P+T)}{g_S(P)} \right)^m = \frac{g_S(P+T)^m}{g_S(P)^m} = \frac{f_S(mP+T)}{f_S(mP)} = \frac{f_S(mP)}{f_S(mP)} = 1$$

(denn $mT = O$). Also ist obige Funktion konstant, und ihr Wert ist eine m te Einheitswurzel. Wir setzen

$$e_m(S, T) = \frac{g_S(P+T)}{g_S(P)}$$

für jeden Punkt $P \in E$, für den die rechte Seite definiert ist.

- (1) Für $T_1, T_2 \in E[m]$ gilt mit passendem $P \in E$

$$\begin{aligned} e_m(S, T_1)e_m(S, T_2) &= \frac{g_S(P+T_1)}{g_S(P)} \frac{g_S((P+T_1)+T_2)}{g_S(P+T_1)} \\ &= \frac{g_S(P+(T_1+T_2))}{g_S(P)} = e_m(S, T_1+T_2). \end{aligned}$$

DEF
Weil-
Paarung



A. Weil
1906–1998
Foto © MFO

Für die andere Relation sei h eine Funktion mit Nullstellen in S_1 und S_2 und Polen in O und $S_1 + S_2$. Dann ist $f_{S_1+S_2}h^m = cf_{S_1}f_{S_2}$ mit einer Konstanten $c \neq 0$. Es folgt $g_{S_1+S_2} \cdot (h \circ [m]) = c'g_{S_1}g_{S_2}$ und daraus

$$\begin{aligned} e_m(S_1, T)e_m(S_2, T) &= \frac{g_{S_1}(P+T)}{g_{S_1}(P)} \frac{g_{S_2}(P+T)}{g_{S_2}(P)} \\ &= \frac{g_{S_1+S_2}(P+T)h(mP+mT)}{g_{S_1+S_2}(P)h(mP)} = e_m(S_1+S_2, T), \end{aligned}$$

denn $mT = O$.

(2) Sei $Q \in E(\bar{K})$ mit $mQ = T$. Das Produkt

$$f_T(P)f_T(P+T)f_T(P+2T) \cdots f_T(P+(m-1)T)$$

ist konstant (denn alle Nullstellen und Pole heben sich weg). Damit ist auch die Funktion

$$P \mapsto g_T(P)g_T(P+Q)g_T(P+2Q) \cdots g_T(P+(m-1)Q)$$

konstant (denn ihre m te Potenz ist im Wesentlichen das f_T -Produkt). Wenn wir $P+Q$ einsetzen, haben wir

$$\begin{aligned} g_T(P)g_T(P+Q)g_T(P+2Q) \cdots g_T(P+(m-1)Q) \\ = g_T(P+Q)g_T(P+2Q) \cdots g_T(P+(m-1)Q)g_T(P+mQ), \end{aligned}$$

also $g_T(P) = g_T(P+mQ) = g_T(P+T)$ und damit $e_m(T, T) = 1$.

(3) Das können wir hier nicht beweisen. Die Surjektivität von e_m folgt dann so: Wenn e_m nicht surjektiv wäre, dann wäre das Bild eine echte Untergruppe von μ_m , also von der Form μ_d mit einem echten Teiler d von m . Dann würde für alle $S, T \in E[m]$ gelten, dass $e_m(dS, T) = e_m(S, T)^d = 1$ ist. Die Nicht-Ausgeartettheit von e_m impliziert dann $dS = O$, also $S \in E[d]$. Damit ergäbe sich die absurde Inklusion $E[m] \subset E[d]$, also muss die Annahme falsch sein.

(4) Wir können die g_S so wählen, dass $\sigma(g_S) = g_{\sigma(S)}$ ist (das erfordert etwas Überlegung). Dann folgt

$$e_m(\sigma(S), \sigma(T)) = \frac{g_{\sigma(S)}(\sigma(P) + \sigma(T))}{g_{\sigma(S)}(\sigma(P))} = \sigma\left(\frac{g_S(P+T)}{g_S(P)}\right) = \sigma(e_m(S, T)).$$

(5) ist nicht allzu schwer (Übung).

(6) beweisen wir hier nicht. □

11.3. Folgerung. Sei E eine elliptische Kurve über K . Sei $\mu(K)$ die aus allen Einheitswurzeln in K bestehende Untergruppe von K^\times . Wir setzen voraus, dass $\mu(K)$ endlich ist.

FOLG
Struktur
der Torsion

Dann gilt für jede endliche Untergruppe G von $E(K)$: $G \cong \mathbb{Z}/d_1\mathbb{Z} \times \mathbb{Z}/d_2\mathbb{Z}$ mit $d_1 \mid d_2$ und $d_1d_2 = \#G$, wobei d_1 ein Teiler von $\#\mu(K)$ ist und nicht von der Charakteristik von K geteilt wird.

Beweis. Sei G eine endliche Untergruppe von $E(K)$ und $\#G = n$. Dann ist $G \subset E[n]$ und $E[n] \subset \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$, also hat G jedenfalls die angegebene Form, und es sind nur noch die Teilbarkeitsaussagen an d_1 zu zeigen. Wäre d_1 ein Vielfaches der Charakteristik p (die dann nicht null ist), dann hätten wir

$$\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z} \subset G \cap E[p] \subset E[p],$$

im Widerspruch zu Satz 11.1. Für die andere Aussage (d_1 teilt $\#\mu(K)$) beachten wir, dass $E[d_1] \subset G \subset E(K)$ ist. Da die Weil-Paarung e_{d_1} surjektiv ist, gibt es $S, T \in E(K)[d_1] = E[d_1]$ mit $e_{d_1}(S, T) = \zeta$, wo $\zeta \in \bar{K}$ eine primitive d_1 -te Einheitswurzel ist. Wenn wir ein Element σ der Galois-Gruppe $\text{Gal}(\bar{K}/K)$ anwenden, bleibt die linke Seite unverändert (da S und T fest bleiben), also liegt ζ schon in K . Es folgt $\zeta \in \mu(K)$ und damit $d_1 = \text{ord}(\zeta) \mid \#\mu(K)$. \square

Die Aussage der Folgerung ist analog zu der bekannten Aussage, dass eine endliche Untergruppe der multiplikativen Gruppe eines Körpers stets zyklisch ist.

Für eine elliptische Kurve E über \mathbb{Q} gilt, dass die Gruppe $E(\mathbb{Q})$ endlich erzeugt ist (Satz von Mordell). Sie hat also die Form $E(\mathbb{Q}) \cong T \oplus \mathbb{Z}^r$ mit einer endlichen abelschen Gruppe T . Da $\mu(\mathbb{Q}) = \{\pm 1\}$ ist, erhalten wir die Aussage, dass T entweder zyklisch oder von der Form $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2d\mathbb{Z}$ ist. Ein berühmtes Resultat von Mazur sagt dann, dass es für zyklisches T genau die Möglichkeiten $\#T \leq 10$ oder $= 12$ gibt. Im anderen Fall muss $d \leq 4$ sein.

12. ELLIPTISCHE KURVEN ÜBER ENDLICHEN KÖRPERN

Einige Spezifika im Zusammenhang mit endlichen Grundkörpern (oder jedenfalls im Fall von null verschiedener Charakteristik) sind schon angedeutet worden. Wir wollen uns jetzt gründlicher mit dieser Situation beschäftigen. Dies geschieht vor allem im Hinblick darauf, dass gerade elliptische Kurven über endlichen Körpern interessante Anwendungen gefunden haben.

Zur Erinnerung folgt hier eine Zusammenstellung der wichtigsten Tatsachen über endliche Körper.

12.1. Satz.

- (1) Die Anzahl der Elemente eines endlichen Körpers ist eine Primzahlpotenz p^f (mit $f \geq 1$).
- (2) Umgekehrt gibt es zu jeder Primzahlpotenz $q = p^f$ bis auf Isomorphie genau einen endlichen Körper \mathbb{F}_q .
- (3) Die Erweiterungen endlicher Körper haben die Form $\mathbb{F}_q \subset \mathbb{F}_{q^n}$; so eine Körpererweiterung ist galoissch mit zyklischer Galois-Gruppe der Ordnung n . Die Galois-Gruppe wird erzeugt vom Frobenius-Automorphismus $x \mapsto x^q$.
- (4) Der algebraische Abschluss von \mathbb{F}_q ist die (aufsteigend filtrierte) Vereinigung $\bar{\mathbb{F}}_q = \bigcup_n \mathbb{F}_{q^n}$. Es gilt

$$\mathbb{F}_q = \{x \in \bar{\mathbb{F}}_q \mid x^q = x\}.$$

- (5) Es gilt

$$\mathbb{F}_q^\times = \{x \in \bar{\mathbb{F}}_q \mid x^{q-1} = 1\} = \mu_{q-1}(\mathbb{F}_q).$$

($\mu_n(K)$ bezeichnet die Gruppe der n -ten Einheitswurzeln in K .)

Elliptische Kurven über endlichen Körpern haben (mindestens) zwei hervorstechende Eigenschaften. Zum einen ist die Gruppe der rationalen Punkte zwangsläufig endlich; ihre Ordnung ist daher ein wichtiges Datum. Zum anderen hat eine solche Kurve stets außer den Multiplikationsendomorphismen auch noch den Frobenius-Endomorphismus. Wie wir gleich sehen werden, gibt es einen Zusammenhang zwischen diesen beiden Dingen.

Eine heuristische Überlegung lässt einen vermuten, dass eine elliptische Kurve über dem endlichen Körper \mathbb{F}_q ungefähr $q + 1$ rationale Punkte haben sollte: Die durchschnittliche Zahl von Lösungen der Gleichung $y^2 = a$, wenn a den Körper \mathbb{F}_q durchläuft, ist 1 (für ungerade Charakteristik). Wenn wir annehmen, dass die Werte eines Polynoms $f(x) = x^3 + a_2x^2 + a_4x + a_6$ annähernd zufällig verteilt sind, dann sollte die Anzahl der Lösungen von $y^2 = f(x)$ etwa q sein; also erwarten wir $\#E(\mathbb{F}_q) \approx q + 1$, wenn E die durch diese Gleichung definierte Kurve ist.

Das stimmt tatsächlich, und man kann die Abweichung sogar sehr genau beschränken. Der folgende Satz wurde zuerst von Hasse bewiesen.

12.2. Satz. Sei E eine elliptische Kurve über dem endlichen Körper \mathbb{F}_q , und sei $\phi \in \text{End}_{\mathbb{F}_q}(E)$ der Frobenius-Endomorphismus $(x, y) \mapsto (x^q, y^q)$.

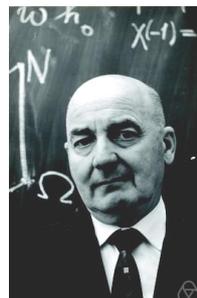
- (1) Sei $t = \phi + \hat{\phi} \in \mathbb{Z}$ die Spur des Frobenius. Dann gilt in $\text{End}_{\mathbb{F}_q}(E)$ die Relation

$$\phi^2 - t\phi + q = 0,$$

und $|t| \leq 2\sqrt{q}$.

SATZ
endliche
Körper

DEF
Frobenius-
Auto-
morphismus



H. Hasse
1898 – 1979
Foto © MFO

SATZ
Frobenius
und Punkt-
anzahl
DEF
Spur des
Frobenius

- (2) Es gilt $\#E(\mathbb{F}_q) = \deg(\phi - 1) = q + 1 - t$. Insbesondere haben wir
- $$|\#E(\mathbb{F}_q) - (q + 1)| \leq 2\sqrt{q}.$$

Beweis.

- (1) In $\text{End}_{\mathbb{F}_q}(E)$ gilt

$$0 = (\phi - \hat{\phi})(\phi - \hat{\phi}) = \phi^2 - (\phi + \hat{\phi})\phi + \phi\hat{\phi} = \phi^2 - t\phi + q,$$

denn $\phi\hat{\phi} = \deg(\phi) = q$.

Für eine rationale Zahl $r/s \in \mathbb{Q}$ gilt

$$\left(\frac{r}{s}\right)^2 - t\frac{r}{s} + q = \frac{1}{s^2}(r^2 - trs + qs^2) = \frac{1}{s^2}\deg(r - s\phi) \geq 0,$$

also hat das Polynom $X^2 - tX + q$ nicht-positive Diskriminante: $t^2 - 4q \leq 0$, d.h. $|t| \leq 2\sqrt{q}$.

- (2) Es gilt

$$\begin{aligned} E(\mathbb{F}_q) &= \{(\xi, \eta) \in E(\bar{\mathbb{F}}_q) \mid \xi = \xi^q, \eta = \eta^q\} \cup \{O\} \\ &= \{P \in E(\bar{\mathbb{F}}_q) \mid \phi(P) = P\} \\ &= \ker(\phi - 1). \end{aligned}$$

Da $\phi - 1$ separabel ist (Lemma 10.15), gilt $\#E(\mathbb{F}_q) = \#\ker(\phi - 1) = \deg(\phi - 1)$ (Satz 10.14). Außerdem ist

$$\deg(\phi - 1) = (\phi - 1)(\hat{\phi} - 1) = \phi\hat{\phi} - (\phi + \hat{\phi}) + 1 = q - t + 1. \quad \square$$

Unter Berücksichtigung von Folgerung 11.3 können wir über die Struktur der Gruppe $E(\mathbb{F}_q)$ also folgende Aussagen machen:

$$E(\mathbb{F}_q) \cong \mathbb{Z}/d\mathbb{Z} \times \mathbb{Z}/d'\mathbb{Z}$$

mit $d \mid q - 1$ und $|d^2d' - (q + 1)| \leq 2\sqrt{q}$.

Es folgt noch ein Ergebnis über den Zusammenhang zwischen Isogenien und der Anzahl der rationalen Punkte.

12.3. Satz. Seien E und E' zwei elliptische Kurven über \mathbb{F}_q . Dann sind die beiden folgenden Aussagen äquivalent:

SATZ
isogene
ell. Kurven

- (1) E und E' sind isogen über \mathbb{F}_q .
(2) $\#E(\mathbb{F}_q) = \#E'(\mathbb{F}_q)$.

Beweis. Wir werden hier nur die Richtung „(1) \Rightarrow (2)“ beweisen. Der Beweis der anderen Richtung erfordert sehr tief liegende Hilfsmittel.

Wir setzen also voraus, es gebe eine (nicht-konstante) über \mathbb{F}_q definierte Isogenie $\psi: E \rightarrow E'$. Wir bezeichnen mit ϕ und ϕ' die Frobenius-Endomorphismen von E und von E' und mit t bzw. t' ihre Spuren. Da die Abbildung $x \mapsto x^q$ mit den vier Grundrechenarten kommutiert und die Elemente von \mathbb{F}_q fest lässt, folgt $\psi \circ \phi = \phi' \circ \psi$. Ebenso gilt $\phi \circ \hat{\psi} = \hat{\psi} \circ \phi'$, woraus wir durch Dualisieren bekommen $\psi \circ \hat{\phi} = \hat{\phi}' \circ \psi$. Zusammen implizieren diese Relationen

$$\psi \circ [t] = \psi \circ \phi + \psi \circ \hat{\phi} = \phi' \circ \psi + \hat{\phi}' \circ \psi = [t'] \circ \psi = \psi \circ [t'].$$

(Die letzte Gleichung folgt, weil ψ ein Homomorphismus ist.) Wir verknüpfen von links mit $\hat{\psi}$ und erhalten die Gleichung

$$\deg(\psi)t = \deg(\psi)t'$$

in $\text{End}(E)$. Da $\deg(\psi) \neq 0$ und $\text{End}(E)$ ein Integritätsbereich der Charakteristik 0 ist (Satz 10.8), folgt $t = t'$, also nach Satz 12.2 auch $\#E(\mathbb{F}_q) = \#E'(\mathbb{F}_q)$. \square

Abschließend möchte ich hier noch bemerken, dass es einen schnellen Algorithmus gibt (polynomial in $\log q$), der die Anzahl der rationalen Punkte bestimmt. Er wurde theoretisch von Schoof entwickelt und von Atkin und Elkies praktikabel gemacht. Seine Grundidee besteht darin, für geeignete Primzahlen ℓ die Restklasse von $t \bmod \ell$ zu bestimmen und daraus auf den Wert von t (und damit von $\#E(\mathbb{F}_q) = q + 1 - t$) zu schließen.

Die Zetafunktion.

Wir haben gesehen, dass die Anzahl der rationalen Punkte einer elliptischen Kurve E über \mathbb{F}_q in engem Zusammenhang steht mit dem Verhalten des Frobenius-Endomorphismus ϕ . Nun können wir E aber auch auffassen als eine elliptische Kurve über \mathbb{F}_{q^n} für $n = 2, 3, 4, \dots$. In diesem Abschnitt wollen wir uns damit beschäftigen, wie die Zahlen

$$\#E(\mathbb{F}_q), \quad \#E(\mathbb{F}_{q^2}), \quad \#E(\mathbb{F}_{q^3}), \quad \dots$$

miteinander zusammenhängen. Dazu führen wir ein Objekt ein, das die Information über diese Zahlen in geeigneter Weise kodiert.

12.4. Definition. Sei C eine glatte projektive Kurve über \mathbb{F}_q . Die *Zetafunktion* von C ist folgende Potenzreihe mit rationalen Koeffizienten:

$$\begin{aligned} Z(C, T) &= \exp \left(\#C(\mathbb{F}_q) T + \frac{\#C(\mathbb{F}_{q^2})}{2} T^2 + \frac{\#C(\mathbb{F}_{q^3})}{3} T^3 + \dots \right) \\ &= \exp \left(\sum_{n=1}^{\infty} \frac{\#C(\mathbb{F}_{q^n})}{n} T^n \right). \end{aligned} \quad \diamond$$

Der Zusammenhang mit der vielleicht naheliegenderen Variante $\sum_{n \geq 1} \#C(\mathbb{F}_{q^n}) T^n$ ist durch die logarithmische Ableitung gegeben:

$$\sum_{n \geq 1} \#C(\mathbb{F}_{q^n}) T^n = T \frac{\frac{d}{dT} Z(C, T)}{Z(C, T)} = T \frac{d}{dT} \log Z(C, T).$$

Der Grund für die etwas umständlich erscheinende Definition der Zetafunktion liegt darin, dass sie in dieser Form eine natürliche Produktentwicklung hat. Dazu betrachten wir die Menge der algebraischen Punkte $C(\mathbb{F}_q) = \bigcup_{n \geq 1} C(\mathbb{F}_{q^n})$. Sie zerfällt in Bahnen unter der Operation des Frobenius-Endomorphismus ϕ . Sei a_d die Anzahl der Bahnen der Länge d . Dann gilt $\#C(\mathbb{F}_{q^n}) = \sum_{d|n} da_d$, und die Zetafunktion schreibt sich als

$$Z(C, T) = \prod_{d=1}^{\infty} (1 - T^d)^{-a_d}$$

(Übung.) Außerdem stellt sich heraus, dass die Zetafunktion in der definierten Form eine besonders einfache Gestalt erhält, wie der folgende Satz zeigt.



R. Schoof
* 1955
Foto © MFO



N.D. Elkies
* 1966
Foto © MFO

DEF
Zetafunktion

12.5. **Satz.** Sei C eine glatte projektive Kurve über \mathbb{F}_q . Dann gilt:

SATZ
Weil-
Vermutungen
für Kurven

- (1) $Z(C, T) \in \mathbb{Q}(T)$ (d.h., $Z(C, T)$ ist die Potenzreihe einer rationalen Funktion).
- (2) $Z(C, 1/(qT)) = q^{1-g}T^{2-2g}Z(C, T)$ (Funktionalgleichung). Dabei ist g das Geschlecht von C ($g = 1$ für elliptische Kurven).
- (3) $Z(C, T) = P(T)/((1 - T)(1 - qT))$ mit einem Polynom $P(T) \in \mathbb{Z}[T]$ vom Grad $2g$, das über \mathbb{C} faktorisiert als

$$P(T) = \prod_{j=1}^g ((1 - \alpha_j T)(1 - \bar{\alpha}_j T))$$

mit $|\alpha_j| = \sqrt{q}$. („Riemannsche Vermutung“)

12.6. **Bemerkungen.**

BEM
Weil-
Vermutungen

- (1) Weil hat seine Vermutungen allgemeiner auch für höherdimensionale projektive Varietäten formuliert. Für Kurven (und abelsche Varietäten) hat er sie selbst auch bewiesen (1949). Die verschiedenen Teile der allgemeinen Vermutung wurden zwischen 1960 und 1973 durch Deligne erledigt.
- (2) Das *Geschlecht* g ist eine wichtige Invariante der Kurve C ; es ist aber nicht einfach zu definieren. Für eine glatte ebene projektive Kurve vom Grad d gilt $g = \frac{1}{2}(d-1)(d-2)$; für elliptische Kurven (die glatte ebene projektive Kurven vom Grad 3 sind) gilt also $g = 1$.
- (3) Die Bezeichnung „Riemannsche Vermutung“ für Teil (3) des Satzes kommt von folgender Analogie. Wir setzen $\zeta(C, s) = Z(C, q^{-s})$; dann hat diese Funktion ζ einfache Pole bei $s = 0$ und bei $s = 1$, und alle ihre Nullstellen haben Realteil $\frac{1}{2}$. (Außerdem sagt die Funktionalgleichung, dass $\zeta(C, 1-s) = q^{(g-1)(2s-1)}\zeta(C, s)$ ist, was an die Funktionalgleichung der Riemannschen Zetafunktion erinnert.)



Wir wollen den Satz jetzt für elliptische Kurven beweisen.

Beweis. Sei also E eine elliptische Kurve über \mathbb{F}_q und $\phi \in \text{End}(E)$ der Frobenius-Endomorphismus. Wir hatten gesehen, dass ϕ die Gleichung $X^2 - tX + q = 0$ löst (Satz 12.2), wobei $t = \phi + \hat{\phi}$ die Spur des Frobenius ist. Weiterhin gilt $|t| \leq 2\sqrt{q}$, woraus folgt, dass

$$X^2 - tX + q = (X - \alpha)(X - \bar{\alpha})$$

ist mit $\alpha \in \mathbb{C}$, $|\alpha| = \sqrt{q}$. Außerdem ist

$$X^2 - tX + q = (X - \phi)(X - \hat{\phi}),$$

das heißt, dass wir einen Isomorphismus

$$\mathbb{Z}[\alpha] \longrightarrow \mathbb{Z}[\phi] \subset \text{End}(E), \quad \alpha \longmapsto \phi$$

haben. (Im Falle $\alpha = \pm\sqrt{q}$ verwenden wir dabei, dass $\text{End}(E)$ ein Integritätsbereich ist, siehe Satz 10.8.) Nun gilt

$$\#E(\mathbb{F}_q) = q + 1 - \phi - \hat{\phi} = q + 1 - \alpha - \bar{\alpha},$$

und dann entsprechend (denn ϕ^n ist der Frobenius-Endomorphismus von E über \mathbb{F}_{q^n})

$$\#E(\mathbb{F}_{q^n}) = q^n + 1 - \phi^n - \hat{\phi}^n = q^n + 1 - \alpha^n - \bar{\alpha}^n.$$

Es folgt

$$\begin{aligned}
 Z(E, T) &= \exp \left(\sum_{n=1}^{\infty} \frac{\#E(\mathbb{F}_{q^n})}{n} T^n \right) \\
 &= \exp \left(\sum_{n=1}^{\infty} \frac{(qT)^n}{n} + \sum_{n=1}^{\infty} \frac{T^n}{n} - \sum_{n=1}^{\infty} \frac{(\alpha T)^n}{n} - \sum_{n=1}^{\infty} \frac{(\bar{\alpha}T)^n}{n} \right) \\
 &= \exp \left(\log \frac{1}{1 - qT} + \log \frac{1}{1 - T} - \log \frac{1}{1 - \alpha T} - \log \frac{1}{1 - \bar{\alpha}T} \right) \\
 &= \frac{(1 - \alpha T)(1 - \bar{\alpha}T)}{(1 - T)(1 - qT)} = \frac{1 - tT + qT^2}{(1 - T)(1 - qT)}.
 \end{aligned}$$

Damit ist Teil (1) bewiesen. Teil (2) folgt durch Nachrechnen, und Teil (3) folgt aus der obigen Aussage über α . \square

Die vielleicht erstaunlichste Folgerung aus diesem Satz ist, dass die Anzahl der rationalen Punkte über \mathbb{F}_q einer elliptischen Kurve E bereits *alle* Anzahlen $\#E(\mathbb{F}_{q^n})$ festlegt.

13. FAKTORISIERUNG UND PRIMZAHLTTEST: GRUNDLAGEN

Nachdem wir nun elliptische Kurven kennen gelernt haben und auch ein wenig über die speziellen Eigenschaften elliptischer Kurven über endlichen Körpern Bescheid wissen, können wir uns einige praktische Anwendungen ansehen. Die erste Anwendung wird die Faktorisierung großer Zahlen sein und damit verbunden der Beweis, dass eine große Zahl prim ist. Die Hauptquelle für diesen und die folgenden Abschnitte ist [Co1]. Zuerst müssen wir aber das Problem genauer betrachten.

Eine Vorbemerkung zur praktischen Faktorisierung. Sie ist ein rekursiver Prozess, der sich aus folgenden Teilalgorithmen zusammensetzt.

- Stelle fest, ob eine natürliche Zahl N zusammengesetzt oder höchstwahrscheinlich prim ist.
- Wenn N höchstwahrscheinlich prim ist, beweise, dass N tatsächlich prim (oder aber doch zusammengesetzt) ist.
- Wenn N zusammengesetzt ist, finde einen nicht trivialen Faktor d und mache rekursiv mit d und N/d weiter.

Üblicherweise wird man zunächst durch Probedivision alle hinreichend kleinen Teiler von N finden.

Primzahltest.

Um zu zeigen, dass eine Zahl zusammengesetzt ist, kann man prüfen, ob sie Bedingungen erfüllt, die für alle Primzahlen gelten. Eine Möglichkeit ist der *kleine Satz von Fermat*, dem zufolge für jede Primzahl p und jede ganze Zahl a mit $a \perp p$ gilt:

$$a^{p-1} \equiv 1 \pmod{p}.$$

Wir schreiben $a \perp b$ für die Aussage, dass a und b teilerfremd sind. Das führt zu folgender Definition.

13.1. Definition. Eine ganze Zahl $N > 1$ heißt *Pseudoprimzahl zur Basis a* , wenn $a^{N-1} \equiv 1 \pmod{N}$ ist (das impliziert $a \perp N$).

N heißt *Carmichael-Zahl*, wenn N keine Primzahl, aber Pseudoprimzahl zur Basis a ist für alle $a \perp N$. ◇

Es ist klar, dass eine Primzahl Pseudoprimzahl zur Basis a ist für alle a mit $p \nmid a$. Wir können also eine Zahl N als zusammengesetzt erkennen, wenn wir $1 < a < N$ finden mit $a^{N-1} \not\equiv 1 \pmod{N}$. Hierbei ist wichtig, dass sich der Rest $a^{N-1} \pmod{N}$ effizient berechnen lässt (durch sukzessives Quadrieren und Reduktion mod N ; bei Verwendung der Standard-Methode für die Multiplikation ist die Laufzeit $O((\log N)^3)$). Leider funktioniert dieser Test nicht immer:

13.2. Satz. (Alford, Granville, Pomerance 1994¹)

Es gibt unendlich viele Carmichael-Zahlen.

Wenn N eine Carmichael-Zahl ist, dann gilt stets $a^{N-1} \equiv 1 \pmod{N}$, außer wenn $\text{ggT}(a, N) > 1$, was aber bei zufälliger Wahl von a für großes N extrem unwahrscheinlich ist.

Es gibt allerdings eine Variante, die besser funktioniert. Dazu verschärfen wir die Bedingung in Definition 13.1.

¹W. R. Alford, A. Granville, and C. Pomerance: *There are infinitely many Carmichael numbers*, Annals of Mathematics **139** (1994) 703–722.

DEF
Pseudo-
primzahl
Carmichael-
Zahl

SATZ
unendlich
viele
Carmichael-
Zahlen

13.3. Definition. Sei N eine ungerade natürliche Zahl. Wir schreiben $N-1 = 2^t q$ mit q ungerade. Sei weiter a eine ganze Zahl. Dann heißt N *starke Pseudoprimzahl zur Basis a* , wenn gilt:

$$a^q \equiv 1 \pmod{N} \quad \text{oder} \quad a^{2^e q} \equiv -1 \pmod{N} \text{ für ein } 0 \leq e < t. \quad \diamond$$

DEF
starke
Pseudo-
primzahl

Dass wir N als ungerade voraussetzen, ist keine wesentliche Einschränkung, da wir ja sehr einfach feststellen können, ob N durch 2 teilbar ist.

13.4. Satz.

- (1) *Ist N prim, so ist N starke Pseudoprimzahl zur Basis a für alle $N \nmid a$.*
- (2) *Ist N zusammengesetzt, so ist N starke Pseudoprimzahl zur Basis a für weniger als $N/4$ Zahlen a mit $1 < a < N$.*

SATZ
Miller-Rabin-
Test

Beweis.

- (1) Wenn $N = p$ eine Primzahl ist, dann folgt aus $x^2 \equiv 1 \pmod{p}$, dass $x \equiv \pm 1 \pmod{p}$ ist (denn das Polynom $X^2 - 1$ kann im Körper \mathbb{F}_p höchstens zwei Nullstellen haben). Der kleine Satz von Fermat sagt, dass $a^{p-1} = a^{2^t q} \equiv 1 \pmod{p}$ ist. Es folgt, dass entweder $a^q \equiv 1 \pmod{p}$ ist oder $a^{2^e q} \equiv -1 \pmod{p}$ ist für ein $0 \leq e < t$.

- (2) Wir betrachten zunächst den Homomorphismus

$$(\mathbb{Z}/N\mathbb{Z})^\times \longrightarrow (\mathbb{Z}/N\mathbb{Z})^\times, \quad \bar{a} \longmapsto \bar{a}^{N-1}.$$

Sei $G \subset (\mathbb{Z}/N\mathbb{Z})^\times$ sein Kern. Dann gilt $\#G \leq \#(\mathbb{Z}/N\mathbb{Z})^\times < N$. (N ist genau dann Carmichael-Zahl, wenn $G = (\mathbb{Z}/N\mathbb{Z})^\times$ ist.) Sei weiter $N = p_1^{e_1} \cdots p_k^{e_k}$ die Primfaktorzerlegung von N . Dann sind die Primzahlen p_j ungerade, und

$$(\mathbb{Z}/N\mathbb{Z})^\times \cong (\mathbb{Z}/p_1^{e_1}\mathbb{Z})^\times \times \cdots \times (\mathbb{Z}/p_k^{e_k}\mathbb{Z})^\times$$

ist ein Produkt zyklischer Gruppen gerader Ordnung $(p_j - 1)p_j^{e_j-1}$. G zerlegt sich entsprechend als $G \cong G_1 \times \cdots \times G_k$, wobei G_j zyklisch ist und die Ordnung $\text{ggT}(N-1, (p_j - 1)p_j^{e_j-1})$ hat. Sei $G'_j \subset G_j$ die Untergruppe vom Index 2. Für $a \in \mathbb{Z}$ mit $\bar{a} \in G_j$ gilt dann: Ist $a \pmod{p_j^{e_j}}$ in G'_j , dann ist $a^{(N-1)/2} \equiv 1 \pmod{p_j^{e_j}}$, andernfalls ist $a^{(N-1)/2} \equiv -1 \pmod{p_j^{e_j}}$. Für $\bar{a} \in G$ setzen wir $\varepsilon_j(\bar{a}) = 1$, falls das Bild von \bar{a} in G_j in G'_j liegt, sonst $\varepsilon_j(\bar{a}) = -1$. Dann ist

$$\varepsilon: G \longrightarrow \{\pm 1\}^k, \quad \bar{a} \longmapsto (\varepsilon_1(\bar{a}), \dots, \varepsilon_k(\bar{a}))$$

ein surjektiver Gruppenhomomorphismus. Es gilt

$$a^{(N-1)/2} \equiv \pm 1 \pmod{N} \iff \varepsilon(\bar{a}) = \pm(1, \dots, 1).$$

Es folgt

$$\#\{\bar{a} \in G \mid a^{(N-1)/2} \equiv \pm 1 \pmod{N}\} = 2^{1-k} \#G.$$

Wenn N keine Carmichael-Zahl und keine Primzahlpotenz ist, dann ist $\#G < N/2$ und $k \geq 2$, und es folgt, dass weniger als $2^{-k}N \leq N/4$ Zahlen a die notwendige Bedingung

$$a^{(N-1)/2} \equiv \pm 1 \pmod{N}$$

erfüllen. Falls N eine Carmichael-Zahl ist, dann gilt $k \geq 3$ (Übungsaufgabe), und wir haben das gleiche Resultat. Falls schließlich $N = p^e$ eine Primzahlpotenz ist (mit $e \geq 2$), dann ist

$$\#G = \text{ggT}(p^e - 1, (p - 1)p^{e-1}) = p - 1 < p^e/4,$$

sodass die Behauptung ebenfalls gilt. \square

Dieses Ergebnis führt auf den *Miller-Rabin-Test*.

13.5. Algorithmus. (Miller-Rabin-Test)

Eingabe: $N > 1$ ungerade; $m \geq 1$ (Anzahl der Tests)

Schreibe $N - 1 = 2^t q$ mit q ungerade

Für $j = 1, \dots, m$:

Wähle $1 < a < N$ zufällig und berechne $b := a^q \bmod N$

Falls $b = \pm 1$: nimm das nächste j

Für $e = 1, \dots, t - 1$:

Setze $b := b^2 \bmod N$.

Falls $b = -1$: nimm das nächste j

// (N ist keine starke Pseudoprimzahl zur Basis a)

Ausgabe „ N ist zusammengesetzt“ und Ende

// (N hat alle Tests überstanden)

Ausgabe „ N ist wahrscheinlich prim“ und Ende

Das Ergebnis aus Satz 13.4, Teil (2), sagt uns, dass die Wahrscheinlichkeit, dass eine zusammengesetzte Zahl N als „wahrscheinlich prim“ erkannt wird, kleiner als 4^{-m} ist.

Auf der anderen Seite kann man mit diesem Verfahren niemals *beweisen*, dass N tatsächlich prim ist. Eine Möglichkeit dies zu tun besteht darin, eine geeignete Umkehrung des kleinen Satzes von Fermat zu verwenden.

13.6. Lemma. Sei $N > 0$ eine ganze Zahl und sei p ein Primteiler von $N - 1$. Sei weiter $a_p \in \mathbb{Z}$ mit

LEMMA
Umkehrung
kl. Fermat

$$(13.1) \quad a_p^{N-1} \equiv 1 \pmod{N} \quad \text{und} \quad (a_p^{(N-1)/p} - 1) \perp N.$$

Sei außerdem p^{e_p} die höchste Potenz von p , die $N - 1$ teilt. Dann gilt für jeden (positiven) Teiler d von N

$$d \equiv 1 \pmod{p^{e_p}}.$$

Beweis. Wir können uns auf Primteiler d beschränken. Aus $a_p \perp N$ folgt $d \nmid a_p$ und damit $a_p^{d-1} \equiv 1 \pmod{d}$. Andererseits ist $a_p^{(N-1)/p} \not\equiv 1 \pmod{d}$, da nach Voraussetzung $a_p^{(N-1)/p} - 1$ zu N teilerfremd ist. Sei n die Ordnung von $a_p \bmod d$; dann folgt $n \mid d - 1$, $n \mid N - 1$ (denn $a_p^{N-1} \equiv 1 \pmod{d}$), aber $n \nmid (N - 1)/p$. Aus den letzten beiden Eigenschaften folgt $p^{e_p} \mid n$, aus der ersten dann $p^{e_p} \mid d - 1$. \square

Wenn wir über die Faktorisierung von $N - 1$ gut genug kennen, dann können wir dieses Ergebnis nutzen, um zu beweisen, dass N prim ist.

13.7. Folgerung. Sei $N > 0$ eine ganze Zahl, $N - 1 = F \cdot U$ mit $F \geq \sqrt{N}$, und alle Primteiler von F seien bekannt.

FOLG
Pocklington-
Lehmer-Test

N ist genau dann prim, wenn es für jeden Primteiler p von F eine Zahl $a_p \in \mathbb{Z}$ gibt, die (13.1) erfüllt.

Beweis. Sei zunächst N prim, und sei g eine Primitivwurzel mod N (d.h. sodass (das Bild von) g die Gruppe $(\mathbb{Z}/N\mathbb{Z})^\times$ erzeugt). Dann hat $a_p = g$ die Eigenschaft (13.1).

Seien nun umgekehrt für alle $p \mid F$ Zahlen a_p mit (13.1) gegeben. Aus Prop. 13.6 folgt dann, dass jeder Teiler d von N die Kongruenz $d \equiv 1 \pmod{F}$ erfüllt. Insbesondere ist $d = 1$ oder $d > F \geq \sqrt{N}$. Wenn N zusammengesetzt wäre, hätte N einen nichttrivialen Teiler $\leq \sqrt{N}$, was wir gerade ausgeschlossen haben. Also ist N prim. \square

Aus diesem Ergebnis lässt sich direkt ein Primzahltest ableiten, der *Pocklington-Lehmer-Test*. Er basiert auf der Verwendung der zyklischen Gruppe $(\mathbb{Z}/N\mathbb{Z})^\times$ der Ordnung $N - 1$. Sein Nachteil ist, dass er eine gute Kenntnis der Faktorisierung von $N - 1$ erfordert, was in der Praxis ein großes Hindernis sein kann. Man sieht daran aber übrigens auch, dass es oft notwendig ist, Zahlen zu faktorisieren, wenn man beweisen will, dass eine gegebene Zahl prim ist, was die rekursive Natur des Faktorisierungsproblems noch verstärkt.

Man kann diesen Ansatz variieren, indem man statt \mathbb{F}_N^\times die Untergruppe der Ordnung $N + 1$ von $\mathbb{F}_{N^2}^\times$ benutzt. Dabei braucht man dann Informationen über die Faktorisierung von $N + 1$. Das führt zum Beispiel zum bekannten *Lucas-Lehmer-Test* für Mersennesche Primzahlen $2^p - 1$.

Elliptische Kurven sind hier hilfreich, da sie Gruppen der Ordnung ungefähr N zur Verfügung stellen, aber dabei eine hinreichend große Variationsbreite haben, sodass man gute Chancen hat, eine Gruppe mit genügend faktorisierbarer Ordnung zu finden. Wir werden das im nächsten Abschnitt genauer diskutieren.

Eine Diskussion von Primzahltests wäre nicht vollständig ohne den deterministischen Polynomzeit-Algorithmus von Agrawal, Kayal und Saxena² zu erwähnen. Dieses Resultat löst ein altes Problem, denn bis dahin war kein Verfahren bekannt, dass für eine beliebige natürliche Zahl deterministisch (d.h. ohne Zufallszahlen zu verwenden wie etwa der Miller-Rabin-Test) und in polynomialer Laufzeit feststellt, ob sie prim ist oder zusammengesetzt. Dieser Durchbruch hat sich aus einem Bachelorprojekt der beiden Studenten Kayal und Saxena entwickelt.

Die zu Grunde liegende Idee ist eine Verallgemeinerung des kleinen Satzes von Fermat auf Polynome, die zu einer Charakterisierung von Primzahlen führt: Für jede ganze Zahl $a \perp N$ gilt

$$N \text{ ist prim} \iff (X - a)^N \equiv X^N - a \pmod{N}$$

im Polynomring $\mathbb{Z}[X]$ (d.h., die Kongruenz mod N gilt koeffizientenweise). Die Berechnung der linken Seite ist allerdings viel zu aufwendig. Deshalb betrachtet man statt dessen die Kongruenz

$$(X - a)^N \equiv X^N - a \pmod{\langle N, X^r - 1 \rangle}$$

für geeignete $r \geq 1$. Die drei Autoren konnten zeigen, dass die Gültigkeit der Kongruenz für r und a wie im folgenden Algorithmus hinreichend dafür ist, dass N eine Primzahlpotenz ist.

²Manindra Agrawal, Neeraj Kayal, Nitin Saxena. *PRIMES is in P*, Annals of Mathematics **160** (2004), no. 2, 781–793.

13.8. **Algorithmus.** (AKS-Primzahltest)**Eingabe:** $N > 1$.Wenn N eine echte Potenz ist, gib aus „zusammengesetzt“; Stop.Finde das kleinste $r \geq 1$ mit $\text{ord}_r(N) > (\log_2 N)^2$.Wenn $1 < \text{ggT}(a, N) < N$ für ein $1 \leq a \leq r$, gib aus „zusammengesetzt“; Stop.Wenn $N \leq r$, gib aus „prim“; Stop.Für $a = 1, \dots, \lfloor \sqrt{\varphi(r)} \log_2 N \rfloor$:Wenn $(X - a)^N \not\equiv X^N - a \pmod{\langle N, X^r - 1 \rangle}$:

gib aus „zusammengesetzt“; Stop.

Gib aus „prim“; Stop.

Hier bezeichnet $\text{ord}_r(N)$ die Ordnung von N in der multiplikativen Gruppe $(\mathbb{Z}/r\mathbb{Z})^\times$, und $\varphi(r)$ ist die Eulersche φ -Funktion, also die Ordnung dieser Gruppe.

Außerdem konnten sie zeigen, dass die Zahl r genügend klein ist, damit die Laufzeit durch ein Polynom in $\log N$ beschränkt werden kann (ursprünglich $O((\log N)^{12})$; diese Abschätzung wurde zwischenzeitlich aber verbessert).

Allerdings sind probabilistische Algorithmen wie der, den wir im nächsten Abschnitt beschreiben werden, in der Praxis immer noch schneller.

Faktorisierung.

Nach der Frage, wie man feststellen kann, ob eine Zahl prim ist, betrachten wir jetzt die Faktorisierung. Hier haben wir eine Zahl N gegeben, von der wir wissen, dass sie zusammengesetzt ist (zum Beispiel weil sie den Miller-Rabin-Test nicht bestanden hat). Das Ziel ist, einen nichttrivialen Teiler d von N zu finden.

13.9. **Definition.** Wir nennen eine ganze Zahl B -glatt, wenn alle ihre Primteiler $\leq B$ sind. Die Zahl heißt B -potenzglatt, wenn alle Primzahlpotenzen, die sie teilen, $\leq B$ sind.

DEF
 B -glatt
 B -potenzglatt
 \diamond

Wir haben beim Pocklington-Lehmer-Test gesehen, dass er eine Art Glattheitsvoraussetzung an $N - 1$ benötigt. Der nun folgende Faktorisierungsalgorithmus hat eine ähnliche Einschränkung: Er findet nur Teiler, wenn es Primteiler p von N gibt, sodass $p - 1$ B -potenzglatt ist.

Die Idee ist wie folgt. Wir wählen eine Schranke B und eine ganze Zahl a . Wenn N einen Primteiler p hat, sodass $p - 1$ B -potenzglatt ist, dann ist $p - 1$ ein Teiler von $L(B) = \text{kgV}(1, 2, \dots, B)$. Aus dem kleinen Satz von Fermat folgt $a^{L(B)} \equiv 1 \pmod{p}$, also ist

$$\text{ggT}(a^{L(B)} - 1, N) > 1.$$

Dieser ggT ist also ein Teiler > 1 von N , und mit etwas Glück ist der Teiler auch $< N$. In der Praxis wird man der Reihe nach $a^{L(1)} \pmod{N}$, $a^{L(2)} \pmod{N}$, \dots , $a^{L(B)} \pmod{N}$ berechnen (durch sukzessives Potenzieren mod N mit $L(n+1)/L(n)$, was entweder 1 ist oder eine Primzahl q ; letzteres, wenn $n + 1 = q^e$ eine Potenz von q ist) und jeweils den ggT überprüfen.

Dieser Algorithmus stammt von Pollard (dem wir auch noch einige andere Faktorisierungsalgorithmen verdanken). Wie wir die Schranke B wählen, hängt hauptsächlich davon ab, wie viel Zeit wir zu investieren gewillt sind.

13.10. Beispiel. Wir betrachten $N = 119$. Erst einmal stellen wir fest, dass N zusammengesetzt ist: $N - 1 = 118 = 2 \cdot 59$, und mit $a = 2$ im Miller-Rabin-Test finden wir $a^{59} \equiv 25 \pmod{119}$ und $a^{118} \equiv 30 \pmod{119}$, sodass N den Test nicht besteht.

BSP
Pollard
 $p - 1$

Jetzt wollen wir einen Teiler von N finden. Wir nehmen wieder $a = 2$. Dann erhalten wir:

$$\begin{aligned} a^{L(2)} &= a^2 \equiv 4 \pmod{119}, & \text{ggT}(3, 119) &= 1, \\ a^{L(3)} &= a^6 \equiv 64 \pmod{119}, & \text{ggT}(63, 119) &= 7, \end{aligned}$$

und wir haben einen Teiler gefunden: $119 = 7 \cdot 17$. ♣

Man kann auch diesen Algorithmus modifizieren, sodass er eine Gruppe der Ordnung $p + 1$ verwendet; dann findet man Teiler p , sodass $p + 1$ B -potenzglatt ist. Wenn man mit der multiplikativen Gruppe eines endlichen Körpers arbeiten will, ist man aber auf diese beiden Möglichkeiten eingeschränkt, wenn man nicht wesentlich größere Gruppen (mit etwa p^2 oder noch mehr Elementen) verwenden möchte, was aber selten etwas bringt.

An dieser Stelle kommen nun elliptische Kurven ins Spiel, denn eine elliptische Kurve über \mathbb{F}_p stellt einem ebenfalls eine abelsche Gruppe der Ordnung ungefähr p zur Verfügung; der genaue Wert der Gruppenordnung variiert aber in einem Intervall um $p + 1$ herum, und die Chancen stehen gut, dass sich in diesem Bereich eine B -potenzglatte Zahl findet.

Bevor wir uns aber der Verwendung von elliptischen Kurven zuwenden, möchte ich noch etwas auf andere Faktorisierungsmethoden eingehen.

Eine davon basiert auf dem *Geburtstagsparadox*. Die Idee ist wie folgt. Sei N die zu faktorisierte Zahl. Wir betrachten eine einfach auszuwertende Funktion $f: \mathbb{Z}/N\mathbb{Z} \rightarrow \mathbb{Z}/N\mathbb{Z}$, zum Beispiel $f(x) = x^2 + 1$. Wir nehmen an, dass sich f bezüglich Iteration im Wesentlichen wie eine zufällige Abbildung verhält. Wir wählen $x_0 \in \mathbb{Z}/N\mathbb{Z}$ und berechnen $x_1 = f(x_0)$, $x_2 = f(x_1)$, usw. Wenn p ein Primteiler von N ist, dann wird die Folge $(x_n \pmod{p})$ irgendwann periodisch; es gibt also n und $m \geq 1$ mit $x_{n+m} \equiv x_n \pmod{p}$. Mit etwas Glück gilt diese Relation nicht für alle Primteiler von N , und wir erhalten einen nichttrivialen Teiler von N mittels $\text{ggT}(x_{n+m} - x_n, N)$.

Um die Anzahl der Vergleiche in einem vernünftigen Rahmen zu halten, kann man parallel zu (x_n) die Folge (x_{2^n}) berechnen und dann jeweils $\text{ggT}(x_{2^n} - x_n, N)$ berechnen. (Verbesserungen sind hier möglich.) Man kann erwarten (aber es ist nicht sicher), einen Teiler in Zeit $O(\sqrt{p}(\log N)^2)$ zu finden, wenn p der kleinste Primteiler von N ist.

13.11. Beispiel. Sei wieder $N = 119$. Wir nehmen $f(x) = x^2 + 1$ und $x_0 = 1$. Wir berechnen

BSP
Pollard ρ

n	0	1	2	3	4
x_n	1	2	5	26	82
$x_n \pmod{7}$	1	2	5	5	5
$x_n \pmod{17}$	1	2	5	9	14

und finden $\text{ggT}(x_2 - x_1, 119) = 1$, $\text{ggT}(x_4 - x_2, 119) = 7$. ♣

Die meisten modernen Faktorisierungsmethoden (aber zum Beispiel nicht die Methode, die mit elliptischen Kurven arbeitet) basieren auf der folgenden Idee: Sei

N eine ungerade zusammengesetzte natürliche Zahl mit wenigstens zwei verschiedenen Primteilern. (Wir können schnell feststellen, ob N eine Potenz ist, also ist das keine Einschränkung.) Wenn wir zwei ganze Zahlen x und y finden mit $x^2 \equiv y^2 \pmod{N}$, dann ist mit Wahrscheinlichkeit $\geq 1/2$ (unter der Annahme, dass x und y zufällig aus den Restklassen mod N gewählt sind) $\text{ggT}(x - y, N)$ ein nichttrivialer Teiler von N . Der Grund dafür ist, dass für jeden Primteiler p von N $x \equiv \varepsilon_p y \pmod{p^{v_p(N)}}$ gilt mit $\varepsilon_p = \pm 1$. Diese Vorzeichen sind voneinander unabhängig, und wir bekommen einen nichttrivialen Teiler, sobald nicht alle Vorzeichen übereinstimmen.

Man versucht also, Kongruenzen der Form $x^2 \equiv y^2 \pmod{N}$ zu generieren. Dazu legt man eine Schranke B fest und betrachtet die Primzahlen q_1, q_2, \dots, q_k , die kleiner als B sind. Die Menge $\{-1, q_1, \dots, q_k\}$ heißt die *Faktorbasis*. Man versucht dann, Relationen der Form

$$x^2 \equiv (-1)^{e_0} q_1^{e_1} \cdots q_k^{e_k} \pmod{N}$$

zu bekommen. Hat man genügend viele davon gesammelt, kann man (durch lineare Algebra über \mathbb{F}_2) Teilmengen dieser Relationen finden, sodass das Produkt der rechten Seiten ein Quadrat wird. Das Produkt der linken Seiten ist in jedem Fall ein Quadrat, und man hat eine Kongruenz der gewünschten Art. Die verschiedenen Methoden unterscheiden sich in der Art und Weise, wie sie die ursprünglichen Relationen erzeugen.

Eine Methode verwendet Kettenbrüche. Für $k = 1, 2, \dots$ berechnet man den Anfang der Kettenbruchentwicklung von \sqrt{kN} und daraus die ersten Näherungsbrüche r/s . Man weiß, dass dann $t = r^2 - s^2 kN$ vergleichsweise klein ist, sodass man hoffen kann, dass sich t über der Faktorbasis faktorisieren lässt. Beachte, dass $r^2 \equiv t \pmod{N}$ ist.

13.12. Beispiel. Sei wieder $N = 119$. Die ersten Näherungsbrüche für $\sqrt{119}$ sind

$$\frac{10}{1}, \quad \frac{11}{1}, \quad \frac{109}{10}, \quad \dots$$

Wir erhalten die Relationen

$$\begin{aligned} 10^2 &\equiv -19 = (-1) \cdot 19 && \pmod{119} \\ 11^2 &\equiv 2 = 2 && \pmod{119} \end{aligned}$$

(die folgenden liefern keine neue Information). Für $\sqrt{2 \cdot 119}$ finden wir die Näherungen 15 und $31/2$ und daraus

$$\begin{aligned} 15^2 &\equiv -13 = (-1) \cdot 13 && \pmod{119} \\ 31^2 &\equiv 9 = 3^2 && \pmod{119}, \end{aligned}$$

und diese letzte Relation führt zum Teiler $\text{ggT}(31 - 3, 119) = 7$. ♣

Beim *Quadratischen Sieb* benutzt man Polynome wie

$$Q(x) = (\lfloor \sqrt{N} \rfloor + x)^2 - N,$$

um relativ kleine Zahlen zu erzeugen, die mod N zu Quadraten kongruent sind. Für die (ansonsten recht aufwendige) Faktorisierung dieser Zahlen kann man verwenden, dass die Teilbarkeit von $Q(a)$ durch p nur von der Restklasse von $a \pmod{p}$ abhängt. Dadurch lassen sich die Primfaktoren aus der Faktorbasis sehr schnell

BSP
Faktorisierung
mit Ketten-
brüchen

aus allen Werten $Q(a)$, $-B < a < B$, entfernen, und man kann die bestimmen, die vollständig faktorisieren.

Diese Methode hat, wenn man sie optimiert, eine erwartete Laufzeit der Größenordnung $O(e^{c\sqrt{\log N \log \log N}})$; sie ist mit die beste Methode, die verfügbar ist. Das *Zahlkörpersieb*, das auf ähnlichen Ideen beruht, aber in einem algebraischen Zahlkörper rechnet, hat sogar eine (vermutete) Komplexität von $O(e^{c\sqrt{\log N (\log \log N)^2}})$, wird aber wegen der komplizierteren Rechnungen erst in einem Bereich schneller, der schon an der Grenze des Machbaren liegt.

14. FAKTORISIERUNG UND PRIMZAHLETEST MIT ELLIPTISCHEN KURVEN

Um die nachfolgenden Resultate ordentlich formulieren zu können, brauchen wir den Begriff einer elliptischen Kurve über $\mathbb{Z}/N\mathbb{Z}$. Ganz allgemein können wir elliptische Kurven über einem (kommutativen) Ring R (mit 1) betrachten. Sie sind genau so definiert, wie über einem Körper; die einzige Schwierigkeit ist, sich zu überlegen, wie die projektive Ebene über R aussieht. Die richtige Definition ist

$$\mathbb{P}^2(R) = \{(\xi, \eta, \zeta) \in R^3 \mid R \cdot \xi + R \cdot \eta + R \cdot \zeta = R\} / \sim,$$

wobei die Äquivalenz \sim wieder gegeben ist durch

$$(\xi, \eta, \zeta) \sim (\xi', \eta', \zeta') \iff \exists \lambda \in R^\times : (\xi', \eta', \zeta') = \lambda \cdot (\xi, \eta, \zeta).$$

Der wesentliche Punkt ist also, dass „ $\neq 0$ “ ersetzt wird durch „invertierbar“ bzw. „relativ prim“. Mit dieser Definition der projektiven Ebene lassen sich alle Begriffe übertragen. Eine elliptische Kurve E über R ist dann gegeben durch eine Weierstraß-Gleichung mit Koeffizienten in R (sodass die Diskriminante invertierbar ist).

Wir bemerken noch, dass ein Ringhomomorphismus $\phi: R \rightarrow S$ eine Abbildung $\mathbb{P}^2(R) \rightarrow \mathbb{P}^2(S)$ induziert, die mit allen Konstruktionen verträglich ist. Wenn wir eine elliptische Kurve E über R haben, dann liefert Anwenden von ϕ auf die Koeffizienten der Gleichung für E eine elliptische Kurve E' über S , und wir erhalten eine Abbildung $E(R) \rightarrow E'(S)$. (Wir haben das im Grunde schon gesehen in dem Fall, dass $R \subset S$ eine Körpererweiterung ist.)

Wir werden das anwenden für $R = \mathbb{Z}/N\mathbb{Z}$. Da wir in jedem Fall kleine Primfaktoren durch Probedivision abspalten können, können wir voraussetzen, dass $N \perp 6$, d.h. dass 6 in $\mathbb{Z}/N\mathbb{Z}$ invertierbar ist. In diesem Fall lässt sich eine lange Weierstraß-Gleichung wieder transformieren in eine kurze Weierstraß-Gleichung (affin geschrieben)

$$E: y^2 = x^3 + ax + b$$

mit $a, b \in \mathbb{Z}/N\mathbb{Z}$, sodass $4a^3 + 27b^2 \in (\mathbb{Z}/N\mathbb{Z})^\times$ ist.

Es ist allerdings nicht mehr so klar, ob $E(\mathbb{Z}/N\mathbb{Z})$ eine Gruppe ist. Wir können aber so tun als ob und für die Addition und die Berechnung von Vielfachen in $E(\mathbb{Z}/N\mathbb{Z})$ mit denselben Formeln wie über einem Körper arbeiten. Dabei werden lediglich die vier Grundrechenarten verwendet. Das einzige, was dann schief gehen kann, ist, dass einmal durch ein Element a geteilt werden soll, das zwar $\neq 0$, aber trotzdem nicht invertierbar ist. In diesem Fall liefert die dabei nötige Berechnung des ggT von a und N einen nichttrivialen Teiler von N , und wir sind fertig. Deswegen können wir annehmen, dass die Berechnungen alle durchführbar sind.

Primzahltest.

Wir betrachten zuerst wieder das Problem, zu beweisen, dass N prim ist. Das folgende Resultat steht in Analogie zu Lemma 13.6.

14.1. Lemma. *Sei $N > 1$ eine ganze Zahl mit $N \perp 6$ und E eine elliptische Kurve über $\mathbb{Z}/N\mathbb{Z}$. Seien weiter $P \in E(\mathbb{Z}/N\mathbb{Z})$ ein Punkt, m eine ganze Zahl, und $q > (\sqrt[4]{N} + 1)^2$ ein Primteiler von m , sodass gilt*

$$(14.1) \quad m \cdot P = O \quad \text{und} \quad \frac{m}{q} \cdot P = (\xi : \eta : \zeta) \quad \text{mit} \quad \zeta \in (\mathbb{Z}/N\mathbb{Z})^\times.$$

Dann ist N prim.

LEMMA
Kriterium
für prim

Beweis. Angenommen, N ist nicht prim; dann gibt es einen Primteiler p von N mit $p \leq \sqrt{N}$. Der kanonische Homomorphismus $\mathbb{Z}/N\mathbb{Z} \rightarrow \mathbb{Z}/p\mathbb{Z} = \mathbb{F}_p$ führt E in eine elliptische Kurve E' über \mathbb{F}_p über; P' sei das Bild von P . Dann ist die Ordnung n von P' (in $E'(\mathbb{F}_p)$) ein Teiler von m , aber kein Teiler von m/q (denn $(m/q) \cdot P' \neq O$, da $\zeta \bmod N$ invertierbar ist, also auch $\bmod p$ nicht verschwindet). Es folgt, dass q diese Ordnung n teilt. Andererseits gilt aber

$$q \mid n \mid \#E'(\mathbb{F}_p) \leq p + 1 + 2\sqrt{p} = (1 + \sqrt{p})^2 \leq (1 + \sqrt[4]{N})^2 < q,$$

ein Widerspruch. \square

Um zu sehen, dass ein darauf gegründeter Algorithmus auch tatsächlich für jede Primzahl funktioniert, brauchen wir noch eine Umkehrung.

14.2. Lemma. Sei $N > 3$ eine Primzahl und E eine elliptische Kurve über $\mathbb{Z}/N\mathbb{Z}$. Sei $m = \#E(\mathbb{Z}/N\mathbb{Z})$ und sei q ein Primteiler von m mit $q > (\sqrt[4]{N} + 1)^2$. Dann gibt es einen Punkt $P \in E(\mathbb{Z}/N\mathbb{Z})$, der (14.1) erfüllt.

LEMMA
Existenz
geeigneter
Punkte

Beweis. Zunächst gilt natürlich für jeden Punkt $P \in E(\mathbb{Z}/N\mathbb{Z})$, dass $m \cdot P = O$ ist. Da N prim ist, bedeutet die zweite Bedingung einfach $(m/q) \cdot P \neq O$. Wir nehmen an, kein Punkt erfülle die zweite Bedingung, d.h., $(m/q) \cdot E(\mathbb{Z}/N\mathbb{Z}) = O$. Wir wissen, dass $E(\mathbb{Z}/N\mathbb{Z}) \cong \mathbb{Z}/d\mathbb{Z} \times \mathbb{Z}/dd'\mathbb{Z}$ ist; es folgt dann $dd' \mid m/q$, also

$$m = \#E(\mathbb{Z}/N\mathbb{Z}) = d^2 d' \leq (dd')^2 \leq (m/q)^2,$$

daher

$$N + 6\sqrt{N} + 1 < (\sqrt[4]{N} + 1)^4 < q^2 \leq m \leq (\sqrt{N} + 1)^2 = N + 2\sqrt{N} + 1,$$

ein Widerspruch. \square

Daraus ergibt sich folgender Algorithmus von *Goldwasser* und *Kilian*.



S. Goldwasser
* 1958

14.3. Algorithmus.

0. Gegeben sei eine (große) natürliche Zahl N , die sehr wahrscheinlich prim ist (insbesondere ist N prim zu 6).
1. Wir wählen zufällige Zahlen a und b in $\mathbb{Z}/N\mathbb{Z}$ mit $4a^3 + 27b^2 \in (\mathbb{Z}/N\mathbb{Z})^\times$. Sei E die durch $y^2 = x^3 + ax + b$ gegebene elliptische Kurve über $\mathbb{Z}/N\mathbb{Z}$.
2. Wir berechnen $m = \#E(\mathbb{Z}/N\mathbb{Z})$ mit dem Polynomzeit-Algorithmus von Schoof-Elkies-Atkin. (Wenn dabei etwas schief geht, dann wissen wir, dass N nicht prim ist.)
3. Wir faktorisieren $m = u \cdot q$ durch Probedivision (mit einer vernünftigen Schranke), wobei u das Produkt der kleinen Primteiler ist.
Dann prüfen wir, ob $(\sqrt[4]{N} + 1)^2 < q \leq m/2$ ist und ob q den Miller-Rabin-Test besteht. Ist dies nicht der Fall, dann versuchen wir es mit einer neuen elliptischen Kurve (Schritt 1.).
4. Wir wählen zufällig Zahlen $x \in \mathbb{Z}/N\mathbb{Z}$, bis das Jacobi-Symbol $\left(\frac{x^3 + ax + b}{N}\right)$ den Wert 0 oder 1 hat. Dann finden wir $y \in \mathbb{Z}/N\mathbb{Z}$ mit $y^2 = x^3 + ax + b$. (Wenn der Algorithmus zum Wurzelziehen versagt, beweist das, dass N nicht prim ist.) Sei $P = (x : y : 1) \in E(\mathbb{Z}/N\mathbb{Z})$.
5. Wir testen, dass $m \cdot P = O$ ist. Ist das nicht der Fall (oder tritt bei der Rechnung ein Fehler auf), dann ist N nicht prim.

6. Wenn $u \cdot P = O$ ist, dann suchen wir einen neuen Punkt auf E (Schritt 4.). Ansonsten ist $u \cdot P = (\xi : \eta : \zeta)$ mit $\zeta \neq 0$. Entweder ist ζ nicht invertierbar; dann ist N nicht prim, oder ζ ist invertierbar, dann ist N prim nach Lemma 14.1, falls q prim ist.
7. Um den Beweis abzuschließen, wenden wir den Algorithmus rekursiv auf q an (bis q klein genug ist, um direkt als prim erkannt zu werden). Stellt sich dabei q als zusammengesetzt heraus, beginnen wir mit einer neuen Kurve von vorn (Schritt 1.).

Man kann zeigen, dass dieser Algorithmus eine erwartete Laufzeit von $O((\log N)^{12})$ hat (unter vernünftigen Annahmen über die Verteilung von Primzahlen in kurzen Intervallen). Für praktische Zwecke ist der Exponent allerdings noch zu groß. Der wesentliche Flaschenhals ist die Bestimmung von $\#E(\mathbb{Z}/N\mathbb{Z})$.

Es gibt eine Variante des Algorithmus (von *Atkin* und *Morain*), die im Wesentlichen spezielle elliptische Kurven konstruiert (solche, deren Endomorphismenring bekannt ist), für die die Zahl m vorher bekannt ist. Dieser Algorithmus ist implementiert worden und ist in der Lage, routinemäßig 1000-stellige Zahlen auf Primalität zu testen (allerdings sind meine Erfahrungen mit Magma in diesem Punkt etwas gemischt). Die schnellsten Varianten dieser Art von Test haben eine heuristische Laufzeit von $O((\log N)^{4+\varepsilon})$ für beliebig kleines $\varepsilon > 0$. In der Praxis sind sie mindestens so gut wie ein anderer schneller Primzahltest (der mit sogenannten Jacobi-Summen arbeitet und ziemlich viel algebraische Zahlentheorie benutzt; seine Komplexität ist $O((\log N)^{c \log \log \log N})$ und damit etwas schlechter als polynomial).

Bemerkt werden sollte auch noch, dass der Goldwasser-Kilian- oder Atkin-Morain-Test gegenüber dem Jacobi-Summen-Test den Vorteil hat, dass er ein *Zertifikat* für die Primalität von N liefert: Mit den Daten E , P , m , q (und dem Zertifikat dafür, dass q prim ist) kann man sich mit Hilfe von Lemma 14.1 sehr schnell davon überzeugen, dass N tatsächlich prim ist.

Adleman und *Huang* haben mit ähnlichen Ideen (unter Verwendung von Kurven vom Geschlecht 2) einen Algorithmus konstruiert, dessen Laufzeit beweisbar polynomial ist; er ist aber (bisher) nicht praktikabel. Dieser Algorithmus ist probabilistisch (wie der von Goldwasser und Kilian); das theoretische Ergebnis, dass es einen (probabilistischen) Polynomzeit-Algorithmus für den Primzahltest gibt, ist inzwischen durch das bessere Resultat von Agrawal, Kayal und Saxena ersetzt worden.

Faktorisierung.

Zur Faktorisierung einer Zahl N (von der wir bereits wissen, dass sie zusammengesetzt ist, z.B. weil sie den Miller-Rabin-Test nicht bestanden hat) kann man genauso vorgehen wie beim $p-1$ -Algorithmus. Statt der multiplikativen Gruppe verwendet man dabei aber die Gruppe der rationalen Punkte einer elliptischen Kurve.

Sei also E eine elliptische Kurve über $\mathbb{Z}/N\mathbb{Z}$ und $P \in E(\mathbb{Z}/N\mathbb{Z})$ ein Punkt. Sei weiter p ein Primteiler von N . Dann haben wir die elliptische Kurve E' über \mathbb{F}_p (durch Reduktion mod p der Gleichung von E) und die kanonische Abbildung $E(\mathbb{Z}/N\mathbb{Z}) \rightarrow E'(\mathbb{F}_p)$. Sei m die Ordnung des Bildes P' von P in E' . Wir nehmen an, m sei B -potenzglatt. Dann ist $L(B) \cdot P' = O$ auf E' . Normalerweise wird die Ordnung des Bildes von P auf den Reduktionen von E modulo anderer Primteiler von N nicht B -potenzglatt sein, und das heißt, dass $L(B) \cdot P$ einerseits nicht der

Punkt O ist, andererseits aber in projektiven Koordinaten die Form $(\xi : \eta : \zeta)$ hat, wo ζ nicht invertierbar ist (denn $\zeta \bmod p$ verschwindet). In diesem Fall ist entweder der ggT von ζ mit N oder der ggT von ξ mit N ein nicht-trivialer Faktor von N .

In der Praxis wird bereits vorher im Verlauf der Rechnung die Situation eintreten, dass eine Division nicht durchführbar ist, weil der Divisor zwar $\neq 0$, aber trotzdem nicht invertierbar ist. In diesem Fall hat man einen nicht-trivialen Faktor gefunden; er wird von der erweiterten ggT-Berechnung geliefert, die versucht, das Inverse des Divisors zu finden.

Außerdem wird man die Kurve so wählen, dass sie einen bekannten Punkt P enthält, denn man kann modulo N keine Quadratwurzeln berechnen (ohne dass man die Faktorisierung von N schon kennt). Man setzt also etwa $P = (1, 1)$ und wählt eine Gleichung der Form

$$y^2 = x^3 + Ax - A \quad \text{oder} \quad y^2 = x^3 + Ax^2 + Bx - (A + B).$$

Man kann auch parallel mit mehreren Kurven arbeiten und abbrechen, sobald eine der Rechnungen erfolgreich ist.

Wir erhalten folgenden Algorithmus.

14.4. Algorithmus.

Eingabe: N (die zu faktorisierende Zahl mit $N \perp 6$)
 B (Parameter wie oben), m (Anzahl Kurven)

1. Für $i = 1, \dots, m$ wiederhole Schritte 2 bis 5.
2. Wähle $A \in \{1, \dots, N - 1\}$ zufällig
 und setze $d_1 = \text{ggT}(A, N)$, $d_2 = \text{ggT}(4A + 27, N)$.
3. (*Diskriminante invertierbar?*)
 Wenn $d_1 > 1$, gib d_1 als Faktor aus; Stop.
 Wenn $1 < d_2 < N$, gib d_2 als Faktor aus; Stop.
 Wenn $d_2 = N$, gehe zu Schritt 2.
4. Setze $E: y^2 = x^3 + \bar{A}x - \bar{A}$ über $\mathbb{Z}/N\mathbb{Z}$ und $P = (\bar{1}, \bar{1}) \in E(\mathbb{Z}/N\mathbb{Z})$.
5. (*Berechnung von $L(B) \cdot P$*)
 Für $p \in \{\text{Primzahlen} \leq B\}$, setze $P = p^{\lfloor \log_p B \rfloor} \cdot P$.
 Dabei verwenden wir die für elliptische Kurven über einem Körper geltenden Formeln. Wenn im Verlauf der Rechnung ein von null verschiedenes, aber nicht invertierbares Element $\bar{d} \in \mathbb{Z}/N\mathbb{Z}$ auftaucht, gib $\text{ggT}(d, N)$ als Faktor aus; Stop.
6. Gib aus „Kein Faktor gefunden“; Stop.

Die Effizienz des Verfahrens hängt davon ab, wie viele B -potenzglatte Zahlen es in der Gegend von p gibt. Wenn wir

$$\ell(x) = e^{\sqrt{\log x \log \log x}}$$

setzen, dann gilt Folgendes.

14.5. Satz. (Canfield, Erdős, Pomerance)³ Die Dichte von $\ell(x)^a$ -potenzglatten Zahlen in der Nähe von x beträgt etwa $\ell(x)^{-1/(2a)}$.

SATZ
 Dichte
 potenzglatter
 Zahlen

³E.R. Canfield, P. Erdős, C. Pomerance: *On a problem of Oppenheim concerning "factorisation numerorum"*, J. Number Theory **17** (1983), no. 1, 1–28.

Wenn wir also Primfaktoren bis zu einer Größe von etwa M finden wollen, dann setzen wir $B = \ell(M)^a$. Wir müssen dann etwa $\ell(M)^{1/(2a)}$ Kurven ausprobieren, bis wir eine passende gefunden haben. Für jede dieser Kurven müssen wir die Multiplikation $L(B) \cdot P$ durchführen. Dafür brauchen wir $O(\log L(B))$ Operationen (Additionen oder Verdopplungen) auf der Kurve, von der jede einen Aufwand von höchstens $O((\log N)^2)$ erfordert; wir werden diesen Faktor jedoch vernachlässigen. Wir brauchen also eine Abschätzung von $\log L(B)$.

14.6. **Satz.** Für $B \rightarrow \infty$ gilt $\log L(B) \sim B$.

SATZ
Asymptotik
von $\log L(B)$

Diese Aussage ist äquivalent zum Primzahlsatz, der sagt, dass für die Anzahl $\pi(x)$ der Primzahlen $\leq x$ die asymptotische Beziehung $\pi(x) \sim x/\log x$ gilt.

Die Rechenzeit für jede Kurve ist also $O(B) = O(\ell(M)^a)$. Insgesamt ergibt sich eine Größenordnung von $\ell(M)^{a+1/(2a)}$. Das wird minimal für $a = 1/\sqrt{2}$ bei einer (erwarteten) Rechenzeit von ungefähr $O(\ell(M)^{\sqrt{2}})$. Hier zeigt sich eine schöne Eigenschaft dieser Methode: Die Rechenzeit hängt hauptsächlich von der Größe der Primfaktoren ab, die man finden möchte. Man kann sie also gut verwenden, um kleine bis mittelgroße Primfaktoren zu finden (und wenn man Glück hat, ist das, was übrigbleibt, schon prim, was man schnell feststellen kann). Im schlimmsten Fall hat man $M = \sqrt{N}$, und die Rechenzeit ist etwa $O(\ell(N))$. Insbesondere ist die Rechenzeit *subexponentiell* in $\log N$, d.h., sie wächst langsamer als jede Funktion $e^{c \log N} = N^c$ (mit $c > 0$). Sie ist im schlimmsten Fall vergleichbar mit dem Quadratischen Sieb, das aber um einen großen konstanten Faktor schneller ist.

Im Vergleich zu anderen Methoden wie etwa dem Quadratischen Sieb hat die hier vorgestellte auch den Vorteil, nur wenig Speicherplatz zu benötigen. Auf der anderen Seite sind andere Verfahren in der Praxis schneller, wenn N ein Produkt zweier etwa gleich großer Primzahlen ist (bei vergleichbarer theoretischer Komplexität), oder auch von besserer theoretischer Komplexität wie $\exp(C \sqrt[3]{\log N (\log \log N)^2})$ beim Zahlkörpersieb.

Hier ist ein Vergleich der verschiedenen Komplexitätsklassen:

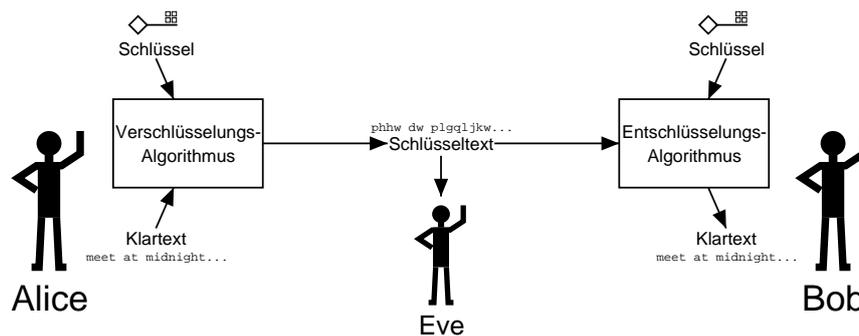
N	\sqrt{N}	$\sqrt[4]{N}$	$e^{\sqrt{\log N \log \log N}}$	$e^{\sqrt[3]{\log N (\log \log N)^2}}$	$(\log N)^5$	$(\log N)^{12}$
1000	10	3,2	38,6	19,2	$1,6 \cdot 10^4$	$1,2 \cdot 10^{10}$
10^6	1000	31,6	413	96,3	$5,0 \cdot 10^5$	$4,8 \cdot 10^{13}$
10^{10}	10^5	316	4910	444	$6,5 \cdot 10^6$	$2,2 \cdot 10^{16}$
10^{20}	10^{10}	10^5	$6 \cdot 10^5$	6460	$2,1 \cdot 10^8$	$9 \cdot 10^{19}$
10^{50}	10^{25}	$3 \cdot 10^{12}$	$1,4 \cdot 10^{10}$	$9 \cdot 10^5$	$2,0 \cdot 10^{10}$	$5 \cdot 10^{24}$
10^{100}	10^{50}	10^{25}	$2,3 \cdot 10^{15}$	$1,7 \cdot 10^8$	$6,4 \cdot 10^{11}$	$2,2 \cdot 10^{28}$
10^{200}	10^{100}	10^{50}	$1,2 \cdot 10^{23}$	$1,7 \cdot 10^{11}$	$2,1 \cdot 10^{13}$	$9 \cdot 10^{31}$
10^{500}	10^{250}	10^{125}	$1,3 \cdot 10^{39}$	$5 \cdot 10^{16}$	$2,0 \cdot 10^{15}$	$5 \cdot 10^{36}$
10^{1000}	10^{500}	10^{250}	10^{58}	$2,8 \cdot 10^{22}$	$6,5 \cdot 10^{16}$	$2,2 \cdot 10^{40}$

Faktorisierungsalgorithmen, die in Computeralgebrasystemen implementiert sind, verwenden in der Regel verschiedene Methoden nacheinander. Üblicherweise beginnt man mit Probedivision durch Primzahlen aus einer gegebenen Liste. Dann prüft man, ob der verbleibende Faktor prim ist. Wenn nicht, kann man die $p-1$ - und die $p+1$ -Methode verwenden (mit nicht zu großem B). Anschließend bietet sich die Elliptische-Kurven-Methode an, um mäßig große Faktoren zu finden (20–30 Stellen oder so). Wenn noch zerlegbare Zahlen übrig sind, kommt eine Version

des Quadratischen Siebs (MPQS: Multiple Polynomial Quadratic Sieve) zum Einsatz. Das Zahlkörpersieb ist für die Verwendung „im Alltag“ noch nicht effizient und robust genug implementiert.

15. KRYPTOGRAPHIE: GRUNDLAGEN

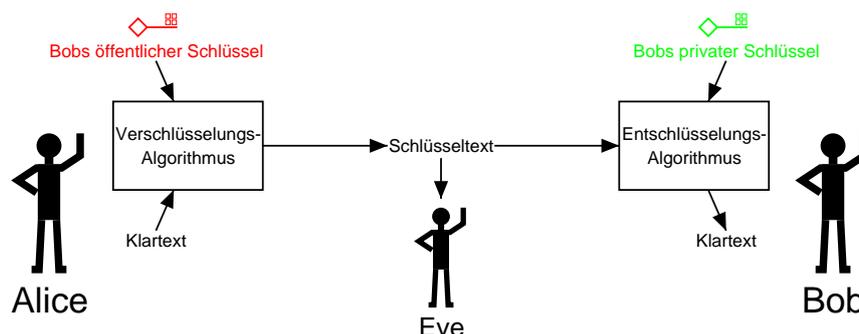
Die Grundaufgabe der Kryptographie besteht darin, eine geheime Nachricht („Klartext“) sicher vom Sender („Alice“) zum Empfänger („Bob“) zu bringen, obwohl der Übertragungskanal (von „Eve“) abgehört werden kann. Die Nachricht muss also so verschlüsselt werden (in einen „Schlüsseltext“), dass sie von möglichen Lauschern nicht rekonstruiert werden kann. Klassischerweise verwendet man Verfahren, die einen geheimzuhaltenden Schlüssel verwenden, und zwar sowohl zum Verschlüsseln als auch zum Entschlüsseln der Nachricht. Man spricht auch von *symmetrischen* Verschlüsselungsverfahren:



Der Vorteil dieser Methoden ist, dass sie üblicherweise sehr effizient sind, man also große Mengen an Information schnell übertragen kann. Aktuell gibt es zum Beispiel als Standard ein AES genanntes Verfahren.

Der Nachteil ist, dass Alice und Bob sich vorher auf einen gemeinsamen Schlüssel geeinigt haben müssen, was im Fall, dass sie bisher noch nicht miteinander kommuniziert haben, auf Schwierigkeiten stößt, denn die dafür nötige Kommunikation muss ja ebenfalls geheim bleiben. Diese Situation tritt zum Beispiel regelmäßig ein, wenn Geschäfte über das Internet abgewickelt werden sollen. Ein weiterer Nachteil der symmetrischen Verfahren ist, dass für jedes *Paar* von Teilnehmern ein eigener Schlüssel generiert werden muss, was bei einer zentralen Erzeugung und Verteilung der Schlüssel (etwa in einem militärischen Kontext) bei wachsender Teilnehmerzahl zu einem nicht mehr beherrschbaren Aufwand führt.

Es sind also neue Ideen gefragt. Ein möglicher Ansatz besteht darin, für Ver- und Entschlüsselung *verschiedene* Schlüssel zu verwenden. Dabei kann der Schlüssel zur Verschlüsselung der Nachrichten an einen bestimmten Teilnehmer öffentlich bekannt sein und heißt dementsprechend *öffentlicher Schlüssel*, während der zur Entschlüsselung benötigte Schlüssel nur dem Empfänger bekannt ist: sein *privater Schlüssel*. Solche Verfahren heißen *asymmetrisch* oder auch *Public-Key-Verfahren*.



Dies stellt allerdings höhere Anforderungen an die Ver- und Entschlüsselungsmethoden. Es darf ja nicht (jedenfalls nicht ohne unvertretbar hohen Aufwand) möglich sein, aus dem Schlüsseltext und dem zum Verschlüsseln benutzten öffentlichen Schlüssel den Klartext zu rekonstruieren. Mathematisch braucht man, was man eine „one-way trapdoor function“ nennt, also eine Funktion, die sich leicht berechnen, aber nur sehr schwer invertieren lässt (one-way), wobei letzteres aber wiederum unter Zuhilfenahme einer Zusatzinformation (trapdoor) ebenfalls leicht möglich ist. Das bekannteste dieser Verfahren ist RSA (nach den Initialen der Erfinder Rivest, Shamir und Adleman). Es beruht darauf, dass das Faktorisieren hinreichend großer ganzer Zahlen offenbar sehr schwierig ist. Es funktioniert wie folgt.

15.1. **Beispiel.** Das RSA-Kryptosystem:

BSP
RSA

1. Wähle zwei große Primzahlen $p \neq q$ und setze $N = pq$.
2. Wähle $1 < e < (p - 1)(q - 1)$ zufällig und teilerfremd zu $\text{kgV}(p - 1, q - 1)$.
3. Berechne d mit $de \equiv 1 \pmod{\text{kgV}(p - 1, q - 1)}$.
4. **Öffentlicher Schlüssel:** (N, e) ,
Privater Schlüssel: d .
5. **Verschlüsselung:**
 $\{0, 1, \dots, N - 1\} \ni m \mapsto c = (m^e \pmod N) \in \{0, 1, \dots, N - 1\}$.
6. **Entschlüsselung:**
 $\{0, 1, \dots, N - 1\} \ni c \mapsto m = (c^d \pmod N) \in \{0, 1, \dots, N - 1\}$. ♣

Dass das Verfahren funktioniert, liegt am kleinen Satz von Fermat:

$$m^{de} = m \cdot (m^{p-1})^a \equiv m \pmod p$$

und ebenso

$$m^{de} = m \cdot (m^{q-1})^b \equiv m \pmod q,$$

wobei wir $de = 1 + a(p - 1) = 1 + b(q - 1)$ gesetzt haben. Es folgt $m^{de} \equiv m \pmod N$.

Die Sicherheit des RSA-Verfahrens beruht auf der Schwierigkeit, d aus e und N zu berechnen, wenn die Primteiler p und q nicht bekannt sind. Dies ist etwa so schwer, wie diese Primteiler zu finden, wie das folgende Lemma zeigt.

15.2. **Lemma.** Seien $p, q > 2$ zwei Primzahlen, $N = pq$ und weiter $d, e \in \mathbb{Z}$ mit $de \equiv 1 \pmod{\text{kgV}(p - 1, q - 1)}$. Wir schreiben $de - 1 = 2^t u$ mit u ungerade. Dann ist für mindestens 50% der ganzen Zahlen $1 \leq a < N$ mit $a \perp N$ eine der Zahlen

LEMMA
RSA und
Faktorisierung

$$\text{ggT}(a^u - 1, N), \quad \text{ggT}(a^{2^u} - 1, N), \quad \dots, \quad \text{ggT}(a^{2^{t-1}u} - 1, N)$$

ein nichttrivialer Teiler von N .

Beweis. Es ist $2^t u$ ein Vielfaches von $p - 1$ und von $q - 1$, und u ist ungerade und daher kein Vielfaches von $p - 1$. Also ist

$$e_p := \min\{e \geq 0 : p - 1 \mid 2^e u\} \in \{1, 2, \dots, t\}$$

und ebenso für das entsprechend definierte e_q . Es folgt, dass $a^{2^{e_p} u} \equiv 1 \pmod p$ ist für alle $a \perp N$; also ist $a^{2^{e_p-1} u} \equiv \pm 1 \pmod p$, und beide Möglichkeiten kommen gleich häufig vor. Die entsprechende Aussage gilt für q und e_q .

Ist $e_p > e_q$, dann gilt für die Hälfte aller a :

$$a^{2^{e_p-1} u} \equiv -1 \pmod p \quad \text{und} \quad a^{2^{e_p-1} u} \equiv 1 \pmod q \quad \implies \quad \text{ggT}(a^{2^{e_p-1} u} - 1, N) = q.$$

Analog erhalten wir den Teiler p als ggT in der Hälfte aller Fälle, wenn $e_p < e_q$ ist. Im Fall $e_p = e_q =: e$ gilt für je ein Viertel der möglichen a

$$a^{2^{e-1}u} \equiv \varepsilon_p \pmod{p} \quad \text{und} \quad a^{2^{e-1}u} \equiv \varepsilon_q \pmod{q}$$

mit jeder möglichen Kombination von Vorzeichen $\varepsilon_p, \varepsilon_q = \pm 1$. (Nach dem Chinesischen Restsatz sind die Restklassen von a modulo p und modulo q unabhängig voneinander.) Für die Hälfte der Restklassen bekommen wir unterschiedliche Vorzeichen; damit ist $\text{ggT}(a^{2^{e-1}u} - 1, N)$ entweder p oder q . \square

Kennen wir also sowohl den Verschlüsselungsexponenten e als auch den Entschlüsselungsexponenten d , dann finden wir nach im Schnitt höchstens zwei Versuchen die beiden Primfaktoren von N mit einer Rechnung, deren Komplexität mit der Ver-/Entschlüsselung vergleichbar ist.

Wie wir gesehen haben, gibt es inzwischen Faktorisierungsalgorithmen subexponentieller Komplexität. Das bedeutet in der Praxis, dass man relativ lange Schlüssel benutzen muss, um ein sicheres Verfahren zu erhalten. Das wirkt sich natürlich negativ auf die Effizienz aus.

Ein anderes Verfahren beruht auf der Schwierigkeit, *diskrete Logarithmen* in multiplikativen Gruppen zu berechnen.

15.3. Definition. Sei $G = \langle g \rangle$ eine (multiplikativ geschriebene) endliche zyklische Gruppe mit gegebenem Erzeuger g . Dann lässt sich jedes Element $h \in G$ schreiben als $h = g^a$, und wir nennen die Zahl a (die modulo der Ordnung von G eindeutig bestimmt ist) den *diskreten Logarithmus* von h zur Basis g . \diamond

DEF
diskreter
Logarithmus

Eine Anwendung ist der *Diffie-Hellman-Schlüsselaustausch*. Hierbei wird nicht eine Nachricht verschlüsselt, sondern die beiden beteiligten Parteien erzeugen ein gemeinsames Geheimnis, das dann zum Beispiel als Schlüssel für ein symmetrisches Verfahren dienen kann.

15.4. Beispiel. Der Schlüsselaustausch nach Diffie und Hellman:

BSP
Diffie-Hellman

1. Man einigt sich auf eine endliche zyklische Gruppe G mit Erzeuger g .
Das ursprüngliche Verfahren verwendet $G = \mathbb{F}_p^\times$ für eine große Primzahl p .
2. Alice wählt eine zufällige Zahl a und berechnet $A = g^a$.
Bob wählt eine zufällige Zahl b und berechnet $B = g^b$.
3. Alice sendet A an Bob. Bob sendet B an Alice.
4. Alice berechnet $s = B^a$. Bob berechnet $s = A^b$. \clubsuit

Wegen $A^b = (g^a)^b = g^{ab} = g^{ba} = (g^b)^a = B^a$ berechnen beide tatsächlich dasselbe Element $s \in G$. Um aus der abgehörten Kommunikation, also den Daten G, g, A, B , das Geheimnis s zu bestimmen, muss man das sogenannte Diffie-Hellman-Problem lösen. Das ist sicher dann möglich, wenn man diskrete Logarithmen in G berechnen kann, denn dann bekommt man zum Beispiel a als Logarithmus von A und kann dann wie Alice $s = B^a$ berechnen. Es wird vermutet, dass beide Probleme (Diffie-Hellman und diskreter Logarithmus) vergleichbar schwer sind.

Man kann die dem Schlüsselaustausch zugrunde liegende Idee auch direkt zum Verschlüsseln benutzen.

15.5. **Beispiel.** Das Kryptosystem nach El Gamal:

BSP
El Gamal

1. Alice und Bob einigen sich auf eine Gruppe G der Ordnung n mit Erzeuger g .
2. Bob wählt eine zufällige Zahl $b \in \mathbb{Z}/n\mathbb{Z}$.
3. **Privater Schlüssel:** b ,
Öffentlicher Schlüssel: $h = g^b$.
4. **Verschlüsselung:**
Alice wählt ein zufälliges $a \in \mathbb{Z}/n\mathbb{Z}$ und berechnet aus dem Klartext $m \in G$ das Paar $(r, s) = (g^a, h^a \cdot m)$.
5. **Entschlüsselung:** Bob berechnet $m = r^{-b} \cdot s$. ♣

Ursprünglich wurden diese Verfahren für $G = \mathbb{F}_p^\times$ vorgeschlagen. Es sind dann aber im Lauf der Zeit Algorithmen für diskrete Logarithmen in multiplikativen Gruppen von endlichen Körpern entwickelt worden, die eine mit den besten Faktorisierungsalgorithmen vergleichbare Komplexität haben. Das Sicherheitsniveau bei gegebener Schlüssellänge ist demnach mit dem des RSA-Verfahrens vergleichbar.

Bevor wir uns ansehen, wie man hier elliptische Kurven gewinnbringend einsetzen kann, möchte ich noch ein wenig auf Algorithmen für diskrete Logarithmen eingehen.

Wir haben also eine (endliche) zyklische Gruppe G mit Erzeuger g und bekannter Ordnung $n = \#G$ gegeben, dazu ein Element $h \in G$, und wir wollen $a \in \mathbb{Z}/n\mathbb{Z}$ bestimmen mit $h = g^a$.

ALGO
Durchprobieren

-
1. Setze $x := 1_G$.
 2. Für $a = 0, 1, \dots, n - 1$ führe Schritte 3 und 4 aus.
 3. Wenn $h = x$, dann gib a aus; Stop.
 4. Setze $x := x \cdot g$.
-

Es ist klar, dass die erwartete Laufzeit (ausgedrückt in der Anzahl der Operationen in G) hier von der Ordnung n und damit exponentiell in der Größe $O(\log n)$ der Eingabedaten ist.

Es ist auch klar, dass jeder andere (vernünftige) Algorithmus besser ist als dieser.

Eine Verbesserungsmöglichkeit besteht darin, dass man nicht ein Element mit allen Elementen von G vergleicht, sondern eine Übereinstimmung in zwei etwa gleich großen Mengen sucht. Diese Idee ist verwandt mit dem Geburtstagsparadox; sie führt auf den folgenden Algorithmus.

ALGO
Baby-Step-
Giant-Step

-
1. Sei $m := \lceil \sqrt{n} \rceil$ und $\gamma := g^m$.
 2. Berechne $\gamma^0, \gamma^1, \dots, \gamma^{m-1}$ und speichere die Paare (j, γ^j) in einer Tabelle T .
 3. Für $r = 0, 1, \dots, m - 1$ führe Schritte 4 und 5 aus.
 4. Berechne $k := hg^{-r}$ und prüfe, ob es einen Eintrag (j, k) in T gibt.
 5. Falls der Eintrag existiert, gib $jm + r$ aus; Stop.

Dieser Ansatz basiert auf folgender Überlegung: Es gilt $n \leq m^2$, also ist $a \leq n - 1 < m^2$; wir können also schreiben

$$a = qm + r$$

mit $q \leq m - 1$ und $0 \leq r \leq m - 1$. Wir haben $h = g^a$ genau dann, wenn gilt

$$hg^{-r} = (g^m)^q.$$

Wir berechnen also zuerst alle möglichen Werte der rechten Seite (in Schritt 2) und dann alle möglichen Werte der linken Seite (in Schritt 4), bis wir eine Übereinstimmung finden. Die Tabelle T muss so organisiert sein, dass man einen Eintrag leicht über seine zweite Komponente finden kann. Dafür eignen sich zum Beispiel Hashtabellen sehr gut.

Die Komplexität ist $O(\sqrt{n})$ Operationen in G . Das ist immer noch exponentiell in $\log n$, aber schon wesentlich besser als das einfache Durchprobieren. Der Nachteil dieses Verfahrens ist, dass es auch $O(\sqrt{n})$ Speicherplatz braucht, um die Tabelle T abzulegen. Das kann für großes n zu Problemen führen.

Das folgende Verfahren beruht auf einer ähnlichen Idee, kommt aber mit recht wenig Speicherplatz aus. Wir benötigen eine Funktion

$$f = (f_1, f_2): G \rightarrow \mathbb{Z} \times \mathbb{Z},$$

die „hinreichend zufällig“ ist. Zum Beispiel kann man einige Bits aus der internen Darstellung der Gruppenelemente extrahieren und den verschiedenen Bitmustern vorher zufällig gewählte ganze Zahlen als Werte von f_1 und f_2 zuordnen. Vier oder fünf Bits sind normalerweise ausreichend. Wir definieren dann (abhängig von den Eingabedaten G, g, h)

$$F: G \longrightarrow G, \quad z \longmapsto z \cdot g^{f_1(z)} \cdot h^{f_2(z)}.$$

Wenn $z = g^a \cdot h^b$ ist, dann ist $F(z) = g^{a+f_1(z)} \cdot h^{b+f_2(z)}$. Wir wählen noch eine (relativ große) Zahl M .

ALGO
Pollard-
Lambda

1. Wähle $x_0, y_0, x'_0, y'_0 \in \mathbb{Z}$ zufällig und setze $z_0 := g^{x_0} \cdot h^{y_0}$ und $z'_0 := g^{x'_0} \cdot h^{y'_0}$.
Initialisiere eine leere Tabelle T .
2. Für $m = 1, 2, \dots$ führe Schritte 3 bis 6 aus.
3. Setze $z_m := F(z_{m-1})$, $(x_m, y_m) := (x_{m-1}, y_{m-1}) + f(z_{m-1})$.
4. Wenn T einen Eintrag (x, y, z_m) enthält und $y - y_m$ modulo n invertierbar ist, dann berechne eine Lösung a von

$$a(y - y_m) \equiv x_m - x \pmod{n}$$

und gib a aus; Stop.

Wenn $y - y_m$ nicht modulo n invertierbar ist, gehe zu Schritt 1.

5. Setze $z'_m := F(z'_{m-1})$, $(x'_m, y'_m) := (x'_{m-1}, y'_{m-1}) + f(z'_{m-1})$.
6. Wenn m durch M teilbar ist, dann speichere (x'_m, y'_m, z'_m) in T .

Wir berechnen hier also zwei Folgen $z_m = g^{x_m} \cdot h^{y_m}$ und $z'_m = g^{x'_m} \cdot h^{y'_m}$ in G und versuchen eine Kollision $z_m = z'_{m'}$ zu finden. In diesem Fall haben wir die Relation

$$g^{x_m} \cdot h^{y_m} = g^{x'_{m'}} \cdot h^{y'_{m'}} \implies g^{a(y'_{m'} - y_m)} = h^{y'_{m'} - y_m} = g^{x_m - x'_{m'}},$$

und wenn $y'_{m'} - y_m$ modulo der Gruppenordnung n invertierbar ist, können wir nach dem diskreten Logarithmus a auflösen. Wenn wir die Kongruenz nicht eindeutig lösen können, können wir neue Anfangswerte nehmen (und eventuell auch die Funktion f ändern). Falls $y'_{m'} \not\equiv y_m \pmod{n}$, bekommen wir immerhin partielle Information über a , die wir im weiteren Verlauf nutzen können. In kryptographischen Anwendungen ist die Gruppenordnung n aber meistens eine Primzahl, sodass dieser Fall nicht eintreten kann.

Dieser Algorithmus wird auch die „Methode der zahmen und wilden Kängurus“ genannt. Das zahme Känguru hüpft durch die Gruppe (Folge (z'_m)) und gräbt nach jeweils M Sprüngen ein Loch. Das wilde Känguru hüpft ebenfalls durch G (Folge (z_m)). Irgendwann wird es auf die Spur des zahmen Kängurus treffen und dann spätestens nach $M - 1$ weiteren Sprüngen in einem Loch gefangen werden.

Ähnlich wie bei der Pollard-Rho-Methode zur Faktorisierung kann man zeigen, dass (bei zufällig gewählter Funktion f) man nach erwarteten $O(\sqrt{n})$ Schritten eine Kollision erhält. Die zeitliche Komplexität ist demnach $O(\sqrt{n} + M)$, und der Speicherplatzbedarf ist $O(\sqrt{n}/M)$. Man kann den Speicherplatz also fast konstant halten, ohne die Größenordnung der Laufzeit zu verschlechtern. Insbesondere kann man den Parameter M an den verfügbaren Speicherplatz anpassen.

Pohlig-Hellman-Reduktion.

Wenn die Gruppenordnung n keine Primzahl ist und ihre Primfaktorzerlegung bekannt ist, dann lässt sich die Berechnung von diskreten Logarithmen in G reduzieren auf die Berechnung von diskreten Logarithmen in Gruppen der Ordnung p , wo p die Primteiler von n durchläuft. Dieser Ansatz geht auf Pohlig und Hellman⁴ zurück.

Sei $n = p_1^{e_1} \cdots p_k^{e_k}$. Im ersten Schritt reduzieren wir das Problem auf die Berechnung von diskreten Logarithmen in Untergruppen von G der Ordnung $p_j^{e_j}$ (für $j = 1, \dots, k$). Dazu beachten wir, dass G für jedes j eine eindeutige solche Untergruppe besitzt, nämlich

$$G_j = \{\gamma \in G \mid \gamma^{p_j^{e_j}} = 1_G\} = \{\gamma^{c_j} \mid \gamma \in G\}$$

mit $c_j = n/p_j^{e_j}$. Wir haben $h^{c_j}, g^{c_j} \in G_j$ und $h^{c_j} = (g^{c_j})^a$. Wenn wir den diskreten Logarithmus von h^{c_j} zur Basis g^{c_j} in G_j berechnen, erhalten wir also $a \pmod{p_j^{e_j}}$. Mit dem Chinesischen Restsatz können wir aus diesen Informationen a berechnen.

Jetzt nehmen wir an, G habe Primzahlpotenzordnung $n = p^e$. Wir bestimmen zunächst $a \pmod{p}$. Dazu beachten wir wie oben, dass

$$G' = \{\gamma^{p^{e-1}} \mid \gamma \in G\}$$

die Untergruppe der Ordnung p von G ist. Wir berechnen den diskreten Logarithmus von $h^{p^{e-1}}$ zur Basis $g^{p^{e-1}}$ in G' ; das liefert $a \pmod{p}$. Sei etwa $a \equiv a_0 \pmod{p}$. Dann liegt hg^{-a_0} in der Untergruppe

$$G'' = \{\gamma^p \mid \gamma \in G\} = \{\gamma \in G \mid \gamma^{p^{e-1}} = 1_G\}$$

der Ordnung p^{e-1} , die von g^p erzeugt wird. Wir berechnen rekursiv den diskreten Logarithmus a' von hg^{-a_0} zur Basis g^p . Dann gilt

$$hg^{-a_0} = g^{a'p} \quad \implies \quad h = g^{a_0 + a'p},$$

⁴G.C. Pohlig, M.E. Hellman: *An improved algorithm for computing logarithms over GF(p) and its cryptographic significance*, IEEE Trans. Information Theory **IT-24**, 106–110 (1978).

also ist $a = a_0 + a'p$.

Kombiniert man die Pohlig-Hellman-Reduktion mit Pollard-Rho oder Baby-Step-Giant-Step, dann reduziert sich die Komplexität im Wesentlichen auf $O(\sqrt{p})$, wobei p der größte Primteiler von $n = \#G$ ist.

Für kryptographische Anwendungen ist man natürlich daran interessiert, dass diskrete Logarithmen nur schwer zu bestimmen sind. Daher wird man hierfür Gruppen verwenden, deren Ordnung eine Primzahl (oder jedenfalls bis auf einen kleinen Faktor prim) ist.

Die bisher beschriebenen Algorithmen sind *generisch*, d.h. auf jede beliebige Gruppe G anwendbar (solange wir in der Gruppe rechnen können, also Produkte und Inverse berechnen und Elemente vergleichen). Ich möchte jetzt noch ein Verfahren beschreiben, das speziell auf $G = \mathbb{F}_p^\times$ zugeschnitten ist. Dafür wählen wir eine Schranke B und setzen $F_B = \{p \mid p \text{ Primzahl}, p \leq B\}$; diese Menge F_B heißt wieder die *Faktorbasis*. g ist in diesem Fall eine Primitivwurzel mod p .

ALGO
Index
Calculus

-
1. Initialisiere eine leere Liste L .
 2. Wiederhole Schritte 3 und 4 solange, bis $\#L \geq \#F_B + 10$ ist.
 3. Wähle $x \in \{1, \dots, p-2\}$ zufällig und berechne $y = g^x \bmod p$.
 4. Falls y B -glatt ist, schreibe $y = \prod_{q \in F_B} q^{e_q}$ und speichere $(x, (e_q)_{q \in F_B})$ in L .
 5. Löse das folgende lineare Gleichungssystem über $\mathbb{Z}/(p-1)\mathbb{Z}$ in den Unbekannten $a_q, q \in F_B$:
Für jeden Eintrag $(x, (e_q)_{q \in F_B})$ in L haben wir die Gleichung

$$x = \sum_{q \in F_B} e_q a_q.$$

6. (Hier gilt $q \equiv g^{a_q} \bmod p$ für alle $q \in F_B$)
Wiederhole Schritte 7 und 8 bis zum Erfolg.
 7. Wähle zufällig $x \in \{0, \dots, p-2\}$ und berechne $y = g^x h \bmod p$.
 8. Falls y B -glatt ist, schreibe $y = \prod_{q \in F_B} q^{e_q}$
und gib $\sum_{q \in F_B} e_q a_q - x$ als Lösung aus.
-

Hier werden (ähnlich wie beim Quadratischen Sieb) erst einmal Relationen zwischen g und den Primzahlen in der Faktorbasis produziert. Diese werden dann dazu benutzt, die diskreten Logarithmen dieser Primzahlen zu bestimmen. Anschließend wird diese Information dazu genutzt, das ursprüngliche Problem zu lösen. Wenn man häufiger diskrete Logarithmen in derselben Gruppe \mathbb{F}_p^\times berechnen muss, dann kann man natürlich das Ergebnis von Schritt 5 abspeichern und dann jeweils gleich mit Schritt 6 beginnen.

Die Komplexitätsanalyse beruht wieder auf dem Satz 14.5 von Canfield, Erdős und Pomerance. Bei optimaler Wahl von B ergibt sich eine Laufzeit von $O(e^{c\sqrt{\log p \log \log p}})$, vergleichbar mit dem Quadratischen Sieb. Man kann auch das Zahlkörpersieb auf die Berechnung diskreter Logarithmen anpassen und bekommt dann wieder eine Komplexität von $O(e^{c\sqrt{\log x (\log \log x)^2}})$.

16. KRYPTOGRAPHIE: ELLIPTISCHE KURVEN

Ähnlich wie die Verwendung von elliptischen Kurven es uns erlaubt, die $(p-1)$ -Methode zum Faktorisieren wesentlich flexibler zu machen, indem wir die multiplikative Gruppe \mathbb{F}_p^\times durch eine Gruppe $E(\mathbb{F}_p)$ ersetzen, können wir auch in den kryptographischen Anwendungen statt einer multiplikativen Gruppe die Gruppe der \mathbb{F}_q -rationalen Punkte auf einer elliptischen Kurve benutzen. Die Verfahren bleiben die gleichen, wie sie oben für allgemeine zyklische Gruppen beschrieben wurden. Der einzige Unterschied ist, dass die Gruppe additiv geschrieben wird. Wir erhalten demnach folgende Versionen.

Zunächst muss eine elliptische Kurve E über einem endlichen Körper \mathbb{F}_q fixiert werden, zusammen mit einem Punkt $P \in E(\mathbb{F}_q)$, dessen Ordnung eine hinreichend große Primzahl n ist. Wir arbeiten mit der Gruppe $G = \langle P \rangle$.

16.1. **Beispiel.** Diffie-Hellman-Schlüsselaustausch mit elliptischen Kurven:

BSP
Diffie-Hellman
mit ell. Kurve

- (1) Alice wählt eine zufällige Zahl a und berechnet $A = a \cdot P$.
Bob wählt eine zufällige Zahl b und berechnet $B = b \cdot P$.
- (2) Alice sendet A an Bob. Bob sendet B an Alice.
- (3) Alice berechnet $S = a \cdot B$. Bob berechnet $S = b \cdot A$.



16.2. **Beispiel.** El Gamal-Verschlüsselung mit elliptischen Kurven:

BSP
El Gamal
mit ell. Kurve

- (1) Bob wählt eine zufällige Zahl $b \in \mathbb{Z}/n\mathbb{Z}$.
- (2) **Privater Schlüssel:** b ,
Öffentlicher Schlüssel: $B = b \cdot P$.
- (3) **Verschlüsselung:**
Alice wählt ein zufälliges $a \in \mathbb{Z}/n\mathbb{Z}$
und berechnet aus dem Klartext $M \in G$ das Paar $(R, S) = (a \cdot P, a \cdot B + M)$.
- (4) **Entschlüsselung:** Bob berechnet $M = S - b \cdot R$.



Es gibt noch weitere Verfahren, etwa zur digitalen Unterschrift oder Authentifizierung.

Warum ist es vorteilhaft, statt mit multiplikativen Gruppen mit elliptischen Kurven zu arbeiten? Wir haben gesehen, dass diskrete Logarithmen in multiplikativen Gruppen in subexponentieller Zeit berechnet werden können. Das bedeutet in der Praxis, dass man relativ große Schlüssellängen (mehrere 1000 Bit) verwenden muss, um ausreichende Sicherheit zu erreichen. Das hat natürlich Auswirkungen auf die Effizienz der Ver- und Entschlüsselung und führt dazu, dass das System wesentlich langsamer arbeitet als symmetrische Verfahren. Außerdem ist es schwer, solche Systeme auf Hardware mit sehr beschränkten Ressourcen, wie zum Beispiel Smartcards, zu implementieren.

Der große Vorteil von elliptischen Kurven ist nun, dass (jedenfalls bisher) kein Algorithmus zur Berechnung von diskreten Logarithmen auf elliptischen Kurven bekannt ist, der auf beliebige elliptische Kurven anwendbar ist und schneller als die generischen Algorithmen (mit Komplexität $O(\sqrt{n})$) wäre. Das bedeutet, dass man bei Verwendung von elliptischen Kurven mit wesentlich kürzeren Schlüssellängen auskommt (wenige 100 Bit). Dadurch ist die Ver- und Entschlüsselung einerseits schneller als bei vergleichbar sicheren Verfahren, die auf Faktorisierung

oder diskreten Logarithmen in multiplikativen Gruppen beruhen (obwohl die einzelne Gruppenoperation aufwendiger ist als etwa eine Multiplikation). Außerdem wird weniger Speicherplatz benötigt, sodass sich diese Verfahren gut für Smartcards oder ähnliche Anwendungen eignen. Es ist gut möglich, dass Sie in Ihrem Geldbeutel eine (oder mehrere) elliptische Kurve(n) mit sich herumtragen!

Es gibt allerdings Angriffsmöglichkeiten in bestimmten Situationen. Wir haben bereits gesehen, dass Pohlig-Hellman-Reduktion die Berechnung von diskreten Logarithmen vereinfacht, wenn die Ordnung von G nicht prim ist.

Ein Angriff, die sogenannte Frey-Rück-Attacke⁵ beruht auf der (mit der Weil-Paarung verwandten) *Tate-Paarung*. Sei dazu E eine elliptische Kurve über einem endlichen Körper \mathbb{F}_q und n eine zu q teilerfremde Zahl. Die Tate-Paarung ist eine Abbildung

$$\langle \cdot, \cdot \rangle_{\text{Tate}}: E(\mathbb{F}_q)/nE(\mathbb{F}_q) \times E(\mathbb{F}_q)[n] \longrightarrow \mathbb{F}_q^\times / \mathbb{F}_q^{\times n}.$$

Um $\langle P + nE(\mathbb{F}_q), Q \rangle_{\text{Tate}}$ zu berechnen, sei $F_Q \in \mathbb{F}_q(E)$ eine rationale Funktion auf E , die in Q eine n -fache Nullstelle und in O einen n -fachen Pol hat. Wir schreiben $P = P_1 - P_2$ mit $\{P_1, P_2\} \cap \{Q, O\} = \emptyset$. Dann ist

$$\langle P, Q \rangle_{\text{Tate}} = \langle P + nE(\mathbb{F}_q), Q \rangle_{\text{Tate}} = \frac{F_Q(P_1)}{F_Q(P_2)} \cdot \mathbb{F}_q^{\times n}.$$

Man kann zeigen, dass diese Definition nicht von der Wahl von F_Q (das ist leicht, denn die möglichen Wahlen unterscheiden sich nur durch Skalierung), der Wahl des Repräsentanten P oder der Darstellung von P als Differenz von P_1 und P_2 abhängt. Die Tate-Paarung ist bilinear (im gleichen Sinne wie bei der Weil-Paarung). Wenn $q \equiv 1 \pmod n$ ist, dann ist die Tate-Paarung auch nicht-ausgeartet. (Im anderen Fall $q \not\equiv 1 \pmod n$ hat $\mathbb{F}_q^\times / \mathbb{F}_q^{\times n}$ Ordnung kleiner als n , sodass die Paarung ausgeartet sein muss.)

Wir brauchen noch ein Lemma:

16.3. Lemma. *Sei E eine elliptische Kurve über \mathbb{F}_q und $P \in E(\mathbb{F}_q)$ ein Punkt der primen Ordnung $n \perp q$. Sei $l \geq 1$ die kleinste Zahl, sodass $q^l \equiv 1 \pmod n$ ist. Wenn $l > 1$ ist, dann gilt $E[n] \subset E(\mathbb{F}_{q^l})$.*



G. Frey
* 1944
Foto © MFO

LEMMA
Definitionskörper von n -Torsionspunkten

Beweis. Da $n \perp q$ ist, ist nach Satz 11.1 (1) $E[n] = E(\bar{\mathbb{F}}_q)[n] \cong \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$. Wir können P zu einer Basis (P, Q) von $E[n]$ ergänzen. Sei $M \in \text{GL}(2, \mathbb{Z}/n\mathbb{Z})$ die Matrix von $\phi|_{E[n]}$ bezüglich dieser Basis, wo ϕ der Frobenius-Endomorphismus von E über \mathbb{F}_q ist. Da $P \in E(\mathbb{F}_q)$ ist, gilt $\phi(P) = P$, also hat M die Form

$$M = \begin{pmatrix} 1 & a \\ 0 & b \end{pmatrix}.$$

Es ist $\zeta = e_n(P, Q)$ eine primitive n -te Einheitswurzel in $\bar{\mathbb{F}}_q$ (nach Satz 11.2 (3) ist e_n nicht-ausgeartet). Aus der Verträglichkeit von e_n mit der Operation der absoluten Galois-Gruppe von $\bar{\mathbb{F}}_q$ (Satz 11.2 (4)) folgt

$$\zeta^b = e_n(P, aP + bQ) = e_n(\phi(P), \phi(Q)) = \phi(\zeta) = \zeta^q,$$

also ist $b = q$ in $\mathbb{Z}/n\mathbb{Z}$. Die Matrix von $\phi^l|_{E[n]}$ ist dann

$$M^l = \begin{pmatrix} 1 & a' \\ 0 & q^l \end{pmatrix} = \begin{pmatrix} 1 & a' \\ 0 & 1 \end{pmatrix},$$

⁵G. Frey, H.-G. Rück: *A remark concerning m -divisibility and the discrete logarithm in the divisor class group of curves*, Math. Comp. **62**, 865–874 (1994)

denn $q^l = 1$ in $\mathbb{Z}/n\mathbb{Z}$ nach Definition von l . Ist $l > 1$, dann ist $b = q \neq 1$ in $\mathbb{Z}/n\mathbb{Z}$, also hat M die beiden verschiedenen Eigenwerte 1 und q ; es folgt, dass M und damit auch M^l diagonalisierbar ist. Das impliziert aber $a' = 0$; damit ist ϕ^l die Identität auf $E[n]$. Das bedeutet aber gerade, dass die Elemente von $E[n]$ in $E(\mathbb{F}_{q^l})$ liegen. \square

Sei jetzt also E eine elliptische Kurve über \mathbb{F}_q und $P \in E(\mathbb{F}_q)$ ein Punkt der (primen) Ordnung n mit $n \perp q$. Sei l wie im Lemma. Falls $l > 1$ ist, dann gibt es nach dem Lemma einen Punkt $P' \in E(\mathbb{F}_{q^l})[n]$, der nicht in $\langle P \rangle$ liegt. Unter diesen Umständen gilt

$$\langle P, P' \rangle_{\text{Tate}} \neq 1.$$

Wir haben den Isomorphismus

$$\alpha: \mathbb{F}_{q^l}^\times / \mathbb{F}_{q^l}^{\times n} \longrightarrow \mu_n(\mathbb{F}_{q^l}), \quad a \cdot \mathbb{F}_{q^l}^{\times n} \longmapsto a^{(q^l-1)/n}.$$

Um jetzt den diskreten Logarithmus von $Q \in \langle P \rangle$ zu berechnen, bestimmen wir

$$r = \alpha(\langle P, P' \rangle_{\text{Tate}}) \quad \text{und} \quad s = \alpha(\langle Q, P' \rangle_{\text{Tate}}).$$

Aus $Q = aP$ und der Bilinearität der Tate-Paarung folgt $s = r^a$. Die Bestimmung von a entspricht also der Berechnung eines diskreten Logarithmus in (der Untergruppe der Ordnung n von) $\mathbb{F}_{q^l}^\times$. Wenn l nicht zu groß ist, sind die dafür verfügbaren subexponentiellen Algorithmen schneller als die generischen Algorithmen für $\langle P \rangle$. In der Praxis sollte man E und P so wählen, dass $l > 20$ ist.

Falls $l = 1$ ist, dann können wir direkt in $E(\mathbb{F}_q)$ arbeiten. In diesem Fall ist (unter der Voraussetzung $n^2 \nmid \#E(\mathbb{F}_q)$) $\langle P, P \rangle_{\text{Tate}}$ nicht trivial, und wir können wie oben verfahren, aber mit $P' = P$. Wir reduzieren dann auf einen diskreten Logarithmus in \mathbb{F}_q^\times , der wesentlich leichter zu berechnen ist, als mit den generischen Methoden. Der Fall $q \equiv 1 \pmod n$ ist also unbedingt zu vermeiden.

Es wurde vorgeschlagen, Kurven E über \mathbb{F}_p mit $\#E(\mathbb{F}_p) = p$ zu verwenden, da diese gegen den eben beschriebenen Angriff immun sind. Es hat sich aber bald herausgestellt, dass sich diskrete Logarithmen auf diesen Kurven noch viel einfacher berechnen lassen. Dazu wählt man eine elliptische Kurve \tilde{E} über \mathbb{Q}_p (dem Körper der p -adischen Zahlen), sodass sich ihre Gleichung mod p auf die von E reduziert. Nach dem Henselschen Lemma kann man auch die Punkte P und Q zu Punkten $\tilde{P}, \tilde{Q} \in \tilde{E}(\mathbb{Q}_p)$ hochheben. Die Punkte $p\tilde{P}$ und $p\tilde{Q}$ liegen im „Kern der Reduktion“, das ist die Untergruppe $\tilde{E}_1(\mathbb{Q}_p)$ von $\tilde{E}(\mathbb{Q}_p)$, deren Elemente die Punkte sind, deren Reduktion mod p gerade der Ursprung $O \in E(\mathbb{F}_p)$ ist. Das sind genau $O \in \tilde{E}(\mathbb{Q}_p)$ und die Punkte (ξ, η) , für die $v_p(\xi/\eta)$ positiv und $v_p(\xi)$ negativ ist. Wir schreiben noch $\tilde{E}_2(\mathbb{Q}_p)$ für die Untergruppe der Punkte mit $v_p(\xi/\eta) \geq 2$ (zusammen mit O). Dann gibt es Isomorphismen

$$\begin{array}{ccccc} E(\mathbb{F}_p) & \longrightarrow & \tilde{E}_1(\mathbb{Q}_p)/\tilde{E}_2(\mathbb{Q}_p) & \longrightarrow & \mathbb{Z}/p\mathbb{Z} \\ Q & \longmapsto & p\tilde{Q} & & \\ & & R & \longmapsto & \frac{x}{py}(R) \pmod p \end{array}$$

falls $p\tilde{P} \notin \tilde{E}_2(\mathbb{Q}_p)$ ist. Falls diese Bedingung nicht erfüllt ist, wähle man eine andere Kurve \tilde{E} .

Die Berechnung des diskreten Logarithmus in $E(\mathbb{F}_p)$ wird auf diese Weise zurückgeführt auf die Berechnung des diskreten Logarithmus in der additiven Gruppe $\mathbb{Z}/p\mathbb{Z}$. Diese ist aber völlig trivial mit dem erweiterten euklidischen Algorithmus zu bewerkstelligen.

Es gibt weitere Angriffsmöglichkeiten, wenn der Körper die Ordnung $q = p^m$ hat, wobei m eine zusammengesetzte Zahl ist. Deshalb wird empfohlen, entweder eine Kurve über einem Körper \mathbb{F}_p zu verwenden, oder eine Kurve über einem Körper \mathbb{F}_{2^p} , wobei jeweils p eine Primzahl ist. Dabei ist in jedem Fall sicherzustellen, dass die Kurve nicht mit einer der oben beschriebenen Methoden angreifbar ist.

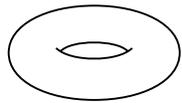
Zur Berechnung der Ordnung $\#E(\mathbb{F}_q)$ wird (für $q = p$) der Algorithmus von Schoof-Elkies-Atkin verwendet; für $q = 2^p$ gibt es einen sehr effizienten Algorithmus von Satoh. Alternativ kann man die Gruppenordnung vorgeben und elliptische Kurven konstruieren, die diese vorgegebene Ordnung haben. Dazu verwendet man Kurven mit komplexer Multiplikation, die über einem geeigneten algebraischen Zahlkörper definiert sind, und reduziert sie modulo einer geeigneten Primzahl.

17. ELLIPTISCHE FUNKTIONEN

Wir betrachten jetzt elliptische Kurven über dem Körper \mathbb{C} der komplexen Zahlen. Für das Folgende orientieren wir uns an Kapitel VI im Buch [Si1] von Silverman.

Sei E eine elliptische Kurve über \mathbb{C} . Die Menge $E(\mathbb{C})$ ihrer Punkte wird beschrieben durch eine Gleichung in zwei Variablen. Wenn wir uns für einen Moment auf den affinen Teil beschränken, dann haben wir eine Teilmenge von \mathbb{C}^2 , die durch eine Gleichung beschrieben wird. Nach dem Satz über implizite Funktionen können wir lokal die Gleichung nach einer der beiden Variablen holomorph auflösen (die zu erfüllende Bedingung bedeutet gerade, dass E glatt ist), sodass $E(\mathbb{C})$ lokal aussieht wie ein Stück der komplexen Ebene \mathbb{C} . Das gilt für jeden affinen Teil von $E(\mathbb{C}) \subset \mathbb{P}^2(\mathbb{C})$, sodass $E(\mathbb{C})$ insgesamt zu einer eindimensionalen komplexen Mannigfaltigkeit, einer sogenannten *Riemannschen Fläche* wird, ähnlich wie die Lösungsmenge einer oder mehrerer Gleichungen im \mathbb{R}^n unter geeigneten Voraussetzungen eine (differenzierbare) reelle Mannigfaltigkeit bildet. Ebenso wie man auf differenzierbaren reellen Mannigfaltigkeiten Analysis treiben kann, kann man Funktionentheorie auf Riemannschen Flächen betreiben. Für eine anschauliche Kurzeinführung in diese Materie seien die letzten Kapitel des Buches von Jänich [Jae] empfohlen.

Diese Fläche $E(\mathbb{C})$ ist topologisch ein Torus. Da man einen Torus auch bekommt, wenn man in der Ebene Punkte identifiziert, deren Differenz in einem gegebenen Gitter (das ist eine additive Untergruppe, die von zwei linear unabhängigen Elementen erzeugt wird) liegt, ist das Hauptresultat dieses Kapitels vielleicht nicht mehr ganz so überraschend. Es sagt nämlich Folgendes.



17.1. Satz. *Zu jeder elliptischen Kurve E über \mathbb{C} gibt es ein Gitter $\Lambda \subset \mathbb{C}$ und einen Isomorphismus $\mathbb{C}/\Lambda \cong E(\mathbb{C})$ als Gruppen und als Riemannsche Flächen. Umgekehrt gibt es zu jedem Gitter $\Lambda \subset \mathbb{C}$ eine elliptische Kurve E über \mathbb{C} mit dieser Eigenschaft.*

SATZ
 $\mathbb{C}/\Lambda \cong E(\mathbb{C})$

Dabei ist \mathbb{C}/Λ eine Gruppe als Faktorgruppe der additiven Gruppe von \mathbb{C} nach dem Normalteiler Λ . Gleichzeitig ist \mathbb{C}/Λ aber auch eine Riemannsche Fläche, da eine kleine Umgebung von $z + \Lambda \in \mathbb{C}/\Lambda$ genauso aussieht wie eine kleine Umgebung von $z \in \mathbb{C}$. (Die Topologie auf \mathbb{C}/Λ ist die sogenannte Quotiententopologie. Sei $\pi: \mathbb{C} \rightarrow \mathbb{C}/\Lambda$ die kanonische Abbildung; dann ist $U \subset \mathbb{C}/\Lambda$ genau dann offen, wenn das Urbild $\pi^{-1}(U) \subset \mathbb{C}$ offen ist. Diese Topologie ist die feinste, die π stetig macht.)

Wir beginnen mit dem zweiten Teil des obigen Satzes; wir gehen also aus von einem gegebenen Gitter Λ . Zunächst die fällige Definition.

17.2. Definition. Ein Gitter $\Lambda \subset \mathbb{C}$ ist eine additive Untergruppe der Form

$$\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2,$$

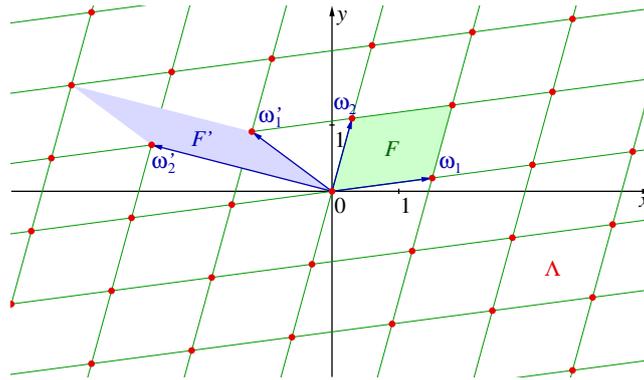
wobei $\omega_1, \omega_2 \in \mathbb{C}$ über \mathbb{R} linear unabhängig sind (also eine Basis des \mathbb{R} -Vektorraums \mathbb{C} bilden).

DEF
Gitter
Fundamental-
parallelogramm

Ein *Fundamentalparallelogramm* für Λ ist eine Menge der Form

$$F = \{a + t_1\omega_1 + t_2\omega_2 \mid t_1, t_2 \in [0, 1]\}$$

mit $a, \omega_1, \omega_2 \in \mathbb{C}$, wo ω_1, ω_2 eine Basis für Λ bilden. ◇



Beachte, dass so ein Gitter viele Basen hat: mit ω_1, ω_2 ist zum Beispiel auch $2\omega_1 + 3\omega_2, \omega_1 + 2\omega_2$ eine Basis.

17.3. Lemma. Sei $\Lambda \subset \mathbb{C}$ ein Gitter, $\pi: \mathbb{C} \rightarrow \mathbb{C}/\Lambda$ die kanonische Abbildung und F ein Fundamentalparallelogramm für Λ . Dann ist $\pi: F \rightarrow \mathbb{C}/\Lambda$ surjektiv. Insbesondere ist \mathbb{C}/Λ kompakt.

LEMMA
 \mathbb{C}/Λ ist kompakt

Beweis. Sei F durch a, ω_1, ω_2 gegeben wie in der Definition, und sei $z + \Lambda \in \mathbb{C}/\Lambda$ ein Element. Wir müssen zeigen, dass es ein $w \in F$ gibt mit $z + \Lambda = w + \Lambda$, d.h., $z - w \in \Lambda$. Dazu schreiben wir alles in der \mathbb{R} -Basis ω_1, ω_2 von \mathbb{C} : $z = z_1\omega_1 + z_2\omega_2$, $a = a_1\omega_1 + a_2\omega_2$ mit $z_1, z_2, a_1, a_2 \in \mathbb{R}$. Sei $n_1 = \lfloor z_1 - a_1 \rfloor$, $n_2 = \lfloor z_2 - a_2 \rfloor$; dann gilt $z = (a + t_1\omega_1 + t_2\omega_2) + (n_1\omega_1 + n_2\omega_2)$ mit $0 \leq t_1, t_2 < 1$, also $z \in F + \Lambda$, was zu zeigen war.

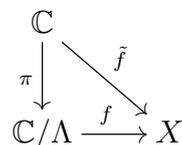
Da F (als abgeschlossene und beschränkte Menge oder als Bild des kompakten Einheitsquadrats unter der stetigen Abbildung $(t_1, t_2) \mapsto a + t_1\omega_1 + t_2\omega_2$) kompakt ist, ist auch \mathbb{C}/Λ als Bild von F unter der stetigen Abbildung π kompakt. \square

Da (wie man sich ähnlich wie in obigem Beweis leicht überlegt) die einzigen Paare von Elementen eines Fundamentalparallelogramms F , die modulo Λ kongruent sind, auf gegenüberliegenden Seiten von F liegen, kann man den topologischen Raum \mathbb{C}/Λ erhalten, indem man so ein F an den zwei Paaren gegenüberliegender Seiten zusammenklebt. Bei der ersten Verklebung entsteht ein Zylinder, bei der zweiten dann ein Torus.

Wir interessieren uns nun für meromorphe Funktionen auf der kompakten Riemannschen Fläche \mathbb{C}/Λ . Wir brauchen dafür aber nicht die Funktionentheorie auf Riemannschen Flächen zu entwickeln, denn es lässt sich alles auf die bekannte Funktionentheorie auf \mathbb{C} zurückführen. Dass das so ist, liegt an der folgenden Bemerkung.

17.4. Bemerkung. Sei $\Lambda \subset \mathbb{C}$ ein Gitter und $\pi: \mathbb{C} \rightarrow \mathbb{C}/\Lambda$ die kanonische Projektion. Dann liefert $f \mapsto f \circ \pi$ eine Bijektion zwischen Abbildungen $f: \mathbb{C}/\Lambda \rightarrow X$ und Abbildungen $\tilde{f}: \mathbb{C} \rightarrow X$ mit der Eigenschaft, dass $\tilde{f}(z + \omega) = \tilde{f}(z)$ gilt für alle $z \in \mathbb{C}$ und $\omega \in \Lambda$.

BEM
Abbildungen auf \mathbb{C}/Λ



Obiger Bemerkung folgend, betrachten wir statt meromorpher Funktionen auf \mathbb{C}/Λ meromorphe Funktionen auf \mathbb{C} , die bezüglich Λ (doppelt-)periodisch sind. Für die verwendeten Sätze aus der Funktionentheorie sei wiederum auf das Buch von Jänich [Jae] verwiesen.

17.5. Definition. Sei $\Lambda \subset \mathbb{C}$ ein Gitter. Eine *elliptische Funktion* für Λ ist eine meromorphe Funktion $f: \mathbb{C} \rightarrow \mathbb{C}$ mit $f(z + \omega) = f(z)$ für alle $z \in \mathbb{C}, \omega \in \Lambda$. \diamond

DEF
elliptische
Funktion

17.6. Bemerkung.

BEM
ell. Funktionen

(1) Summe, Differenz, Produkt und Quotient zweier elliptischer Funktionen für Λ sind wieder elliptische Funktionen für Λ . Die Menge der elliptischen Funktionen für Λ bildet also einen Körper $\mathcal{M}(\Lambda)$. Dieser Körper enthält \mathbb{C} als die konstanten Funktionen.

(2) Wenn ω_1, ω_2 eine Basis von Λ ist, dann ist die Periodizitätsbedingung äquivalent zu

$$f(z + \omega_1) = f(z + \omega_2) = f(z) \quad \text{für alle } z \in \mathbb{C}.$$

(3) Ist $0 \neq f \in \mathcal{M}(\Lambda)$, dann gibt es ein Fundamentalparallelogramm F für Λ , sodass f auf dem Rand ∂F keine Nullstellen oder Pole hat. (Zum Beweis betrachte ein F_0 mit $a = 0$. In der kompakten Menge $2F$ hat f nur endlich viele Nullstellen und Pole (sie können sich wegen des Identitätssatzes nicht häufen), also sind von allen $F = a + F_0$ mit $a \in F_0$ nur endlich viele ausgeschlossen.) \spadesuit

Nun könnte man fragen, warum wir nicht erst einmal holomorphe Funktionen betrachten. Die Antwort ist, dass es keine (interessanten) gibt.

17.7. Lemma. *Eine holomorphe elliptische Funktion ist konstant.*

LEMMA
holomorphe
ell. Fkt.
sind konstant

Beweis. Sei $f \in \mathcal{M}(\Lambda)$ holomorph und F ein Fundamentalparallelogramm von Λ . Dann ist f als stetige Funktion auf der kompakten Menge F beschränkt. Nach Lemma 17.3 gilt aber $f(\mathbb{C}) = f(F)$, also ist f überhaupt beschränkt und damit nach dem Satz von Liouville konstant. \square

Eine nicht-konstante elliptische Funktion muss also mindestens einen Pol (modulo Λ) haben. Das ist aber nicht die einzige Einschränkung.

17.8. Satz. *Seien $\Lambda \subset \mathbb{C}$ ein Gitter und $0 \neq f \in \mathcal{M}(\Lambda)$ eine elliptische Funktion. Sei weiter F ein Fundamentalparallelogramm für Λ , sodass f auf dem Rand ∂F keine Nullstellen oder Pole hat (siehe Bemerkung 17.6 (3)). Dann gilt:*

SATZ
Nullstellen
und Pole
von ell. Fkt.

$$(1) \sum_{z \in F} \text{res}_z(f) = 0;$$

$$(2) \sum_{z \in F} \text{ord}_z(f) = 0;$$

$$(3) \sum_{z \in F} z \cdot \text{ord}_z(f) \in \Lambda.$$

Dabei ist $\text{res}_z(f)$ das Residuum von f in z und $\text{ord}_z(f)$ die Ordnung (d.h. Vielfachheit der Nullstelle oder negativ genommene Vielfachheit des Pols) von f in z . Da f in F nur endlich viele Nullstellen und Pole hat, sind die Summen endlich.

Beweis. Der Beweis besteht jeweils in der Anwendung des Residuensatzes auf ein Integral $\int_{\partial F} g(z) dz$ mit einer geeigneten Funktion g . Der Integrationsweg ist dabei der entgegen dem Uhrzeigersinn orientierte Rand von F . Wenn F durch a, ω_1, ω_2 gegeben ist und (ω_1, ω_2) eine orientierte Basis ist (d.h., es ist $\text{Im}(\omega_2/\omega_1) > 0$), dann gilt also

$$(17.1) \quad \int_{\partial F} g(z) dz = \left(\int_a^{a+\omega_1} - \int_{a+\omega_2}^{a+\omega_1+\omega_2} \right) g(z) dz - \left(\int_a^{a+\omega_2} - \int_{a+\omega_1}^{a+\omega_1+\omega_2} \right) g(z) dz \\ = \int_a^{a+\omega_1} (g(z) - g(z + \omega_2)) dz - \int_a^{a+\omega_2} (g(z) - g(z + \omega_1)) dz.$$

(1) Wir setzen $g = f$ in (17.1). Es folgt

$$2\pi i \sum_{z \in F} \text{res}_z(f) = \int_{\partial F} f(z) dz = 0$$

wegen der Periodizität von f .

- (2) Wir setzen $g = f'/f \in \mathcal{M}(\Lambda)$. Dann gilt $\text{res}_z(g) = \text{ord}_z(f)$, und wie eben folgt $\sum_{z \in F} \text{ord}_z(f) = 0$.
- (3) Hier nehmen wir $g(z) = z f'(z)/f(z)$; dann ist $\text{res}_z(g) = z \cdot \text{ord}_z(f)$. Allerdings ist g nicht mehr Λ -periodisch, sondern es gilt

$$g(z) - g(z + \omega) = -\omega f'(z)/f(z)$$

für $z \in \mathbb{C}$, $\omega \in \Lambda$. Mit (17.1) haben wir also

$$2\pi i \sum_{z \in F} z \cdot \text{ord}_z(f) = -\omega_2 \int_a^{a+\omega_1} \frac{f'(z)}{f(z)} dz + \omega_1 \int_a^{a+\omega_2} \frac{f'(z)}{f(z)} dz.$$

Da f auf ∂F keine Pole oder Nullstellen hat, gibt es eine holomorphe Funktion h auf einer Umgebung der Strecke $[a, a + \omega_1]$ mit $\exp(h(z)) = f(z)$ auf dieser Umgebung. Es gilt dann $h' = f'/f$, also haben wir

$$\int_a^{a+\omega_1} \frac{f'(z)}{f(z)} dz = h(a + \omega_1) - h(a),$$

und das muss $2\pi i$ mal eine ganze Zahl sein, weil

$$\exp(h(a + \omega_1) - h(a)) = f(a + \omega_1)/f(a) = 1$$

ist. Das andere Integral behandelt man ebenso. Insgesamt folgt

$$2\pi i \sum_{z \in F} z \cdot \text{ord}_z(f) = -2\pi i \omega_2 n_2 + 2\pi i \omega_1 n_1$$

mit $n_1, n_2 \in \mathbb{Z}$, was zur Behauptung äquivalent ist. \square

Teil (2) von Satz 17.8 zeigt, dass die folgende Definition sinnvoll ist.

17.9. Definition. Sei $0 \neq f \in \mathcal{M}(\Lambda)$. Dann heißt die Anzahl der Pole von f (mit Vielfachheit gezählt) in einem (jedem) Fundamentalparallelogramm, auf dessen Rand keine Pole oder Nullstellen von f liegen, die *Ordnung* von f . Die Ordnung ist auch gleich der Anzahl der Nullstellen (mit Vielfachheit gezählt). \diamond

DEF
Ordnung
einer ell. Fkt.

Aus Satz 17.8 folgt, dass es keine elliptischen Funktionen der Ordnung 1 gibt, denn ein einfacher Pol hat immer ein nicht-verschwindendes Residuum. Die einfachsten elliptischen Funktionen werden also Ordnung 2 haben und entweder einen zweifachen Pol mit Residuum 0 in $z + \Lambda$ mit $2z \in \Lambda$ oder zwei einfache Pole bei $z + \Lambda$ und $-z + \Lambda$ mit entgegengesetzten Residuen haben. Wir werden bald so eine Funktion der Ordnung 2 konstruieren.

Man kann sich allgemeiner die Frage stellen, wann es zu vorgegebenen Ordnungen $\text{ord}_z(f)$, $z \in F$, eine elliptische Funktion gibt. Die Teile (2) und (3) von Satz 17.8 geben notwendige Bedingungen. Wir werden sehen, dass diese Bedingungen auch hinreichend sind.

Wir haben gesehen, dass ein möglicher Kandidat für eine einfachste nicht-konstante elliptische Funktion einen Pol zweiter Ordnung mit Residuum null in jedem Gitterpunkt hat. So eine Funktion sollte also zum Beispiel in jedem $\omega \in \Lambda$ den Hauptteil $(z - \omega)^{-2}$ haben.

Man wäre jetzt vielleicht versucht,

$$(17.2) \quad f(z) = \sum_{\omega \in \Lambda} \frac{1}{(z - \omega)^2}$$

zu setzen; diese Funktion wäre offensichtlich Λ -periodisch. Dummerweise konvergiert aber die Reihe für kein $z \in \mathbb{C}$ absolut, denn für festes z ist der Term, über den summiert wird, von der Größenordnung ω^{-2} , und $\sum_{\omega \in \Lambda \setminus \{0\}} |\omega|^{-2}$ ist divergent (Übung). Also muss man etwas tun, um die Konvergenz zu erzwingen: Man zieht einfach ω^{-2} vom Term unter der Summe ab und setzt

$$\wp(z) = \wp_\Lambda(z) = \frac{1}{z^2} + \sum_{\omega \in \Lambda \setminus \{0\}} \left(\frac{1}{(z - \omega)^2} - \frac{1}{\omega^2} \right).$$

Das ist die sogenannte *Weierstraßsche \wp -Funktion*. Der Term in der Summe ist jetzt

DEF
Weierstraßsche
 \wp -Funktion

$$\frac{1}{(z - \omega)^2} - \frac{1}{\omega^2} = \frac{\omega^2 - (z - \omega)^2}{\omega^2(z - \omega)^2} = \frac{2z\omega - z^2}{\omega^2(z - \omega)^2} = O(|\omega|^{-3}),$$

gleichmäßig für z in einem Kompaktum in $\mathbb{C} \setminus \Lambda$. Da $\sum_{\omega \in \Lambda \setminus \{0\}} |\omega|^{-3} < \infty$ ist (Übung), konvergiert die Reihe gegen eine in \mathbb{C} meromorphe Funktion mit Polen nur in den Gitterpunkten von Λ mit Hauptteil $(z - \omega)^{-2}$. (Diese Methode mit den „Konvergenz erzeugenden Summanden“ funktioniert übrigens in ähnlicher Weise immer: Zu einem Gebiet $U \subset \mathbb{C}$ gibt es stets eine auf U meromorphe Funktion mit vorgeschriebenen Hauptteilen in U , solange sich die vorgeschriebenen Pole in U nicht häufen. Das ist der Inhalt des *Satzes von Mittag-Leffler*.)

Nun haben wir also eine meromorphe Funktion \wp mit den richtigen Polen. Ist \wp auch elliptisch? Heuristisch würde man das vielleicht erwarten, denn wir haben zu unserem Λ -periodischen Ansatz (17.2) nur eine Konstante addiert. Allerdings verhalten sich unendliche Reihen ja oft kontra-intuitiv; deswegen sollten wir uns nach einem ordentlichen Beweis umsehen. Man kann das relativ einfach anhand der Reihe sehen — für $\lambda \in \Lambda$ ergibt die Differenz $\wp(z + \lambda) - \wp(z)$ eine absolut konvergente Reihe, die nicht von z abhängt und deren Summe null ist. Eleganter

ist aber der folgende üblicherweise angewandte Trick. Da die Reihe für \wp kompakt konvergent ist, dürfen wir gliedweise differenzieren und erhalten

$$\wp'(z) = -2 \sum_{\omega \in \Lambda} \frac{1}{(z - \omega)^3},$$

woraus man sofort sieht, dass \wp' elliptisch ist. Da die Ableitung der Differenz $\wp(z + \omega) - \wp(z)$ also verschwindet, muss die Differenz konstant sein. Nun verwendet man, dass \wp eine gerade Funktion ist (d.h., $\wp(-z) = \wp(z)$); das ist aus der Definition offensichtlich) und wertet obige Differenz bei $-\omega/2$ aus:

$$\text{const.} = \wp(-\omega/2 + \omega) - \wp(-\omega/2) = \wp(\omega/2) - \wp(\omega/2) = 0.$$

Wir bezeichnen den Teilkörper der geraden Funktionen im Körper $\mathcal{M}(\Lambda)$ der elliptischen Funktionen mit $\mathcal{M}(\Lambda)^+$ und die Teilmenge der ungeraden Funktionen mit $\mathcal{M}(\Lambda)^-$. Wir haben mit den obigen Überlegungen bereits den ersten Teil des folgenden Satzes gezeigt.

17.10. Satz.

SATZ
Eigenschaften
von \wp_Λ

- (1) $\wp_\Lambda \in \mathcal{M}(\Lambda)^+$ ist eine gerade elliptische Funktion der Ordnung 2;
 $\wp'_\Lambda \in \mathcal{M}(\Lambda)^-$ ist eine ungerade elliptische Funktion der Ordnung 3.
- (2) $\mathcal{M}(\Lambda)^+ = \mathbb{C}(\wp_\Lambda)$, d.h., jede gerade elliptische Funktion ist eine rationale Funktion von \wp_Λ . Hat $f \in \mathcal{M}(\Lambda)^+$ Pole nur in Λ , so ist f sogar ein Polynom in \wp_Λ .
- (3) $\mathcal{M}(\Lambda) = \mathcal{M}(\Lambda)^+(\wp'_\Lambda)$ ist eine quadratische Körpererweiterung, und es gilt

$$(\wp'_\Lambda)^2 = 4 \wp_\Lambda^3 - 60 G_4(\Lambda) \wp_\Lambda - 140 G_6(\Lambda).$$

Insbesondere ist $\mathcal{M}(\Lambda) = \mathbb{C}(\wp_\Lambda, \wp'_\Lambda)$.

Dabei sei für $k \geq 3$ definiert

$$G_k(\Lambda) = \sum_{\omega \in \Lambda \setminus \{0\}} \omega^{-k}.$$

Es ist leicht zu sehen, dass $G_k(\Lambda) = 0$ ist für k ungerade. Interessant sind also nur $G_4(\Lambda)$, $G_6(\Lambda)$, $G_8(\Lambda)$ usw.

Beweis.

- (1) Siehe oben. Da \wp (modulo Λ) nur einen Pol der Ordnung 2 hat, hat \wp Ordnung 2 als elliptische Funktion. Als Ableitung einer geraden Funktion ist \wp' ungerade. Dass \wp' Ordnung 3 hat, sieht man wie für \wp .
- (2) Die Hauptteile einer Funktion in $\mathcal{M}(\Lambda)^+$ haben (modulo Λ) die Gestalt

$$\sum_{n=1}^N a_n \left(\frac{1}{(z - w)^n} + \frac{(-1)^n}{(z + w)^n} \right),$$

falls $2w \notin \Lambda$, bzw.

$$\sum_{n=1}^N a_n \frac{1}{(z - w)^{2n}},$$

falls $2w \in \Lambda$ (d.h. $w \equiv -w \pmod{\Lambda}$). Auf der anderen Seite gilt, dass der einzige Hauptteil mod Λ von $\wp'(w)/(\wp(z) - \wp(w))$ (für $2w \notin \Lambda$) gerade

$$\frac{1}{z - w} - \frac{1}{z + w}$$

ist. Für $2w \in \Lambda$, $w \notin \Lambda$ hat man

$$\frac{1}{(z-w)^2}$$

als Hauptteil von $\wp''(w)/(2(\wp(z) - \wp(w)))$; für $w \in \Lambda$ hat \wp selbst den richtigen Hauptteil. Man sieht, dass man durch eine endliche Linearkombination von Termen $1/(\wp(z) - \wp(w))^n$ bzw. $\wp(z)^n$ eine gerade elliptische Funktion g bekommen kann, die dieselben Hauptteile wie eine gegebene gerade elliptische Funktion f hat. Die Differenz $f - g$ ist dann holomorph, also konstant. Damit ist die erste Aussage bewiesen. Die zweite folgt ebenfalls, da für den Hauptteil bei 0 nur Potenzen von \wp selbst benötigt werden.

- (3) Sei $f \in \mathcal{M}(\Lambda)$; dann kann man $f = f_+ + f_-$ schreiben als Summe der geraden Funktion $f_+(z) = (f(z) + f(-z))/2 \in \mathcal{M}(\Lambda)^+$ und der ungeraden Funktion $f_-(z) = (f(z) - f(-z))/2 \in \mathcal{M}(\Lambda)^-$. Da \wp' ungerade ist, ist f_-/\wp' gerade. Nach Teil (2) gibt es also rationale Funktionen r_1 und r_2 , sodass $f = r_1(\wp) + r_2(\wp)\wp'$ ist. Folglich ist $1, \wp'$ eine Basis von $\mathcal{M}(\Lambda)$ über $\mathcal{M}(\Lambda)^+ = \mathbb{C}(\wp)$, was die erste Aussage beweist.

Da \wp' ungerade ist, ist $(\wp')^2$ gerade; außerdem hat $(\wp')^2$ Pole nur in Λ . Daher muss $(\wp')^2$ ein Polynom in \wp sein. Um die genaue Relation herauszufinden, brauchen wir die Laurent-Reihe von \wp bei 0.

17.11. Lemma. Die Laurent-Reihe von $\wp_\Lambda(z)$ bei $z = 0$ ist gegeben durch

$$\wp_\Lambda(z) = z^{-2} + \sum_{n=1}^{\infty} (2n+1)G_{2n+2}(\Lambda) z^{2n}.$$

LEMMA
Laurent-Reihe
von \wp

Beweis. Für $|z| < |\omega|$ haben wir

$$(z - \omega)^{-2} - \omega^{-2} = \omega^{-2} \left((1 - z/\omega)^{-2} - 1 \right) = \sum_{n=1}^{\infty} (n+1) \omega^{-(n+2)} z^n.$$

In die definierende Gleichung für \wp eingesetzt und aufsummiert, ergibt das das Ergebnis. (Beachte, dass $G_{2n+1}(\Lambda) = 0$ ist!) \square

Die angegebene Gleichung für $(\wp')^2$ ergibt sich nun durch Abgleichen des Hauptteils und des konstanten Terms der Laurent-Reihe von $(\wp')^2$ wie oben. \square

18. GITTER UND ELLIPTISCHE KURVEN

In Satz 17.8 haben wir Einschränkungen an die Null- und Polstellen (mit ihren Vielfachheiten) einer elliptischen Funktion kennengelernt: Die Summe der Null- und Polstellenordnungen (letztere negativ gerechnet), erstreckt über ein Fundamentalparallelogramm, muss verschwinden, und die Summe der Null- und Polstellen selbst (mit den entsprechenden Vielfachheiten) muss in \mathbb{C}/Λ ebenfalls verschwinden. In diesem Abschnitt wollen wir nun zeigen, dass das die einzigen Einschränkungen sind. Dazu führen wir eine weitere Funktion ein, die uns als multiplikativer Baustein für elliptische Funktionen dienen kann.

18.1. Definition. Die *Weierstraßsche Sigma-Funktion* (zum Gitter Λ) ist definiert als

DEF
Weierstraßsche
 σ - und ζ -Fkt.

$$\sigma(z) = \sigma_\Lambda(z) = z \prod_{\omega \in \Lambda \setminus \{0\}} \left(1 - \frac{z}{\omega}\right) \exp\left(\frac{z}{\omega} + \frac{z^2}{2\omega^2}\right);$$

ihre logarithmische Ableitung

$$\zeta(z) = \zeta_\Lambda(z) = \sigma'(z)/\sigma(z) = \frac{1}{z} + \sum_{\omega \in \Lambda \setminus \{0\}} \left(\frac{1}{z - \omega} + \frac{1}{\omega} + \frac{z}{\omega^2}\right)$$

heißt *Weierstraßsche Zeta-Funktion*. ◇

Die Weierstraßsche Zeta-Funktion ist etwas anderes als die *Riemannsche Zeta-Funktion* $\zeta(s) = \sum_{n=1}^\infty n^{-s}$!



Hier sind die wichtigsten Eigenschaften der Sigma-Funktion.

18.2. Lemma. σ ist eine ganze Funktion mit einfachen Nullstellen genau in den Gitterpunkten von Λ . Zu $\omega \in \Lambda$ gibt es Konstanten $a_\omega, b_\omega \in \mathbb{C}$, sodass

LEMMA
Eigenschaften
von σ

$$\sigma(z + \omega) = e^{a_\omega + b_\omega z} \sigma(z)$$

gilt für alle $z \in \mathbb{C}$.

Beweis. Es ist

$$\begin{aligned} \left(1 - \frac{z}{\omega}\right) \exp\left(\frac{z}{\omega} + \frac{z^2}{2\omega^2}\right) &= \left(1 - \frac{z}{\omega}\right) \left(1 + \frac{z}{\omega} + \frac{z^2}{\omega^2} + O(\omega^{-3})\right) \\ &= 1 + O(\omega^{-3}); \end{aligned}$$

also konvergiert das Produkt absolut und gleichmäßig auf kompakten Mengen (die O -Konstante hängt stetig von z ab). Damit ist σ eine ganze Funktion mit einfachen Nullstellen genau da, wo einer der Faktoren verschwindet, also genau in den Punkten von Λ .

Für die zweite Aussage beachten wir, dass das Glied in der Reihe für $\zeta(z)$ von der Größenordnung $O(\omega^{-3})$ ist; also konvergiert die Reihe absolut und gleichmäßig in kompakten Teilmengen von $\mathbb{C} \setminus \Lambda$. Wir dürfen also gliedweise differenzieren und sehen, dass $\zeta' = -\wp$ ist. Da \wp elliptisch ist, gibt es also ein $b_\omega \in \mathbb{C}$ mit

$$\zeta(z + \omega) = \zeta(z) + b_\omega \quad \text{für alle } z \in \mathbb{C}.$$

Wir betrachten jetzt

$$f(z) = \frac{\sigma(z + \omega)}{e^{b_\omega z} \sigma(z)}.$$

Es gilt

$$\frac{f'(z)}{f(z)} = \frac{\sigma'(z + \omega)}{\sigma(z + \omega)} - \frac{\sigma'(z)}{\sigma(z)} - b_\omega = \zeta(z + \omega) - \zeta(z) - b_\omega = 0,$$

also ist $f(z) = e^{a_\omega}$ konstant. □

18.3. Bemerkungen.

BEM

- (1) Die Faktoren $\exp(z/\omega + \frac{1}{2}(z/\omega)^2)$ in der Produktformel für $\sigma(z)$ heißen (aus naheliegenden Gründen) *Konvergenz erzeugende Faktoren*. Sie dienen dazu, den Faktor auf die Form $1 + O(\omega^{-3})$ zu bringen. Derselbe Trick funktioniert auch ganz allgemein und liefert den Beweis des *Weierstraßschen Produktsatzes*: Ist in einem Gebiet $D \subset \mathbb{C}$ eine Folge von paarweise verschiedenen Punkten $a_n \in D$ mit Vielfachheiten $v_n \in \mathbb{N}$ gegeben, sodass sich (a_n) nirgends in D häuft, so gibt es eine auf D holomorphe Funktion, deren Nullstellen genau die a_n mit Vielfachheit v_n sind. Als Konvergenz erzeugende Faktoren muss man dabei

$$\exp(z/a_n + \frac{1}{2}(z/a_n)^2 + \dots + \frac{1}{k_n}(z/a_n)^{k_n})$$

mit geeigneten k_n nehmen. (Der Term in der Exponentialfunktion ist der Anfang der Potenzreihe von $\log(1 - z/a_n)^{-1}$.)

- (2) Eine Funktion, die ein Transformationsverhalten wie σ aufweist, heißt auch eine *Thetafunktion* bezüglich Λ . Dieser Begriff lässt sich auf Gitter in \mathbb{C}^n verallgemeinern und spielt eine wichtige Rolle in der Theorie der *abelschen Varietäten*, das sind höherdimensionale Analoga von elliptischen Kurven, nämlich projektive Varietäten, die eine Gruppenstruktur haben.
- (3) Aus $\sigma(z + (\omega_1 + \omega_2)) = \sigma((z + \omega_1) + \omega_2)$ bekommt man

$$b_{\omega_1 + \omega_2} = b_{\omega_1} + b_{\omega_2} \quad \text{und} \\ a_{\omega_1 + \omega_2} \equiv a_{\omega_1} + b_{\omega_1}\omega_2 + a_{\omega_2} \pmod{2\pi i\mathbb{Z}}.$$

b definiert also einen Homomorphismus $\Lambda \rightarrow \mathbb{C}$, während das Paar (a, b) einen sogenannten 1-Kozykel $\Lambda \rightarrow \mathbb{C}/2\pi i\mathbb{Z} \times \mathbb{C}$ definiert, wobei Λ auf $\mathbb{C}/2\pi i\mathbb{Z} \times \mathbb{C}$ operiert durch $(\alpha, \beta)^\omega = (\alpha + \beta\omega, \beta)$. ♠

Die Sigma-Funktion stellt uns das nötige Material zur Verfügung, um den Satz von Abel-Jacobi zu beweisen.

18.4. Satz. Sei $\Lambda \subset \mathbb{C}$ ein Gitter, sei F ein Fundamentalparallelogramm für Λ , seien $w_1, \dots, w_n \in F$ paarweise verschieden und $m_1, \dots, m_n \in \mathbb{Z}$ mit

SATZ
Satz von
Abel-Jacobi

$$\sum_{j=1}^n m_j = 0 \quad \text{und} \quad \sum_{j=1}^n m_j w_j = \lambda \in \Lambda.$$

Dann gibt es $f \in \mathcal{M}(\Lambda)$ mit $\text{ord}_{w_j}(f) = m_j$ für alle j und $\text{ord}_z(f) = 0$ für alle $z \in F \setminus \{w_1, \dots, w_n\}$.

Das liefert eine Umkehrung zu Satz 17.8.

Beweis. Da σ genau eine einfache Nullstelle modulo Λ hat, liegt folgender Ansatz nahe:

$$f_0(z) = \prod_{j=1}^n \sigma(z - w_j)^{m_j}.$$

Diese Funktion f_0 hat dann jedenfalls die richtigen Null- und Polstellen. Ist f_0 auch elliptisch? Dazu sei $\omega \in \Lambda$. Es gilt

$$\begin{aligned} f_0(z + \omega) &= \prod_{j=1}^n \sigma(z + \omega - w_j)^{m_j} \\ &= \exp\left(a_\omega \sum_{j=1}^n m_j + b_\omega \sum_{j=1}^n m_j(z - w_j)\right) f_0(z) \\ &= \exp(-b_\omega \lambda) f_0(z). \end{aligned}$$

f_0 ist also nur fast elliptisch. Auf der anderen Seite ist aber $\sigma(z)/\sigma(z - \lambda)$ eine ganze Funktion ohne Nullstellen, und es gilt

$$\frac{\sigma(z + \omega)}{\sigma(z + \omega - \lambda)} = \exp(b_\omega \lambda) \frac{\sigma(z)}{\sigma(z - \lambda)};$$

also leistet

$$f(z) = f_0(z) \frac{\sigma(z)}{\sigma(z - \lambda)}$$

das Gewünschte. (Der Korrekturfaktor $\sigma(z)/\sigma(z - \lambda)$ ist von der Form ae^{bz} .) \square

Wir können jetzt die eine Hälfte des Uniformisierungssatzes beweisen.

18.5. Satz. Sei $\Lambda \subset \mathbb{C}$ ein Gitter und E die durch die affine Gleichung

$$(18.1) \quad y^2 = x^3 - 15G_4(\Lambda)x - 35G_6(\Lambda)$$

definierte projektive ebene Kurve. Dann ist E eine elliptische Kurve, und die Abbildung

$$\phi: \mathbb{C}/\Lambda \ni z \mapsto (\wp(z) : \frac{1}{2}\wp'(z) : 1) \in E(\mathbb{C})$$

liefert einen Isomorphismus von \mathbb{C}/Λ und $E(\mathbb{C})$ als Gruppen und Riemannschen Flächen.

SATZ
Ell. Kurve
zu Gitter

Beweis. Nach Satz 17.10, Teil (3), ist ϕ jedenfalls eine wohldefinierte Abbildung (in der Nähe von 0 muss man die projektiven Koordinaten mit z^3 multiplizieren, um den Pol von \wp' zu eliminieren; dann sieht man, dass $\phi(0) = O$ ist).

Wir fixieren eine Basis ω_1, ω_2 von Λ und setzen $\omega_3 = \omega_1 + \omega_2$. Da \wp eine gerade elliptische Funktion der Ordnung 2 ist, sind die Werte

$$e_1 = \wp(\omega_1/2), \quad e_2 = \wp(\omega_2/2), \quad e_3 = \wp(\omega_3/2)$$

paarweise verschieden (denn $\wp(z - \omega_j/2)$ ist ebenfalls gerade, also hat $\wp - e_j$ eine doppelte Nullstelle in $\omega_j/2$ und keine weiteren Nullstellen mod Λ). Weiterhin gilt

$$\wp'(\omega_j/2) = 0 \quad \text{für } j = 1, 2, 3,$$

da \wp' ungerade ist. Es folgt

$$y^2 = (x - e_1)(x - e_2)(x - e_3);$$

also verschwindet die Diskriminante von (18.1) nicht, und es liegt eine elliptische Kurve vor.

Wir zeigen als nächstes, dass ϕ bijektiv ist. Zur Injektivität: Sei $\phi(z_1) = \phi(z_2)$. Wir betrachten $f(z) = \wp(z) - \wp(z_1)$, eine gerade elliptische Funktion der Ordnung 2. Wenn z_1 und z_2 modulo Λ verschieden sind, dann sind z_1 und z_2 genau die beiden einfachen Nullstellen von f . Es folgt $z_2 \equiv -z_1 \pmod{\Lambda}$, also $\wp'(z_2) = -\wp'(z_1)$, und das ist ungleich null, da $2z_1 \equiv z_1 - z_2 \notin \Lambda$. Es folgt $\phi(z_1) \neq \phi(z_2)$, Widerspruch.

Also muss $z_1 \equiv z_2 \pmod{\Lambda}$ gelten. Dass 0 der einzige Wert mit Bild O ist, ist klar. Zur Surjektivität: Sei $(x, y) \in E(\mathbb{C})$ (O liegt natürlich im Bild). Die Funktion $f(z) = \wp(z) - x$ hat Ordnung 2, also mindestens eine Nullstelle z_0 . Wegen Gleichung (18.1) gilt dann $\frac{1}{2}\wp'(z_0) = \pm y$, also $(x, y) = \phi(z_0)$ oder $(x, y) = \phi(-z_0)$.

Da \wp und \wp' meromorph sind, ist ϕ eine holomorphe Abbildung zwischen Riemannschen Flächen. Genauer: Eine Abbildung zwischen Riemannschen Flächen ist holomorph, wenn sie „holomorph bezüglich Karten“ ist. Die Karten für \mathbb{C}/Λ kommen von der kanonischen Projektion $\pi: \mathbb{C} \rightarrow \mathbb{C}/\Lambda$. Die Karten für $E(\mathbb{C})$ sind gegeben durch Projektion auf eine Koordinate in einem passenden affinen Teil der Kurve. Im üblichen affinen Teil bekommt man als Funktion bezüglich der Karten entweder \wp oder $\frac{1}{2}\wp'$ auf einem Gebiet in $\mathbb{C} \setminus \Lambda$; diese Funktionen sind holomorph. In einer Umgebung von O liefert x/y eine Karte; die zugehörige Funktion bezüglich Karten ist dann $2\wp/\wp'$, und diese Funktion ist holomorph in einer Umgebung von 0, weil der dreifache Pol von \wp' den doppelten Pol von \wp „kürzt“ (und tatsächlich in eine Nullstelle verwandelt). Da ϕ bijektiv ist, ist ϕ sogar biholomorph, also ein Isomorphismus Riemannscher Flächen.

Es bleibt zu zeigen, dass ϕ auch mit den Gruppenstrukturen verträglich ist. Da wir bereits gesehen haben, dass ϕ bijektiv ist, reicht es aus zu zeigen, dass für alle Punkte $P_1, P_2, P_3 \in E(\mathbb{C})$ gilt

$$P_1 + P_2 + P_3 = O \implies \phi^{-1}(P_1) + \phi^{-1}(P_2) + \phi^{-1}(P_3) = 0.$$

Im Fall $P_1 = O$ (z.B.) ist das klar, denn

$$\phi(-z) = (\wp(z), -\frac{1}{2}\wp'(z)) = -(\wp(z), \frac{1}{2}\wp'(z)) = -\phi(z).$$

Seien also o.B.d.A. alle $P_j = (x_j, y_j)$ von O verschieden. Dann liegen sie auf einer Geraden, die nicht parallel zur y -Achse ist:

$$y_j = a x_j + b \quad \text{für } j = 1, 2, 3$$

mit geeigneten $a, b \in \mathbb{C}$. Sei $f(z) = \frac{1}{2}\wp'(z) - a\wp(z) - b$. Dann ist f eine elliptische Funktion der Ordnung 3 (einziger Pol ist bei 0, mit Polordnung 3); ihre drei Nullstellen sind also (mit Vielfachheit) gerade die $\phi^{-1}(P_j)$. Nach Satz 17.8, Teil (3), haben wir also

$$\phi^{-1}(P_1) + \phi^{-1}(P_2) + \phi^{-1}(P_3) - 3 \cdot 0 = 0 \in \mathbb{C}/\Lambda;$$

das ist die Behauptung. □

Wir wollen jetzt sehen, wie man Isogenien zwischen elliptischen Kurven, die durch Gitter gegeben sind, beschreiben kann. Dazu brauchen wir folgende Aussage.

18.6. Lemma. *Seien $\Lambda, \Lambda' \subset \mathbb{C}$ zwei Gitter. Eine Abbildung $f: \mathbb{C}/\Lambda \rightarrow \mathbb{C}/\Lambda'$ ist genau dann holomorph, wenn es eine holomorphe Funktion $\tilde{f}: \mathbb{C} \rightarrow \mathbb{C}$ gibt, sodass $f \circ \pi = \pi' \circ \tilde{f}$ ist; dabei sind $\pi: \mathbb{C} \rightarrow \mathbb{C}/\Lambda$ und $\pi': \mathbb{C} \rightarrow \mathbb{C}/\Lambda'$ die kanonischen Projektionen:*

LEMMA
holomorphe
Abbildung
 $\mathbb{C}/\Lambda \rightarrow \mathbb{C}/\Lambda'$

$$\begin{array}{ccc} \mathbb{C} & \xrightarrow{\tilde{f}} & \mathbb{C} \\ \pi \downarrow & & \downarrow \pi' \\ \mathbb{C}/\Lambda & \xrightarrow{f} & \mathbb{C}/\Lambda' \end{array}$$

Beweis. Die Richtung „ \Leftarrow “ ist klar: Wenn \tilde{f} wie angegeben existiert, dann liefert das überall die holomorphe Funktion bezüglich der Karten von \mathbb{C}/Λ und \mathbb{C}/Λ' .

Für die Richtung „ \Rightarrow “ sei $f: \mathbb{C}/\Lambda \rightarrow \mathbb{C}/\Lambda'$ holomorph. Dann ist $f \circ \pi: \mathbb{C} \rightarrow \mathbb{C}/\Lambda'$ holomorph; für jeden Punkt $z \in \mathbb{C}$ als Repräsentant eines Punktes in \mathbb{C}/Λ und jedes $z' \in \mathbb{C}$ mit $f(\pi(z)) = \pi'(z')$ gibt es dann eine holomorphe Funktion $\tilde{f}_{z,z'}$, die eine Umgebung U von z in eine Umgebung U' von z' abbildet und für die $\pi' \circ \tilde{f}_{z,z'} = f \circ \pi$ auf U gilt. Da \mathbb{C}/Λ kompakt ist, können wir die Umgebung U stets so wählen, dass sie die offene Kreisscheibe $B(z, r)$ mit Radius r um z enthält; hierbei ist $r > 0$ unabhängig von z . Wir fixieren jetzt ein z' mit $f(\pi(0)) = \pi'(z')$ und betrachten die Menge

$$M = \{R > 0 \mid \exists \tilde{f}_R: B(0, R) \rightarrow \mathbb{C}: \tilde{f}_R \text{ holomorph, } \pi' \circ \tilde{f}_R = f \circ \pi|_{B(0,R)}\}.$$

Nach unserer obigen Überlegung ist $r \in M$. Sind $R, R' \in M$ mit $R < R'$, dann stimmen \tilde{f}_R und $\tilde{f}_{R'}$ auf einer Umgebung der Null überein; nach dem Identitätssatz folgt $\tilde{f}_{R'}|_{B(0,R)} = \tilde{f}_R$. Wir zeigen jetzt, dass M unbeschränkt ist; dann können wir \tilde{f} definieren als $\tilde{f}(z) = \tilde{f}_R(z)$ für alle z mit $|z| < R$.

Sei $R \in M$. Wir zeigen, dass $R + r/2 \in M$ ist; daraus folgt die Behauptung. Wir können $B(0, R + r/2)$ überdecken durch $B(0, R)$ und endlich viele Kreisscheiben $B_j = B(z_j, r)$ vom Radius r , die Mittelpunkte $z_j \in B(0, R)$ haben. Wir haben dann die Funktion $\tilde{f}_{z_j, \tilde{f}_R(z_j)}$ auf B_j , die auf $B(0, R) \cap B_j$ mit \tilde{f}_R übereinstimmt. Wir können also \tilde{f}_R auf $U_j = (B(0, R) \cup B_j) \cap B(0, R + r/2)$ fortsetzen. Die Vereinigung der U_j ist $B(0, R + r/2)$, und je zwei dieser Fortsetzungen stimmen auf ihrem gemeinsamen Definitionsbereich überein. Daher erhalten wir die gewünschte Funktion auf $B(0, R + r/2)$ als Vereinigung (der Graphen) dieser Fortsetzungen. \square

Der entscheidende Punkt ist hier, dass \mathbb{C} einfach zusammenhängend ist. Daraus folgt, dass jede holomorphe Abbildung $h: \mathbb{C} \rightarrow \mathbb{C}/\Lambda'$ sich zu einer holomorphen Abbildung $\tilde{h}: \mathbb{C} \rightarrow \mathbb{C}$ „hochheben“ lässt; d.h., sodass $h = \pi' \circ \tilde{h}$ gilt.

Wir können diese Abbildungen recht genau beschreiben.

18.7. Satz. *Seien $\Lambda, \Lambda' \subset \mathbb{C}$ zwei Gitter. Dann haben wir die folgende Bijektion:*

$$\begin{aligned} \{\alpha \in \mathbb{C} \mid \alpha\Lambda \subset \Lambda'\} &\longleftrightarrow \{f: \mathbb{C}/\Lambda \rightarrow \mathbb{C}/\Lambda' \mid f \text{ holomorph, } f(0) = 0\} \\ \alpha &\longmapsto (f_\alpha: z + \Lambda \mapsto \alpha z + \Lambda') \end{aligned}$$

SATZ
hol. Abb.
 $\mathbb{C}/\Lambda \rightarrow \mathbb{C}/\Lambda'$

Beweis. Die Abbildung ist jedenfalls wohldefiniert, denn wir können $\tilde{f}_\alpha(z) = \alpha z$ nehmen; das induziert eine wohldefinierte Abbildung $\mathbb{C}/\Lambda \rightarrow \mathbb{C}/\Lambda'$, denn aus $z' = z + \lambda$ mit $\lambda \in \Lambda$ folgt $\alpha z' = \alpha z + \alpha \lambda$ mit $\alpha \lambda \in \alpha\Lambda \subset \Lambda'$.

Die Abbildung ist injektiv: Seien α, β in der linken Menge mit $f_\alpha = f_\beta$. Für $z \in \mathbb{C}$ gilt dann $\alpha z - \beta z \in \Lambda'$. Da Λ' diskret und $z \mapsto (\alpha - \beta)z$ stetig ist, muss $(\alpha - \beta)z$ konstant sein. Einsetzen von $z = 0$ und $z = 1$ zeigt $\alpha = \beta$.

Die Abbildung ist surjektiv: Sei $f: \mathbb{C}/\Lambda \rightarrow \mathbb{C}/\Lambda'$ holomorph mit $f(0) = 0$. Dann gibt es $\tilde{f}: \mathbb{C} \rightarrow \mathbb{C}$ mit $f \circ \pi = \pi' \circ \tilde{f}$; wir können $\tilde{f}(0) = 0$ annehmen. Sei jetzt $\lambda \in \Lambda$. Dann gilt für alle $z \in \mathbb{C}$, dass $\tilde{f}(z + \lambda) - \tilde{f}(z) \in \Lambda'$ ist. Weil die Funktion links stetig und Λ' diskret ist, muss $\tilde{f}(z + \lambda) - \tilde{f}(z)$ konstant sein. Es folgt $\tilde{f}'(z + \lambda) = \tilde{f}'(z)$, d.h., \tilde{f}' ist eine holomorphe elliptische Funktion bezüglich Λ . Nach Lemma 17.7 ist \tilde{f}' konstant, also ist $\tilde{f}(z) = \alpha z$ mit $\alpha \in \mathbb{C}$ (denn $\tilde{f}(0) = 0$). Für $\lambda \in \Lambda$ gilt $\alpha \lambda = \tilde{f}(\lambda) \in \Lambda'$, also ist $f = f_\alpha$ mit α aus der Menge links. \square

Das nächste Resultat bringt den Zusammenhang mit Isogenien.

18.8. Satz. *Seien $\Lambda, \Lambda' \subset \mathbb{C}$ zwei Gitter und seien E, E' die zugehörigen elliptische Kurven über \mathbb{C} wie in Satz 18.5. Jede Isogenie $\phi: E \rightarrow E'$ induziert eine holomorphe Abbildung $f_\phi: \mathbb{C}/\Lambda \cong E(\mathbb{C}) \rightarrow E'(\mathbb{C}) \cong \mathbb{C}/\Lambda'$ mit $f_\phi(0) = 0$. Die so definierte Abbildung*

$$\begin{aligned} \{\phi: E \rightarrow E' \text{ Isogenie}\} &\longrightarrow \{f: \mathbb{C}/\Lambda \rightarrow \mathbb{C}/\Lambda' \mid f \text{ holomorph, } f(0) = 0\} \\ \phi &\longmapsto f_\phi \end{aligned}$$

ist bijektiv.

Beweis. Als Morphismus zwischen projektiven ebenen Kurven ist ϕ durch rationale Funktionen der Koordinaten gegeben; diese sind offensichtlich holomorph auf $E(\mathbb{C})$. Die Isomorphismen zwischen \mathbb{C}/Λ und $E(\mathbb{C})$ und analog für E' sind biholomorph, also ist f_ϕ eine holomorphe Abbildung, und die Abbildung im Satz ist wohldefiniert. Sie ist injektiv, weil rationale Abbildungen auf Kurven durch ihre Werte auf den Punkten über einem algebraisch abgeschlossenen Körper eindeutig bestimmt sind.

Wir zeigen noch die Surjektivität. Sei $f: \mathbb{C}/\Lambda \rightarrow \mathbb{C}/\Lambda'$ holomorph mit $f(0) = 0$. Nach Satz 18.7 ist $f = f_\alpha$ für ein $\alpha \in \mathbb{C}$ mit $\alpha\Lambda \subset \Lambda'$. In projektiven Koordinaten ist dann die von f induzierte Abbildung $E(\mathbb{C}) \rightarrow E'(\mathbb{C})$ gegeben durch

$$(\wp_\Lambda(z) : \frac{1}{2}\wp'_\Lambda(z) : 1) \longmapsto (\wp_{\Lambda'}(\alpha z) : \frac{1}{2}\wp'_{\Lambda'}(\alpha z) : 1).$$

Wir müssen daher zeigen, dass $\wp_{\Lambda'}(\alpha z)$ und $\wp'_{\Lambda'}(\alpha z)$ als rationale Funktionen in \wp_Λ und \wp'_Λ geschrieben werden können. Für $\lambda \in \Lambda$ ist $\alpha\lambda \in \Lambda'$ und daher

$$\wp_{\Lambda'}(\alpha(z + \lambda)) = \wp_{\Lambda'}(\alpha z + \alpha\lambda) = \wp_{\Lambda'}(\alpha z)$$

und analog für $\wp'_{\Lambda'}(\alpha z)$. Nach Satz 17.10 sind also

$$\wp_{\Lambda'}(\alpha z), \wp'_{\Lambda'}(\alpha z) \in \mathcal{M}(\Lambda) = \mathbb{C}(\wp_\Lambda, \wp'_\Lambda)$$

wie gewünscht. □

18.9. Folgerung. *Seien E, E' elliptische Kurven über \mathbb{C} assoziiert zu Gittern Λ, Λ' . E und E' sind genau dann isomorph (über \mathbb{C}), wenn es $\alpha \in \mathbb{C}$ gibt mit $\alpha\Lambda = \Lambda'$.*

FOLG
Isomorphie
über \mathbb{C}

Beweis. Das folgt unmittelbar aus den Sätzen 18.7 und 18.8. □

18.10. Definition. Wir schreiben E_Λ für die elliptische Kurve

$$E_\Lambda: y^2 = x^3 - 15G_4(\Lambda)x - 35G_6(\Lambda),$$

die zu einem Gitter $\Lambda \subset \mathbb{C}$ gehört.

Sei \mathcal{L} die Menge aller Gitter in \mathbb{C} . Auf \mathcal{L} operiert die multiplikative Gruppe \mathbb{C}^\times durch $\alpha \cdot \Lambda = \alpha\Lambda$. Wir schreiben $[\Lambda]$ für die Bahn von Λ unter dieser Operation.

Sei \mathcal{E} die Menge aller Isomorphieklassen von elliptischen Kurven über \mathbb{C} . ◇

Die obige Folgerung kann so interpretiert werden, dass wir eine *injektive* Abbildung

$$\mathcal{L}/\mathbb{C}^\times \longrightarrow \mathcal{E}, \quad [\Lambda] \longmapsto (\text{Isomorphieklasse von } E_\Lambda)$$

bekommen. Der Uniformisierungssatz 17.1 sagt, dass das sogar eine *Bijektion* ist, dass es also zu jeder elliptischen Kurve ein passendes Gitter gibt. Das werden wir im nächsten Abschnitt beweisen.

DEF
 $E_\Lambda, \mathcal{L}, \mathcal{E}$

19. MODULFORMEN UND MODULFUNKTIONEN

Die grobe Idee für den Beweis der anderen Richtung des Uniformisierungssatzes besteht darin, dass wir zeigen wollen, dass $j(E_\Lambda)$ jede komplexe Zahl als Wert für ein geeignetes Gitter Λ annimmt. Dazu müssen wir aber ein wenig ausholen.

Wir führen zunächst die obere Halbebene ein.

19.1. Definition. Wir setzen $\mathbb{H} := \{z \in \mathbb{C} \mid \text{Im}(z) > 0\}$ und nennen \mathbb{H} die *obere Halbebene*.

DEF
obere
Halbebene

Ist $\tau \in \mathbb{H}$, dann sind 1 und τ \mathbb{R} -linear unabhängig, also ist $\mathbb{Z} + \mathbb{Z}\tau$ ein Gitter.

19.2. Lemma. Die Abbildung

$$\mathbb{H} \longrightarrow \mathcal{L} \longrightarrow \mathcal{L}/\mathbb{C}^\times, \quad \tau \longmapsto \mathbb{Z} + \mathbb{Z}\tau \longmapsto [\mathbb{Z} + \mathbb{Z}\tau]$$

ist surjektiv.

LEMMA
Parametri-
sierung
der Gitter

Beweis. Sei $\Lambda \subset \mathbb{C}$ ein Gitter. Wir müssen zeigen, dass es $\alpha \in \mathbb{C}^\times$ und $\tau \in \mathbb{H}$ gibt mit $\Lambda = \alpha(\mathbb{Z} + \mathbb{Z}\tau)$. Dazu sei ω_1, ω_2 eine Basis von Λ . Nach eventuellem Ersetzen von ω_2 durch $-\omega_2$ können wir annehmen, dass $\text{Im}(\omega_2/\omega_1) > 0$ ist (da ω_1 und ω_2 \mathbb{R} -linear unabhängig sind, ist der Imaginärteil nicht null). Dann tun es $\alpha = \omega_1$ und $\tau = \omega_2/\omega_1$. \square

Die obige Abbildung ist allerdings nicht injektiv, denn die Wahl einer anderen Basis von Λ im Beweis wird in der Regel zu einem anderen $\tau \in \mathbb{H}$ führen. Genauer gilt Folgendes:

19.3. Lemma. Seien $\tau, \tau' \in \mathbb{H}$. Es gilt genau dann $[\mathbb{Z} + \mathbb{Z}\tau] = [\mathbb{Z} + \mathbb{Z}\tau']$, wenn es $a, b, c, d \in \mathbb{Z}$ gibt mit $ad - bc = 1$, sodass

LEMMA
äquivalente
Gitter

$$\tau' = \frac{a\tau + b}{c\tau + d}$$

ist.

Beweis. Die Aussage „ $[\mathbb{Z} + \mathbb{Z}\tau] = [\mathbb{Z} + \mathbb{Z}\tau']$ “ bedeutet, dass es $\alpha \in \mathbb{C}^\times$ gibt, sodass $\alpha, \alpha\tau'$ eine Basis von $\mathbb{Z} + \mathbb{Z}\tau$ ist. Die Basen von $\mathbb{Z} + \mathbb{Z}\tau$ haben die Form $a\tau + b, c\tau + d$ mit $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{GL}(2, \mathbb{Z})$, also $a, b, c, d \in \mathbb{Z}$ und $ad - bc = \pm 1$.

„ \Rightarrow “: Aus der obigen Überlegung ergibt sich $\tau' = (a\tau + b)/(c\tau + d)$. Es bleibt zu zeigen, dass $ad - bc = 1$ ist und nicht -1 . Dazu betrachten wir den Imaginärteil von τ' :

$$\begin{aligned} 0 < \text{Im}(\tau') &= \text{Im}\left(\frac{a\tau + b}{c\tau + d}\right) = \frac{\text{Im}((a\tau + b)(c\bar{\tau} + d))}{|c\tau + d|^2} \\ &= \frac{\text{Im}(ac|\tau|^2 + ad\tau + bc\bar{\tau} + bd)}{|c\tau + d|^2} = \frac{(ad - bc) \text{Im}(\tau)}{|c\tau + d|^2} \end{aligned}$$

wegen $\text{Im}(\tau) > 0$ folgt $ad - bc > 0$.

„ \Leftarrow “: Mit $\alpha = c\tau + d$ ist $\alpha, \alpha\tau'$ eine Basis von $\mathbb{Z} + \mathbb{Z}\tau$. \square

Die Rechnung zum Imaginärteil zeigt, dass für $\tau \in \mathbb{H}$ und $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{SL}(2, \mathbb{Z})$ auch $(a\tau + b)/(c\tau + d) \in \mathbb{H}$ ist. Wie man leicht nachprüft, erhält man so eine Operation von $\text{SL}(2, \mathbb{Z})$ von links auf \mathbb{H} . Dabei operiert $-I = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ trivial.

19.4. **Definition.** Wir definieren die *Modulgruppe* Γ als

$$\Gamma = \frac{\mathrm{SL}(2, \mathbb{Z})}{\langle -I \rangle}.$$

DEF
Modulgruppe
 Γ

Wir schreiben die Elemente von Γ als Matrizen, müssen dabei aber beachten, dass zwei Elemente von Γ auch dann gleich sind, wenn sich die Matrizen durch Multiplikation mit -1 voneinander unterscheiden.

Wir setzen außerdem

$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in \Gamma \quad \text{und} \quad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \Gamma. \quad \diamond$$

Es gilt dann $S^2 = (ST)^3 = -I = 1$ in Γ .

Die Aussagen von Lemma 19.2 und Lemma 19.3 kann man dann wie folgt zusammenfassen:

19.5. **Folgerung.** Die Abbildung aus Lemma 19.2 induziert eine Bijektion

$$\Gamma \backslash \mathbb{H} \longrightarrow \mathcal{L} / \mathbb{C}^\times.$$

FOLG
 $\mathcal{L} / \mathbb{C}^\times = \Gamma \backslash \mathbb{H}$

Wir betrachten jetzt Funktionen $\mathcal{L} \rightarrow \mathbb{C}$, die „homogen“ sind.

19.6. **Definition.** Sei $k \in \mathbb{Z}$. Eine Funktion $f: \mathcal{L} \rightarrow \mathbb{C}$ heißt eine *Gitterfunktion vom Gewicht k* , wenn für alle $\Lambda \in \mathcal{L}$ und alle $\alpha \in \mathbb{C}^\times$ gilt $f(\alpha\Lambda) = \alpha^{-k}f(\Lambda)$. \diamond

DEF
Gitterfunktion

Eine Gitterfunktion von ungeradem Gewicht muss die Nullfunktion sein, da $f(\Lambda) = f(-\Lambda) = (-1)^k f(\Lambda)$ ist.

19.7. **Beispiel.** Für $k \geq 4$ gerade ist G_k eine Gitterfunktion vom Gewicht k , denn

BSP
Gitter-
funktionen

$$G_k(\alpha\Lambda) = \sum_{0 \neq \omega \in \alpha\Lambda} \omega^{-k} = \sum_{0 \neq \omega \in \Lambda} (\alpha\omega)^{-k} = \alpha^{-k} \sum_{0 \neq \omega \in \Lambda} \omega^{-k} = \alpha^{-k} G_k(\Lambda). \quad \clubsuit$$

Jede Gitterfunktion f vom Gewicht k induziert eine Funktion \tilde{f} auf der oberen Halbebene mittels $\tilde{f}(\tau) = f(\mathbb{Z} + \mathbb{Z}\tau)$.

19.8. **Lemma.** Für jedes $k \in \mathbb{Z}$ ist die Abbildung $f \mapsto \tilde{f}$ eine Bijektion

LEMMA
Gitter-
funktionen

$$\begin{aligned} &\{\text{Gitterfunktionen vom Gewicht } k\} \\ &\longrightarrow \{\phi: \mathbb{H} \rightarrow \mathbb{C} \mid \forall \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma \forall \tau \in \mathbb{H}: \phi\left(\frac{a\tau+b}{c\tau+d}\right) = (c\tau+d)^k \phi(\tau)\}. \end{aligned}$$

Beweis. Sei zunächst f eine Gitterfunktion vom Gewicht k . Für $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$ und $\tau \in \mathbb{H}$ ist dann

$$\begin{aligned} \tilde{f}\left(\frac{a\tau+b}{c\tau+d}\right) &= f\left(\mathbb{Z} + \mathbb{Z}\frac{a\tau+b}{c\tau+d}\right) = f\left(\frac{1}{c\tau+d}(\mathbb{Z}(c\tau+d) + \mathbb{Z}(a\tau+b))\right) \\ &= f\left(\frac{1}{c\tau+d}(\mathbb{Z} + \mathbb{Z}\tau)\right) = (c\tau+d)^k f(\mathbb{Z} + \mathbb{Z}\tau) = (c\tau+d)^k \tilde{f}(\tau). \end{aligned}$$

Sei jetzt umgekehrt ϕ ein Element der rechten Menge. Sei ω_1, ω_2 eine Basis eines Gitters Λ mit $\text{Im}(\omega_2/\omega_1) > 0$. Wir zeigen, dass es eine eindeutige Gitterfunktion f vom Gewicht k gibt mit $\tilde{f} = \phi$. Für ein solches f muss gelten

$$f(\Lambda) = f(\mathbb{Z}\omega_1 + \mathbb{Z}\omega_2) = f\left(\omega_1\left(\mathbb{Z} + \mathbb{Z}\frac{\omega_2}{\omega_1}\right)\right) = \omega_1^{-k} \tilde{f}\left(\frac{\omega_2}{\omega_1}\right) = \omega_1^{-k} \phi\left(\frac{\omega_2}{\omega_1}\right).$$

Wir müssen also nachweisen, dass diese Definition sinnvoll ist, also nur von Λ und nicht von der Wahl der Basis abhängt. Sei also ω'_1, ω'_2 eine weitere Basis von Λ mit $\text{Im}(\omega'_2/\omega'_1) > 0$. Dann gibt es $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{SL}(2, \mathbb{Z})$ mit

$$\omega'_1 = d\omega_1 + c\omega_2, \quad \omega'_2 = b\omega_1 + a\omega_2.$$

Es ist dann

$$\begin{aligned} \omega'^{-k}_1 \phi\left(\frac{\omega'_2}{\omega'_1}\right) &= (d\omega_1 + c\omega_2)^{-k} \phi\left(\frac{b\omega_1 + a\omega_2}{d\omega_1 + c\omega_2}\right) \\ &= \omega_1^{-k} \left(c\frac{\omega_2}{\omega_1} + d\right)^{-k} \phi\left(\frac{a\frac{\omega_2}{\omega_1} + b}{c\frac{\omega_2}{\omega_1} + d}\right) = \omega_1^{-k} \phi\left(\frac{\omega_2}{\omega_1}\right). \end{aligned}$$

Also ist f wohldefiniert. □

Ist f eine Gitterfunktion (egal welchen Gewichts), dann gilt

$$\tilde{f}(\tau + 1) = \tilde{f}(T \cdot \tau) = \tilde{f}(\tau) \quad \text{für alle } \tau \in \mathbb{H}.$$

Sei $\mathbb{D} = B(0, 1) \subset \mathbb{C}$ die offene Einheitskreisscheibe und $\mathbb{D}' = B(0, 1) \setminus \{0\}$. Wir schreiben $q(\tau) = e^{2\pi i \tau}$. Dann bildet q die obere Halbebene \mathbb{H} auf \mathbb{D}' ab.

19.9. Lemma. *Wir haben die folgende Bijektion:*

$$\begin{aligned} \{\phi: \mathbb{D}' \rightarrow \mathbb{C}\} &\longrightarrow \{\psi: \mathbb{H} \rightarrow \mathbb{C} \mid \forall \tau \in \mathbb{H}: \psi(\tau + 1) = \psi(\tau)\} \\ \phi &\longmapsto \phi \circ q \end{aligned}$$

LEMMA
periodische
Funktionen

Wir erhalten ebenfalls eine Bijektion, wenn wir beide Mengen auf holomorphe Funktionen einschränken.

Beweis. Wegen $q(\tau + 1) = e^{2\pi i \tau} e^{2\pi i} = q(\tau)$ hat $\phi \circ q$ die geforderte Periodizität; die Abbildung ist also wohldefiniert, und da q holomorph ist, bildet sie holomorphe Funktionen auf holomorphe Funktionen ab. Sei umgekehrt ψ aus der rechten Menge. Wir wollen zeigen, dass ψ ein eindeutiges Urbild ϕ hat. Für so ein ϕ muss gelten $\phi(e^{2\pi i \tau}) = \psi(\tau)$, also kann es höchstens ein Urbild geben. Wie oben zeigen wir, dass ϕ durch diese Relation sinnvoll definiert werden kann. Seien also $\tau, \tau' \in \mathbb{H}$ mit $q(\tau) = q(\tau')$. Dann ist $\tau' - \tau \in \mathbb{Z}$, woraus $\psi(\tau') = \psi(\tau)$ folgt. Ist $z_0 \in \mathbb{D}'$, dann gibt es in einer Umgebung von z_0 einen holomorphen Logarithmus \log . In dieser Umgebung gilt dann $\phi(z) = \psi((2\pi i)^{-1} \log z)$. Das zeigt, dass ϕ holomorph ist, wenn ψ holomorph ist. □

Insbesondere gibt es also zu jeder Gitterfunktion f eine Funktion $\phi: \mathbb{D}' \rightarrow \mathbb{C}$, sodass $\tilde{f} = \phi \circ q$ ist.

19.10. Definition. Sei $k \in \mathbb{Z}$. Eine *Modulform vom Gewicht k* ist eine Funktion $\psi: \mathbb{H} \rightarrow \mathbb{C}$, sodass $\psi = \tilde{f}$ ist für eine Gitterfunktion f vom Gewicht k und sodass die Funktion $\phi: \mathbb{D}' \rightarrow \mathbb{C}$ mit $\psi = \phi \circ q$ zu einer holomorphen Funktion auf \mathbb{D} fortgesetzt werden kann. ψ heißt eine *Spitzenform*, wenn die Fortsetzung von ϕ in 0 verschwindet.

DEF
Modulform
Spitzenform

Wir schreiben $\mathcal{M}(k)$ für den \mathbb{C} -Vektorraum der Modulformen vom Gewicht k und $\mathcal{S}(k)$ für den Untervektorraum der Spitzenformen vom Gewicht k . \diamond

19.11. Bemerkungen.

BEM

- (1) Nach dem Riemannsches Hebbarkeitssatz ist die Bedingung an ϕ dazu äquivalent, dass ψ holomorph ist und $\psi(\tau)$ für $\text{Im}(\tau) \rightarrow \infty$ beschränkt bleibt (bzw. $\lim_{\text{Im}(\tau) \rightarrow \infty} \psi(\tau) = 0$ ist für Spitzenformen).
- (2) Eine auf \mathbb{D} holomorphe Funktion hat eine auf ganz \mathbb{D} konvergente Taylorreihe in 0. Es ist also

$$\psi(\tau) = \sum_{n=0}^{\infty} a_n q^n$$

mit $a_n \in \mathbb{C}$, wobei $q = q(\tau) = e^{2\pi i \tau}$ ist. Diese Reihenentwicklung heißt die *q-Entwicklung* von ψ . Die zusätzliche Bedingung für eine Spitzenform ist dann $a_0 = 0$. Wir schreiben auch $\psi(i\infty)$ für den „Wert“ a_0 von ψ im Grenzübergang $\text{Im}(\tau) \rightarrow \infty$.

DEF
q-Entwicklung
 $\psi(i\infty)$

- (3) Das Produkt einer Modulform vom Gewicht k und einer Modulform vom Gewicht k' ist eine Modulform vom Gewicht $k + k'$. Ist eine der beiden Formen eine Spitzenform, dann gilt das auch für das Produkt. \spadesuit

19.12. Beispiele. Sei $k \geq 4$ gerade. Die Funktion \tilde{G}_k ist eine Modulform vom Gewicht k ; sie heißt die *k-te Eisenstein-Reihe*. Um das zu sehen, müssen wir noch zeigen, dass \tilde{G}_k holomorph ist und für $\text{Im}(\tau) \rightarrow \infty$ beschränkt bleibt. Es ist

BSP
Eisenstein-
Reihen

$$\begin{aligned} \tilde{G}_k(\tau) &= G_k(\mathbb{Z} + \mathbb{Z}\tau) = \sum_{m,n \in \mathbb{Z}, (m,n) \neq (0,0)} \frac{1}{(m + n\tau)^k} \\ &= 2 \sum_{m=1}^{\infty} \frac{1}{m^k} + 2 \sum_{n=1}^{\infty} \sum_{m=-\infty}^{\infty} \frac{1}{(m + n\tau)^k}. \end{aligned}$$



G. Eisenstein
1823–1852

Es ist $|m + n\tau|^2 = (m + n \text{Re}(\tau))^2 + n^2 \text{Im}(\tau)^2$. Für $\text{Im}(\tau) \geq \varepsilon > 0$ gilt dann

$$\begin{aligned} \sum_{n=1}^{\infty} \sum_{m \in \mathbb{Z}} \frac{1}{|m + n\tau|^k} &\leq \sum_{n=1}^{\infty} \sum_{m \in \mathbb{Z}} \frac{1}{((m + n \text{Re}(\tau))^2 + n^2 \varepsilon^2)^{k/2}} \\ &\leq 2 \sum_{n=1}^{\infty} \sum_{m'=0}^{\infty} \frac{1}{(m'^2 + \varepsilon^2 n^2)^{k/2}} < \infty, \end{aligned}$$

und wir sehen, dass die Reihe absolut und gleichmäßig für $\text{Im}(\tau) \geq \varepsilon$ konvergiert. (Dabei ist $m' = \lfloor m + n \text{Re}(\tau) \rfloor$ für $m + n \text{Re}(\tau) \geq 0$ und $m' = \lfloor -(m + n \text{Re}(\tau)) \rfloor$ sonst.) Es folgt

$$\lim_{\text{Im}(\tau) \rightarrow \infty} \tilde{G}_k(\tau) = 2 \sum_{m=1}^{\infty} \frac{1}{m^k} + 2 \sum_{n=1}^{\infty} \sum_{m=-\infty}^{\infty} \lim_{\text{Im}(\tau) \rightarrow \infty} \frac{1}{(m + n\tau)^k} = 2\zeta(k),$$

denn die Terme der zweiten Summe konvergieren gegen null. Hier ist $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$ die Riemannsche Zetafunktion. Insbesondere ist

$$\tilde{G}_4(i\infty) = 2\zeta(4) = \frac{\pi^4}{45} \quad \text{und} \quad \tilde{G}_6(i\infty) = 2\zeta(6) = \frac{2\pi^6}{945}.$$

Da $\zeta(k) > 1$ ist, ist \tilde{G}_k keine Spitzenform. Es folgt (für $k \geq 4$ gerade)

$$\mathcal{M}(k) = \mathbb{C}\tilde{G}_k \oplus \mathcal{S}(k). \quad \clubsuit$$

19.13. Beispiel. Die Diskriminante von $E_{\mathbb{Z}+\mathbb{Z}\tau}$ ist

BSP
 Δ

$$\Delta(\tau) = -16(4(-15\tilde{G}_4(\tau))^3 + 27(-35\tilde{G}_6(\tau))^2) = 10\,800(20\tilde{G}_4(\tau)^3 - 49\tilde{G}_6(\tau)^2).$$

Δ ist eine Spitzenform vom Gewicht 12, denn \tilde{G}_4^3 und \tilde{G}_6^2 sind Modulformen vom Gewicht 12, und

$$\Delta(i\infty) = 10\,800\pi^{12} \left(\frac{20}{45^3} - \frac{49 \cdot 2^2}{945^2} \right) = 0.$$

Da $E_{\mathbb{Z}+\mathbb{Z}\tau}$ eine elliptische Kurve ist, verschwindet Δ auf \mathbb{H} nicht. \clubsuit

Wir wollen jetzt die q -Entwicklung von \tilde{G}_k bestimmen. Dazu brauchen wir das folgende Resultat aus der Funktionentheorie:

19.14. Lemma. Für $z \in \mathbb{C} \setminus \mathbb{Z}$ konvergiert die Reihe

LEMMA
 $\pi \cot \pi z$

$$\frac{1}{z} + \sum_{m=1}^{\infty} \left(\frac{1}{z-m} + \frac{1}{z+m} \right)$$

absolut und lokal gleichmäßig gegen $\pi \cot \pi z$.

Beweis. Der Term unter der Summe ist $(2z)/(z^2 - m^2)$; für z in einer kompakten Teilmenge von $\mathbb{C} \setminus \mathbb{Z}$ wird die Reihe also von $c \sum_{m=1}^{\infty} m^{-2}$ majorisiert. Das zeigt die absolute und lokal gleichmäßige Konvergenz. Die Grenzfunktion ist somit eine meromorphe Funktion auf \mathbb{C} mit einfachen Polen mit Residuum 1 genau bei allen ganzen Zahlen. Die Funktion $z \mapsto \pi \cot \pi z$ hat dieselben Eigenschaften. Die Differenz ist also eine ganze Funktion. Man prüft nach, dass beide Funktionen für $|\operatorname{Im}(z)| \rightarrow \infty$ beschränkt bleiben; das gilt dann auch für die Differenz. Nach dem Satz von Liouville muss die Differenz konstant sein. Da beide Funktionen ungerade sind, ist die Differenz null. \square

19.15. Folgerung. Für $k \geq 2$ ist die q -Entwicklung der periodischen Funktion

FOLG

$$z \mapsto \sum_{m=-\infty}^{\infty} \frac{1}{(z-m)^k}$$

auf \mathbb{H} gegeben durch

$$\frac{(-2\pi i)^k}{(k-1)!} \sum_{n=1}^{\infty} n^{k-1} q^n.$$

Beweis. Wegen der lokal gleichmäßigen Konvergenz dürfen wir die Reihe für $\pi \cot \pi z$ beliebig oft termweise differenzieren. Nach dem ersten Ableiten können wir die Terme $-(z - m)^{-2}$ und $-(z + m)^{-2}$ beliebig umordnen, da die resultierende Reihe absolut konvergiert. Das ergibt mit $dq = 2\pi i q dz$

$$\begin{aligned} \sum_{m=-\infty}^{\infty} \frac{1}{(z - m)^k} &= \frac{(-1)^{k-1}}{(k-1)!} \left(\frac{d}{dz}\right)^{k-1} \pi \cot \pi z = \frac{(-1)^k \pi i}{(k-1)!} \left(\frac{d}{dz}\right)^{k-1} \frac{1+q}{1-q} \\ &= \frac{(-1)^k \pi i}{(k-1)!} \left(\frac{d}{dz}\right)^{k-1} \left(1 + 2 \sum_{n=1}^{\infty} q^n\right) = \frac{(-2\pi i)^k}{(k-1)!} \left(q \frac{d}{dq}\right)^{k-1} \sum_{n=1}^{\infty} q^n \\ &= \frac{(-2\pi i)^k}{(k-1)!} \sum_{n=1}^{\infty} n^{k-1} q^n. \quad \square \end{aligned}$$

19.16. Folgerung. Für $k \geq 4$ gerade ist

$$\tilde{G}_k(\tau) = 2\zeta(k) \left(1 + C_k \sum_{n=1}^{\infty} \sigma_{k-1}(n) q^n\right),$$

FOLG
 q -Entwicklung
 der Eisenstein-
 Reihen

wobei

$$C_k = \frac{(-1)^{k/2} (2\pi)^k}{\zeta(k) (k-1)!}$$

und

$$\sigma_m(n) = \sum_{d|n} d^m$$

die Summe der m -ten Potenzen der Teiler von n ist.

Beweis. Wir rechnen unter Verwendung von Folgerung 19.15:

$$\begin{aligned} \tilde{G}_k(\tau) &= 2\zeta(k) + 2 \sum_{n=1}^{\infty} \sum_{m=-\infty}^{\infty} \frac{1}{(n\tau + m)^k} \\ &= 2\zeta(k) \left(1 + \frac{1}{\zeta(k)} \frac{(-2\pi i)^k}{(k-1)!} \sum_{n=1}^{\infty} \sum_{l=1}^{\infty} l^{k-1} q^{ln}\right) \\ &= 2\zeta(k) \left(1 + \frac{(-1)^{k/2} (2\pi)^k}{\zeta(k) (k-1)!} \sum_{n=1}^{\infty} \left(\sum_{d|n} d^{k-1}\right) q^n\right) \\ &= 2\zeta(k) \left(1 + C_k \sum_{n=1}^{\infty} \sigma_{k-1}(n) q^n\right). \quad \square \end{aligned}$$

19.17. Bemerkung. Die Zahlen C_k sind von null verschiedene rationale Zahlen. Für $k \geq 2$ gerade ist nämlich

BEM
 Bernoulli-
 Zahlen

$$\zeta(k) = -(-1)^{k/2} \frac{(2\pi)^k}{2 \cdot k!} B_k,$$

wobei B_k die Bernoulli-Zahlen sind, die durch die Potenzreihenentwicklung

$$\frac{z}{e^z - 1} = \sum_{n=0}^{\infty} \frac{B_n}{n!} z^n$$

definiert werden können. Es ergibt sich

$$C_k = -\frac{2k}{B_k}. \quad \spadesuit$$

Die Aussage über die Werte der Riemannschen Zetafunktion an positiven geraden ganzen Zahlen lässt sich ebenfalls mithilfe der Funktion $\pi \cot \pi z$ beweisen. Für $k \geq 2$ gerade sei $f_k(z) = (\pi \cot \pi z)/z^k$. Für $n \in \mathbb{Z} \setminus \{0\}$ ist dann $\operatorname{res}_{z=n} f_k(z) = |n|^{-k}$. Sei $0 < \varepsilon < \frac{1}{2}$. Dann ist $|\pi \cot \pi z| \leq c(\varepsilon)$ für alle $z \in \mathbb{C}$ mit $\min\{|z - n| \mid n \in \mathbb{Z}\} \geq \varepsilon$; dabei ist $c(\varepsilon)$ eine nur von ε abhängige Konstante. Es folgt, dass f_k auf dem Kreis γ_N mit Radius $N + \frac{1}{2}$ um 0 (mit $N \in \mathbb{Z}_{>0}$) durch $c(N + \frac{1}{2})^{-k}$ beschränkt ist. Daraus bekommen wir

$$2\zeta(k) + \operatorname{res}_{z=0} f_k(z) = \lim_{N \rightarrow \infty} \sum_{n=-N}^N \operatorname{res}_{z=n} f_k(z) = \lim_{N \rightarrow \infty} \frac{1}{2\pi i} \int_{\gamma_N} f_k(z) dz = 0,$$

denn

$$\left| \int_{\gamma_N} f_k(z) dz \right| \leq 2\pi(N + \frac{1}{2}) \cdot c(\frac{1}{2})(N + \frac{1}{2})^{-k} \xrightarrow{N \rightarrow \infty} 0.$$

Es folgt $\zeta(k) = -\frac{1}{2} \operatorname{res}_{z=0} (\pi \cot \pi z)/z^k$. Das Residuum ist gerade der Koeffizient von z^{k-1} in der Laurent-Reihe von $\pi \cot \pi z$ um 0, also π^k mal der Koeffizient von z^{k-1} in der Laurent-Reihe von \cot . Nun ist

$$\cot z = \frac{\cos z}{\sin z} = \frac{\frac{1}{2}(e^{iz} + e^{-iz})}{\frac{1}{2i}(e^{iz} - e^{-iz})} = i \frac{e^{2iz} + 1}{e^{2iz} - 1} = i + \frac{1}{z} \cdot \frac{2iz}{e^{2iz} - 1} = i + \sum_{n=0}^{\infty} \frac{B_n}{n!} (2i)^n z^{n-1}.$$

Es folgt

$$\zeta(k) = -\frac{1}{2} \pi^k \frac{B_k}{k!} (2i)^k = -\frac{(2\pi i)^k B_k}{2k!}.$$

19.18. **Beispiele.** Aus dem Obigen erhalten wir:

$$\begin{aligned} \tilde{G}_4(\tau) &= \frac{\pi^4}{45} \left(1 + 240 \sum_{n=1}^{\infty} \sigma_3(n) q^n \right) = \frac{\pi^4}{45} (1 + 240(q + 9q^2 + 28q^3 + \dots)) \\ \tilde{G}_6(\tau) &= \frac{2\pi^6}{945} \left(1 - 504 \sum_{n=1}^{\infty} \sigma_5(n) q^n \right) = \frac{2\pi^6}{945} (1 - 504(q + 33q^2 + 244q^3 + \dots)) \end{aligned}$$

BSP
 q -Entwicklung
 von \tilde{G}_4 , \tilde{G}_6
 und Δ

Daraus ergibt sich

$$\begin{aligned} \Delta(\tau) &= 10\,800(20\tilde{G}_4(\tau)^3 - 49\tilde{G}_6(\tau)^2) \\ &= (2\pi)^{12} \left(\frac{1}{12^3} (1 + 240q + \dots)^3 - \frac{1}{12^3} (1 - 504q - \dots)^2 \right) \\ &= (2\pi)^{12} (q - 24q^2 + \dots). \end{aligned} \quad \clubsuit$$

19.19. **Bemerkung.** Für die q -Entwicklung von Δ gilt die folgende schöne Produktformel:

BEM
 Produktformel
 für Δ

$$\Delta(\tau) = (2\pi)^{12} q \prod_{n=1}^{\infty} (1 - q^n)^{24}.$$

Wir werden sie allerdings hier nicht beweisen. ♠

Wir können die j -Invariante der elliptischen Kurve $E_{\mathbb{Z}+\mathbb{Z}\tau}$ betrachten. Sie liefert uns eine holomorphe Funktion

$$j: \mathbb{H} \rightarrow \mathbb{C},$$

für die $j(\gamma \cdot \tau) = j(\tau)$ gilt für alle $\gamma \in \Gamma$ und $\tau \in \mathbb{H}$. Sie erfüllt die Bedingungen für eine „Modulform vom Gewicht null“ mit einer Ausnahme: Die induzierte Funktion auf \mathbb{D} ist nicht holomorph, sondern nur meromorph mit einem einfachen Pol in 0. Für die q -Entwicklung ergibt sich nämlich

$$j(\tau) = \frac{(720\tilde{G}_4(\tau))^3}{\Delta(\tau)} = \frac{(1 + 240q + 2160q^2 + \dots)^3}{q - 24q^2 + 252q^3 + \dots} = q^{-1} + 744 + 196884q + \dots$$

19.20. Definition. Eine *Modulfunktion* ist eine meromorphe Funktion f auf \mathbb{H} , die unter der Operation von Γ invariant ist, also $f(\gamma \cdot \tau) = f(\tau)$ für alle $\gamma \in \Gamma$, $\tau \in \mathbb{H}$ erfüllt, und zusätzlich die Eigenschaft hat, dass die entsprechende Funktion auf \mathbb{D}' höchstens einen Pol in 0 hat (also auf \mathbb{D} meromorph ist). \diamond

DEF
Modulfunktion

Sind etwa f und $g \neq 0$ Modulformen von selben Gewicht, dann ist f/g eine Modulfunktion.

19.21. Beispiel. Wie wir gerade gesehen haben, ist j eine Modulfunktion. \clubsuit

BSP
Modulfunktion

Wir zeigen nun, dass j auf \mathbb{H} jeden Wert aus \mathbb{C} annimmt. Daraus folgt dann der zweite Teil des Uniformisierungssatzes 17.1.

19.22. Satz. Sei $j_0 \in \mathbb{C}$. Dann gibt es ein $\tau \in \mathbb{H}$ mit $j(\tau) = j_0$.

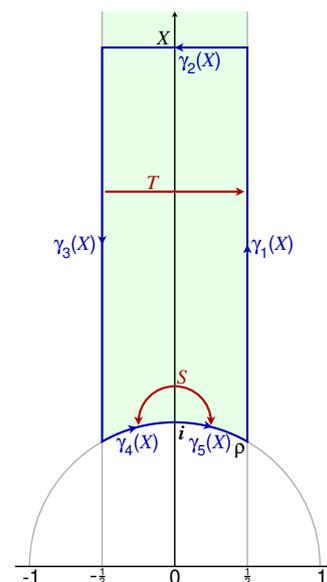
SATZ
 j ist surjektiv

Beweis. Sei X eine große reelle Zahl. Wir werden später $X \rightarrow \infty$ gehen lassen. Sei $\rho = \frac{1}{2}(1 + \sqrt{3}i)$. Sei weiter $\gamma(X)$ die geschlossene Kurve, die aus den folgenden Stücken besteht:

- $\gamma_1(X)$: $[\rho, \frac{1}{2} + Xi]$
- $\gamma_2(X)$: $[\frac{1}{2} + Xi, -\frac{1}{2} + Xi]$
- $\gamma_3(X)$: $[-\frac{1}{2} + Xi, \rho - 1]$
- $\gamma_4(X)$: Einheitskreis im Uhrzeigersinn von $\rho - 1$ bis i
- $\gamma_5(X)$: Einheitskreis im Uhrzeigersinn von i bis ρ

Falls $j(\tau) - j_0$ irgendwo auf γ verschwindet, dann sind wir fertig. Wir können also annehmen, dass das nicht der Fall ist. Dann ist die Anzahl der j_0 -Stellen von j (mit Vielfachheit) im von $\gamma(X)$ umschlossenen Gebiet gegeben durch

$$\frac{1}{2\pi i} \int_{\gamma(X)} \frac{j'(\tau)}{j(\tau) - j_0} d\tau.$$



Wir werden jetzt den Wert dieses Integrals im Limes $X \rightarrow \infty$ berechnen. Aus $j(\tau + 1) = j(\tau)$ und daraus, dass $\gamma_1(X)$ und $\gamma_3(X)$ entgegengesetzt orientiert sind, folgt

$$\int_{\gamma_1(X)} \frac{j'(\tau)}{j(\tau) - j_0} d\tau + \int_{\gamma_3(X)} \frac{j'(\tau)}{j(\tau) - j_0} d\tau = 0.$$

Außerdem gilt $j(-1/\tau) = j(\tau)$. Es folgt $j'(-1/\tau) = \tau^2 j'(\tau)$. $S: \tau \mapsto -1/\tau$ bildet $\gamma_4(X)$ auf die entgegengesetzt orientierte Kurve $\gamma_5(X)$ ab. Daraus ergibt sich:

$$\begin{aligned} \int_{\gamma_4(X)} \frac{j'(\tau)}{j(\tau) - j_0} d\tau &= - \int_{\gamma_5(X)} \frac{j'(-1/\tau)}{j(-1/\tau) - j_0} d(-1/\tau) \\ &= - \int_{\gamma_5(X)} \frac{\tau^2 j'(\tau)}{j(\tau) - j_0} \tau^{-2} d\tau = - \int_{\gamma_5(X)} \frac{j'(\tau)}{j(\tau) - j_0} d\tau; \end{aligned}$$

damit heben sich die Integrale über $\gamma_4(X)$ und $\gamma_5(X)$ ebenfalls weg. Es bleibt das Integral über $\gamma_2(X)$. Wir schreiben J für die meromorphe Funktion auf \mathbb{D} mit $j = J \circ q$. Dann ist

$$\begin{aligned} \frac{1}{2\pi i} \int_{\gamma_2(X)} \frac{j'(\tau)}{j(\tau) - j_0} d\tau &= - \frac{1}{2\pi i} \int_{\partial B(0, e^{-2\pi X})} \frac{2\pi i q J'(q)}{J(q) - j_0} \frac{dq}{2\pi i q} = - \frac{1}{2\pi i} \int_{\partial B(0, e^{-2\pi X})} \frac{J'(q)}{J(q) - j_0} dq \\ &= \#\{\text{Pole von } J \text{ in } B(0, e^{-2\pi X})\} - \#\{\text{Nullstellen von } J - j_0 \text{ in } B(0, e^{-2\pi X})\}. \end{aligned}$$

Wir wissen, dass J einen einfachen Pol in 0 hat. Ist X hinreichend groß, dann liegen keine j_0 -Stellen von J in $B(0, e^{-2\pi X})$, also ist der Wert des Integrals 1. Also muss es (genau) ein τ im Inneren der (oben in der Grafik hellgrün hervorgehobenen) Menge

$$\mathcal{F} = \{\tau \in \mathbb{H} \mid |\operatorname{Re}(\tau)| \leq \frac{1}{2}, |\tau| \geq 1\}$$

geben, sodass $j(\tau) = j_0$ ist. □

19.23. **Folgerung.** Sei

$$E: y^2 = x^3 + ax + b$$

eine elliptische Kurve über \mathbb{C} . Dann gibt es ein Gitter Λ , sodass $E = E_\Lambda$ ist.

FOLG
Gitter zu
ell. Kurve

Beweis. Nach Satz 19.22 gibt es $\tau \in \mathbb{H}$ mit $j(\tau) = j(E)$. Sei $E' = E_{\mathbb{Z} + \mathbb{Z}\tau}$, also

$$E': y^2 = x^3 - 15\tilde{G}_4(\tau)x - 35\tilde{G}_6(\tau).$$

Dann ist $j(E) = j(\tau) = j(E')$. Nach Satz 8.5 sind E und E' isomorph. Ebenfalls nach Satz 8.5 gibt es $\alpha \in \mathbb{C}^\times$ mit

$$\alpha^4 a = -15\tilde{G}_4(\tau) \quad \text{und} \quad \alpha^6 b = -35\tilde{G}_6(\tau).$$

Sei $\Lambda = \mathbb{Z}\alpha + \mathbb{Z}\alpha\tau$. Dann ist

$$G_4(\Lambda) = \alpha^{-4}G_4(\mathbb{Z} + \mathbb{Z}\tau) \quad \text{und} \quad G_6(\Lambda) = \alpha^{-6}G_6(\mathbb{Z} + \mathbb{Z}\tau),$$

also ist $a = -15G_4(\Lambda)$ und $b = -35G_6(\Lambda)$ und damit $E = E_\Lambda$. □

19.24. **Bemerkung.** Da die j -Invarianten von zwei elliptischen Kurven über \mathbb{C} genau dann übereinstimmen, wenn die Kurven isomorph sind, folgt

$$\forall \tau, \tau' \in \mathbb{H}: (j(\tau) = j(\tau')) \iff \exists \gamma \in \Gamma: \gamma \cdot \tau = \tau'.$$

Aus dem Beweis von Satz 19.22 ergibt sich, dass es für jedes $\tau \in \mathbb{H}$ entweder ein $\tau' \in \partial \mathcal{F}$ gibt mit $\tau' \in \Gamma \cdot \tau$ oder *genau ein* $\tau' \in \mathcal{F}^\circ$ mit dieser Eigenschaft. Somit erfüllt \mathcal{F} die Definition eines *Fundamentaltbereichs* für die Operation von Γ auf \mathbb{H} : \mathcal{F} ist abgeschlossen, jedes $\tau \in \mathbb{H}$ liegt in der Γ -Bahn eines Punktes von \mathcal{F} und dieser Punkt ist eindeutig bestimmt, wenn er im Inneren von \mathcal{F} liegt. ♠

Es folgt, dass Γ von S und T erzeugt wird. Man kann nämlich jedes $\tau \in \mathbb{H}$ nur durch Anwenden von S, T und T^{-1} in den Fundamentaltbereich \mathcal{F} bringen. Um das zu sehen, überlegt man sich zunächst, dass die Menge $\{\text{Im}(\gamma \cdot \tau) \mid \gamma \in \Gamma\}$ beschränkt und diskret in $\mathbb{R}_{>0}$ ist. Das folgt aus der Formel $\text{Im}\left(\frac{a\tau+b}{c\tau+d}\right) = |c\tau+d|^{-2} \text{Im}(\tau)$ und daraus, dass die Menge der von null verschiedenen Gitterpunkte von $\mathbb{Z} + \mathbb{Z}\tau$ vom Nullpunkt weg beschränkt und diskret ist. Es gibt dann ein Element $\gamma_0 \in \langle S, T \rangle$ mit $\text{Im}(\gamma_0 \cdot \tau)$ maximal. Für $n \in \mathbb{Z}$ geeignet und $\gamma = T^n \gamma_0$ gilt dann $|\text{Re}(\gamma \cdot \tau)| \leq \frac{1}{2}$ (und $\text{Im}(\gamma \cdot \tau) = \text{Im}(\gamma_0 \cdot \tau)$). Wäre $|\gamma \cdot \tau| < 1$, dann hätten wir den Widerspruch

$$\text{Im}(S\gamma \cdot \tau) = |\gamma \cdot \tau|^{-2} \text{Im}(\gamma \cdot \tau) > \text{Im}(\gamma \cdot \tau).$$

Also ist $\gamma \cdot \tau \in \mathcal{F}$ mit $\gamma \in \langle S, T \rangle$.

Sei jetzt $\gamma \in \Gamma$ beliebig. Nach dem oben Gesagten gibt es $\gamma' \in \langle S, T \rangle$ mit $\gamma' \cdot (\gamma \cdot (2i)) \in \mathcal{F}$. Da $2i \in \mathcal{F}^\circ$ ist, muss $\gamma' \gamma \cdot (2i) = 2i$ sein. Sei $\gamma' \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Aus der Formel für den Imaginärteil folgt $c = 0$ und $|d| = 1$. Dann muss $\gamma' \gamma$ eine Potenz von T sein, und da T keine Fixpunkte hat, ist $\gamma' \gamma = 1$ und damit $\gamma = \gamma'^{-1} \in \langle S, T \rangle$.

Etwas konstruktiver kann man so vorgehen: Solange $\tau \notin \mathcal{F}$ ist, ersetzen wir zunächst τ durch $\tau' = T^n \cdot \tau$, sodass $|\text{Re}(\tau')| \leq \frac{1}{2}$ ist. Ist dann $|\tau'| < 1$, ersetzen wir τ durch $S \cdot \tau'$. Der Imaginärteil von τ wird bei jedem Durchlauf durch diese Schleife größer; daraus folgt, dass das Verfahren abbrechen muss.

19.25. **Bemerkung.** Seien

$$R_1 = \{z \in \mathbb{H} \mid \text{Re}(z) = \frac{1}{2}, \text{Im}(z) > \text{Im}(\rho)\} \quad \text{und} \\ R_2 = \{z \in \mathbb{H} \mid 0 < \text{Re}(z) < \frac{1}{2}, |z| = 1\}.$$

Mit einer Integration ähnlich wie im Beweis von Satz 19.22 kann man für Modulformen $0 \neq f \in \mathcal{M}(k)$ Folgendes zeigen:

$$\frac{1}{3} \text{ord}_\rho f + \frac{1}{2} \text{ord}_i f + \sum_{z \in R_1 \cup R_2 \cup \mathcal{F}^\circ} \text{ord}_z f + \text{ord}_{i_\infty} f = \frac{k}{12}.$$

Dabei ist $\text{ord}_{i_\infty} f$ als die Verschwindungsordnung in 0 der zugehörigen Funktion auf \mathbb{D} zu verstehen. Daraus kann man recht weit reichende Folgerungen über $\mathcal{M}(k)$ und $\mathcal{S}(k)$ ziehen:

- (1) $\mathcal{M}(k) = \{0\}$ für $k < 0$ und $\mathcal{M}(0) = \mathbb{C}$.
- (2) $\mathcal{M}(2) = \{0\}$, $\mathcal{M}(4) = \mathbb{C}\tilde{G}_4$, $\mathcal{M}(6) = \mathbb{C}\tilde{G}_6$, $\mathcal{M}(8) = \mathbb{C}\tilde{G}_4^2$, $\mathcal{M}(10) = \mathbb{C}\tilde{G}_4\tilde{G}_6$.
- (3) $\mathcal{S}(k) = \{0\}$ für $k < 12$ und $\mathcal{S}(12) = \mathbb{C}\Delta$.
- (4) $\phi_k: \mathcal{M}(k) \rightarrow \mathcal{S}(k+12)$, $f \mapsto f\Delta$, ist ein Isomorphismus von \mathbb{C} -Vektorräumen.
- (5) $\mathcal{M}(k)$ ist endlich-dimensional. Genauer gilt für $k \geq 0$ gerade

$$\dim \mathcal{M}(k) = \begin{cases} 1 + \left\lfloor \frac{k}{12} \right\rfloor, & k \not\equiv 2 \pmod{12}, \\ \left\lfloor \frac{k}{12} \right\rfloor, & k \equiv 2 \pmod{12}. \end{cases}$$

BEM
Fundamentalbereich von Γ

BEM
 $\mathcal{M}(k)$

- (6) Der Ring $\bigoplus_{k \geq 0} \mathcal{M}(k)$ ist isomorph zum Polynomring $\mathbb{C}[\tilde{G}_4, \tilde{G}_6]$ (genauer: Die offensichtliche Abbildung vom Polynomring in den Ring der Modulformen ist ein Isomorphismus).
- (7) Die Modulfunktionen sind genau die rationalen Funktionen in j . ♠

20. DIE RATIONALE TORSIONSUNTERGRUPPE

Im verbleibenden Teil der Vorlesung wollen wir uns mit der Gruppe $E(\mathbb{Q})$ der rationalen Punkte einer elliptischen Kurve über \mathbb{Q} befassen. Als ersten Schritt untersuchen wir die Torsionsuntergruppe $E(\mathbb{Q})_{\text{tors}}$ genauer. Im Folgenden sei

$$E: y^2 = x^3 + ax + b$$

eine elliptische Kurve, gegeben durch eine kurze Weierstraß-Gleichung mit Koeffizienten $a, b \in \mathbb{Z}$. Unser erstes Ziel ist zu zeigen, dass jeder nicht-triviale (also $\neq O$) Punkt endlicher Ordnung in $E(\mathbb{Q})$ ganzzahlige Koordinaten haben muss.

Ist $P = (\xi, \eta) \in E(\mathbb{Q})$ ein Punkt, sodass ξ oder η nicht ganzzahlig ist, dann gibt es eine Primzahl p mit $v_p(\xi) < 0$ oder $v_p(\eta) < 0$ (d.h., p teilt den Nenner einer der Koordinaten). Ist $v_p(\xi) < 0$, dann ist der Term ξ^3 in der rechten Seite $\xi^3 + a\xi + b$ der in P ausgewerteten Kurvengleichung der mit der kleinsten p -adischen Bewertung; es folgt

$$2v_p(\eta) = v_p(\eta^2) = v_p(\xi^3 + a\xi + b) = 3v_p(\xi),$$

und wir sehen, dass *beide* Bewertungen negativ sind und dass es $e \geq 1$ gibt mit $v_p(\xi) = -2e$, $v_p(\eta) = -3e$. Ist $v_p(\eta) < 0$, dann muss die Bewertung der rechten Seite ebenfalls negativ sein, woraus wieder $v_p(\xi) < 0$ folgt.

Der Punkt hat also Koordinaten, die „groß“ sind in der p -adischen Metrik (es ist $|\xi|_p = p^{-v_p(\xi)}$, also ist $|\xi|_p \geq p^2$, $|\eta|_p \geq p^3$), und liegt daher „nahe bei“ dem Punkt O im Unendlichen. Das motiviert uns, die Gleichung von E in einem affinen Teil zu betrachten, der den Punkt O enthält. Dafür bietet sich der affine Teil von \mathbb{P}^2 an, in dem $y \neq 0$ ist. Wir schreiben also unsere Punkte in der Form $(z : 1 : w)$ (also mit $z = x/y$ und $w = 1/y$). Die Gleichung von E hat in diesen Koordinaten die Form

$$E: w = z^3 + aw^2z + bw^3,$$

und der Punkt O hat die Koordinaten $(z, w) = (0, 0)$. Für einen Punkt, dessen Koordinaten durch p teilbare Nenner haben, haben wir dann

$$v_p(z(P)) = v_p(x(P)/y(P)) = v_p(x(P)) - v_p(y(P)) = -2e - (-3e) = e$$

und $v_p(w(P)) = v_p(1/y(P)) = 3e$ mit e wie oben. Ist das umgekehrt für einen Punkt P erfüllt, dann folgt aus der Gleichung, dass $v_p(w(P)) = 3v_p(z(P))$ ist, und wir bekommen daraus

$$\begin{aligned} v_p(y(P)) &= -v_p(w(P)) = -3v_p(z(P)) && \text{und} \\ v_p(x(P)) &= v_p(z(P)) + v_p(y(P)) = -2v_p(z(P)). \end{aligned}$$

Die uns interessierenden Punkte sind also genau die, für die (die Zähler von) $z(P)$ und $w(P)$ durch p teilbar sind.

20.1. Definition. Sei $E: y^2 = x^3 + ax + b$ wie oben und sei p eine Primzahl. Für $e \geq 1$ definieren wir

$$\begin{aligned} E_p^{(e)}(\mathbb{Q}) &= \{P \in E(\mathbb{Q}) \mid v_p(z(P)) \geq e, v_p(w(P)) \geq 3e\} \\ &= \{P \in E(\mathbb{Q}) \mid v_p(x(P)) \leq -2e, v_p(y(P)) \leq -3e\} \cup \{O\} \end{aligned}$$

und nennen $E_p^{(e)}(\mathbb{Q})$ den e -ten Kern der Reduktion modulo p von $E(\mathbb{Q})$. $E_p^{(1)}(\mathbb{Q})$ heißt auch einfach der Kern der Reduktion modulo p von $E(\mathbb{Q})$. \diamond

Der Grund für diese Bezeichnung wird später klar werden.

Wir beweisen zunächst ein Lemma. Wenn wir im Folgenden für rationale Zahlen α und β schreiben $\alpha \equiv \beta \pmod{p^e}$, dann meinen wir damit, dass $v_p(\alpha), v_p(\beta) \geq 0$

DEF
Kern der
Reduktion

sind und $v_p(\alpha - \beta) \geq p^e$ ist. (D.h., p teilt weder den Nenner von α noch den von β und p^e teilt den Zähler von $\alpha - \beta$.)

20.2. Lemma. Seien $P = (\zeta : 1 : \omega)$ und $P' = (\zeta' : 1 : \omega')$ zwei Punkte in $E_p^{(e)}(\mathbb{Q})$.

LEMMA
Punkte
in $E_p^{(e)}(\mathbb{Q})$

- (1) $\zeta' = \zeta \implies P' = P$.
- (2) Die Gerade durch P und P' (Tangente an E in P im Fall $P = P'$) hat eine Gleichung der Form $w = sz + t$ mit $v_p(s) \geq 2e$, $v_p(t) \geq 3e$.

Beweis. Wir bilden die Differenz aus den beiden Gleichungen

$$\omega' = \zeta'^3 + a\omega'^2\zeta' + b\omega'^3 \quad \text{und} \quad \omega = \zeta^3 + a\omega^2\zeta + b\omega^3$$

und stellen die Terme geeignet um. Das ergibt

$$(\omega' - \omega)(1 - a(\omega' + \omega)\zeta - b(\omega'^2 + \omega'\omega + \omega^2)) = (\zeta' - \zeta)(\zeta'^2 + \zeta'\zeta + \zeta^2 + a\omega'^2).$$

- (1) Ist $\zeta' = \zeta$, dann ist die rechte Seite null. Da der zweite Faktor auf der linken Seite p -adische Bewertung 0 hat und damit insbesondere $\neq 0$ ist, muss $\omega' = \omega$ sein.
- (2) Im Fall $P' \neq P$ ist nach (1) $\zeta' \neq \zeta$, und es folgt $s = (\omega' - \omega)/(\zeta' - \zeta)$. Wir erhalten

$$s \underbrace{(1 - a(\omega' + \omega)\zeta - b(\omega'^2 + \omega'\omega + \omega^2))}_{v_p=0} = \underbrace{(\zeta'^2 + \zeta'\zeta + \zeta^2 + a\omega'^2)}_{v_p \geq 2e}$$

und damit $v_p(s) \geq 2e$ und $v_p(t) = v_p(\omega - s\zeta) \geq 3e$.

Im Fall $P' = P$ erhalten wir durch implizites Differenzieren oder durch Grenzübergang im Ausdruck oben analog

$$s(1 - 2a\omega\zeta - 3b\omega^2) = 3\zeta^2 + a\omega^2$$

und können ebenso schließen. □

Der folgende Satz fasst die wichtigsten Eigenschaften von $E_p^{(e)}(\mathbb{Q})$ zusammen.

20.3. Satz. Sei $E: y^2 = x^3 + ax + b$ eine elliptische Kurve über \mathbb{Q} mit $a, b \in \mathbb{Z}$, sei p eine Primzahl und sei $e \in \mathbb{Z}_{\geq 1}$.

SATZ
Kern der
Reduktion

- (1) $E_p^{(e)}(\mathbb{Q})$ ist eine Untergruppe von $E(\mathbb{Q})$.
- (2) Ist $P \in E_p^{(1)}(\mathbb{Q})$ mit $v_p(z(P)) = e$ und ist $0 \neq m \in \mathbb{Z}$, dann gilt

$$v_p(z(mP)) = e + v_p(m).$$

- (3) $E_p^{(1)}(\mathbb{Q})_{\text{tors}} = \{O\}$.

Beweis.

- (1) Es ist klar, dass $O \in E_p^{(e)}(\mathbb{Q})$ ist. Wegen $-(z : 1 : w) = (-z : 1 : -w)$ ist $E_p^{(e)}(\mathbb{Q})$ unter Negation abgeschlossen. Seien jetzt $P = (\zeta : 1 : \omega)$ und $P' = (\zeta' : 1 : \omega')$ Punkte in $E_p^{(e)}(\mathbb{Q})$. Dann ist $P + P' = -P''$, wobei P'' der dritte Schnittpunkt der Geraden durch P und P' mit E ist. Nach Lemma 20.2

hat diese Gerade die Form $w = sz + t$ mit $v_p(s) \geq 2e$, $v_p(t) \geq 3e$. Einsetzen der Geradengleichung in die Gleichung zwischen w und z für E ergibt

$$(1 + as^2 + bs^3)z^3 + (2ast + 3bs^2t)z^2 + \dots = 0.$$

Ist $\zeta'' = z(P'')$ und $\omega'' = w(P'')$, dann folgt

$$\zeta + \zeta' + \zeta'' = -\frac{2ast + 3bs^2t}{1 + as^2 + bs^3},$$

also

$$\begin{aligned} v_p(\zeta'') &= v_p\left(-\frac{2ast + 3bs^2t}{1 + as^2 + bs^3} - \zeta - \zeta'\right) \\ &\geq \min\{v_p(2ast + 3bs^2t) - v_p(1 + as^2 + bs^3), v_p(\zeta), v_p(\zeta')\} \\ &\geq \min\{5e, e, e\} = e \end{aligned}$$

und dann auch $v_p(\omega'') = v_p(s\zeta'' + t) \geq 3e$, also $P'' \in E_p^{(e)}(\mathbb{Q})$ wie gewünscht.

(2) Zunächst folgt aus der Rechnung oben im Beweis von (1) die Kongruenz

$$z(P + P') = -\zeta'' \equiv z(P) + z(P') \pmod{p^{5e}}.$$

Induktion liefert dann $z(mP) \equiv mz(P) \pmod{p^{5e}}$. Wir zeigen die Behauptung jetzt durch Induktion über $k = v_p(m)$.

$k = 0$: Dann ist $v_p(z(mP)) = v_p(mz(P)) = e = e + k$.

$k > 0$: Es ist $z(pP) \equiv pz(P) \pmod{p^{5e}}$, also $v_p(z(pP)) = e + 1$. Aus der Induktionsvoraussetzung für pP ergibt sich dann

$$v_p(z(mP)) = v_p(z(\frac{m}{p} \cdot pP)) = v_p(z(pP)) + v_p(m/p) = (e + 1) + (k - 1) = e + k.$$

(3) Sei $P \in E_p^{(1)}(\mathbb{Q})_{\text{tors}}$. Wäre $P \neq O$, dann wäre $e = v_p(z(P)) \in \mathbb{Z}_{\geq 1}$. Da P ein Torsionspunkt ist, gibt es $m \in \mathbb{Z}_{\geq 2}$ mit $mP = O$. Aus (2) folgt dann aber $v_p(z(mP)) = e + v_p(m) < \infty$, also $mP \neq O$. Dieser Widerspruch zeigt, dass $P = O$ sein muss. \square

Der Beweis funktioniert völlig analog, wenn man statt über \mathbb{Q} über dem Körper \mathbb{Q}_p der p -adischen Zahlen arbeitet. In diesem Fall kann man wegen der Vollständigkeit von \mathbb{Q}_p und mithilfe der Theorie der formalen Gruppen (siehe [Si1, Ch. IV]) die Gruppe $E^{(1)}(\mathbb{Q}_p)$ recht genau beschreiben: Ist E durch eine kurze Weierstraß-Gleichung gegeben, dann ist $E^{(1)}(\mathbb{Q}_p)$ isomorph zur additiven Gruppe $p\mathbb{Z}_p$ und der Isomorphismus bildet $E^{(e)}(\mathbb{Q}_p)$ auf $p^e\mathbb{Z}_p$ ab.

20.4. Bemerkung. Ist E durch eine lange Weierstraß-Gleichung

$$E: y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$$

mit $a_1, \dots, a_6 \in \mathbb{Z}$ gegeben statt durch eine kurze wie in Satz 20.3, dann bleibt Lemma 20.2 gültig. Man bekommt mehr Terme in den Gleichungen, aber das ändert nichts an den Abschätzungen für die p -adische Bewertung von s und t . Im Beweis von Satz 20.3 ergeben sich folgende Änderungen: Die Kongruenz für $z(P + P')$ gilt in der schwächeren Form

$$z(P + P') \equiv z(P) + z(P') \pmod{p^{2e}} \quad \text{bzw.} \quad \pmod{p^{3e}}, \quad \text{wenn } a_1 = 0 \text{ ist.}$$

Damit bleibt Aussage (1) allgemein gültig. Aussage (2) gilt allgemein für $p \nmid m$, während man für beliebiges m zusätzlich $e \geq 2$ oder $a_1 = 0$ voraussetzen muss. Da für den Beweis nur die p -adische Bewertung relevant ist, kann man die Voraussetzung an die Ganzzahligkeit der Koeffizienten abschwächen zu $v_p(a_j) \geq 0$. Ist $p \neq 2$, dann erhält man durch quadratisches Ergänzen eine Gleichung für eine

BEM
Satz 20.3
für lange
Weierstraß-
Gleichungen

isomorphe Kurve E' mit $a_1 = a_3 = 0$ und $v_p(a_j) \geq 0$ für $j = 2, 4, 6$, sodass der zugehörige Isomorphismus $E_p^{(e)}(\mathbb{Q})$ auf $E_p'^{(e)}(\mathbb{Q})$ abbildet (die x -Koordinate bleibt gleich, also bleibt auch ihre p -adische Bewertung gleich). Für $p \geq 3$ bleiben somit die Aussagen (2) und (3) gültig (da sie für E' gelten). Für $p = 2$ bekommen wir statt (3) die schwächeren Aussagen $E_2^{(2)}(\mathbb{Q})_{\text{tors}} = \{O\}$ und $E_2^{(1)}(\mathbb{Q})_{\text{tors}} \subset E(\mathbb{Q})[2]$.

Dass es nicht besser geht, zeigt das Beispiel

$$E: y^2 + xy + y = x^3 + x^2 - 110x - 880;$$

diese elliptische Kurve hat den Punkt $(\frac{51}{4}, -\frac{55}{8})$ der Ordnung 2, der in $E_2^{(1)}(\mathbb{Q})$ liegt. ♠

Teil (3) des Satzes liefert die Ganzzahligkeit der Torsionspunkte. Wir reichern das noch mit einer weiteren Aussage an, die insbesondere zeigt, dass $E(\mathbb{Q})_{\text{tors}}$ endlich ist.

20.5. Satz. Sei $E: y^2 = x^3 + ax + b$ eine elliptische Kurve über \mathbb{Q} mit $a, b \in \mathbb{Z}$ und sei $O \neq P = (\xi, \eta) \in E(\mathbb{Q})_{\text{tors}}$. Dann sind $\xi, \eta \in \mathbb{Z}$ und entweder ist $\eta = 0$ oder η^2 teilt $4a^3 + 27b^2$.

SATZ
Satz von
Nagell-Lutz

Beweis. Wäre $\xi \notin \mathbb{Z}$ oder $\eta \notin \mathbb{Z}$, dann gäbe es eine Primzahl p und $e \geq 1$ mit $v_p(\xi) = -2e, v_p(\eta) = -3e$, also

$$P \in E_p^{(e)}(\mathbb{Q}) \cap E(\mathbb{Q})_{\text{tors}} \subset E_p^{(1)}(\mathbb{Q})_{\text{tors}} = \{O\}$$

nach Satz 20.3 (3), ein Widerspruch zur Voraussetzung $P \neq O$. Also sind ξ und η ganze Zahlen. Wir nehmen jetzt an, dass $\eta \neq 0$ ist und müssen $\eta^2 \mid 4a^3 + 27b^2$ zeigen. Aus $\eta \neq 0$ folgt $2P \neq O$. Mit P ist auch $2P$ ein Torsionspunkt, hat also nach dem eben Gezeigten ganzzahlige Koordinaten. Wir zeigen allgemeiner die folgende Aussage:

Sind sowohl $P = (\xi, \eta)$ als auch $2P$ ganzzahlige Punkte auf E , dann ist η^2 ein Teiler von $4a^3 + 27b^2$.

Zum Beweis rechnet man nach, dass

$$x(2P) = \frac{\xi^4 - 2a\xi^2 - 8b\xi + a^2}{4(\xi^3 + a\xi + b)}$$

ist. Aus $x(2P) \in \mathbb{Z}$ folgt dann, dass $\eta^2 = \xi^3 + a\xi + b$ ein Teiler von $\xi^4 - 2a\xi^2 - 8b\xi + a^2$ ist. Setzt man ξ in die Relation

$$(3x^2 + 4a)(x^4 - 2ax^2 - 8bx + a^2) - (3x^3 - 5ax - 27b)(x^3 + ax + b) = 4a^3 + 27b^2$$

ein, dann folgt die Behauptung, da η^2 beide Terme auf der linken Seite teilt. □

20.6. Folgerung. Sei E eine elliptische Kurve über \mathbb{Q} . Dann ist die Torsionsuntergruppe $E(\mathbb{Q})_{\text{tors}}$ endlich.

FOLG
 $E(\mathbb{Q})_{\text{tors}}$ ist
endlich

Beweis. E ist isomorph zu einer Kurve, die durch eine kurze Weierstraß-Gleichung gegeben ist. Durch Skalieren von x und y können wir erreichen, dass die Gleichung ganzzahlige Koeffizienten hat. Sei ohne Beschränkung der Allgemeinheit E schon von dieser Form. Der Satz 20.5 von Nagell und Lutz zeigt, dass es nur endlich viele Möglichkeiten für die y -Koordinate eines Torsionspunkts $P \neq O$ gibt (nämlich $y = 0$ und die $y \in \mathbb{Z}$ mit $y^2 \mid 4a^3 + 27b^2 \neq 0$; man beachte, dass $\Delta(E) = -16(4a^3 + 27b^2)$ ist). Für jedes gegebene y gibt es aber höchstens drei mögliche Werte von x , sodass $(x, y) \in E$ ist. □

Aus dem Satz von Nagell-Lutz bekommt man einen Algorithmus, mit dem man die Gruppe $E(\mathbb{Q})_{\text{tors}}$ konkret bestimmen kann. Allerdings muss man dafür die Diskriminante von E faktorisieren. Zuerst bestimmt man die Punkte mit $y = 0$ (das sind genau die Punkte der Ordnung 2). Aus der Faktorisierung von $4a^3 + 27b^2$ gewinnt man die Liste der η mit $\eta^2 \mid 4a^3 + 27b^2$; für jedes solche η bestimmt man die ganzzahligen Nullstellen ξ von $x^3 + ax + b - \eta^2$, und für jeden so gefundenen ganzzahligen Punkt $P = (\xi, \eta)$ berechnet man $2P, 4P, 8P, \dots$, bis man entweder den Punkt O erhält oder einen Punkt bekommt, der in dieser Folge schon aufgetreten ist (dann ist $P \in E(\mathbb{Q})_{\text{tors}}$) oder einen Punkt, der nicht ganzzahlig ist (dann ist $P \notin E(\mathbb{Q})_{\text{tors}}$).

20.7. Beispiel. Wir betrachten die elliptische Kurve

$$E: y^2 = x^3 - x + 1$$

mit $a = -1$ und $b = 1$. Dann ist $4a^3 + 27b^2 = 23$. Die möglichen y -Koordinaten von nichttrivialen Torsionspunkten sind somit $y = -1, 0, 1$. Das ergibt die Kandidaten

$$(-1, \pm 1), \quad (0, \pm 1), \quad (1, \pm 1).$$

Wir berechnen $2P, 4P, \dots$ für jeden dieser Punkte (bis auf Negation):

$$2 \cdot (-1, 1) = (3, -5), \quad 2 \cdot (3, -5) = \left(\frac{19}{25}, -\frac{103}{125}\right) \notin E(\mathbb{Q})_{\text{tors}}$$

$$2 \cdot (0, 1) = \left(\frac{1}{4}, -\frac{7}{8}\right) \notin E(\mathbb{Q})_{\text{tors}}$$

$$2 \cdot (1, 1) = (-1, 1) \notin E(\mathbb{Q})_{\text{tors}}$$

Das zeigt $E(\mathbb{Q})_{\text{tors}} = \{O\}$.

BSP
Torsions-
gruppe



20.8. Beispiel. Wir betrachten die elliptische Kurve

$$E: y^2 = x^3 - 1386747x + 368636886.$$

Wir berechnen

$$4a^3 + 27b^2 = -6998115764183040000 = -2^{16} \cdot 3^{20} \cdot 5^4 \cdot 7^2.$$

Die möglichen y -Koordinaten von nichttrivialen Torsionspunkten sind also 0 und die Teiler von $2^8 \cdot 3^{10} \cdot 5^2 \cdot 7$; davon gibt es $2 \cdot (8+1) \cdot (10+1) \cdot (2+1) \cdot (1+1) = 1188$. Für jeden dieser Teiler finden wir alle Punkte in $E(\mathbb{Q})$ mit dieser y -Koordinate. Das ergibt die folgende Liste:

$$(147, \pm 12960), \quad (1227, \pm 22680), \quad (-285, \pm 27216), \\ (-933, \pm 29160), \quad (2307, \pm 97200), \quad (8787, \pm 816480).$$

Dazu kommen noch drei Punkte mit $y = 0$:

$$(-1293, 0), \quad (282, 0), \quad (1011, 0).$$

Diese letzten drei Punkte haben Ordnung 2. Die übrigen Punkte verdoppeln wir wiederholt, bis wir entweder einen Torsionspunkt erhalten oder einen Punkt, der nicht ganzzahlig ist.

$$2 \cdot (147, 12960) = (2307, 97200) \\ 2 \cdot (2307, 97200) = (1011, 0) \in E(\mathbb{Q})_{\text{tors}} \\ 2 \cdot (1227, 22680) = (2307, -97200) \in E(\mathbb{Q})_{\text{tors}} \\ 2 \cdot (-285, 27216) = (1011, 0) \in E(\mathbb{Q})_{\text{tors}} \\ 2 \cdot (-933, 29160) = (2307, -97200) \in E(\mathbb{Q})_{\text{tors}} \\ 2 \cdot (8787, 816480) = (2307, 97200) \in E(\mathbb{Q})_{\text{tors}}$$

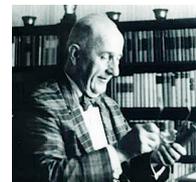
BSP
Torsions-
gruppe

Das zeigt, dass alle Punkte, die wir gefunden haben, tatsächlich Torsionspunkte sind. Außerdem sieht man, dass $E(\mathbb{Q})_{\text{tors}} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/8\mathbb{Z}$ ist mit Erzeugern (z.B.) $(282, 0)$ und $(147, 12960)$. ♣

Im nächsten Abschnitt werden wir sehen, wie man $E(\mathbb{Q})_{\text{tors}}$ bestimmen kann, ohne dass man die Diskriminante faktorisieren muss.

Wir werden im weiteren Verlauf der Vorlesung den Satz von Mordell behandeln (allerdings nicht in allen Fällen beweisen), der besagt, dass die Gruppe $E(\mathbb{Q})$ endlich erzeugt ist. Daraus folgt dann auch, dass $E(\mathbb{Q})_{\text{tors}}$ endlich ist.

Es gibt einen weiteren Satz, der von Siegel bewiesen wurde⁶ und (u.a.) besagt, dass es auf einer elliptischen Kurve über \mathbb{Q} stets nur endlich viele ganzzahlige Punkte gibt. Zusammen mit Satz 20.3 (3) folgt dann auch wieder, dass $E(\mathbb{Q})_{\text{tors}}$ endlich ist. Allerdings ist der Satz von Siegel sehr tieflegend (und benutzt den Satz von Mordell).



C.L. Siegel
1896–1981
Foto © MFO

⁶C.L. Siegel, *Über einige Anwendungen diophantischer Approximationen*, Abhandlungen Akad. Berlin 1929, No. 1, 70 S. (1929).

21. GUTE UND SCHLECHTE REDUKTION

Wenn wir eine elliptische Kurve

$$E: y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$$

mit Koeffizienten $a_1, a_2, a_3, a_4, a_6 \in \mathbb{Z}$ gegeben haben, dann können wir für eine Primzahl p die Kurve

$$\bar{E}: y^2 + \bar{a}_1xy + \bar{a}_3y = x^3 + \bar{a}_2x^2 + \bar{a}_4x + \bar{a}_6$$

betrachten, wobei $\bar{a} \in \mathbb{F}_p$ für die Restklasse von $a \bmod p$ steht. Ist p kein Teiler der Diskriminante $\Delta(E)$, dann ist $\Delta(\bar{E}) = \overline{\Delta(E)} \neq 0$, und \bar{E} ist eine elliptische Kurve über \mathbb{F}_p .

Nun kann man statt E auch eine über \mathbb{Q} isomorphe Kurve E' betrachten, deren Gleichung ebenfalls ganzzahlige Koeffizienten hat. Es gilt dann $\Delta(E') = u^{12}\Delta(E)$ mit einem $u \in \mathbb{Q}^\times$. Insbesondere ist es möglich, dass p zwar ein Teiler von $\Delta(E)$ ist, aber $\Delta(E')$ nicht teilt. Das motiviert die folgende Definition.

21.1. Definition. Sei E eine elliptische Kurve über \mathbb{Q} . Eine Weierstraß-Gleichung

$$E': y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$$

heißt eine *minimale Weierstraß-Gleichung für E* , wenn E' über \mathbb{Q} zu E isomorph ist, $a_1, a_2, a_3, a_4, a_6 \in \mathbb{Z}$ sind und $|\Delta(E')|$ unter allen solchen Gleichungen minimal ist. $\Delta(E')$ heißt dann die *minimale Diskriminante von E* . \diamond

DEF
minimale
Weierstraß-
Gleichung

Die Minimalitäts-Bedingung ist äquivalent zu „ $\Delta(E')$ teilt die Diskriminante jeder ganzzahligen Weierstraß-Gleichung einer elliptischen Kurve, die isomorph zu E ist“. Die Minimalität gilt also auch in Bezug auf Teilbarkeit. Die minimale Diskriminante ist eindeutig bestimmt (Übung).

21.2. Definition. Sei E eine elliptische Kurve über \mathbb{Q} ; wir nehmen an, dass E durch eine minimale Weierstraß-Gleichung gegeben ist. Sei p eine Primzahl. Wir sagen, E habe *gute Reduktion bei p* , wenn p die (minimale) Diskriminante $\Delta(E)$ nicht teilt; die elliptische Kurve \bar{E} wie oben heißt dann die *Reduktion von $E \bmod p$* . Anderenfalls sagen wir, E habe *schlechte Reduktion bei p* . \diamond

DEF
gute/schlechte
Reduktion

Wir werden uns jetzt überlegen, dass es auch eine Beziehung zwischen den Gruppen $E(\mathbb{Q})$ und $\bar{E}(\mathbb{F}_p)$ gibt. Dazu zeigen wir erst einmal, dass man überhaupt eine vernünftige Abbildung $\mathbb{P}^2(\mathbb{Q}) \rightarrow \mathbb{P}^2(\mathbb{F}_p)$ hat.

21.3. Lemma. Sei $P = (\xi : \eta : \zeta) \in \mathbb{P}^2(\mathbb{Q})$. Dann können wir P schreiben als $P = (\xi' : \eta' : \zeta')$ mit $\xi', \eta', \zeta' \in \mathbb{Z}$, sodass $\text{ggT}(\xi', \eta', \zeta') = 1$ ist. Sei p eine Primzahl. Dann ist die Abbildung

$$\rho_p: \mathbb{P}^2(\mathbb{Q}) \longrightarrow \mathbb{P}^2(\mathbb{F}_p), \quad P \longmapsto (\bar{\xi} : \bar{\eta} : \bar{\zeta})$$

(wobei $\mathbb{Z} \rightarrow \mathbb{F}_p, a \mapsto \bar{a}$, der kanonische Epimorphismus ist) wohldefiniert.

Ist $G \subset \mathbb{P}^2_{\mathbb{Q}}$ eine Gerade, dann ist $\rho_p(G(\mathbb{Q})) = \bar{G}(\mathbb{F}_p)$ mit einer Geraden $\bar{G} \subset \mathbb{P}^2_{\mathbb{F}_p}$.

LEMMA
Reduktion
von Punkten
in $\mathbb{P}^2(\mathbb{Q})$

Beweis. Wir können die Koordinaten von P zunächst mit einem gemeinsamen Nenner skalieren; dann erhalten wir eine Darstellung von P mit ganzzahligen Koordinaten. Dann skalieren wir mit dem Inversen des ggTs dieser Koordinaten und erhalten eine Darstellung wie angegeben. Skalieren der Koordinaten mit $\lambda \in \mathbb{Z}$

skaliert den ggT mit $|\lambda|$; es folgt, dass diese Darstellung bis auf Skalierung mit -1 eindeutig ist.

Der Punkt $(\bar{\xi}' : \bar{\eta}' : \bar{\zeta}') \in \mathbb{P}^2(\mathbb{F}_p)$ ist dann wohldefiniert, denn wenigstens eine der Koordinaten ist nicht null (wegen $p \nmid \text{ggT}(\xi', \eta', \zeta')$), und die Koordinaten sind eindeutig bis auf Skalierung (mit -1). Damit ist auch ρ_p wohldefiniert.

Sei jetzt $G: ax + by + cz = 0$ eine Gerade in $\mathbb{P}_{\mathbb{Q}}^2$. Analog zu den Koordinaten von P können wir die Geradengleichung so skalieren, dass $a, b, c \in \mathbb{Z}$ sind und $\text{ggT}(a, b, c) = 1$ ist. Sei dann $\bar{G}: \bar{a}x + \bar{b}y + \bar{c}z = 0$ in $\mathbb{P}_{\mathbb{F}_p}^2$. Einsetzen von ξ', η', ζ' und Reduzieren mod p zeigt, dass $\rho_p(G(\mathbb{Q})) \subset \bar{G}(\mathbb{F}_p)$ ist. Ist umgekehrt $(\bar{\xi} : \bar{\eta} : \bar{\zeta}) \in \bar{G}(\mathbb{F}_p)$, dann seien zunächst ξ', η', ζ' beliebige ganze Zahlen in den entsprechenden Restklassen. Es ist dann $a\xi' + b\eta' + c\zeta' = pd$ mit $d \in \mathbb{Z}$. Wegen $\text{ggT}(a, b, c) = 1$ gibt es $r, s, t \in \mathbb{Z}$ mit $ra + sb + tc = 1$. Mit

$$(\xi, \eta, \zeta) := (\xi' - prd, \eta' - psd, \zeta' - ptd)$$

gilt dann $P := (\xi : \eta : \zeta) \in G(\mathbb{Q})$ und $\rho_p(P) = (\bar{\xi} : \bar{\eta} : \bar{\zeta})$. □

21.4. Folgerung. Sei E eine elliptische Kurve über \mathbb{Q} , gegeben durch eine Gleichung mit ganzzahligen Koeffizienten. Sei p eine Primzahl mit $p \nmid \Delta(E)$ und sei \bar{E} die durch Reduktion modulo p aus E entstehende elliptische Kurve über \mathbb{F}_p . Dann ist die Einschränkung von ρ_p auf $E(\mathbb{Q})$ ein Gruppenhomomorphismus

FOLG
Reduktions-
homomor-
phismus

$$\rho_{p,E}: E(\mathbb{Q}) \longrightarrow \bar{E}(\mathbb{F}_p).$$

Es gilt dann $\ker \rho_{p,E} = E_p^{(1)}(\mathbb{Q})$.

Die letzte Aussage erklärt die Bezeichnung „Kern der Reduktion“ für $E_p^{(1)}(\mathbb{Q})$.

Beweis. Einsetzen von passend skalierten ganzzahligen Koordinaten in die Gleichung von E und Reduktion mod p zeigt, dass $\rho_p(E(\mathbb{Q})) \subset \bar{E}(\mathbb{F}_p)$ ist; damit ist $\rho_{p,E}$ jedenfalls als Abbildung wohldefiniert. Es bleibt zu zeigen, dass $\rho_{p,E}$ ein Gruppenhomomorphismus ist. Es ist klar, dass $\rho_{p,E}(O) = O$ ist. Seien $P_1, P_2, P_3 \in E(\mathbb{Q})$ mit $P_1 + P_2 + P_3 = O$. Dann sind P_1, P_2, P_3 die drei Schnittpunkte von E mit einer Geraden G (mit passender Vielfachheit, wenn Punkte zusammenfallen). Aus Lemma 21.3 folgt, dass $\rho_{p,E}(P_1), \rho_{p,E}(P_2), \rho_{p,E}(P_3)$ die drei Schnittpunkte von \bar{E} mit der Geraden \bar{G} sind. Also ist $\rho_{p,E}(P_1) + \rho_{p,E}(P_2) + \rho_{p,E}(P_3) = O$ in $\bar{E}(\mathbb{F}_p)$. Es folgt, dass $\rho_{p,E}$ ein Gruppenhomomorphismus ist.

Für den Beweis der letzten Aussage seien die projektiven Koordinaten von $P \in E(\mathbb{Q})$ als teilerfremde ganze Zahlen gewählt. Dann gilt

$$\begin{aligned} P = (\xi : \eta : \zeta) \in \ker \rho_{p,E} &\iff \rho_{p,E}(P) = O = (0 : 1 : 0) \\ &\iff p \mid \xi, \quad p \nmid \eta, \quad p \mid \zeta \\ &\stackrel{(*)}{\iff} v_p(\eta/\zeta) < 0 \\ &\iff P \in E_p^{(1)}(\mathbb{Q}). \end{aligned}$$

In der Äquivalenz $(*)$ ist die Richtung „ \Rightarrow “ trivial. Die Umkehrung ergibt sich daraus, dass aus $v_p(\eta/\zeta) < 0$ folgt, dass es $e \geq 1$ gibt mit $v_p(\xi/\zeta) = -2e$ und $v_p(\eta/\zeta) = -3e$ (durch Vergleichen der Bewertungen der verschiedenen Terme in der Gleichung von E ; siehe auch den vorigen Abschnitt). Da $\min\{v_p(\xi), v_p(\eta), v_p(\zeta)\} = 0$ ist, muss dann $v_p(\xi) = e, v_p(\eta) = 0$ und $v_p(\zeta) = 3e$ sein ($e = \infty$ ist dabei möglich; dann ist $P = O$). □

Mit unserem Wissen aus dem vorigen Abschnitt können wir den Reduktionshomomorphismus nutzen, um Aussagen über $E(\mathbb{Q})_{\text{tors}}$ zu treffen.

21.5. Satz. *Sei*

$$E: y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$$

eine elliptische Kurve über \mathbb{Q} mit $a_1, a_2, a_3, a_4, a_6 \in \mathbb{Z}$. Ist p eine Primzahl mit $p \nmid \Delta(E)$ und gilt $a_1 = 0$ oder $p \geq 3$, dann ist

$$\rho_{p,E}|_{E(\mathbb{Q})_{\text{tors}}} : E(\mathbb{Q})_{\text{tors}} \rightarrow \bar{E}(\mathbb{F}_p)$$

injektiv.

SATZ
Reduktion
der Torsion

Beweis. Unter den angegebenen Voraussetzungen gilt

$$\ker \rho_{p,E}|_{E(\mathbb{Q})_{\text{tors}}} = \ker \rho_{p,E} \cap E(\mathbb{Q})_{\text{tors}} = E_p^{(1)}(\mathbb{Q}) \cap E(\mathbb{Q})_{\text{tors}} \stackrel{(*)}{=} \{O\};$$

daraus folgt die Behauptung. Wir haben die Gleichheit (*) für kurze Weierstraß-Gleichungen in Satz 20.3 bewiesen. Für den allgemeinen Fall siehe Bemerkung 20.4. \square

Wir können also $E(\mathbb{Q})_{\text{tors}}$ als eine Untergruppe von $\bar{E}(\mathbb{F}_p)$ realisieren für jede Primzahl p ($p \geq 3$, falls $a_1 \neq 0$) mit $p \nmid \Delta(E)$. Es gibt unendlich viele solche Primzahlen. Da $\bar{E}(\mathbb{F}_p)$ endlich ist, erhalten wir einen weiteren Beweis der Aussage, dass $E(\mathbb{Q})_{\text{tors}}$ endlich ist.

Wir können Satz 21.5 aber auch dafür nutzen, mit wenig Aufwand eine gute Schranke für $\#E(\mathbb{Q})_{\text{tors}}$ zu bekommen, denn nach dem Satz von Lagrange muss $\#E(\mathbb{Q})_{\text{tors}}$ ein Teiler von $\#\bar{E}(\mathbb{F}_p)$ sein.

21.6. Beispiel. Sei $E: y^2 = x^3 - x + 1$ mit $\Delta(E) = -2^4 \cdot 23$. Wir können Satz 21.5 also z.B. mit $p = 3$ und $p = 5$ anwenden. Es folgt

$$\#E(\mathbb{Q})_{\text{tors}} \mid \#\bar{E}(\mathbb{F}_3) = 7 \quad \text{und} \quad \#E(\mathbb{Q})_{\text{tors}} \mid \#\bar{E}(\mathbb{F}_5) = 8,$$

also muss $E(\mathbb{Q})_{\text{tors}}$ trivial sein. Da $E(\mathbb{Q})$ affine Punkte enthält (z.B. $(1, 1)$), folgt, dass $E(\mathbb{Q})$ unendlich ist. (Tatsächlich ist $E(\mathbb{Q}) \cong \mathbb{Z}$; die Gruppe wird von $(1, 1)$ erzeugt.) \clubsuit

BSP
 $E(\mathbb{Q})_{\text{tors}}$
ist trivial

21.7. Beispiel. Sei

$$E: y^2 + xy + y = x^3 + x^2 - 70x - 279;$$

es ist $\Delta(E) = -2 \cdot 19^5$. Wir erhalten folgende Tabelle:

p	3	5	7	11	13	17	23	29	31	37
$\#\bar{E}(\mathbb{F}_p)$	5	10	5	10	15	15	25	35	40	40

Es folgt, dass $\#E(\mathbb{Q})_{\text{tors}} \in \{1, 5\}$ ist. Man kann das Polynom aufstellen, dessen Nullstellen die x -Koordinaten der Punkte der Ordnung 5 sind:

$$\begin{aligned} &5x^{12} + 25x^{11} - 4284x^{10} - 112875x^9 - 904395x^8 + 1848750x^7 + 97164150x^6 \\ &+ 1084824520x^5 + 7387397375x^4 + 28604803425x^3 + 39626137350x^2 \\ &- 77025287125x - 228943289601 \end{aligned}$$

und überprüfen, dass es keine ganzzahligen Nullstellen hat. Also ist tatsächlich $E(\mathbb{Q})_{\text{tors}} = \{O\}$.

BSP
Schranke
nicht scharf

Es gibt einen Grund, dass die Schranke für $\#E(\mathbb{Q})_{\text{tors}}$ hier nicht scharf ist. E ist nämlich isogen zu der elliptischen Kurve

$$E': y^2 + xy + y = x^3 + x^2 + 1,$$

für die $\#E'(\mathbb{Q})_{\text{tors}} = 5$ gilt. Aus Satz 21.5 folgt dann $5 \mid \#\bar{E}'(\mathbb{F}_p)$ für alle $p \geq 3$, $p \neq 19$. Nach Satz 12.3 (das ist die einfache Richtung, die wir auch bewiesen haben) gilt $\#\bar{E}(\mathbb{F}_p) = \#\bar{E}'(\mathbb{F}_p)$, denn \bar{E} und \bar{E}' sind ebenfalls isogen. Daraus folgt

$$\text{ggT}(\{\#\bar{E}(\mathbb{F}_p) \mid p \geq 3 \text{ prim, } p \neq 19\}) = 5,$$

also *kann* die Schranke nicht scharf sein. ♣

Aus der Schranke von Hasse für die Anzahl der \mathbb{F}_p -rationalen Punkte auf einer elliptischen Kurve (Satz 12.2) ergibt sich die Abschätzung

$$\#E(\mathbb{Q})_{\text{tors}} \leq \min\{[(\sqrt{p} + 1)^2] \mid p \geq 3 \text{ prim, } p \nmid \Delta(E)\}.$$

Ist die Torsionsuntergruppe also „groß“, dann muss E schlechte Reduktion bei allen „kleinen“ Primzahlen haben. (Für $2 \nmid \Delta(E)$ bekommt man die Abschätzung $\#E(\mathbb{Q})_{\text{tors}} \leq 10$.)

Allerdings hat Barry Mazur in den 1970er Jahren gezeigt, dass die Torsionsuntergruppe von $E(\mathbb{Q})$ nicht beliebig groß werden kann:⁷

21.8. Satz. *Sei E eine elliptische Kurve über \mathbb{Q} . Dann ist*

$$E(\mathbb{Q})_{\text{tors}} \cong \mathbb{Z}/d\mathbb{Z} \quad \text{mit } 1 \leq d \leq 10 \text{ oder } d = 12,$$

oder

$$E(\mathbb{Q})_{\text{tors}} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2d\mathbb{Z} \quad \text{mit } 1 \leq d \leq 4.$$

Jede dieser möglichen Strukturen tritt für unendlich viele paarweise nicht isomorphe elliptische Kurven über \mathbb{Q} auf.

Da ± 1 die einzigen Einheitswurzeln in \mathbb{Q} sind, folgt aus der Existenz der Weil-Paarung, dass $E(\mathbb{Q})_{\text{tors}}$ eine der beiden oben angegebenen Formen haben muss (für irgend ein $d \geq 1$).

Die zweite Aussage ist nicht schwer zu zeigen, denn man kann explizite Familien von solchen Kurven angeben, die von einem Parameter abhängen.

Um die Möglichkeit der Existenz eines Punktes $P \in E(\mathbb{Q})$ der Ordnung d zu untersuchen, kann man (für $d \geq 4$) die folgende „Normalform“ benutzen:

$$E: y^2 + uxy + vy = x^3 + vx^2$$

mit $u, v \in \mathbb{Q}$; der Punkt P ist dabei $P = (0, 0)$. Die Bedingung, dass P Ordnung d hat, ergibt eine Gleichung $P_d(u, v) = 0$ mit einem Polynom P_d . Diese Gleichung definiert eine affine ebene Kurve, deren rationale Punkte entweder zu Paaren (E, P) gehören, wobei E eine elliptische Kurve über \mathbb{Q} und $P \in E(\mathbb{Q})$ ein Punkt der Ordnung d ist, oder zu einer Kurve führen, die nicht glatt (und damit keine elliptische Kurve) ist. Es gibt eine glatte projektive (nicht unbedingt ebene) Kurve $X_1(d)$, die über \mathbb{Q} definiert ist und zu dieser Kurve birational ist. Die Aussage im Satz oben folgt dann daraus, dass die rationalen Punkte von $X_1(d)$ für d eine Primzahl ≥ 11 und für $d = 14, 15, 16, 18, 20, 21, 24, 25, 27, 35, 49$ alle zu nicht-glatten Kurven E gehören.

⁷B. Mazur, *Modular curves and the Eisenstein ideal*, Inst. Hautes Études Sci. Publ. Math. **47**, 33–186 (1978).



B. Mazur
* 1937

Foto © MFO

SATZ
Satz von
Mazur

Zum Beispiel ist $X_1(11)$ selbst eine elliptische Kurve, und man kann zeigen, dass sie genau fünf rationale Punkte hat, die aber alle nicht zu einer elliptischen Kurve mit einem Punkt der Ordnung 11 gehören. Für großes d wird das Geschlecht der Kurve $X_1(d)$ allerdings ebenfalls groß, und der Beweis im allgemeinen Fall erfordert sehr tief liegende Hilfsmittel.

22. DER SATZ VON MORDELL

Unser nächstes Ziel ist es, den folgenden Satz (jedenfalls in einem Spezialfall) zu beweisen.

22.1. Satz. *Sei E eine elliptische Kurve über \mathbb{Q} . Dann ist die Gruppe $E(\mathbb{Q})$ eine endlich erzeugte abelsche Gruppe.*

SATZ
Satz von
Mordell



L.J. Mordell
1888 – 1972
Foto © MFO

Dieser Satz wurde von Mordell bewiesen.⁸ Die Aussage wurde von Weil einige Jahre später in seiner Doktorarbeit auf Jacobische Varietäten von algebraischen Kurven über beliebigen algebraischen Zahlkörpern (also endlichen Körpererweiterungen von \mathbb{Q}) verallgemeinert.⁹ (Elliptische Kurven sind ihre eigenen Jacobischen Varietäten.) Daher wird der Satz meistens als *Satz von Mordell-Weil* bezeichnet. (Offenbar war Mordell selbst darüber nicht sehr glücklich.)

Wir werden das Pferd gewissermaßen von hinten aufzäumen und den Satz von Mordell auf einige andere Aussagen zurückführen.

22.2. Satz. *Seien G eine (additiv geschriebene) abelsche Gruppe, $m \in \mathbb{Z}_{\geq 2}$ und $h: G \rightarrow \mathbb{R}_{\geq 0}$ eine Abbildung mit folgenden Eigenschaften:*

SATZ
abelsche
Gruppe ist
endlich
erzeugt

- (1) G/mG ist endlich.
- (2) Für jedes $B > 0$ ist $\{g \in G \mid h(g) \leq B\}$ endlich.
- (3) Es gibt $C > 0$, sodass für alle $g \in G$ gilt $h(mg) \geq m^2h(g) - C$.
- (4) Für jedes $g \in G$ gibt es $c_g > 0$, sodass für alle $g' \in G$ gilt $h(g+g') \leq 2h(g') + c_g$.

Dann ist G endlich erzeugt.

22.3. Definition. Eine Abbildung $h: G \rightarrow \mathbb{R}_{\geq 0}$ mit den Eigenschaften (2), (3) (für ein $m \geq 2$) und (4) heißt eine *Höhenfunktion* auf G . ◇

DEF
Höhen-
funktion

Beweis. Seien $g_i \in G$ für $i = 1, \dots, k$ Repräsentanten der wegen (1) endlich vielen Restklassen in G/mG und sei

$$\gamma = \frac{C + \max\{c_{-g_i} \mid i = 1, \dots, k\}}{m^2 - 2} > 0.$$

Wir zeigen, dass G von der (nach (2)) endlichen Menge

$$M = \{g_i \mid i = 1, \dots, k\} \cup \{g \in G \mid h(g) \leq \gamma\}$$

erzeugt wird.

Wir nehmen dafür an, dass das nicht der Fall ist. Dann gibt es Elemente $g \in G$ mit $g \notin \langle M \rangle$. Da es nach (2) nur endlich viele Elemente $g' \in G$ mit $h(g') \leq h(g)$ gibt, können wir annehmen, dass g unter allen Gegenbeispielen minimale Höhe $h(g)$ hat. Es ist dann jedenfalls $g \notin M$, woraus $h(g) > \gamma$ folgt. Sei nun g_i der gewählte Repräsentant der Restklasse $g + mG \in G/mG$. Dann gibt es $g' \in G$ mit $g = g_i + mg'$. Aus den Eigenschaften (3) und (4) folgt

$$2h(g) + c_{-g_i} \geq h(g - g_i) = h(mg') \geq m^2h(g') - C$$

⁸L.J. Mordell, *On the rational solutions of the indeterminate equations of the third and fourth degrees*, Cambr. Phil. Soc. Proc. **21**, 179–192 (1922).

⁹A. Weil, *L'arithmétique sur les courbes algébriques*, Acta Math. **52**, 281–315 (1929).

und daraus

$$h(g') \leq h(g) - \frac{(m^2 - 2)h(g) - (C + c_{-g_i})}{m^2} \leq h(g) - \frac{m^2 - 2}{m^2}(h(g) - \gamma) < h(g).$$

Da g ein Gegenbeispiel mit minimaler Höhe war, ist $g' \in \langle M \rangle$. Dann folgt aber auch $g = g_i + mg' \in \langle M \rangle$ (denn $g_i \in M$). Das ist ein Widerspruch, also muss die Annahme falsch gewesen sein, und es folgt wie gewünscht, dass $G = \langle M \rangle$ ist. \square

Die Aussage (1) heißt im Kontext von elliptischen Kurven (oder abelschen Varietäten), also mit $G = E(\mathbb{Q})$, auch *schwacher Satz von Mordell-(Weil)*. Es ist klar, dass das eine notwendige Bedingung dafür ist, dass G endlich erzeugt ist. Sie ist aber nicht hinreichend, wie man an Beispielen wie den additiven Gruppen \mathbb{Q} oder \mathbb{R} sieht. Diesem schwachen Satz von Mordell (mit $m = 2$) werden wir uns später zuwenden. In diesem Abschnitt werden wir zeigen, dass eine Höhenfunktion auf $E(\mathbb{Q})$ existiert.

22.4. Definition. Sei E eine elliptische Kurve über \mathbb{Q} . Wir definieren

$$h: E(\mathbb{Q}) \longrightarrow \mathbb{R}_{\geq 0}, \quad P \longmapsto \begin{cases} 0, & P = O, \\ \log \max\{|u|, |v|\}, & x(P) = u/v \text{ mit } u \perp v. \end{cases} \quad \diamond$$

DEF
Höhen-
funktion
auf $E(\mathbb{Q})$

Hier ist \log der natürliche Logarithmus. Anschaulich ist $h(P)$ ein Maß dafür, wie viel Platz man braucht, um (die x -Koordinate von) P hinzuschreiben. Wenn wir $x(O) = 1/0$ setzen, dann brauchen wir die Fallunterscheidung in der Definition nicht. Das ist mit der Interpretation der x -Koordinate als Morphismus $x: E \rightarrow \mathbb{P}^1$, $(\xi : \eta : \zeta) \mapsto (\xi : \zeta)$, kompatibel.

Wir zeigen nun, dass h eine Höhenfunktion auf $E(\mathbb{Q})$ ist.

22.5. Satz. Sei $E: y^2 = x^3 + ax + b$ eine elliptische Kurve über \mathbb{Q} mit $a, b \in \mathbb{Z}$. Dann ist h wie in Definition 22.4 eine Höhenfunktion auf $E(\mathbb{Q})$ (für $m = 2$).

SATZ
Höhen-
funktion
auf $E(\mathbb{Q})$

Beweis. Wir zeigen die drei relevanten Eigenschaften aus Satz 22.2.

- (2) Ist $P \in E(\mathbb{Q})$ mit $h(P) \leq B$, dann ist entweder $P = O$ oder $x(P) = u/v$ mit $u, v \in \mathbb{Z}$, $|u|, |v| \leq e^B$. Es gibt also nur endlich viele Möglichkeiten für die x -Koordinate von P , und zu jeder x -Koordinate gibt es höchstens zwei y -Koordinaten, sodass es insgesamt nur endlich viele solche Punkte geben kann.
- (3) Es ist zu zeigen, dass es $C > 0$ gibt mit

$$h(2P) \geq 4h(P) - C \quad \text{für alle } P \in E(\mathbb{Q}).$$

Dafür verwenden wir die Verdopplungsformel

$$x(2P) = \frac{x(P)^4 - 2ax(P)^2 - 8bx(P) + a^2}{4(x(P)^3 + ax(P) + b)}.$$

Wenn wir $x(P) = u/v$ mit $u \perp v$ schreiben, dann ist

$$(22.1) \quad x(2P) = \frac{u^4 - 2au^2v^2 - 8buv^3 + a^2v^4}{4(u^3 + auv^2 + bv^3)v} =: \frac{F_1(u, v)}{F_2(u, v)}$$

mit $F_1(u, v), F_2(u, v) \in \mathbb{Z}$ und daher

$$h(2P) = \log \max\{|F_1(u, v)|, |F_2(u, v)|\} - \log v,$$

wobei g der ggT von $F_1(u, v)$ und $F_2(u, v)$ ist. Nun gilt

$$\begin{aligned} G_1(u, v) \cdot F_1(u, v) + G_2(u, v) \cdot F_2(u, v) &= 4(4a^3 + 27b^2)u^7 \quad \text{und} \\ H_1(u, v) \cdot F_1(u, v) + H_2(u, v) \cdot F_2(u, v) &= 4(4a^3 + 27b^2)v^7 \end{aligned}$$

mit

$$\begin{aligned} G_1(u, v) &= (16a^3 + 108b^2)u^3 - 4a^2bu^2v + (12a^4 + 88ab^2)uv^2 + (12a^3b + 96b^3)v^3, \\ G_2(u, v) &= a^2bu^3 + (5a^4 + 32ab^2)u^2v + (26a^3b + 192b^3)uv^2 - (3a^5 + 24a^2b^2)v^3, \\ H_1(u, v) &= (12u^2 + 16av^2)v, \\ H_2(u, v) &= -(3u^3 - 5auv^2 - 27bv^3). \end{aligned}$$

Aus $u \perp v$ folgt $g \mid 4(4a^3 + 27b^2) = -\Delta(E)/4 \neq 0$. Aus der Dreiecksungleichung folgt, dass es eine Konstante A gibt mit

$$\max\{|G_1(u, v)|, |G_2(u, v)|, |H_1(u, v)|, |H_2(u, v)|\} \leq A \max\{|u|, |v|\}^3.$$

Daraus schließen wir, dass

$$\max\{|u|, |v|\}^7 \leq \frac{2A}{4|4a^3 + 27b^2|} \max\{|u|, |v|\}^3 \max\{|F_1(u, v)|, |F_2(u, v)|\}$$

ist. Daraus und mit $|g| \leq 4|4a^3 + 27b^2|$ folgt schließlich $h(2P) \geq 4h(P) - C$ mit $C = \log(2A)$.

- (4) Schließlich müssen wir zeigen, dass es für jeden Punkt $P \in E(\mathbb{Q})$ eine Konstante $c_P > 0$ gibt mit $h(P + Q) \leq 2h(Q) + c_P$ für alle $Q \in E(\mathbb{Q})$. Im Fall $P = O$ ist die Aussage klar, und sie gilt für $Q = O$, wenn $c_P \geq h(P)$ ist. Ebenso gilt die Behauptung für $Q = -P$, und für $Q = P$ gilt sie mit $c_P \geq h(2P) - 2h(P)$. Wir können also $P \neq O$ und $Q \neq O, P, -P$ annehmen. Wir schreiben $P = (\xi_P : \eta_P : \zeta_P)$ und $Q = (\xi_Q : \eta_Q : \zeta_Q)$ mit $\xi_P, \zeta_P, \xi_Q, \zeta_Q \in \mathbb{Z}$ und $\xi_P \perp \zeta_P, \xi_Q \perp \zeta_Q$. Dabei sind η_P und η_Q dabei im Allgemeinen *nicht* ganzzahlig. Aus der Gleichung

$$\eta_Q^2 \zeta_Q = \xi_Q^3 + a\xi_Q \zeta_Q^2 + b\zeta_Q^3$$

folgt durch Multiplikation mit ζ_Q aber, dass $\eta_Q \zeta_Q \in \mathbb{Z}$ und

$$(22.2) \quad |\eta_Q \zeta_Q| \leq \sqrt{1 + |a| + |b|} \max\{|\xi_Q|, |\zeta_Q|\}^2$$

ist (und ebenso für P). Es ist

$$\begin{aligned} x(P + Q) &= \left(\frac{y(P) - y(Q)}{x(P) - x(Q)} \right)^2 - x(P) - x(Q) \\ &= \frac{(\xi_P \xi_Q + a\zeta_P \zeta_Q)(\xi_P \zeta_Q + \zeta_P \xi_Q) + 2b\zeta_P^2 \zeta_Q^2 - 2\eta_P \zeta_P \eta_Q \zeta_Q}{(\xi_P \zeta_Q - \zeta_P \xi_Q)^2} =: \frac{Z}{N}. \end{aligned}$$

Zähler Z und Nenner N dieses Bruchs sind ganze Zahlen. Aus (22.2) und mit der Dreiecksungleichung erhalten wir die Abschätzung

$$|Z|, |N| \leq 4(1 + |a| + |b|) \max\{|\xi_P|, |\zeta_P|\}^2 \max\{|\xi_Q|, |\zeta_Q|\}^2.$$

Es folgt

$$\begin{aligned} h(P + Q) &\leq \log \max\{|Z|, |N|\} \\ &\leq \log(4(1 + |a| + |b|)) + 2 \log \max\{|\xi_P|, |\zeta_P|\} + 2 \log \max\{|\xi_Q|, |\zeta_Q|\} \\ &= \log(4(1 + |a| + |b|)) + 2h(P) + 2h(Q); \end{aligned}$$

die gewünschte Aussage gilt also mit

$$c_P = \max\{\log(4(1 + |a| + |b|)) + 2h(P), h(2P) - 2h(P)\}.$$

Wir bemerken noch, dass man aus (22.1) in ähnlicher Weise die Abschätzung

$$h(2P) \leq 4h(P) + \log \max\{(1 + |a|)^2 + 8|b|, 4(1 + |a| + |b|)\}$$

bekommt. Das ergibt

$$c_P = 2h(P) + \log \max\{(1 + |a|)^2 + 8|b|, 4(1 + |a| + |b|)\}. \quad \square$$

Damit reduziert sich der Beweis des Satzes 22.1 von Mordell auf den Beweis des schwachen Satzes von Mordell in der Form, dass $E(\mathbb{Q})/2E(\mathbb{Q})$ endlich ist. Damit werden wir uns in den nächsten Abschnitten befassen.

22.6. Bemerkung. Die Existenz einer Höhenfunktion auf $E(\mathbb{Q})$ liefert einen weiteren Beweis dafür, dass $E(\mathbb{Q})_{\text{tors}}$ endlich ist. Ist nämlich $P \in E(\mathbb{Q})_{\text{tors}}$, dann ist die Menge $\{2^n P \mid n \geq 0\}$ endlich; sei $H = \max\{h(2^n P) \mid n \geq 0\}$ und sei n so gewählt, dass $h(2^n P) = H$ ist. Eigenschaft (3) liefert

BEM
 $\#E(\mathbb{Q})_{\text{tors}}$
 ist endlich

$$H \geq h(2^{n+1}P) \geq 4h(2^n P) - C = 4H - C \implies H \leq \frac{C}{3}$$

und damit $h(P) \leq H \leq C/3$. Es folgt

$$E(\mathbb{Q})_{\text{tors}} \subset \{P \in E(\mathbb{Q}) \mid h(P) \leq C/3\}.$$

und die rechts stehende Menge ist wegen Eigenschaft (2) endlich.

Da man die Elemente der rechts stehenden Menge im Prinzip explizit bestimmen kann (man betrachtet alle $\xi = u/v$ mit $|u|, |v| \leq e^{C/3}$ und stellt fest, welche davon x -Koordinaten von Punkten in $E(\mathbb{Q})$ sind), liefert das einen weiteren (allerdings nicht sehr effizienten) Algorithmus für die Bestimmung von $E(\mathbb{Q})_{\text{tors}}$. Man beachte, dass $e^{C/3} = \sqrt[3]{2A}$ ist mit A wie im Beweis von Satz 22.5; A ergibt sich als explizites Polynom in $|a|$ und $|b|$. ♠

Man kann mit ähnlichen Argumenten wie im Beweis von Satz 22.5 die folgende stärkere Aussage zeigen:

22.7. Satz. Sei E eine elliptische Kurve über \mathbb{Q} . Dann gibt es eine Konstante c_E , die nur von (der Gleichung von) E abhängt, sodass für alle Punkte $P, Q \in E(\mathbb{Q})$ gilt

SATZ
 angenäherte
 Parallelo-
 gramm-
 gleichung

$$|h(P + Q) + h(P - Q) - 2h(P) - 2h(Q)| \leq c_E.$$

Insbesondere gilt auch

$$|h(2P) - 4h(P)| \leq c_E.$$

Das kann man dann für die folgende Konstruktion benutzen.

22.8. Satz. Sei E eine elliptische Kurve über \mathbb{Q} und sei $P \in E(\mathbb{Q})$. Dann existiert

SATZ
 kanonische
 Höhe

$$\hat{h}(P) := \hat{h}_E(P) := \lim_{n \rightarrow \infty} \frac{h(2^n P)}{4^n} \in \mathbb{R}_{\geq 0}.$$

Die Funktion $\hat{h}: E(\mathbb{Q}) \rightarrow \mathbb{R}_{\geq 0}$ hat folgende Eigenschaften:

(1) Es gibt γ_E mit $|\hat{h}(P) - h(P)| \leq \gamma_E$ für alle $P \in E(\mathbb{Q})$.

(2) Für jedes $B \geq 0$ ist $\{P \in E(\mathbb{Q}) \mid \hat{h}(P) \leq B\}$ endlich.

(3) Für alle $P, Q \in E(\mathbb{Q})$ gilt

$$\hat{h}(P + Q) + \hat{h}(P - Q) = 2\hat{h}(P) + 2\hat{h}(Q).$$

(4) Für alle $P \in E(\mathbb{Q})$ und alle $m \in \mathbb{Z}$ gilt $\hat{h}(mP) = m^2 \hat{h}(P)$.

- (5) Für alle $P \in E(\mathbb{Q})$ gilt $\hat{h}(P) = 0 \iff P \in E(\mathbb{Q})_{\text{tors}}$.
- (6) Ist $\phi: E \rightarrow E'$ ein Isomorphismus, dann gilt $\hat{h}_E = \hat{h}_{E'} \circ \phi$.
- (7) \hat{h} induziert eine positiv definite quadratische Form auf $E(\mathbb{Q}) \otimes_{\mathbb{Z}} \mathbb{R}$.

22.9. Definition. Die Funktion \hat{h}_E heißt die *kanonische Höhe* auf E . ◇ DEF

„Kanonisch“ deshalb, weil sie nicht von der konkreten Gleichung von E abhängt (und außerdem algebraisch schöne Eigenschaften hat). kanonische Höhe

Die kanonische Höhe ist ebenfalls eine Höhenfunktion auf $E(\mathbb{Q})$, sogar mit noch besseren Eigenschaften als eine beliebige Höhenfunktion.

Aus (4) und (1) folgt dann übrigens auch, dass h eine Höhenfunktion auf $E(\mathbb{Q})$ ist für jedes $m \geq 2$.

Beweis. Aus Satz 22.7 folgt, dass

$$\left| \frac{h(2^{n+1}P)}{4^{n+1}} - \frac{h(2^n P)}{4^n} \right| = \frac{1}{4^{n+1}} |h(2 \cdot 2^n P) - 4h(2^n P)| \leq \frac{c_E}{4^{n+1}}$$

ist. Daraus folgt für $n \geq m$

$$\left| \frac{h(2^n P)}{4^n} - \frac{h(2^m P)}{4^m} \right| \leq \sum_{k=m}^{n-1} \left| \frac{h(2^{k+1}P)}{4^{k+1}} - \frac{h(2^k P)}{4^k} \right| \leq c_E \sum_{k=m}^{\infty} \frac{1}{4^{k+1}} = \frac{c_E}{3 \cdot 4^m},$$

also ist $(4^{-n}h(2^n P))_{n \geq 0}$ eine Cauchy-Folge, und der Grenzwert existiert.

- (1) Die Überlegung oben zeigt insbesondere, dass $|4^{-n}h(2^n P) - h(P)| \leq c_E/3$ ist. Für $n \rightarrow \infty$ erhalten wir die Behauptung (mit $\gamma_E = c_E/3$).
- (2) Aus $\hat{h}(P) \leq B$ folgt mit (1), dass $h(P) \leq B + \gamma_E$ ist. Die Behauptung folgt also aus der entsprechenden Eigenschaft von h .
- (3) Nach Satz 22.7 ist

$$\left| \frac{h(2^n(P+Q))}{4^n} + \frac{h(2^n(P-Q))}{4^n} - 2\frac{h(2^n P)}{4^n} - 2\frac{h(2^n Q)}{4^n} \right| \leq \frac{c_E}{4^n}.$$

Die Behauptung folgt für $n \rightarrow \infty$.

- (4) Das folgt aus (3) durch Induktion.
- (5) Ist $P \in E(\mathbb{Q})_{\text{tors}}$, dann ist $\{2^n P \mid n \geq 0\}$ endlich, also $h(2^n P)$ beschränkt. Aus der Definition von \hat{h} folgt dann $\hat{h}(P) = 0$. Ist umgekehrt $\hat{h}(P) = 0$, dann ist nach (4) $\hat{h}(mP) = 0$ für alle $m \in \mathbb{Z}$. Nach (2) ist $\{Q \in E(\mathbb{Q}) \mid \hat{h}(Q) = 0\}$ endlich, also hat P nur endlich viele verschiedene Vielfache. Das bedeutet $P \in E(\mathbb{Q})_{\text{tors}}$.
- (6) Man überlegt sich ähnlich wie im Beweis von Satz 22.5, dass $h_{E'}(\phi(P)) \leq h_E(P) + c_\phi$ ist. Umgekehrt gilt auch $h_E(\phi^{-1}(Q)) \leq h_{E'}(Q) + c_{\phi^{-1}}$. Der Unterschied zwischen h_E und $h_{E'} \circ \phi$ ist also beschränkt. Wie oben folgt die Behauptung durch einen geeigneten Grenzübergang.
- (7) Zunächst einmal folgt aus (3) und (5), dass \hat{h} eine quadratische Form auf (der freien abelschen Gruppe) $E(\mathbb{Q})/E(\mathbb{Q})_{\text{tors}}$ induziert. Wir bekommen dann eine quadratische Form auf dem Vektorraum $V_{\mathbb{Q}} := E(\mathbb{Q}) \otimes_{\mathbb{Z}} \mathbb{Q}$, die außerhalb von $\mathbf{0}$ stets echt positive Werte hat. Es folgt, dass \hat{h} auf $V := E(\mathbb{Q}) \otimes_{\mathbb{Z}} \mathbb{R} = V_{\mathbb{Q}} \otimes_{\mathbb{Q}} \mathbb{R}$ positiv semidefinit ist. Aus (2) folgt, dass es $\varepsilon > 0$ gibt, sodass die Torsionspunkte die einzigen $P \in E(\mathbb{Q})$ sind mit $\hat{h}(P) < \varepsilon$.

Sei $\mathbf{0} \neq \mathbf{x} \in V$, dann liegt \mathbf{x} in $V' := \langle P_1, \dots, P_n \rangle_{\mathbb{R}}$ für endlich viele Punkte $P_1, \dots, P_n \in E(\mathbb{Q})$. Das Bild von $\langle P_1, \dots, P_n \rangle_{\mathbb{Z}} \subset E(\mathbb{Q})$ in V' ist ein Gitter Λ mit $\Lambda \otimes_{\mathbb{Z}} \mathbb{R} = V'$; wir können $\dim V' = n$ annehmen. Wäre $\hat{h}(\mathbf{x}) = 0$, dann wäre \hat{h} auf V' in einer geeigneten Basis gegeben durch $x_1^2 + \dots + x_m^2$ mit $m < n$. Insbesondere wäre die Menge $\{\mathbf{y} \in V' \mid \hat{h}(\mathbf{y}) < \varepsilon\}$ eine zentralsymmetrische konvexe Menge mit unendlichem Volumen, die aber Λ nur im Nullpunkt schneidet. Das ist ein Widerspruch zum Gitterpunktsatz von Minkowski (siehe auch das Skript zur Vorlesung „Diophantische Gleichungen“). Also ist $\hat{h}(\mathbf{x}) > 0$. \square

23. DER SCHWACHE SATZ VON MORDELL

Wir wollen im Folgenden den schwachen Satz von Mordell in der Form

$$E(\mathbb{Q})/2E(\mathbb{Q}) \text{ ist endlich}$$

beweisen unter der Voraussetzung, dass $E(\mathbb{Q})[2] \neq \{O\}$ ist. Es gibt also einen rationalen Punkt T der Ordnung 2 auf E . Wir können die x -Koordinate unserer gegebenen kurzen Weierstraß-Gleichung so verschieben, dass $T = (0, 0)$ ist. Dann ist E gegeben durch eine Gleichung der Form

$$(23.1) \quad E: y^2 = x(x^2 + ax + b).$$

Weil E eine elliptische Kurve ist, gilt dann $b \neq 0$ und $a^2 - 4b \neq 0$. Dann hat E eine Isogenie ϕ vom Grad 2 auf die Kurve

$$E': y^2 = x(x^2 - 2ax + (a^2 - 4b)),$$

die zusammen mit der dualen Isogenie gegeben ist durch

$$\begin{aligned} \phi: E &\longrightarrow E', & (x, y) &\longmapsto \left(\frac{y^2}{x^2}, \frac{b - x^2}{x^2} y \right) = \left(\frac{x^2 + ax + b}{x}, \frac{b - x^2}{x^2} y \right) \quad \text{und} \\ \hat{\phi}: E' &\longrightarrow E, & (x, y) &\longmapsto \left(\frac{y^2}{4x^2}, \frac{a^2 - 4b - x^2}{8x^2} y \right). \end{aligned}$$

Es ist $\ker(\phi) = \{O, T\}$ und $\ker(\hat{\phi}) = \{O', T'\}$, wobei $O' = (0 : 1 : 0) \in E'(\mathbb{Q})$ und $T' = (0, 0) \in E'(\mathbb{Q})$ sind. Siehe Beispiel 10.12.

23.1. Lemma. *Seien E, E', ϕ und $\hat{\phi}$ wie oben. Sind die Gruppen $E(\mathbb{Q})/\hat{\phi}(E'(\mathbb{Q}))$ und $E'(\mathbb{Q})/\phi(E(\mathbb{Q}))$ beide endlich, dann ist auch $E(\mathbb{Q})/2E(\mathbb{Q})$ endlich.*

LEMMA
 $E'(\mathbb{Q})/\phi(E(\mathbb{Q}))$
endlich
genügt

Beweis. Es ist $2E(\mathbb{Q}) = (\hat{\phi} \circ \phi)(E(\mathbb{Q})) \subset \hat{\phi}(E'(\mathbb{Q}))$, also haben wir einen kanonischen Epimorphismus

$$\alpha: \frac{E(\mathbb{Q})}{2E(\mathbb{Q})} \longrightarrow \frac{E(\mathbb{Q})}{\hat{\phi}(E'(\mathbb{Q}))}.$$

Offensichtlich ist $\ker(\alpha) = \hat{\phi}(E'(\mathbb{Q}))/2E(\mathbb{Q})$. Der Homomorphismus

$$\beta: \frac{E'(\mathbb{Q})}{\phi(E(\mathbb{Q}))} \longrightarrow \frac{E(\mathbb{Q})}{2E(\mathbb{Q})}, \quad P' + \phi(E(\mathbb{Q})) \longmapsto \hat{\phi}(P') + 2E(\mathbb{Q})$$

ist wohldefiniert (denn $\hat{\phi}(\phi(E(\mathbb{Q}))) \subset 2E(\mathbb{Q})$), und $\text{im}(\beta) = \ker(\alpha)$. Es folgt

$$\# \frac{E(\mathbb{Q})}{2E(\mathbb{Q})} = \# \ker(\alpha) \cdot \# \text{im}(\alpha) = \# \text{im}(\beta) \cdot \# \frac{E(\mathbb{Q})}{\hat{\phi}(E'(\mathbb{Q}))} \leq \# \frac{E'(\mathbb{Q})}{\phi(E(\mathbb{Q}))} \cdot \# \frac{E(\mathbb{Q})}{\hat{\phi}(E'(\mathbb{Q}))}$$

und daraus dann die Behauptung. □

Für den Beweis des schwachen Satzes von Mordell genügt es also zu zeigen, dass $E(\mathbb{Q})/\hat{\phi}(E'(\mathbb{Q}))$ in der obigen Situation endlich ist; die Endlichkeit der anderen Gruppe $E'(\mathbb{Q})/\phi(E(\mathbb{Q}))$ ergibt sich genauso (bis auf eine Skalierung von x und y ist $E = E''$).

Wir schreiben im Folgenden

$$\mathbb{Q}^{\times 2} := \{\alpha^2 \mid \alpha \in \mathbb{Q}^{\times}\}.$$

23.2. **Lemma.** Die Abbildung

LEMMA
 $x(P)$ mod
 Quadrate

$$\delta: E(\mathbb{Q}) \longrightarrow \frac{\mathbb{Q}^\times}{\mathbb{Q}^{\times 2}}, \quad P \longmapsto \begin{cases} \mathbb{Q}^{\times 2}, & P = O, \\ b \cdot \mathbb{Q}^{\times 2}, & P = T, \\ x(P) \cdot \mathbb{Q}^{\times 2}, & \text{sonst} \end{cases}$$

ist ein Gruppenhomomorphismus; es ist $\ker(\delta) = \hat{\phi}(E'(\mathbb{Q}))$.

Beweis. Wir müssen zeigen, dass $\delta(P + Q) = \delta(P) \cdot \delta(Q)$ ist für $P, Q \in E(\mathbb{Q})$. Das ist klar für $P = O$ oder $Q = O$ und auch für $P = Q = T$ und für $P + Q = O$ (denn $\delta(-P) = \delta(P)$). Ist (z.B.) $Q = T$ und $P \neq T$, dann folgt die Behauptung aus $x(P + T) = b/x(P)$, denn

$$\delta(P + T) = \frac{b}{x(P)} \cdot \mathbb{Q}^{\times 2} = x(P) \cdot b \cdot \mathbb{Q}^{\times 2} = \delta(P) \cdot \delta(T).$$

Sind $P, Q \notin \{O, T\}$, aber $P + Q = T$, dann ist $Q = -P + T$, und die Behauptung folgt aus derselben Rechnung. Wir können also annehmen, dass P, Q und $P + Q$ verschieden von O und T sind. Dann sind $P, Q, R := -(P + Q)$ die Schnittpunkte einer Geraden $y = \lambda x + \mu$ mit E . Es folgt

$$x(x^2 + ax + b) - (\lambda x + \mu)^2 = (x - x(P))(x - x(Q))(x - x(R));$$

Einsetzen von $x = 0$ zeigt, dass $x(P)x(Q)x(R) = \mu^2 \in \mathbb{Q}^{\times 2}$ ist (beachte, dass alle drei x -Koordinaten von null verschieden sind). Das ist äquivalent zu

$$\delta(P + Q) = \delta(R) = \delta(P) \cdot \delta(Q).$$

Wir zeigen jetzt, dass $\hat{\phi}(E'(\mathbb{Q})) \subset \ker(\delta)$ ist. Sei $P' \in E'(\mathbb{Q})$. Ist $P' \in \ker(\hat{\phi}) = \{O', T'\}$, dann ist $\delta(\hat{\phi}(P')) = \delta(O) = \mathbb{Q}^{\times 2}$. Ist $x(\hat{\phi}(P')) = 0$, dann muss $y(P') = 0$ sein und $P' \neq T'$. Es folgt, dass $x^2 - 2ax + a^2 - 4b$ über \mathbb{Q} in Linearfaktoren zerfällt; insbesondere ist $a^2 - (a^2 - 4b) = 4b$ und damit b ein Quadrat. Damit ist $\delta(\hat{\phi}(P')) = \delta(T) = \mathbb{Q}^{\times 2}$. Wir nehmen jetzt an, dass $x(\hat{\phi}(P')) \notin \{0, \infty\}$ ist. Die explizite Form von $\hat{\phi}$ oben zeigt, dass $x(\hat{\phi}(P'))$ dann ein Quadrat $\neq 0$ ist, also ist $\delta(\hat{\phi}(P')) = \mathbb{Q}^{\times 2}$.

Es bleibt die umgekehrte Inklusion zu zeigen. Sei also $P \in \ker(\delta)$. Ist $P = O$, dann ist $P \in \hat{\phi}(E'(\mathbb{Q}))$. Ist $P = T$, dann ist b ein Quadrat, woraus folgt (s.o.), dass es einen Punkt $P' = (\xi, 0) \neq T'$ in $E'(\mathbb{Q})$ gibt; es ist dann $\hat{\phi}(P') = T$. Ist $P \notin \{O, T\}$, dann ist $x(P) = \alpha^2$ ein von null verschiedenes Quadrat. Aus der Gleichung von E folgt, dass dann $\alpha^4 + a\alpha^2 + b$ ein Quadrat ist (nämlich $(y(P)/\alpha)^2$). Wir suchen einen Punkt $P' \in E'(\mathbb{Q})$ mit $\hat{\phi}(P') = P$. Wir setzen $y = 2\alpha x$ in der Gleichung von E' ; nach Kürzen von x erhalten wir die Gleichung

$$x^2 - 2(a + 2\alpha^2)x + a^2 - 4b = 0.$$

Es ist

$$(a + 2\alpha^2)^2 - (a^2 - 4b) = 4(a\alpha^2 + \alpha^4 + b) = \left(\frac{2y(P)}{\alpha}\right)^2,$$

also hat die Gleichung eine Lösung $\xi \in \mathbb{Q}$, und $P' = (\xi, 2\alpha\xi)$ ist ein rationaler Punkt auf E' . Dann ist $\hat{\phi}(P') \in E(\mathbb{Q})$ ein Punkt mit derselben x -Koordinate wie P , also gilt entweder $P = \hat{\phi}(P')$ oder $P = -\hat{\phi}(P') = \hat{\phi}(-P')$. \square

23.3. Folgerung. Seien $E, E', \phi, \hat{\phi}$ wie oben und δ wie in Lemma 23.2. Hat δ endliches Bild, dann ist auch $E(\mathbb{Q})/\hat{\phi}(E'(\mathbb{Q}))$ endlich.

FOLG
 $\text{im}(\delta)$ endlich
 genügt

Beweis. Aus Lemma 23.2 folgt, dass δ einen Isomorphismus

$$\frac{E(\mathbb{Q})}{\hat{\phi}(E'(\mathbb{Q}))} = \frac{E(\mathbb{Q})}{\ker(\delta)} \xrightarrow{\cong} \text{im}(\delta)$$

induziert. □

Die Idee für das weitere Vorgehen wird nun sein, eine endliche obere Schranke für $\text{im}(\delta)$ zu finden, also eine endliche Untergruppe $S_{\hat{\phi}} \subset \mathbb{Q}^{\times}/\mathbb{Q}^{\times 2}$ mit $\text{im}(\delta) \subset S_{\hat{\phi}}$.

Die Gruppe $\mathbb{Q}^{\times}/\mathbb{Q}^{\times 2}$ selbst ist unendlich. Sie ist ein \mathbb{F}_2 -Vektorraum abzählbar unendlicher Dimension mit Basis $(-1) \cdot \mathbb{Q}^{\times 2}$ und $p \cdot \mathbb{Q}^{\times 2}$ für jede Primzahl p (das folgt leicht aus der eindeutigen Primfaktorzerlegung in \mathbb{Z}). Insbesondere enthält jede Nebenklasse einen eindeutigen Repräsentanten, der eine quadratfreie ganze Zahl ist.

Wir wollen jetzt zeigen, dass diese quadratfreie ganze Zahl, die $\delta(P)$ repräsentiert, nur gewisse Primteiler haben kann.

Sei dazu $P \in E(\mathbb{Q})$ mit $P \notin \{O, T\}$. Aus den Überlegungen am Anfang von Abschnitt 20 ergibt sich, dass es ganze Zahlen r, s, t gibt mit $r \perp t$ und $s \perp t$, sodass $P = (r/t^2, s/t^3)$ ist. Eingesetzt in die Gleichung (23.1) von E erhalten wir nach Beseitigung der Nenner

$$s^2 = r(r^2 + art^2 + bt^4);$$

außerdem ist $\delta(P) = r \cdot \mathbb{Q}^{\times 2}$. Wir nehmen an, dass a und b ganzzahlig sind.

Sei p ein Primteiler von r mit $v_p(r)$ ungerade. Weil $v_p(s^2) = 2v_p(s)$ gerade ist, muss p ein Teiler von $r^2 + art^2 + bt^4$ sein. Es folgt

$$p \mid \text{ggT}(r, r^2 + art^2 + bt^4) = \text{ggT}(r, bt^4) = \text{ggT}(r, b) \mid b,$$

weil $r \perp t$ ist. Wir sehen, dass $\delta(P)$ in der Untergruppe von $\mathbb{Q}^{\times}/\mathbb{Q}^{\times 2}$ liegt, die von den Klassen von -1 und den Primteilern von b erzeugt wird. Das gilt auch für $P = O$ (klar) und $P = T$ (denn $\delta(T) = b \cdot \mathbb{Q}^{\times 2}$).

23.4. Satz. Sei E eine elliptische Kurve über \mathbb{Q} mit $E(\mathbb{Q})[2] \neq \{O\}$. Dann ist $E(\mathbb{Q})$ endlich erzeugt.

SATZ
 Satz von
 Mordell

Beweis. E ist isomorph zu einer elliptischen Kurve der Form (23.1); wir können also annehmen, dass E diese Form hat. Dann ist $b \neq 0$ und $a^2 - 4b \neq 0$. Nach geeigneter Skalierung von x und y können wir $a, b \in \mathbb{Z}$ annehmen. Dann sind die Untergruppen

$$H = \langle (-1) \cdot \mathbb{Q}^{\times 2}, p \cdot \mathbb{Q}^{\times 2} \mid p \text{ prim}, p \mid b \rangle \quad \text{und} \\ H' = \langle (-1) \cdot \mathbb{Q}^{\times 2}, p \cdot \mathbb{Q}^{\times 2} \mid p \text{ prim}, p \mid a^2 - 4b \rangle$$

von $\mathbb{Q}^{\times}/\mathbb{Q}^{\times 2}$ endlich. Wir haben in der Diskussion vor dem Satz gesehen, dass $\text{im}(\delta) \subset H$ ist; genauso ergibt sich $\text{im}(\delta') \subset H'$, wenn $\delta': E'(\mathbb{Q}) \rightarrow \mathbb{Q}^{\times}/\mathbb{Q}^{\times 2}$ die analoge Abbildung ist. Es folgt mit Folgerung 23.3, dass $E(\mathbb{Q})/\hat{\phi}(E'(\mathbb{Q}))$ und $E'(\mathbb{Q})/\phi(E(\mathbb{Q}))$ beide endlich sind. Aus Lemma 23.1 folgt dann, dass $E(\mathbb{Q})/2E(\mathbb{Q})$ endlich ist, d.h., der schwache Satz von Mordell für E und $m = 2$. Nach Satz 22.5 existiert eine Höhenfunktion auf $E(\mathbb{Q})$. Die Behauptung folgt somit aus Satz 22.2. □

Eine endlich erzeugte abelsche Gruppe G ist isomorph zu $G_{\text{tors}} \times \mathbb{Z}^r$ mit einem $r \in \mathbb{Z}_{\geq 0}$; dabei ist die Torsionsuntergruppe G_{tors} endlich (für $G = E(\mathbb{Q})$ wissen wir das bereits).

23.5. Definition. Sei E eine elliptische Kurve über \mathbb{Q} mit $E(\mathbb{Q}) \cong E(\mathbb{Q})_{\text{tors}} \times \mathbb{Z}^r$. **DEF**
Dann heißt $\text{rk}(E(\mathbb{Q})) := r \in \mathbb{Z}_{\geq 0}$ der *Rang von $E(\mathbb{Q})$* . \diamond **Rang**

23.6. Lemma. Sei E eine elliptische Kurve über \mathbb{Q} mit Rang r . Dann ist **LEMMA**
$$r = \dim_{\mathbb{F}_2} \frac{E(\mathbb{Q})}{2E(\mathbb{Q})} - \dim_{\mathbb{F}_2} E(\mathbb{Q})[2].$$
 Rang und $E(\mathbb{Q})/2E(\mathbb{Q})$

Beweis. Es ist

$$\frac{E(\mathbb{Q})}{2E(\mathbb{Q})} \cong \frac{E(\mathbb{Q})_{\text{tors}}}{2E(\mathbb{Q})_{\text{tors}}} \times \left(\frac{\mathbb{Z}}{2\mathbb{Z}} \right)^r.$$

Sei $\mu: E(\mathbb{Q})_{\text{tors}} \rightarrow E(\mathbb{Q})_{\text{tors}}, T \mapsto 2T$. Es ist

$$2E(\mathbb{Q})_{\text{tors}} = \text{im}(\mu) \cong \frac{E(\mathbb{Q})_{\text{tors}}}{\ker(\mu)} = \frac{E(\mathbb{Q})_{\text{tors}}}{E(\mathbb{Q})[2]},$$

also $\#E(\mathbb{Q})[2] = \#(E(\mathbb{Q})_{\text{tors}}/2E(\mathbb{Q})_{\text{tors}})$. Die Behauptung folgt aus

$$\dim_{\mathbb{F}_2} \frac{E(\mathbb{Q})}{2E(\mathbb{Q})} = \dim_{\mathbb{F}_2} \frac{E(\mathbb{Q})_{\text{tors}}}{2E(\mathbb{Q})_{\text{tors}}} + r = \dim_{\mathbb{F}_2} E(\mathbb{Q})[2] + r. \quad \square$$

23.7. Definition. Sei $n \in \mathbb{Z}$ mit $n \neq 0$. Wir schreiben $\omega(n)$ für die Anzahl der **DEF**
verschiedenen Primteiler von n . \diamond **$\omega(n)$**

23.8. Folgerung. Seien $a, b \in \mathbb{Z}$ mit $b, a^2 - 4b \neq 0$, und sei **FOLG**
$$E: y^2 = x(x^2 + ax + b).$$
 Schranke für r

Dann ist E eine elliptische Kurve über \mathbb{Q} , und es gilt

$$\text{rk}(E(\mathbb{Q})) \leq \omega(b) + \omega(a^2 - 4b).$$

Beweis. Aus dem Beweis von Satz 23.4 ergibt sich

$$\dim \text{im}(\delta) \leq \dim H = 1 + \omega(b) \quad \text{und} \quad \dim \text{im}(\delta') \leq \dim H' = 1 + \omega(a^2 - 4b).$$

Weiterhin ist

$$\dim \frac{E(\mathbb{Q})}{\hat{\phi}(E'(\mathbb{Q}))} = \dim \text{im}(\delta) \quad \text{und} \quad \dim \frac{E'(\mathbb{Q})}{\phi(E(\mathbb{Q}))} = \dim \text{im}(\delta').$$

Es ist $\dim E(\mathbb{Q})[2] \in \{1, 2\}$ (immer ≤ 2 und > 0 wegen der Existenz von T der Ordnung 2). Im Fall $\dim E(\mathbb{Q})[2] = 2$ erhalten wir

$$\begin{aligned} \text{rk}(E(\mathbb{Q})) &= \dim \frac{E(\mathbb{Q})}{2E(\mathbb{Q})} - \dim E(\mathbb{Q})[2] \\ &\leq \dim \frac{E(\mathbb{Q})}{\hat{\phi}(E'(\mathbb{Q}))} + \dim \frac{E'(\mathbb{Q})}{\phi(E(\mathbb{Q}))} - 2 \\ &\leq (1 + \omega(b)) + (1 + \omega(a^2 - 4b)) - 2 = \omega(b) + \omega(a^2 - 4b). \end{aligned}$$

Im Fall $\dim E(\mathbb{Q})[2] = 1$ ist $T' \notin \phi(E(\mathbb{Q}))$, denn die Urbilder von T' unter ϕ sind gerade die beiden anderen Punkte ($\neq T$) der Ordnung 2 auf E . Die Nebenklasse

$T' + \phi(E(\mathbb{Q}))$ ist dann ein nichttriviales Element des Kerns der Abbildung β , die wir im Beweis von Lemma 23.1 betrachtet haben. Es folgt, dass

$$\dim \ker(\alpha) = \dim \operatorname{im}(\beta) \leq \dim \frac{E'(\mathbb{Q})}{\phi(E(\mathbb{Q}))} - 1$$

ist. Damit ist dann wieder (mit α, β wie im Beweis von Lemma 23.1)

$$\begin{aligned} \operatorname{rk}(E(\mathbb{Q})) &= \dim \frac{E(\mathbb{Q})}{2E(\mathbb{Q})} - \dim E(\mathbb{Q})[2] \\ &= \dim \operatorname{im}(\alpha) + \dim \ker(\alpha) - 1 \\ &\leq \dim \frac{E(\mathbb{Q})}{\hat{\phi}(E'(\mathbb{Q}))} + \dim \frac{E'(\mathbb{Q})}{\phi(E(\mathbb{Q}))} - 2 \\ &\leq \omega(b) + \omega(a^2 - 4b). \end{aligned} \quad \square$$

Die Schranke, die wir hier bewiesen haben, ist leider nie scharf. Im nächsten Abschnitt werden wir überlegen, wie man sie verbessern kann.

23.9. Bemerkung. Was kann man tun, wenn E keinen rationalen Punkt der Ordnung 2 hat? Sei $T = (\xi, 0) \in E[2]$ (wir nehmen an, dass in der Gleichung von E die Koeffizienten a_1 und a_3 null sind), dann ist $K = \mathbb{Q}(\xi)$ eine kubische Körpererweiterung von \mathbb{Q} und $T \in E(K)[2]$. Man kann dann E über K betrachten und mit dem im Wesentlichen gleichen Beweis zeigen, dass $E(K)/2E(K)$ endlich ist. Dazu braucht man zwei grundlegende Ergebnisse aus der algebraischen Zahlentheorie (Endlichkeit der Klassenzahl, endliche Erzeugtheit der Einheitengruppe), mit deren Hilfe man wieder endliche Untergruppen H und H' von $K^\times/K^{\times 2}$ bekommt, die $\operatorname{im}(\delta)$ bzw. $\operatorname{im}(\delta')$ enthalten. Man muss dann noch zeigen, dass die natürliche Abbildung $E(\mathbb{Q})/2E(\mathbb{Q}) \rightarrow E(K)/2E(K)$ endlichen Kern hat. (Sie muss nicht injektiv sein, denn es ist möglich, dass ein Punkt $P \in E(\mathbb{Q})$, der in $E(\mathbb{Q})$ nicht durch 2 teilbar ist, in $E(K)$ durch 2 teilbar wird.) Das kann man folgendermaßen sehen:

BEM
Der allgemeine Fall

Indem wir K eventuell vergrößern, können wir annehmen, dass K eine endliche Galois-Erweiterung von \mathbb{Q} ist (ist $E: y^2 = f(x)$, dann kann man den Zerfällungskörper von f nehmen). Der Kern der betrachteten Abbildung kann dadurch nur größer werden. Sei $\phi: E(\mathbb{Q}) \rightarrow E(K)/2E(K)$ die kanonische Abbildung. Sei jetzt $P \in \ker(\phi)$. Dann gibt es $Q \in E(K)$ mit $P = 2Q$. Wir definieren eine Abbildung $\delta_P: \operatorname{Gal}(K/\mathbb{Q}) \rightarrow E[2]$ durch $\delta_P(\sigma) = \sigma(Q) - Q$. Es gilt

$$2(\sigma(Q) - Q) = \sigma(2Q) - 2Q = \sigma(P) - P = 0$$

(denn $P \in E(\mathbb{Q})$), also ist $\delta_P(\sigma) \in E[2]$. Die Abbildung δ_P hängt von der Wahl von Q ab; wir fixieren für jedes P ein Q . Das liefert uns eine Abbildung

$$\delta: \ker(\phi) \longrightarrow \operatorname{Abb}(\operatorname{Gal}(K/\mathbb{Q}), E[2]), \quad P \longmapsto \delta_P.$$

Gilt jetzt $\delta_P = \delta_{P'}$, dann ist also für alle $\sigma \in \operatorname{Gal}(K/\mathbb{Q})$

$$\sigma(Q) - Q = \sigma(Q') - Q' \implies \sigma(Q - Q') = Q - Q';$$

das bedeutet, dass $R = Q - Q' \in E(\mathbb{Q})$ ist. Es folgt

$$P = 2Q = 2(Q' + R) = 2Q' + 2R = P' + 2R,$$

also ist $P + 2E(\mathbb{Q}) = P' + 2E(\mathbb{Q})$. Repräsentanten von verschiedenen Elementen des Kerns von $E(\mathbb{Q})/2E(\mathbb{Q}) \rightarrow E(K)/2E(K)$ werden also unter δ auf verschiedene

Abbildungen $\text{Gal}(K/\mathbb{Q}) \rightarrow E[2]$ abgebildet. Es folgt

$$\# \ker \left(\frac{E(\mathbb{Q})}{2E(\mathbb{Q})} \rightarrow \frac{E(K)}{2E(K)} \right) \leq \# \text{Abb}(\text{Gal}(K/\mathbb{Q}), E[2]) = 4^{[K:\mathbb{Q}]} < \infty.$$

Statt den Kern dieser Abbildung zu beschränken, kann man auch die Definition von h geeignet verallgemeinern, sodass sie auch für $E(K)$ anwendbar ist. Dann zeigt man direkt, dass $E(K)$ endlich erzeugt ist, woraus die Behauptung auch für $E(\mathbb{Q})$ folgt (denn $E(\mathbb{Q}) \subset E(K)$). Die Schwierigkeit hierbei ist, dass man im Allgemeinen keine (bis auf Vorzeichen) eindeutige Darstellung der x -Koordinate als „gekürzter Bruch“ mehr hat.

Eine Alternative, die keine algebraische Zahlentheorie benötigt, geht so vor, dass man eine injektive Abbildung δ auf $E(\mathbb{Q})/2E(\mathbb{Q})$ definiert, deren Werte Äquivalenzklassen von homogenen Polynomen vom Grad 4 in zwei Variablen sind (bezüglich einer geeigneten Äquivalenzrelation). Man zeigt dann, dass nur endlich viele solche Klassen als Bilder in Frage kommen. Dafür muss man die sogenannte Invariantentheorie dieser Quartiken genauer studieren. Siehe [Cre]. ♠

24. EINE BESSERE SCHRANKE FÜR DEN RANG

Aus den Überlegungen im vorigen Abschnitt ergibt sich, dass wir den Rang einer elliptischen Kurve E über \mathbb{Q} mit $E(\mathbb{Q})[2] \neq \{O\}$ bestimmen können, wenn wir die Bilder von δ und δ' bestimmen können: Der Beweis von Folgerung 23.8 zeigt, dass

$$(24.1) \quad \text{rk}(E(\mathbb{Q})) = \dim \text{im}(\delta) + \dim \text{im}(\delta') - 2$$

ist. Das Problem ist, dass es nicht so einfach ist, die Bilder von δ und δ' exakt zu bestimmen. (Es gibt tatsächlich bisher kein Verfahren, von dem man beweisen konnte, dass es das immer leistet.)

Sei wieder

$$E: y^2 = x(x^2 + ax + b)$$

mit $a, b \in \mathbb{Z}$, $b, a^2 - 4b \neq 0$. Wir werden der Einfachheit halber im Folgenden einfach d schreiben, wenn wir $d\mathbb{Q}^{\times 2}$ meinen.

Wir wissen, dass $\text{im}(\delta) \subset H$ ist, wobei H die von (den Quadratklassen von) -1 und den Primteilern von b erzeugte Untergruppe von $\mathbb{Q}^{\times}/\mathbb{Q}^{\times 2}$ ist. Die Aufgabe, $\text{im}(\delta)$ zu bestimmen, ist also äquivalent dazu, für jedes Element $d \in H$ zu entscheiden, ob es $P \in E(\mathbb{Q})$ gibt mit $\delta(P) = d$, also mit $x(P) = dt^2$ für $t \in \mathbb{Q}^{\times}$ (wir lassen die Fälle $P = O$ und $P = T$ beiseite, da wir ihre Bilder 1 und b unter δ kennen). Eingesetzt in die Gleichung von E bekommen wir

$$y^2 = dt^2(d^2t^4 + adt^2 + b)$$

oder

$$\left(\frac{y}{dt}\right)^2 = dt^4 + at^2 + \frac{b}{d}.$$

Wenn wir d als quadratfrei wählen, dann ist $b/d \in \mathbb{Z}$, da d bis aufs Vorzeichen ein Produkt von Primteilern von b ist. Wenn wir $t = u/v$ mit ganzen Zahlen u, v schreiben und $w = yv^3/(du)$ setzen, dann bekommen wir

$$(24.2) \quad w^2 = du^4 + au^2v^2 + \frac{b}{d}v^4.$$

Jeder Punkt $P \in E(\mathbb{Q})$ mit $\delta(P) = d$ führt also zu einer Lösung von (24.2) in ganzen Zahlen u, v, w mit $u \perp v$. Das stimmt auch noch für $P = O$ und $P = T$: Im ersten Fall haben wir die Lösung $(u, v, w) = (1, 0, 1)$ mit $d = 1$, im zweiten Fall die Lösung $(u, v, w) = (0, 1, 1)$ mit $d = b$.

Die Existenz einer ganzzahligen Lösung mit $u \perp v$ ist äquivalent zur Existenz einer rationalen Lösung mit $(u, v) \neq (0, 0)$, denn mit (u, v, w) ist auch $(\lambda u, \lambda v, \lambda^2 w)$ eine Lösung. Wenn wir d durch $s^2 d$ ersetzen mit $s \in \mathbb{Q}^{\times}$, dann bekommen wir eine äquivalente Gleichung, indem wir u und w mit s skalieren. Die Lösbarkeit von Gleichung (24.2) hängt also tatsächlich nur von der Quadratklasse von d ab.

Haben wir umgekehrt eine Lösung (u, v, w) von (24.2) mit $(u, v) \neq (0, 0)$, dann ist

$$P = \left(d \frac{u^2}{v^2}, d \frac{uw}{v^3}\right) \in E(\mathbb{Q})$$

(im Fall $v = 0$ setzen wir $P = O$) mit $\delta(P) = d$.

Wie können wir nun entscheiden, ob (24.2) lösbar ist? Falls es eine Lösung gibt, können wir (jedenfalls im Prinzip) eine finden. Nachzuweisen, dass es keine Lösung gibt, ist im Allgemeinen schwieriger. Eine Möglichkeit dafür ist, die Gleichung modulo n zu betrachten mit geeignetem $n \geq 2$. Hat die Kongruenz

$$w^2 \equiv du^4 + au^2v^2 + \frac{b}{d}v^4 \pmod{n}$$

keine Lösung in \mathbb{Z} mit $\text{ggT}(u, v, n) = 1$, dann ist (24.2) nicht nichttrivial lösbar. Aus dem Chinesischen Restsatz folgt, dass es genügt, den Fall zu betrachten, dass n eine Primzahlpotenz ist. Es ist auch möglich, dass die rechte Seite der Gleichung stets negativ ist; dann kann es ebenfalls keine Lösung geben.

24.1. **Beispiel.** Wir betrachten die elliptische Kurve

$$E: y^2 = x^3 + x = x(x^2 + 1).$$

Die isogene Kurve ist

$$E': y^2 = x(x^2 - 4).$$

Wir haben die folgenden Schranken für die Bilder von δ und von δ' :

$$\text{im}(\delta) \subset H = \langle -1 \rangle, \quad \text{im}(\delta') \subset H' = \langle -1, 2 \rangle.$$

Für $-1 \in H$ erhalten wir aus (24.2) die Gleichung

$$w^2 = -u^4 - v^4,$$

die nicht einmal eine reelle Lösung mit $(u, v) \neq (0, 0)$ hat. Es folgt $-1 \notin \text{im}(\delta)$ und damit $\text{im}(\delta) = \{1\}$. Mit (24.1) erhalten wir

$$0 \leq \text{rk}(E(\mathbb{Q})) = \dim \text{im}(\delta) + \dim \text{im}(\delta') - 2 \leq 0 + 2 - 2 = 0$$

und daraus $\text{rk}(E(\mathbb{Q})) = 0$. Es ist demnach

$$E(\mathbb{Q}) = E(\mathbb{Q})_{\text{tors}} = \{O, (0, 0)\};$$

die zweite Gleichheit ergibt sich z.B. aus dem Satz 20.5 von Nagell und Lutz.

(Es ergibt sich, dass $\text{im}(\delta') = H'$ sein muss. Tatsächlich ist

$$\delta'(0, 0) = -1, \quad \delta'(2, 0) = 2, \quad \delta'(-2, 0) = -2.)$$



24.2. **Beispiel.** Diesmal betrachten wir

$$E: y^2 = x(x^2 + 10x + 8), \quad E': y^2 = x(x^2 - 20x + 68).$$

Es ist

$$\text{im}(\delta) \subset H = \langle -1, 2 \rangle \quad \text{und} \quad \text{im}(\delta') \subset H' = \langle -1, 2, 17 \rangle.$$

Wir wissen, dass $\langle 2 \rangle \subset \text{im}(\delta)$ ist, denn $\delta(T) = 8 = 2$ (modulo Quadrate). Es genügt also, (z.B.) $d = -1$ zu betrachten. Das ergibt die Gleichung

$$w^2 = -u^4 + 10u^2v^2 - 8v^4,$$

die die Lösung $(u, v, w) = (1, 1, 1)$ hat. Es ist also $-1 = \delta(-1, 1) \in \text{im}(\delta)$ und somit $\text{im}(\delta) = H$.

Auf der anderen Seite haben wir $\langle 17 \rangle \subset \text{im}(\delta')$, denn $\delta'(T') = 68 = 17$. Wir betrachten $d = -1$; das ergibt

$$w^2 = -u^4 - 20u^2v^2 - 68v^4.$$

Die rechte Seite ist stets negativ und somit gibt es keine Lösung. Dasselbe Argument funktioniert für jedes $d < 0$; das zeigt

$$\text{im}(\delta') \subset \langle 2, 17 \rangle.$$

Nun betrachten wir $d = 2$. Die Gleichung ist

$$w^2 = 2u^4 - 20u^2v^2 + 34v^4.$$

Sie wird gelöst von $(u, v, w) = (1, 1, 4)$. Es folgt $\text{im}(\delta') = \langle 2, 17 \rangle$ und damit dann

$$\text{rk}(E(\mathbb{Q})) = \dim \text{im}(\delta) + \dim \text{im}(\delta') - 2 = 2 + 2 - 2 = 2.$$



BSP
Bestimmung
von $E(\mathbb{Q})$

BSP
Bestimmung
von $\text{rk}(E(\mathbb{Q}))$

Wir sehen, dass man häufig Werte von d ausschließen kann, weil sie zu negativ definiten Quartiken auf der rechten Seite von (24.2) führen. Das folgende Lemma gibt dafür Kriterien.

24.3. Lemma. Sei $E: y^2 = x(x^2 + ax + b)$ mit $a, b \in \mathbb{Z}$ eine elliptische Kurve. Ist $a^2 - 4b < 0$ oder ($a \leq 0$ und $b > 0$), dann ist

$$\text{im}(\delta) \subset \langle p \mid p \text{ prim, } p \mid b \rangle.$$

LEMMA
negativ
definite
Quartik

Beweis. Die Aussage ist äquivalent dazu, dass negative d nicht als Werte von δ auftreten können. Sei also $d = -|d| < 0$. Die Quartik auf der rechten Seite von (24.2) ist dann

$$-(|d|u^4 - au^2v^2 + (b/|d|)v^4).$$

Die Diskriminante der quadratischen Form $|d|X^2 - aXY + (b/|d|)Y^2$ ist $a^2 - 4|d|(b/|d|) = a^2 - 4b$. Ist sie negativ, dann kann die Klammer oben nur positive Werte annehmen (für $(u, v) \neq (0, 0)$); die Gleichung hat also keine nichttriviale reelle Lösung. Ist $a \leq 0$ und $b > 0$, dann sind alle Terme in der Klammer ≥ 0 , und wir haben ebenfalls keine nichttriviale reelle Lösung. \square

Umgekehrt gilt, dass im Fall $a^2 - 4b > 0$ und ($a > 0$ oder $b < 0$) die Gleichung zu $d < 0$ eine nichttriviale reelle Lösung hat (Übung); wenn man nur die reelle Lösbarkeit betrachtet, ist die Aussage des Lemmas also optimal.

Man kann Lemma 24.3 benutzen (Übung), um die Schranke aus Folgerung 23.8 zu verbessern zu

$$\text{rk}(E(\mathbb{Q})) \leq \omega(b) + \omega(a^2 - 4b) - 1.$$

Dass das aber im Allgemeinen nicht genügt, zeigt das folgende Beispiel.

24.4. Beispiel. Seien

$$E: y^2 = x(x^2 - 15x + 63) \quad \text{und} \quad E': y^2 = x(x^2 + 30x - 27).$$

Es ist $a = -15 < 0$ und $b = 63 > 0$; mit Lemma 24.3 bekommen wir also

$$\text{im}(\delta) \subset \langle 3, 7 \rangle \quad \text{und} \quad \text{im}(\delta') \subset \langle -1, 3 \rangle.$$

Außerdem ist $\langle 7 \rangle \subset \text{im}(\delta)$ und $\langle -3 \rangle \subset \text{im}(\delta')$. Wir versuchen zu entscheiden, ob $3 \in \text{im}(\delta)$ ist. Die entsprechende Gleichung ist

$$w^2 = 3u^4 - 15u^2v^2 + 21v^4$$

mit der Lösung $(u, v, w) = (1, 1, 3)$. Es folgt $\text{im}(\delta) = \langle 3, 7 \rangle$. Als nächstes betrachten wir $d = 3$ für δ' . Die Gleichung ist

$$w^2 = 3u^4 + 30u^2v^2 - 9v^4.$$

Die rechte Seite ist durch 3 teilbar, also ist $w = 3w_1$ mit $w_1 \in \mathbb{Z}$. Wir erhalten die neue Gleichung

$$3w_1^2 = u^4 + 10u^2v^2 - 3v^4 \equiv u^2(u^2 + v^2) \pmod{3}.$$

Da $u^2 + v^2$ nur durch 3 teilbar sein kann, wenn sowohl u als auch v durch 3 teilbar sind, was der Bedingung $u \perp v$ widerspricht, muss $u = 3u_1$ sein mit $u_1 \in \mathbb{Z}$ (und $3 \nmid v$). Einsetzen ergibt

$$w_1^2 = 27u_1^4 + 30u_1^2v^2 - v^4 \equiv -v^4 \pmod{3}.$$

BSP
Bestimmung
von $\text{rk}(E(\mathbb{Q}))$

Das geht nur, wenn v durch 3 teilbar ist, ein Widerspruch. Also hat die Gleichung keine Lösung, und $\text{im}(\delta') = \langle -3 \rangle$. Insgesamt folgt

$$\text{rk}(E(\mathbb{Q})) = \dim \text{im}(\delta) + \dim \text{im}(\delta') - 2 = 2 + 1 - 2 = 1. \quad \clubsuit$$

Diese Beispiele motivieren die folgenden Definitionen.

24.5. Definition. Wir sagen, dass eine Gleichung der Form (24.2) *überall lokal lösbar* ist, wenn sie nichttriviale Lösungen in \mathbb{R} und nichttriviale Lösungen modulo n hat für alle $n \in \mathbb{Z}_{\geq 2}$. \diamond

DEF
überall
lokal
lösbar

„Nichttrivial“ bedeutet dabei über einem Körper, dass $(u, v) \neq (0, 0)$ ist, und modulo n , dass $\text{ggT}(u, v, n) = 1$ ist. Es ist klar, dass eine nichttrivial lösbare Gleichung auch überall lokal lösbar ist.

24.6. Definition. Seien E, E' und $\hat{\phi}$ wie üblich. Dann heißt

$$S_{\hat{\phi}} := \{d \in \mathbb{Q}^\times / \mathbb{Q}^{\times 2} \mid (24.2) \text{ ist überall lokal lösbar}\}$$

DEF
Selmer-
Gruppe

die *Selmer-Gruppe* der Isogenie $\hat{\phi}$. \diamond

Die hier definierte Menge $S_{\hat{\phi}}$ ist tatsächlich eine Untergruppe von $\mathbb{Q}^\times / \mathbb{Q}^{\times 2}$.

Das kann man sich folgendermaßen überlegen. Die Lösbarkeit überall lokal ist äquivalent zur nichttrivialen Lösbarkeit in \mathbb{R} und in \mathbb{Q}_p für alle Primzahlen p . Über jedem Körper $K \supset \mathbb{Q}$ gilt (mit demselben Beweis wie für \mathbb{Q}), dass die Gleichung zu d genau dann in K nichttrivial lösbar ist, wenn es $P \in E(K)$ gibt mit $\delta_K(P) = d$ (mit $\delta_K: E(K) \rightarrow K^\times / K^{\times 2}$). Sind die Gleichungen zu d und zu d' nichttrivial lösbar in K , dann gibt es also $P, P' \in E(K)$ mit $\delta_K(P) = d$ und $\delta_K(P') = d'$. Weil δ_K ein Homomorphismus ist, folgt $\delta_K(P + P') = dd'$, und daraus, dass auch die Gleichung zu dd' in K nichttrivial lösbar ist. Wenn man das verwendet für $K = \mathbb{R}$ und $K = \mathbb{Q}_p$ für alle p , dann sieht man, dass $S_{\hat{\phi}}$ eine Gruppe ist.

24.7. Lemma. *Mit den üblichen Bezeichnungen gilt $S_{\hat{\phi}} \subset H$.*

LEMMA
 $S_{\hat{\phi}} \subset H$

Beweis. Sei $d \notin H$. Dann gibt es eine Primzahl p mit $p \nmid b$ und $p \mid d$; wir schreiben $d = pd'$ mit $p \nmid d'$. Die Gleichung zu d ist äquivalent zu

$$w^2 = d^3u^4 + ad^2u^2v^2 + bdv^4 = p^3d'^3u^4 + p^2ad'^2u^2v^2 + pbd'v^4.$$

Es muss dann $w = pw_1$ mit $w_1 \in \mathbb{Z}$ sein. Das führt auf

$$pw_1^2 = p^2d'^3u^4 + pad'^2u^2v^2 + bd'v^4.$$

Da $p \nmid bd'$, muss $v = pv_1$ sein mit $v_1 \in \mathbb{Z}$. Das führt wiederum auf

$$w_1^2 = pd'^3u^4 + p^2ad'^2u^2v_1^2 + p^3bd'v_1^4$$

und dann auf $w_1 = pw_2$ mit $w_2 \in \mathbb{Z}$, also

$$pw_2^2 = d'^3u^4 + pad'^2u^2v_1^2 + p^2bd'v_1^4.$$

Dann muss aber auch u durch p teilbar sein, im Widerspruch zu $u \perp v$. Also gibt es keine nichttriviale Lösung modulo p^4 ; damit ist $d \notin S_{\hat{\phi}}$. \square

Wir zeigen jetzt, dass wir aus $S_{\hat{\phi}}$ und S_{ϕ} eine Schranke für den Rang bekommen.

24.8. **Lemma.** *Im Kontext der obigen Definition gilt*

$$\frac{E(\mathbb{Q})}{\hat{\phi}(E'(\mathbb{Q}))} \cong \text{im}(\delta) \subset S_{\hat{\phi}}.$$

LEMMA
Schranke
für $\text{im}(\delta)$

Insbesondere haben wir die Abschätzung

$$\text{rk}(E(\mathbb{Q})) \leq \dim S_{\hat{\phi}} + \dim S_{\phi} - 2.$$

Beweis. Die Isomorphie $E(\mathbb{Q})/\hat{\phi}(E'(\mathbb{Q})) \cong \text{im}(\delta)$ hatten wir bereits gezeigt. Ist $d \in \text{im}(\delta)$, dann hat die Gleichung zu d eine ganzzahlige Lösung mit $u \perp v$. Diese Lösung ist dann auch eine nichttriviale reelle Lösung und eine nichttriviale Lösung mod n für alle $n \geq 2$. Also ist $d \in S_{\hat{\phi}}$. Die zweite Aussage folgt dann aus (24.1). \square

24.9. **Bemerkung.** Man kann allgemeiner für jede Isogenie $\varphi: E' \rightarrow E$ von elliptischen Kurven über \mathbb{Q} (oder allgemeiner über algebraischen Zahlkörpern) eine Selmer-Gruppe S_{φ} definieren zusammen mit einer Abbildung $\delta: E(\mathbb{Q}) \rightarrow S_{\varphi}$ mit $\ker(\delta) = \varphi(E'(\mathbb{Q}))$. Die Selmer-Gruppe S_{φ} ist endlich und kann (im Prinzip jedenfalls) berechnet werden. (Die Berechenbarkeit im Fall einer Isogenie vom Grad 2 werden wir gleich noch zeigen.) Ist $\varphi = [m]$ die Multiplikation mit $m \geq 2$, dann bekommt man direkt eine Schranke für $\text{rk}(E(\mathbb{Q}))$, nämlich

BEM
Selmer-
Gruppen

$$\text{rk}(E(\mathbb{Q})) \leq \log_m \# \frac{S_{[m]}}{\delta(E(\mathbb{Q})_{\text{tors}})}.$$

Anderenfalls braucht man (Schranken für) $\#S_{\varphi}$ und $\#S_{\hat{\varphi}}$, analog zu Lemma 24.8. In der Praxis kann man $S_{[m]}$ berechnen für $m = 2, 3, 4$ und mit Einschränkungen $m = 8, 9$. Für Isogenien φ kann man S_{φ} berechnen, wenn der Grad von φ nicht zu groß ist.

Für die allgemeine Definition der Selmer-Gruppe S_{φ} setzen wir zunächst für jede Körpererweiterung K von \mathbb{Q}

$$\begin{aligned} Z^1(K, \ker(\varphi)) &= \{ \xi: \text{Gal}_K \rightarrow \ker(\varphi) \mid \forall \sigma, \tau \in \text{Gal}_K: \xi(\sigma\tau) = \sigma(\xi(\tau)) - \xi(\sigma) \}, \\ B^1(K, \ker(\varphi)) &= \{ \xi: \text{Gal}_K \rightarrow \ker(\varphi), \sigma \mapsto \sigma(T) - T \mid T \in \ker(\varphi) \} \quad \text{und} \\ H^1(K, \ker(\varphi)) &= Z^1(K, \ker(\varphi)) / B^1(K, \ker(\varphi)). \end{aligned}$$

Die Elemente von $Z^1(K, \ker(\varphi))$ heißen *1-Kozykel auf Gal_K mit Werten in $\ker(\varphi)$* , die Elemente von $B^1(K, \ker(\varphi))$ heißen *1-Koränder auf Gal_K mit Werten in $\ker(\varphi)$* und $H^1(K, \ker(\varphi))$ ist die *erste Galois-Kohomologiegruppe über K mit Werten in $\ker(\varphi)$* . $Z^1(K, \ker(\varphi))$ ist eine Untergruppe von $\text{Abb}(G_K, \ker(\varphi))$ und $B^1(K, \ker(\varphi))$ ist eine Untergruppe von $Z^1(K, \ker(\varphi))$, sodass $H^1(K, \ker(\varphi))$ als Gruppe wohldefiniert ist. Die Abbildung δ ist dann definiert als

$$\delta: E(\mathbb{Q}) \longrightarrow H^1(\mathbb{Q}, \ker(\varphi)), \quad P \longmapsto [\sigma \mapsto \sigma(Q) - Q],$$

wobei $Q \in E'(\bar{\mathbb{Q}})$ ist mit $\varphi(Q) = P$ und $[\xi]$ für die Nebenklasse von ξ in $H^1(\mathbb{Q}, \ker(\varphi))$ steht. Man prüft leicht nach, dass δ wohldefiniert ist (d.h., die Abbildung $\sigma \mapsto \sigma(Q) - Q$ ist in $Z^1(\mathbb{Q}, \ker(\varphi))$), und dieser Kozykel ändert sich um einen Korand, wenn man Q durch ein anderes Urbild von P ersetzt). Für jedes $v = p$ prim oder $v = \infty$ hat man ein kommutatives Diagramm (mit $\mathbb{Q}_{\infty} := \mathbb{R}$)

$$\begin{array}{ccc} E(\mathbb{Q}) & \xrightarrow{\delta} & H^1(\mathbb{Q}, \ker(\varphi)) \\ \downarrow & & \downarrow r_v \\ E(\mathbb{Q}_v) & \xrightarrow{\delta_v} & H^1(\mathbb{Q}_v, \ker(\varphi)) \end{array}$$

und man setzt

$$S_\varphi := \{\xi \in H^1(\mathbb{Q}, \ker(\varphi)) \mid \forall v: r_v(\xi) \in \text{im}(\delta_v)\}.$$

Es ist dann $\text{im}(\delta) \subset S_\varphi$.

Der Zusammenhang mit $S_{\hat{\phi}}$, wie wir die Gruppe definiert haben, ist wie folgt. Die Operation von Gal_K (für jeden Körper $K \supset \mathbb{Q}$) auf $\ker(\hat{\phi})$ ist trivial. Das bedeutet, dass

$$H^1(K, \ker(\hat{\phi})) \cong \text{Hom}(\text{Gal}_K, \ker(\hat{\phi}))$$

ist. Auf der anderen Seite kann man zeigen, dass

$$\frac{K^\times}{K^{\times 2}} \longrightarrow \text{Hom}(\text{Gal}_K, \{\pm 1\}), \quad \alpha K^{\times 2} \longmapsto \left(\sigma \mapsto \frac{\sigma(\sqrt{\alpha})}{\sqrt{\alpha}}\right)$$

ein Isomorphismus ist. Als „Galois-Modul“ (d.h., als abelsche Gruppe mit Operation von Gal_K) ist $\ker(\hat{\phi})$ isomorph zu $\{\pm 1\}$, sodass wir unser δ als Komposition des ersten Isomorphismus mit dem Inversen des zweiten erhalten. So können wir dann $S_{\hat{\phi}}$ als Untergruppe von $\mathbb{Q}^\times/\mathbb{Q}^{\times 2}$ definieren, und die Bedingung „ $r_v(\xi) \in \text{im}(\delta_v)$ “ oben übersetzt sich in die nichttriviale Lösbarkeit in \mathbb{R} und in \mathbb{Q}_p für alle Primzahlen p der zu $\xi \leftrightarrow d\mathbb{Q}^{\times 2}$ gehörenden Gleichung. ♠

Als nächstes wollen wir uns überlegen, dass man $S_{\hat{\phi}}$ algorithmisch bestimmen kann. Da wir schon wissen, dass $S_{\hat{\phi}} \subset H$ mit einer expliziten endlichen Gruppe H ist, ist diese Aufgabe dazu äquivalent, für ein gegebenes $d \in H$ festzustellen, ob die Gleichung

$$w^2 = du^4 + au^2v^2 + \frac{b}{d}v^4$$

überall lokal lösbar ist. Lemma 24.3 gibt ein Kriterium für die nichttriviale Lösbarkeit in \mathbb{R} im Fall $d < 0$ (für $d > 0$ gibt es stets die Lösung $(u, v, w) = (1, 0, \sqrt{d})$). Es bleibt also zu entscheiden, ob die Gleichung eine nichttriviale Lösung mod p^e hat für alle Primzahlen p und alle $e \geq 1$.

24.10. Lemma. *Seien $a, c, d \in \mathbb{Z}$ und sei p eine ungerade Primzahl mit $p \nmid cd$ und $p \nmid a^2 - 4cd$. Dann hat die Gleichung*

$$w^2 = du^4 + au^2v^2 + cv^4$$

nichttriviale Lösungen mod p^e für alle $e \geq 1$.

LEMMA
Lösbarkeit
für fast
alle p

Beweis. Wir schreiben \bar{n} für das Bild von $n \in \mathbb{Z}$ in \mathbb{F}_p . Sei $E: y^2 = x(x^2 + \bar{a}x + \bar{c}\bar{d})$; E ist eine elliptische Kurve über \mathbb{F}_p . Es gibt eine Isogenie $\hat{\phi}: E' \rightarrow E$ vom Grad 2. Man kann die Abbildung $\delta: E(\mathbb{F}_p) \rightarrow \mathbb{F}_p^\times/\mathbb{F}_p^{\times 2}$ genauso definieren wie über \mathbb{Q} , und man sieht ebenfalls in derselben Weise, dass die nichttriviale Lösbarkeit mod p der Gleichung im Lemma dazu äquivalent ist, dass es einen Punkt $P \in E(\mathbb{F}_p)$ gibt mit $\delta(P) = \bar{d}\mathbb{F}_p^{\times 2}$. Ebenso ist $\ker(\delta) = \hat{\phi}(E'(\mathbb{F}_p))$. Die Behauptung für $e = 1$ ist dann dazu äquivalent, dass δ surjektiv ist. Nun sind $E'(\mathbb{F}_p)$ und $E(\mathbb{F}_p)$ endliche abelsche Gruppen. Nach (der einfachen Richtung von) Satz 12.3 haben sie dieselbe Ordnung. Es folgt

$$\begin{aligned} \#\text{im}(\delta) &= (E(\mathbb{F}_p) : \ker(\delta)) = (E(\mathbb{F}_p) : \hat{\phi}(E'(\mathbb{F}_p))) \\ &= \frac{\#E(\mathbb{F}_p)}{\#\hat{\phi}(E'(\mathbb{F}_p))} = \frac{\#E(\mathbb{F}_p)}{\#E'(\mathbb{F}_p)/\#\ker(\hat{\phi})} = \#\ker(\hat{\phi}) = 2 = \#\frac{\mathbb{F}_p^\times}{\mathbb{F}_p^{\times 2}}, \end{aligned}$$

also ist δ surjektiv.

Es gibt also eine nichttriviale Lösung $(u, v, w) \text{ mod } p$. Wir können u und v so skalieren, dass $\bar{u} = 1$ oder $\bar{v} = 1$ ist. Da es nur auf die Restklassen von u, v, w

mod p ankommt, können wir also annehmen, dass $u = 1$ oder $v = 1$ ist. Ohne Einschränkung sei $v = 1$ (der andere Fall ist symmetrisch). Es ist dann also

$$w^2 \equiv du^4 + au^2 + c \pmod{p}.$$

Ist p kein Teiler von w , dann ist das Bild der rechten Seite in \mathbb{F}_p ein von null verschiedenes Quadrat. Aus dem Henselschen Lemma 24.11 folgt dann, dass die rechte Seite ein Quadrat mod p^e ist für alle $e \geq 1$; das zeigt die Behauptung für $p \nmid w$.

Gilt $p \mid w$, dann ist \bar{u} eine Nullstelle von $\bar{d}x^4 + \bar{a}x^2 + \bar{c}$. Diese Nullstelle ist einfach, da p die Diskriminante $a^2 - 4cd$ von $dx^2 + ax + c$ nicht teilt und $\bar{u} \neq 0$ ist. Wiederum nach dem Henselschen Lemma 24.11 folgt, dass es für jedes $e \geq 1$ ein $u_e \in \mathbb{Z}$ gibt, sodass $du_e^4 + au_e^2 + c \equiv 0 \pmod{p^e}$ ist. Das zeigt die Behauptung auch im Fall $p \mid w$. \square

Im Beweis haben wir folgende Aussage benutzt:

24.11. Satz. *Seien $f \in \mathbb{Z}[x]$ ein Polynom und p eine Primzahl. Sei weiter $u \in \mathbb{Z}$ mit $p \mid f(u)$ und $p \nmid f'(u)$. Dann gibt es für jedes $e \geq 1$ eine ganze Zahl u_e mit $u_e \equiv u \pmod{p}$ und $p^e \mid f(u_e)$. Dabei ist die Restklasse von $u_e \pmod{p^e}$ eindeutig bestimmt.*

SATZ
Henselsches
Lemma

Beweis. Wir zeigen die Aussage durch Induktion. Für $e = 1$ ist sie durch die Voraussetzungen gegeben. Sei also $u_e \in \mathbb{Z}$ mit $u_e \equiv u \pmod{p}$ und $f(u_e) = \alpha p^e$ mit $\alpha \in \mathbb{Z}$. Wir wissen aus der Induktionsannahme, dass u_e modulo p^e eindeutig bestimmt ist. Das bedeutet, dass $u_{e+1} \equiv u_e \pmod{p^e}$ sein muss, also setzen wir $u_{e+1} = u_e + xp^e$. Wir erhalten

$$f(u_{e+1}) = f(u_e + xp^e) \equiv f(u_e) + f'(u_e)xp^e = (\alpha + f'(u_e)x)p^e \pmod{p^{e+1}}.$$

Es ist $f'(u_e) \equiv f'(u) \not\equiv 0 \pmod{p}$, also hat die Kongruenz

$$\alpha + f'(u_e)x \equiv 0 \pmod{p}$$

eine Lösung, die modulo p eindeutig bestimmt ist. Das zeigt die Existenz von u_{e+1} und die Eindeutigkeit modulo p^{e+1} . \square

24.12. Folgerung. *Seien $f \in \mathbb{Z}[x]$ ein Polynom und p eine Primzahl. Seien weiter $u \in \mathbb{Z}$ und $e_0 = v_p(f'(u))$. Gilt $v_p(f(u)) > 2e_0$, dann gibt es für jedes $e > 2e_0$ eine ganze Zahl u_e mit $u_e \equiv u \pmod{p^{e_0+1}}$ und $p^e \mid f(u_e)$.*

FOLG
Henselsches
Lemma
(Variante)

Beweis. Sei $F(x) = p^{-2e_0} f(u + p^{e_0}x) \in \mathbb{Z}[x]$. Die Behauptung folgt aus Satz 24.11 für $(f, p, u) \leftarrow (F, p, 0)$. \square

Lemma 24.10 reduziert das Problem also auf endlich viele Primzahlen p , nämlich $p = 2$, die Primteiler von b und die Primteiler von $a^2 - 4b$. Es bleibt zu zeigen, dass man die nichttriviale Lösbarkeit unserer Gleichung modulo allen Potenzen von p für eine Primzahl p entscheiden kann. Das Henselsche Lemma wird dabei wieder die entscheidende Rolle spielen.

24.13. Lemma. Seien $a, c, d \in \mathbb{Z}$ mit $cd \neq 0$ und $a^2 - 4cd \neq 0$. Sei weiter p eine Primzahl und $e_0 = 2v_p(4cd(a^2 - 4cd)) + 1$. Die Gleichung

$$w^2 = du^4 + au^2v^2 + cv^4$$

hat genau dann nichttriviale Lösungen mod p^e für alle $e \geq 1$, wenn sie eine nichttriviale Lösung mod p^{e_0} hat.

LEMMA
Lösbarkeit
mod p^e

Beweis. Die Richtung „ \Rightarrow “ ist trivial. Für „ \Leftarrow “ können wir wie im Beweis von Lemma 24.10 ohne Einschränkung annehmen, dass unsere nichttriviale Lösung mod p^{e_0} die Form $(u, 1, w)$ hat. Sei $f(x) = dx^4 + ax^2 + c$. Es muss dann $v_p(f(u))$ entweder $< e_0$ und gerade oder $\geq e_0$ sein.

Sei zunächst p ungerade. Im Fall $v_p(f(u)) < e_0$ ist $p^{-v_p(f(u))}f(u)$ ein quadratischer Rest modulo p ; daraus folgt wieder, dass $f(u)$ ein Quadrat modulo p^e ist für alle $e \geq 1$. Im Fall $v_p(f(u)) \geq e_0$ beachten wir die Relation

$$(4adu^2 + 2(a^2 - 4cd))f(u) - (adu^3 + (a^2 - 2cd)u)f'(u) = 2c(a^2 - 4cd).$$

Wäre hier $v_p(f'(u)) \geq e_0/2$, dann würde wegen $v_p(f(u)) \geq e_0$ auch

$$v_p(2cd(a^2 - 4cd)) \geq e_0/2$$

folgen im Widerspruch zur Definition von e_0 . Also muss $v_p(f'(u)) < e_0/2$ und damit $v_p(f(u)) > 2v_p(f'(u))$ sein. Nach Folgerung 24.12 gibt es dann für jedes e ein $u_e \in \mathbb{Z}$ mit $f(u_e) \equiv 0 \pmod{p^e}$; es ist dann $(u, v, w) = (u_e, 1, 0)$ eine nichttriviale Lösung mod p^e .

Sei jetzt $p = 2$. Es gilt immer noch, dass $v_2(f(u))$ entweder $< e_0$ und gerade oder $\geq e_0$ ist. Ist $v_2(f(u)) \leq e_0 - 3$, dann muss $2^{-v_2(f(u))}f(u) \equiv 1 \pmod{8}$ sein; dann ist aber $f(u)$ ein Quadrat mod 2^e für alle $e \geq 1$. Anderenfalls ist $v_2(f(u)) \geq e_0 - 1$. Aus $v_2(f'(u)) \geq (e_0 - 1)/2$ folgt wie oben, dass $v_2(4cd(a^2 - cd)) > e_0/2$ wäre, was der Definition von e_0 widerspricht. Wie oben folgt dann, dass es für jedes e ein u_e gibt mit $f(u_e) \equiv 0 \pmod{2^e}$, was die Existenz von nichttrivialen Lösungen zeigt. \square

24.14. Folgerung. Sei $\hat{\phi}: E' \rightarrow E$ eine Isogenie vom Grad 2 zwischen zwei elliptischen Kurven über \mathbb{Q} . Dann ist die Selmergruppe $S_{\hat{\phi}}$ berechenbar.

FOLG
Berechenbarkeit
der Selmergruppe

Beweis. Wir können annehmen, dass E die Form $y^2 = x(x^2 + ax + b)$ hat mit $a, b \in \mathbb{Z}$ (dann ist $b \neq 0$ und $a^2 - 4b \neq 0$). Sei H die von -1 und den Primteilern von b erzeugte endliche Untergruppe von $\mathbb{Q}^\times/\mathbb{Q}^{\times 2}$. Es ist $S_{\hat{\phi}} \subset H$, also genügt es, für jedes der endlich vielen Elemente von H zu prüfen, ob es in $S_{\hat{\phi}}$ enthalten ist. Die reelle Lösbarkeit der relevanten Gleichung kann man leicht überprüfen. Lemma 24.10 zeigt, dass wir die Lösbarkeit mod p^e nur für endlich viele Primzahlen p prüfen müssen, und Lemma 24.13 reduziert das für jede Primzahl auf ein endliches Problem. \square

Wir erhalten das folgende Verfahren, mit dem wir versuchen können, den Rang $\text{rk}(E(\mathbb{Q}))$ zu bestimmen, wenn $E(\mathbb{Q})[2] \neq \{O\}$ ist. Sei also $E: y^2 = x(x^2 + ax + b)$ mit $a, b \in \mathbb{Z}$ wie üblich.

1. Bestimme die Primteiler von b und $a^2 - 4b$ und daraus H und H' .
2. Für jedes $d \in H$ und jedes $d' \in H'$, stelle fest, ob $d \in S_{\hat{\phi}}$ bzw. $d' \in S_{\hat{\phi}}$ ist.

3. Versuche, für jedes $d \in S_{\hat{\phi}}$ und für jedes $d' \in S_{\phi}$ eine nichttriviale ganzzahlige Lösung der zugehörigen Gleichung zu finden. Seien T und T' die Teilmengen von $S_{\hat{\phi}}$ und von S_{ϕ} , für die das gelingt.
4. Ist $\langle T \rangle = S_{\hat{\phi}}$ und $\langle T' \rangle = S_{\phi}$, dann ist

$$\text{rk}(E(\mathbb{Q})) = \dim S_{\hat{\phi}} + \dim S_{\phi} - 2.$$

In jedem Fall ist

$$\dim \langle T \rangle + \dim \langle T' \rangle - 2 \leq \text{rk}(E(\mathbb{Q})) \leq \dim S_{\hat{\phi}} + \dim S_{\phi} - 2.$$

Es gibt allerdings keine Garantie, dass dieses Verfahren erfolgreich ist: Eine Gleichung der Form $w^2 = du^4 + au^2v^2 + cv^4$, die nichttriviale Lösungen in \mathbb{R} und modulo n hat für alle $n \geq 2$, muss nicht unbedingt auch nichttriviale Lösungen in \mathbb{Z} haben.

24.15. Beispiel. Die Gleichung

$$w^2 = 2u^4 - 34v^4$$

hat offensichtlich nichttriviale Lösungen in \mathbb{R} . Sie hat auch nichttriviale Lösungen modulo p^e für alle Primzahlen p und Exponenten $e \geq 1$. Für $p \neq 2, 17$ folgt das aus Lemma 24.10. Für $p = 17$ ist $(u, v, w) = (1, 0, 6)$ eine nichttriviale Lösung mod 17, die sich mit dem Henselschen Lemma zu einer Lösung $(1, 0, w_e)$ mod 17^e für jedes $e \geq 1$ hochheben lässt. Für $p = 2$ haben wir, dass 17 eine vierte Potenz modulo jeder Potenz von 2 ist: $17 \equiv 3^4 \pmod{2^5}$, und für die Ableitung $f'(x) = 4x^3$ von $f(x) = x^4 - 17$ gilt $v_2(f'(3)) = 2$; die Behauptung folgt aus Folgerung 24.12. Es gibt also immer eine Lösung der Form $(u_e, 1, 0) \pmod{2^e}$.

BSP
Selmer-
Schranke
nicht scharf

Auf der anderen Seite gibt es aber *keine* nichttriviale ganzzahlige Lösung. Um das zu zeigen, nehmen wir an, (u, v, w) sei eine ganzzahlige Lösung mit $u \perp v$. Dann ist jedenfalls $w \neq 0$. Sei p ein ungerader Primteiler von w . Dann kann p weder u noch v teilen, weil p sonst beide teilen müsste im Widerspruch zu $u \perp v$. Wäre $p = 17$, dann müsste aber $p \mid u$ gelten, was nicht möglich ist. Es folgt dann aus $u^4 \equiv 17v^4 \pmod{p}$, dass 17 ein quadratischer Rest mod p ist. Nach dem Quadratischen Reziprozitätsgesetz ist dann (wegen $17 \equiv 1 \pmod{4}$) auch p ein quadratischer Rest mod 17. Da auch -1 und 2 quadratische Reste mod 17 sind, muss w als Produkt von quadratischen Resten ebenfalls ein quadratischer Rest mod 17 sein. Es gibt also $t \in \mathbb{Z}$ mit $w \equiv t^2 \pmod{17}$. Daraus erhalten wir $t^4 \equiv 2u^4 \pmod{17}$, was (da $2 \equiv x^4 \pmod{17}$ keine Lösung hat) $17 \mid t$ und $17 \mid u$ impliziert. Dann gilt aber auch $17 \mid v$ im Widerspruch zu $u \perp v$. Also kann es keine nichttriviale ganzzahlige Lösung geben. (Dieses Beispiel wurde zuerst unabhängig von Lind¹⁰ und Reichardt¹¹ gefunden.)

Die oben betrachtete Gleichung gehört zu $d = 2 \in S_{\hat{\phi}}$, wenn wir das Paar

$$E: y^2 = x(x^2 - 68), \quad E': y^2 = x(x^2 + 272)$$

betrachten. Es ist $H = H' = \langle -1, 2, 17 \rangle$, und man findet

$$S_{\hat{\phi}} = \langle -1, 2, 17 \rangle = H \quad \text{und} \quad S_{\phi} = \langle 17 \rangle,$$

was die Schranke $\text{rk}(E(\mathbb{Q})) \leq 2$ ergibt. Tatsächlich ist aber

$$\text{im}(\delta) = \delta(\langle T \rangle) = \langle -17 \rangle$$

¹⁰Carl-Erik Lind: *Untersuchungen über die rationalen Punkte der ebenen kubischen Kurven vom Geschlecht Eins*, Uppsala: Diss. 97 S. (1940).

¹¹Hans Reichardt: *Einige im Kleinen überall lösbare, im Grossen unlösbare diophantische Gleichungen*, J. reine angew. Math. **184** (1942), 12–18.

und somit $\text{rk}(E(\mathbb{Q})) = 0$, also $E(\mathbb{Q}) = E(\mathbb{Q})_{\text{tors}} = \{O, T\}$. Die Überlegung oben zeigt, dass $2 \notin \text{im}(\delta)$ ist. Analog zeigt man, dass $-2, 34, -34 \notin \text{im}(\delta)$ sind. Es folgt $\text{im}(\delta) \subset \langle -1, 17 \rangle$. Wir zeigen jetzt noch, dass $-1 \notin \text{im}(\delta)$ ist. Die Gleichung dazu ist

$$(24.3) \quad w^2 = -u^4 + 68v^4.$$

Es folgt, dass u und w gerade sein müssen: $u = 2u_1$, $w = 2w_1$ und

$$w_1^2 = -4u_1^4 + 17v^4$$

mit v ungerade. Mit $U = (u_1/v)^2$ gilt dann $(w_1/v^2)^2 + (2U)^2 = 17$. Wir setzen $w_1/v^2 = 1 - \lambda$ und $U = 2 - \lambda t$ und erhalten

$$\lambda(-2 + \lambda - 16t + 4\lambda t^2) = 0.$$

$\lambda = 0$ führt zu keiner Lösung, da $2 = U = (u_1/v)^2$ nicht lösbar ist. Es ist also

$$\lambda = \frac{4 + 16t}{1 + 4t^2}$$

und damit

$$\frac{w_1}{v^2} = \frac{-1 - 16t + 4t^2}{1 + 4t^2} \quad \text{und} \quad \left(\frac{u_1}{v}\right)^2 = U = \frac{2 - 2t - 8t^2}{1 + 4t^2}.$$

Hier ist $t \in \mathbb{Q}$. (Der Grenzfall $t = \infty$ führt auf $U = -2$, was ebenfalls kein Quadrat ist.) Wenn wir $t = r/s$ als gekürzten Bruch schreiben, dann ergibt sich

$$\frac{u_1^2}{v^2} = \frac{-8r^2 - 2rs + 2s^2}{4r^2 + s^2}.$$

Der ggT von Zähler und Nenner rechts muss ein Teiler von $2^3 \cdot 17$ sein, denn

$$\begin{aligned} (-8r + s)(-8r^2 - 2rs + 2s^2) + (18r - 2s)(4r^2 + s^2) &= 2^3 \cdot 17r^3 \quad \text{und} \\ (2r + 4s)(-8r^2 - 2rs + 2s^2) + (4r + 9s)(4r^2 + s^2) &= 17s^3. \end{aligned}$$

Ist der ggT gerade, dann ist s gerade und r ungerade. Ist s nicht durch 4 teilbar, dann ist $v_2(-8r^2 - 2rs + 2s^2) = 2$ und $v_2(4r^2 + s^2) = 3$; das ist wegen v ungerade nicht möglich. Es muss also $s = 4s_1$ sein; das ergibt

$$\frac{-8r^2 - 2rs + 2s^2}{4r^2 + s^2} = \frac{-2r^2 - 2rs_1 + 8s_1^2}{r^2 + 4s_1^2}$$

mit ungeradem Nenner. Je nachdem, ob der ggT durch 17 teilbar ist, erhalten wir eines der beiden Gleichungssysteme

$$\left\{ \begin{array}{l} u_1^2 = -2r^2 - 2rs_1 + 8s_1^2 \\ v^2 = r^2 + 4s_1^2 \end{array} \right\} \quad \text{oder} \quad \left\{ \begin{array}{l} 17u_1^2 = -2r^2 - 2rs_1 + 8s_1^2 \\ 17v^2 = r^2 + 4s_1^2 \end{array} \right\}.$$

Im Fall, dass der ggT ungerade ist, bekommen wir analog

$$\left\{ \begin{array}{l} u_1^2 = -8r^2 - 2rs + 2s^2 \\ v^2 = 4r^2 + s^2 \end{array} \right\} \quad \text{oder} \quad \left\{ \begin{array}{l} 17u_1^2 = -8r^2 - 2rs + 2s^2 \\ 17v^2 = 4r^2 + s^2 \end{array} \right\}.$$

In jedem Fall muss u_1 gerade sein. Wir setzen $u_1 = 2u_2$ und teilen durch 2. Die linke Seite der ersten Gleichung ist dann immer noch gerade, und wir bekommen

$$r(r + s_1) \equiv 0 \pmod{2} \quad \text{bzw.} \quad s(s - r) \equiv 0 \pmod{2}.$$

Im ersten Fall ist r ungerade, im zweiten Fall ist s ungerade. Es folgt, dass s_1 bzw. r ebenfalls ungerade ist. Dann ist aber $r^2 + 4s_1^2 \equiv 5 \pmod{8}$ bzw. $4r^2 + s^2 \equiv 5 \pmod{8}$, im Widerspruch zu $v^2 \equiv 17v^2 \equiv 1 \pmod{8}$. Also hat die Gleichung (24.3) keine nichttrivialen ganzzahligen Lösungen. ♣

Im Prinzip kann man allgemein auf ähnliche Art zu zeigen versuchen, dass ein Element von $S_{\hat{\phi}}$ nicht im Bild von δ liegt. Eine Lösung von $w^2 = du^4 + au^2v^2 + cv^4$ liefert auch eine Lösung von $w^2 = dX^2 + aXY + cY^2$. Hat diese Gleichungen überall lokal Lösungen, dann hat sie auch eine nichttriviale ganzzahlige Lösung (Satz von Hasse-Minkowski); im Beispiel war das $(X, Y, w) = (8, 1, 2)$. Die Lösungen der Gleichung in w, X, Y kann man dann rational parametrisieren (Parameter t im Beispiel); daraus bekommt man dann ein oder mehrere Gleichungssysteme der Art

$$u^2 = Q_1(r, s), \quad v^2 = Q_2(r, s)$$

mit binären quadratischen Formen Q_1 und Q_2 , die man wiederum auf überall lokale Lösbarkeit untersuchen kann. Allerdings gibt es auch Beispiele, bei denen das auch nicht genügt, um den Rang zu bestimmen.

Wenn man die Bilder von δ und von δ' erfolgreich bestimmt hat, dann hat man auch zu jedem $d \in \text{im}(\delta)$ und zu jedem $d' \in \text{im}(\delta')$ einen Punkt $P_d \in E(\mathbb{Q})$ mit $\delta(P_d) = d$ bzw. einen Punkt $Q_{d'} \in E'(\mathbb{Q})$ mit $\delta'(Q_{d'}) = d'$ gefunden. Dann enthält die Menge

$$R = \{P_d \mid d \in \text{im}(\delta)\} + \{\hat{\phi}(Q_{d'}) \mid d' \in \text{im}(\delta')\}$$

ein vollständiges Repräsentantensystem der Nebenklassen von $2E(\mathbb{Q})$ in $E(\mathbb{Q})$. Das folgt aus der exakten Sequenz (d.h., das Bild eines Homomorphismus ist der Kern des nächsten)

$$\frac{E'(\mathbb{Q})}{\phi(E'(\mathbb{Q}))} \longrightarrow \frac{E(\mathbb{Q})}{2E(\mathbb{Q})} \longrightarrow \frac{E(\mathbb{Q})}{\hat{\phi}(E'(\mathbb{Q}))} \longrightarrow 0,$$

die hinter dem Beweis von Lemma 23.1 steckt. Die erste Abbildung muss nicht injektiv sein (der Kern wird erzeugt von $T' + \phi(E(\mathbb{Q}))$ und kann Ordnung 1 oder 2 haben; siehe auch den Beweis von Folgerung 23.8); deswegen können unter den $\hat{\phi}(Q_{d'})$ jeweils zwei Repräsentanten derselben Restklasse auftreten. In jedem Fall bekommen wir aus Satz 22.2 und mit den expliziten Werten für C und c_P aus dem Beweis von Satz 22.5 eine explizite Schranke γ , sodass $E(\mathbb{Q})$ von R zusammen mit allen Punkten $P \in E(\mathbb{Q})$ mit $h(P) \leq \gamma$ erzeugt wird. Auf diese Weise kann man dann die Struktur und Erzeuger von $E(\mathbb{Q})$ bestimmen. (In der Praxis benutzt man schärfere Abschätzungen für die verschiedenen Konstanten und bekommt eine kleinere Schranke γ , aber das Prinzip bleibt dasselbe.)

24.16. Beispiel. Wir betrachten wieder

$$E: y^2 = x(x^2 - 15x + 63) \quad \text{und} \quad E': y^2 = x(x^2 + 30x - 27).$$

aus Beispiel 24.4. Dort hatten wir gezeigt, dass $\text{rk}(E(\mathbb{Q})) = 1$ ist. Genauer hatten wir gesehen, dass $\text{im}(\delta) = \langle 3, 7 \rangle$ und $\text{im}(\delta') = \langle -3 \rangle$ ist. Die explizite Lösung der Gleichung zu $3 \in \text{im}(\delta)$ ergibt den Punkt $P_3 = P = (3, 9) \in E(\mathbb{Q})$. Ein Urbild von 7 ist $P_7 = T = (0, 0)$. Es folgt, dass $P_{21} = P + T = (21, -63)$ ein Urbild von $21 = 3 \cdot 7$ ist. Ein Urbild von $-3 \in \text{im}(\delta')$ ist $T' = (0, 0) \in E'(\mathbb{Q})$, aber das Bild $\hat{\phi}(T') = O$ liefert nichts Neues. Wir können also

$$R = \{O, (0, 0), (3, 9), (21, -63)\}$$

als Repräsentantensystem von $E(\mathbb{Q})/2E(\mathbb{Q})$ wählen. Aus den Beweisen von Satz 22.2 und von Satz 22.5 bekommen wir die Schranke

$$\gamma = 15,518$$

für die Höhe von Punkten, die $E(\mathbb{Q})$ erzeugen. Wir müssen also alle rationalen Punkte $P = (\xi, \eta)$ finden, sodass Zähler und Nenner von ξ durch $\lfloor e^\gamma \rfloor = 5486637$

BSP
Bestimmung
von $E(\mathbb{Q})$

beschränkt sind. Das klingt schlimmer, als es ist: Das Programm `ratpoints` erledigt das in unter einer Sekunde und findet 48 rationale Punkte mit Höhe $\leq \gamma$. Diese Punkte sind alle in der von T und P erzeugten Untergruppe enthalten. Es folgt

$$E(\mathbb{Q}) = \langle T, P \rangle \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}. \quad \clubsuit$$

Sind die Koeffizienten etwas größer, dann ist die Schranke γ , die wir aus Satz 22.2 und Satz 22.5 bekommen, nicht mehr praktikabel. Es gibt deutlich bessere Schranken für $h(Q) - \hat{h}(Q)$ als was sich aus unseren Überlegungen hier ergibt.

24.17. Beispiel. Wir setzen die Betrachtung der Kurve aus dem obigen Beispiel fort. Wenn P kein Erzeuger des freien Teils von $E(\mathbb{Q})$ ist, dann muss $\bar{P} = m\bar{Q}$ sein mit $m \geq 2$ und $Q \in E(\mathbb{Q})$; hier schreiben wir \bar{Q} für das Bild von Q in $E(\mathbb{Q})/E(\mathbb{Q})_{\text{tors}} \cong \mathbb{Z}$. Das ist äquivalent dazu, dass $P = mQ$ oder $P + T = mQ$ ist, denn $E(\mathbb{Q})_{\text{tors}} = \{O, T\}$, wie man z.B. mit dem Satz 20.5 von Nagell und Lutz sieht. Nun sind aber P und $P + T$ nicht durch 2 teilbar, denn sonst wären sie im Bild von $\hat{\phi}$, und dann müssten sie unter δ auf 1 abgebildet werden, was aber nicht der Fall ist. Daher muss m ungerade sein. Da dann $T = mT$ ist, folgt $P = mQ$. Außerdem ist $m \geq 3$.

BSP
Fortsetzung

Es ist $E_1: y^2 = x^3 - 12x + 65$ eine kurze Weierstraß-Gleichung für E . Die Magma-Funktion `SiksekBound` liefert die Schranke

$$\forall Q \in E_1(\mathbb{Q}): h(Q) \leq \hat{h}(Q) + 3,071.$$

Dann wäre für das Bild Q_1 von Q in $E_1(\mathbb{Q})$ (beachte, dass die kanonischen Höhen von P auf E und von dem entsprechenden Punkt auf E_1 übereinstimmen)

$$h(Q_1) \leq \hat{h}(Q_1) + 3,071 = \frac{1}{m^2} \hat{h}(P) + 3,071 \leq \frac{1}{9} \hat{h}(P) + 3,071 \leq 3,082.$$

Das reduziert die Schranke für Zähler und Nenner der x -Koordinate von Q_1 auf 21 und die Anzahl der Punkte, die man untersuchen muss, auf 18. \clubsuit

25. DIE VERMUTUNG VON BIRCH UND SWINNERTON-DYER

Zum Abschluss der Vorlesung möchte ich noch auf die Vermutung von Birch¹² und Swinnerton-Dyer eingehen. Sie besagt im Wesentlichen, dass die Anzahlen $\#\tilde{E}(\mathbb{F}_p)$ der \mathbb{F}_p -rationalen Punkte auf der Reduktion modulo p einer elliptischen Kurve E über \mathbb{Q} für alle bis auf endlich viele p den Rang $\text{rk}(E(\mathbb{Q}))$ festlegen. Die Heuristik dahinter ist, dass es, wenn der Rang groß ist, in gewisser Weise „viele“ rationale Punkte auf E gibt, deren systematisch vorkommende Bilder unter der Reduktion modulo p im Mittel zu einer etwas höheren Anzahl von \mathbb{F}_p -rationalen Punkten führen. Eine relativ elementare Möglichkeit das auszudrücken, besteht darin,

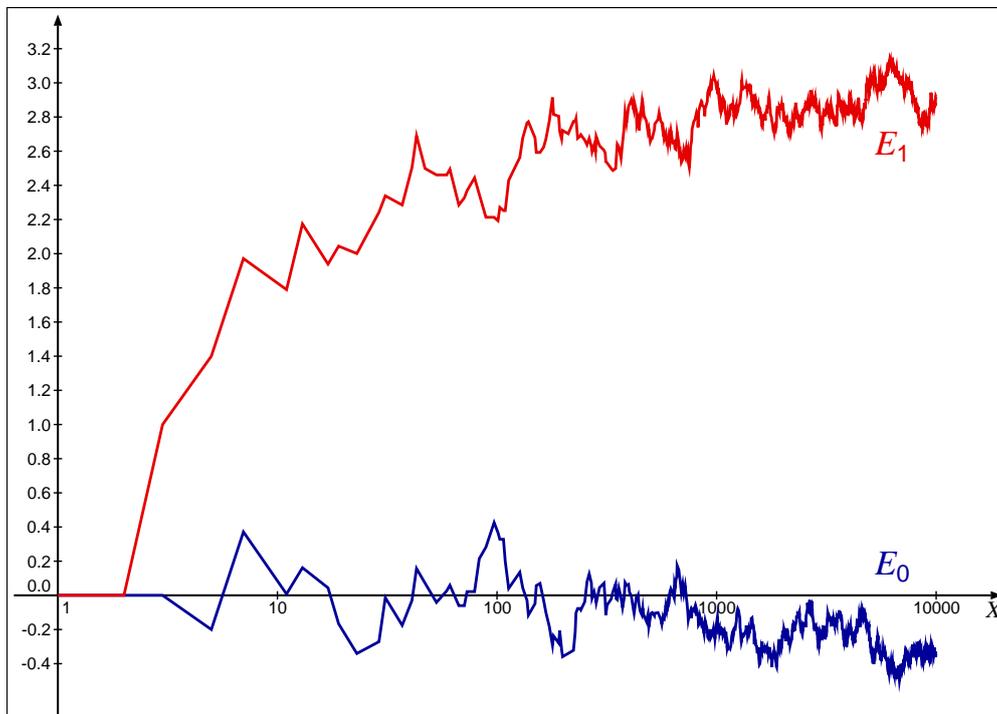
$$\#\tilde{E}(\mathbb{F}_p) = p + 1 + A_p$$

zu schreiben (für Primzahlen p , für die E gute Reduktion hat; vgl. Abschnitt 21); dann ist $|A_p| \leq 2\sqrt{p}$ nach dem Satz 12.2 von Hasse (und $-A_p$ ist die Spur des Frobenius). Um die Abweichung vom Mittel $p+1$ statistisch zu erfassen, summieren wir A_p/p für alle p unterhalb einer Schranke X und betrachten das Verhalten dieser Summe für $X \rightarrow \infty$. Zum Beispiel erhalten wir für die Kurven

$$\begin{aligned} E_0: y^2 &= x^3 - 2x + 1 && \text{mit } \text{rk}(E_0(\mathbb{Q})) = 0 && \text{und} \\ E_1: y^2 &= x^3 - x + 1 && \text{mit } \text{rk}(E_1(\mathbb{Q})) = 1 \end{aligned}$$

das folgende Verhalten von

$$N_E(X) = \sum_{p < X} \frac{A_p}{p} :$$



Man sieht, dass $N_{E_0}(X)$ nahe bei 0 bleibt, während $N_{E_1}(X)$ ein deutliches Wachstum erkennen lässt. Man sieht aber auch, dass das lokale Verhalten dieser Graphen recht erratisch ist.

Um die statistische Tendenz der Zahlen A_p analytisch „schöner“ zu erfassen, betrachtet man statt $N_E(X)$ die sogenannte L -Funktion von E .



B. Birch
* 1931



H.P.F. Sw.-Dyer
1927 – 2018
Foto © MFO

¹²Foto © W. Stein

25.1. Definition. Sei E eine elliptische Kurve über \mathbb{Q} , gegeben durch eine minimale Weierstraß-Gleichung (vgl. Definition 21.1). Das folgende unendliche Produkt über alle Primzahlen p konvergiert für $s \in \mathbb{C}$ mit $\operatorname{Re}(s) > \frac{3}{2}$; die dadurch definierte holomorphe Funktion heißt die (Hasse-Weil-) L -Funktion von E :

DEF
 L -Funktion

$$L(E, s) = \prod_p \frac{1}{1 - a_p p^{-s} + \varepsilon(p) p^{1-2s}}$$

mit

$$\varepsilon(p) = \begin{cases} 0, & \text{falls } p \mid \Delta(E), \\ 1, & \text{sonst} \end{cases}$$

und

$$a_p = \begin{cases} -A_p, & \text{falls } p \nmid \Delta(E), \\ 0, & \text{falls } p \mid c_4(E), c_6(E), \\ \pm 1, & \text{sonst.} \end{cases}$$

Im letzten Fall richtet sich das Vorzeichen danach, ob die beiden Tangentensteigungen im singulären Punkt der modulo p reduzierten Kurve über \mathbb{F}_p definiert sind (+1) oder nicht (-1). \diamond

Im Fall guter Reduktion (also $p \nmid \Delta(E)$) ist a_p die Spur des Frobenius, und der Nenner des entsprechenden Faktors im Produkt ist das „reziproke charakteristische Polynom des Frobenius“ ausgewertet in p^{-s} . Die Bezeichnung erklärt sich dadurch, dass $X^2 - a_p X + p$ das (reduzierte) charakteristische Polynom des Frobenius ϕ als Element des Endomorphismenrings der modulo p reduzierten elliptischen Kurve \tilde{E} ist ($a_p = \phi + \hat{\phi}$ ist die Spur, $p = \deg(\phi) = \phi\hat{\phi}$ die Norm).

Im Fall schlechter Reduktion unterscheidet man zwischen „multiplikativer“ und „additiver“ Reduktion, je nachdem, ob der singuläre Punkt der reduzierten Kurve ein einfacher Doppelpunkt (lokal wie $y^2 = \lambda x^2$ mit $\lambda \neq 0$) oder eine Spitze oder Kuspel ist (lokal wie $y^2 = x^3$). Die übliche Vorschrift, die die Verknüpfung für die Gruppenstruktur definiert, funktioniert auch noch für singuläre Kurven, die durch eine Weierstraß-Gleichung gegeben sind, wenn man den singulären Punkt ausschließt. Im Fall eines einfachen Doppelpunkts erhält man über einem algebraisch abgeschlossenen Grundkörper k eine Gruppe, die zur multiplikativen Gruppe k^\times isomorph ist, während man im Fall einer Spitze eine Gruppe bekommt, die zur additiven Gruppe von k isomorph ist. Das erklärt die Bezeichnungen. Im Fall additiver Reduktion ist der Faktor im Produkt einfach 1. Im Fall multiplikativer Reduktion unterscheidet man noch zwischen zerfallender („split“) und nicht zerfallender („non-split“) multiplikativer Reduktion. Im singulären Punkt hat die reduzierte Kurve zwei Tangenten. Sind deren Steigungen über \mathbb{F}_p definiert, dann ist die multiplikative Reduktion zerfallend, und der Faktor im Produkt ist $1/(1 - p^{-s})$; die Gruppe ist dann \mathbb{F}_p^\times . Im anderen Fall ist der Faktor $1/(1 + p^{-s})$, und die Gruppe ist die Untergruppe der Ordnung $p + 1$ von $\mathbb{F}_{p^2}^\times$. In jedem Fall gilt (immer unter der Voraussetzung, dass die Gleichung von E minimal ist) $\tilde{E}(\mathbb{F}_p) = 1 + p - a_p$.

Die Aussage, dass das Produkt für $\operatorname{Re}(s) > \frac{3}{2}$ konvergiert, folgt aus dem Satz 12.2 von Hasse: Aus $|a_p| \leq 2\sqrt{p}$ folgt für $\operatorname{Re}(s) > \frac{1}{2}$

$$\left| \frac{1}{1 - a_p p^{-s} + \varepsilon(p) p^{1-2s}} - 1 \right| \ll p^{\frac{1}{2} - \operatorname{Re}(s)},$$

und $\sum_p p^{\frac{1}{2}-\text{Re}(s)} < \infty$, sobald $\text{Re}(s) > \frac{3}{2}$ ist. (Ein unendliches Produkt $\prod_j a_j$ konvergiert genau dann (absolut), wenn die unendliche Reihe $\sum_j (a_j - 1)$ (absolut) konvergiert. Das sieht man durch Logarithmieren, denn $\log(1+x) \sim x$ für kleines x .)

Rein formal bekommen wir für $s = 1$

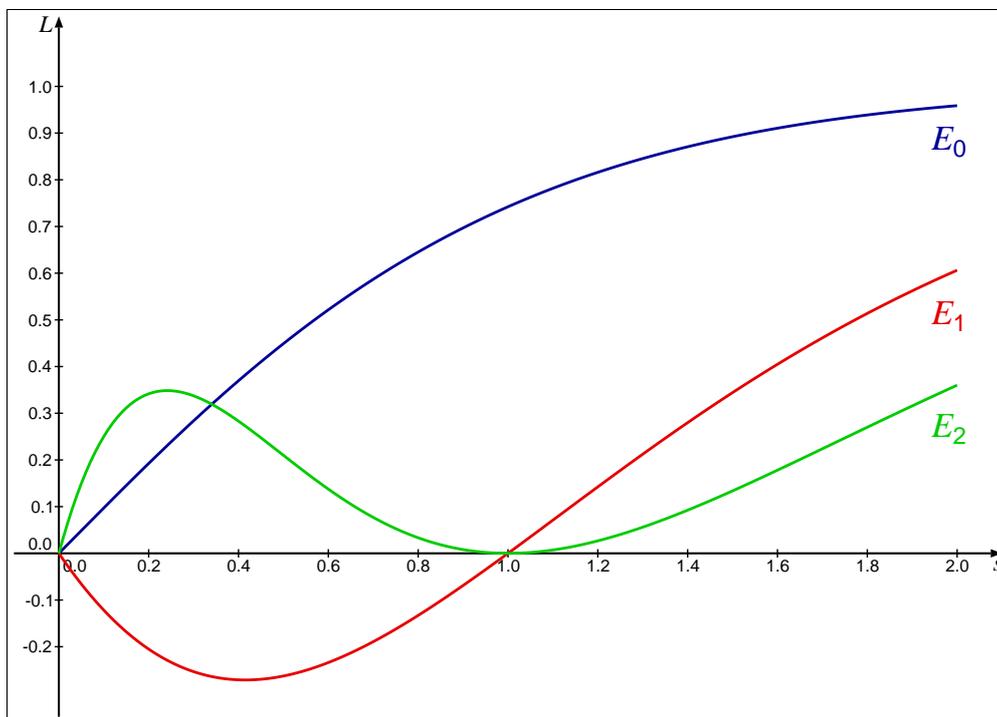
$$\begin{aligned} \log L(E, 1) &\text{ „=“ } \log \prod_p \frac{1}{1 - \frac{a_p}{p} + \frac{\varepsilon(p)}{p}} \\ &= - \sum_p \log \left(1 - \frac{a_p}{p} + \frac{\varepsilon(p)}{p} \right) = - \sum_p \left(\frac{A_p}{p} + O\left(\frac{1}{p}\right) \right), \end{aligned}$$

sodass wir erwarten können, dass das Verhalten von $L(E, s)$ nahe $s = 1$ etwas mit dem Wachstum von $N_E(X)$ für $X \rightarrow \infty$ zu tun hat. Das Problem ist hierbei nur, dass $L(E, s)$ bei $s = 1$ gar nicht definiert ist! Trotzdem haben Birch und Swinnerton-Dyer ca. 1965 folgende Vermutung aufgestellt:

25.2. Vermutung. *Sei E eine elliptische Kurve über \mathbb{Q} . Dann lässt sich die L -Funktion $L(E, s)$ holomorph in eine Umgebung von $s = 1$ fortsetzen, und es gilt $\text{ord}_{s=1} L(E, s) = \text{rk}(E(\mathbb{Q}))$.*

VERM
schwache
BSD-
Vermutung

Für die Kurven E_0 und E_1 und eine weitere Kurve E_2 mit $\text{rk}(E_2(\mathbb{Q})) = 2$ wird das in der folgenden Grafik veranschaulicht, die die Graphen der zugehörigen L -Funktionen auf der positiven reellen Achse zeigt:



Für elliptische Kurven E mit komplexer Multiplikation, also sodass $\text{End}_{\mathbb{Q}}(E)$ echt größer als \mathbb{Z} ist (wie z.B. Kurven der Form $y^2 = x^3 + ax$ oder $y^2 = x^3 + b$), war bekannt, dass ihre L -Funktion übereinstimmt mit einer gewissen anderen L -Funktion (einer sogenannten Hecke- L -Funktion), für die wiederum bewiesen war (Deuring 1941), dass sie eine holomorphe Fortsetzung auf ganz \mathbb{C} hat. Die Vermutung basierte auf vielen numerischen Beispielen für solche Kurven, die auf dem Computer EDSAC 2 in Cambridge gerechnet wurden.

Der erste Teil der Vermutung, der bereits auf Hasse zurückgeht, ist inzwischen bewiesen, sodass der zweite, wesentliche, Teil tatsächlich für alle elliptischen Kurven über \mathbb{Q} eine sinnvolle Aussage ist. Dahinter steckt wiederum, dass man die L -Funktion einer elliptischen Kurve E mit einer anderen Art von L -Funktion identifizieren kann, nämlich der einer Modulform vom Gewicht 2 bezüglich einer geeigneten Untergruppe der Gruppe $\Gamma = \text{PSL}(2, \mathbb{Z})$.

25.3. Definition. Sei $N \in \mathbb{Z}_{\geq 1}$. Wir definieren die Untergruppe $\Gamma_0(N)$ von Γ als

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma \mid c \equiv 0 \pmod{N} \right\}. \quad \diamond$$

DEF
 $\Gamma_0(N)$

(Es gibt auch Untergruppen $\Gamma_1(N)$ und $\Gamma(N)$.)

25.4. Definition. Sei f holomorph auf der oberen Halbebene \mathbb{H} , sei $k \in \mathbb{Z}$ und $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$. Dann setzen wir für $z \in \mathbb{H}$

$$(f|_k \gamma)(z) = (cz + d)^{-k} f(\gamma z). \quad \diamond$$

DEF
 $f|_k \gamma$

Mit dieser Schreibweise gilt, dass f genau dann eine Modulform (für Γ) vom Gewicht k ist, wenn $f|_k \gamma = f$ ist für alle $\gamma \in \Gamma$ und f „in $i\infty$ holomorph“ ist.

Mit diesen Notationen können wir definieren, was eine Modulform vom Gewicht k für $\Gamma_0(N)$ ist.

25.5. Definition. Sei $N \in \mathbb{Z}_{\geq 1}$ und sei $k \in \mathbb{Z}$. Eine *Modulform vom Gewicht k für $\Gamma_0(N)$* ist eine holomorphe Funktion $f: \mathbb{H} \rightarrow \mathbb{C}$, sodass gilt $f|_k \gamma = f$ für alle $\gamma \in \Gamma_0(N)$ und sodass für alle $\gamma \in \Gamma$ die Funktion $f|_k \gamma$ holomorph in $i\infty$ ist. f ist eine *Spitzenform* vom Gewicht k für $\Gamma_0(N)$, wenn zusätzlich $(f|_k \gamma)(i\infty) = 0$ ist für alle $\gamma \in \Gamma$.

DEF
Modulform
für $\Gamma_0(N)$

Die Funktionen $f|_k \gamma$ haben eine q -Entwicklung der Form $\sum_n a_n q^{n/m}$ für ein festes $m \in \mathbb{Z}_{\geq 1}$. So eine Funktion ist dann in $i\infty$ holomorph, wenn $a_n = 0$ ist für alle $n < 0$; dann setzen wir $(f|_k \gamma)(i\infty) := a_0$. Modulformen bzw. Spitzenformen für $\Gamma = \Gamma_0(1)$ im Sinne dieser Definition sind dann genau die früher in Abschnitt 19 definierten Modulformen bzw. Spitzenformen.

Da $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \Gamma_0(N)$ ist für alle N und $(f|_k T)(z) = f(z + 1)$ gilt, sind Modulformen für $\Gamma_0(N)$ periodisch mit Periode 1 und haben deshalb eine q -Entwicklung der üblichen Form

$$f(z) = \sum_{n=0}^{\infty} a_n(f) q^n \quad \text{mit } q = e^{2\pi i z}.$$

Ist f eine Spitzenform, also $a_0(f) = 0$, dann liefert die Mellin-Transformation von f auf der positiven imaginären Halbachse eine Dirichlet-Reihe (bis auf den Faktor $(2\pi)^{-s} \Gamma(s)$):

$$\begin{aligned} \int_0^{\infty} f(it) t^s \frac{dt}{t} &= \sum_{n=1}^{\infty} a_n(f) \int_0^{\infty} e^{-2\pi n t} t^s \frac{dt}{t} \stackrel{2\pi n t \leftarrow u}{=} (2\pi)^{-s} \sum_{n=1}^{\infty} \frac{a_n(f)}{n^s} \int_0^{\infty} e^{-u} u^s \frac{du}{u} \\ &= (2\pi)^{-s} \Gamma(s) \sum_{n=1}^{\infty} \frac{a_n(f)}{n^s} \end{aligned}$$

Hier ist $\Gamma(s)$ die Gammafunktion.

Sei $\mathcal{S}(k, \Gamma_0(N))$ der \mathbb{C} -Vektorraum der Spitzenformen vom Gewicht k für $\Gamma_0(N)$. Für jede ganze Zahl $n \geq 1$ kann man einen Endomorphismus T_n von $\mathcal{S}(k, \Gamma_0(N))$ definieren, einen sogenannten Hecke-Operator. Diese Endomorphismen kommutieren miteinander und erzeugen eine kommutative \mathbb{C} -Unteralgebra des Endomorphismenrings $\text{End}_{\mathbb{C}}(\mathcal{S}(k, \Gamma_0(N)))$. Eine Spitzenform f , die ein simultaner Eigenvektor all dieser Hecke-Operatoren ist, heißt eine (Hecke-)Eigenform. Eine Eigenform f heißt *normiert*, wenn $a_1(f) = 1$ ist. Es ist dann $T_n f = a_n(f) f$. Ist eine normierte Eigenform f nicht von der Form $f(z) = \tilde{f}(dz)$ für ein $\tilde{f} \in \mathcal{S}(k, \Gamma_0(M))$ und ein $d \in \mathbb{Z}_{\geq 1}$ mit $dM \mid N$ und $M < N$, dann heißt f eine *Neuform* von der Stufe („level“) N („neu“, da f nicht von einer niedrigeren, „alten“ Stufe M kommt). Nun gilt Folgendes:

DEF
Eigenform
normiert
Neuform

25.6. Satz. Sei $N \in \mathbb{Z}_{\geq 1}$ und sei $f \in \mathcal{S}(2, \Gamma_0(N))$ eine Neuform mit Koeffizienten $a_n(f) \in \mathbb{Z}$ für alle $n \geq 1$. Dann gibt es eine elliptische Kurve E über \mathbb{Q} , sodass

SATZ
ell. Kurve
zu Neuform

$$L(f, s) := \sum_{n=1}^{\infty} \frac{a_n(f)}{n^s} = L(E, s)$$

ist. Die Funktion $L(f, s)$ lässt sich auf ganz \mathbb{C} holomorph fortsetzen und erfüllt die Funktionalgleichung

$$\Lambda(f, 2 - s) = \pm \Lambda(f, s)$$

(für eines der beiden Vorzeichen) mit

$$\Lambda(f, s) := (2\pi)^{-s} \Gamma(s) N^{s/2} L(f, s).$$

Dabei erhält man durch Entwickeln von $(1 - a_p p^{-s} + \varepsilon(p) p^{1-2s})^{-1}$ in eine formale Potenzreihe in p^{-s} und dann formales Ausmultiplizieren des aus der Produktdarstellung von $L(E, s)$ entstehenden unendlichen Produkts dieser Reihen ebenfalls eine Dirichlet-Reihe.

Insbesondere hat auch $L(E, s)$ dann eine holomorphe Fortsetzung auf ganz \mathbb{C} und erfüllt die Funktionalgleichung. Da isogene elliptische Kurven dieselbe L -Funktion haben (siehe Satz 12.3), ist E hier nur bis auf Isogenie eindeutig bestimmt.

Die Zahl N (die Stufe von f) stimmt dann mit dem *Führer* N_E von E überein. Der Führer hat die folgenden Eigenschaften:

DEF
Führer N_E

- (1) N_E ist ein Teiler der minimalen Diskriminante von E .
- (2) Die Primteiler von N_E sind genau die Primzahlen schlechter Reduktion für E .
- (3) Hat E multiplikative Reduktion bei p , dann ist $v_p(N_E) = 1$.
- (4) Hat E additive Reduktion bei p , dann ist $v_p(N_E) \geq 2$, mit Gleichheit für $p \geq 5$. Es gilt $v_2(N_E) \leq 8$ und $v_3(N_E) \leq 5$; der genaue Wert lässt sich algorithmisch bestimmen.

25.7. Beispiel. Man kann zeigen, dass die Funktion mit q -Entwicklung

BSP
ell. Kurve
zu Neuform

$$\begin{aligned} f(z) &= q \prod_{n=1}^{\infty} ((1 - q^n)(1 - q^{11n}))^2 \\ &= q - 2q^2 - q^3 + 2q^4 + q^5 + 2q^6 - 2q^7 - 2q^9 - 2q^{10} + q^{11} - 2q^{12} + 4q^{13} + \dots \end{aligned}$$

in $\mathcal{S}(2, \Gamma_0(11))$ liegt und dass dieser Raum eindimensional ist. Da 11 prim ist, muss f eine Neuf orm sein. Es gibt eine Isogenieklasse von elliptischen Kurven E mit $N_E = 11$; eine dieser Kurven ist

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

Tatsächlich gilt zum Beispiel

$$\#\tilde{E}(\mathbb{F}_2) = 5 = 2 + 1 - (-2)$$

$$\#\tilde{E}(\mathbb{F}_3) = 5 = 3 + 1 - (-1)$$

$$\#\tilde{E}(\mathbb{F}_5) = 5 = 5 + 1 - 1$$

$$\#\tilde{E}(\mathbb{F}_7) = 10 = 7 + 1 - (-2)$$

$$\#\tilde{E}(\mathbb{F}_{13}) = 10 = 13 + 1 - 4$$

usw. (Die Punkt-Anzahlen sind alle durch 5 teilbar, da $\#E(\mathbb{Q})_{\text{tors}} = 5$ ist.) Man bekommt also gewissermaßen eine explizite Formel für diese Anzahlen! ♣

25.8. Definition. Eine elliptische Kurve E über \mathbb{Q} , für die es eine Neuf orm f gibt mit $L(f, s) = L(E, s)$, heißt *modular*. ◇

DEF
modulare
ell. Kurve

Für modulare elliptische Kurven E ist also der Ausdruck „ $\text{ord}_{s=1} L(E, s)$ “ in Vermutung 25.2 definiert.

Es war lange Zeit eine Vermutung (1958 von Taniyama und Shimura), die recht weit hergeholt schien, dass *alle* elliptischen Kurven über \mathbb{Q} modular sein sollten. Diese Vermutung wurde schließlich Mitte der 1990er Jahre zunächst von Wiles und Taylor für „semistabile“ elliptische Kurven (das sind Kurven, die bei allen Primzahlen p entweder gute oder multiplikative Reduktion haben) und dann 2001 von Breuil, Conrad, Diamond und Taylor¹³ für alle elliptischen Kurven über \mathbb{Q} bewiesen:

25.9. Satz. Sei E eine elliptische Kurve über \mathbb{Q} . Dann ist E modular.

SATZ
Modularitäts-
satz

Die Motivation von Wiles, die Modularität für semistabile elliptische Kurven über \mathbb{Q} zu beweisen, lag darin, dass bekannt war, dass daraus die Fermatsche Vermutung folgt:

25.10. Folgerung. Die Gleichung

$$x^n + y^n = z^n$$

hat keine Lösung in ganzen Zahlen $x, y, z \neq 0$ und $n \geq 3$.

FOLG
Fermatsche
Vermutung

Man kann sich hier auf $n = 4$ oder n eine ungerade Primzahl beschränken. Der Fall $n = 4$ wurde bereits von Fermat selbst bewiesen. Er zeigte die stärkere Aussage, dass die Gleichung

$$(25.1) \quad w^2 = u^4 + v^4$$

keine nichttriviale (das heißt hier $u, v, w \neq 0$) ganzzahlige Lösung hat. Diese Gleichung ist von der Form, wie sie bei der Berechnung einer Selmer-Gruppe auftrat. Die zugehörige elliptische Kurve ist

$$E_4: y^2 = x^3 + x,$$

¹³C. Breuil, B. Conrad, F. Diamond, R. Taylor: *On the modularity of elliptic curves over \mathbb{Q} : Wild 3-adic exercises*, J. Amer. Math. Soc. **14** (2001), 843–939.

für die wir in Beispiel 24.1 gezeigt hatten, dass $E_4(\mathbb{Q}) = \{O, (0, 0)\}$ ist. Jede nichttriviale Lösung (u, v, w) von (25.1) liefert einen rationalen Punkt auf E mit x -Koordinate $(u/v)^2 \neq 0, \infty$; solche Punkte gibt es aber nicht.

Der Fall $n = 3$ wurde möglicherweise auch von Fermat schon bewiesen; in jedem Fall dann von Euler. Die ebene projektive Kurve $X^3 + Y^3 = Z^3$ ist als Kurve isomorph zur elliptischen Kurve

$$E_3: y^2 = x^3 - 432,$$

von der man ebenfalls zeigen kann, dass sie Rang 0 hat. Man sieht dann leicht, dass $E_3(\mathbb{Q}) = \{O, (12, 36), (12, -36)\}$ ist, was zeigt, dass auch die ursprüngliche Kurve nur die drei rationalen Punkte $(1 : -1 : 0)$, $(1 : 0 : 1)$ und $(0 : 1 : 1)$ hat.

Man kann also $n = p \geq 5$ prim annehmen. Die Idee ist nun, dass man zu einer angenommenen ganzzahligen Lösung

$$a^p + b^p = c^p$$

mit $a, b, c \neq 0$ und ohne Einschränkung $\text{ggT}(a, b, c) = 1$ die elliptische Kurve

$$E_{a,b,c}: y^2 = x(x + a^p)(x - b^p)$$

betrachtet, deren Diskriminante $\Delta(E_{a,b,c}) = -16(abc)^{2p}$ ist. Eventuell nach einer Permutation von $(a, b, -c)$ und/oder einem gemeinsamen Vorzeichenwechsel ist die Gleichung von $E_{a,b,c}$ eine minimale Weierstraß-Gleichung. Man betrachtet nun die Operation der absoluten Galois-Gruppe $G_{\mathbb{Q}}$ von \mathbb{Q} auf der p -Torsion $E_{a,b,c}[p]$. Daraus, dass $E_{a,b,c}$ modular und $v_q(\Delta(E_{a,b,c})) \in p\mathbb{Z}$ ist für alle Primzahlen $q \geq 3$, folgt nach einem Ergebnis von Ribet, dass es eine Neuf orm der Stufe 2 und vom Gewicht 2 geben müsste, deren q -Entwicklungs-Koeffizienten modulo p kongruent zu denen der zu $E_{a,b,c}$ gehörenden Neuf orm sind. Es ist aber $\mathcal{S}(2, \Gamma_0(2)) = \{0\}$, also gibt es gar keine Neuf ormen mit Gewicht 2 und Stufe 2. Dieser Widerspruch zeigt, dass es die Lösung, von der wir ausgegangen waren, nicht geben kann.

Eine weitere Schlussfolgerung aus dem Modularitätssatz ist, dass wir *alle* elliptischen Kurven über \mathbb{Q} mit Führer N bekommen, indem wir die Neuf ormen der Stufe N mit ganzzahligen Koeffizienten bestimmen und dann die nach Satz 25.6 zugehörigen elliptischen Kurven finden (für beides gibt es Algorithmen; siehe z.B. [Cre]). Auf diese Weise wurde eine Liste aller elliptischen Kurven E über \mathbb{Q} mit $N_E \leq 500\,000$ erstellt. (Stand Juli 2020; das ist ein laufendes Projekt.)

Bevor wir dazu kommen, was über die Vermutung von Birch und Swinnerton-Dyer bekannt ist, wollen wir erst noch die „starke“ Version der Vermutung formulieren. Darin kommen einige Objekte vor, die wir noch einführen müssen.

25.11. Definition. Sei $E: y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$ eine minimale Weierstraß-Gleichung einer elliptischen Kurve über \mathbb{Q} . Die *reelle Periode* $\Omega(E)$ von E ist das Integral

DEF
reelle
Periode

$$\Omega(E) = \int_{E(\mathbb{R})} \left| \frac{dx}{2y + a_1x + a_3} \right|. \quad \diamond$$

Die Differentialform

$$\omega_E = \frac{dx}{2y + a_1x + a_3} = \frac{dy}{3x^2 + 2a_2x + a_4 - a_1y}$$

heißt auch das *invariante Differential* von E . Sie hat die Eigenschaft, dass sie

DEF
invariantes
Differential

unter der Addition eines beliebigen Punktes von E invariant ist: $\tau_P^* \omega_E = \omega_E$, wenn $\tau_P: E \rightarrow E, Q \mapsto P + Q$ ist.

25.12. Definition. Sei E eine elliptische Kurve über \mathbb{Q} . Sei $r = \text{rk}(E(\mathbb{Q}))$ und seien $P_1, \dots, P_r \in E(\mathbb{Q})$, sodass ihre Bilder in $E(\mathbb{Q})/E(\mathbb{Q})_{\text{tors}} \cong \mathbb{Z}^r$ eine Basis bilden. Dann heißt

DEF
Regulator

$$R(E) = \det(\langle P_i, P_j \rangle)_{1 \leq i, j \leq r}$$

der *Regulator* von E . Hier ist

$$\langle P, Q \rangle = \frac{1}{2}(\hat{h}(P + Q) - \hat{h}(P) - \hat{h}(Q))$$

die zur quadratischen Form \hat{h} gehörende symmetrische Bilinearform. Der Faktor $\frac{1}{2}$ führt zu $\langle P, P \rangle = \hat{h}(P)$. ◇

Der Regulator hängt nicht von der Wahl der Basis ab: Sei G die (Gramsche) Matrix in der Definition und sei $T \in \text{GL}(r, \mathbb{Z})$ die Basiswechselmatrix. Die Determinante von G geht beim Wechsel der Basis über in

$$\det(T^T G T) = \det(T)^2 \det(G) = (\pm 1)^2 \det(G) = \det(G).$$

Für das Folgende brauchen wir den Körper \mathbb{Q}_p der p -adischen Zahlen. Ähnlich wie \mathbb{R} ist er eine Vervollständigung von \mathbb{Q} , aber statt des üblichen Absolutbetrages verwendet man hier den p -adischen Absolutbetrag

$$|x|_p = \begin{cases} 0, & \text{falls } x = 0, \\ p^{-v_p(x)}, & \text{sonst} \end{cases}$$

zur Definition der Metrik ($d(x, y) = |x - y|_p$), bezüglich der man vervollständigt. Siehe zum Beispiel Abschnitt 7 im Skript zur Vorlesung „Diophantische Gleichungen“.

25.13. Definition. Sei E eine durch eine minimale Weierstraß-Gleichung gegebene elliptische Kurve über \mathbb{Q} . Sei p eine Primzahl und sei \tilde{E} die durch die modulo p reduzierte Gleichung definierte (nicht notwendig elliptische) Kurve über \mathbb{F}_p . Wir betrachten E als elliptische Kurve über dem Körper \mathbb{Q}_p der p -adischen Zahlen; dann haben wir die Reduktionsabbildung $\rho: E(\mathbb{Q}_p) \rightarrow \tilde{E}(\mathbb{F}_p)$. Wir definieren

DEF
Tamagawa-Zahl

$$E^{(0)}(\mathbb{Q}_p) = \{P \in E(\mathbb{Q}_p) \mid \rho(P) \in \tilde{E}(\mathbb{F}_p) \text{ ist nicht-singulär}\};$$

dann ist $E^{(0)}(\mathbb{Q}_p)$ eine Untergruppe von endlichem Index in $E(\mathbb{Q}_p)$. Die *Tamagawa-Zahl* von E bei p ist dann der Index

$$c_p(E) = (E(\mathbb{Q}_p) : E^{(0)}(\mathbb{Q}_p)). \quad \diamond$$

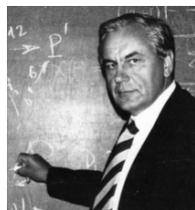
Hat E gute Reduktion bei p , dann ist $E^{(0)}(\mathbb{Q}_p) = E(\mathbb{Q}_p)$ und damit $c_p(E) = 1$. Es sind also nur endlich viele der $c_p(E)$ von 1 verschieden; damit ist das Produkt über alle Primzahlen

$$c(E) = \prod_p c_p(E)$$

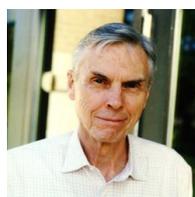
sinnvoll.

Die Tamagawa-Zahlen haben folgende Eigenschaften:

- (1) Hat E bei p gute Reduktion, dann ist $c_p(E) = 1$.
- (2) Hat E bei p schlechte Reduktion, die nicht zerfallend multiplikativ ist, dann ist $c_p(E) \leq 4$.



I.R. Shafarevich
1923 – 2017
Foto © MFO



J.T. Tate
1925 – 2019
Foto © MFO

- (3) Hat E bei p zerfallend multiplikative Reduktion, dann ist $c_p(E) = v_p(\Delta(E)) = -v_p(j(E))$.

Das letzte Objekt, das wir brauchen, ist die *Shafarevich-Tate-Gruppe* (oder auch *Tate-Shafarevich-Gruppe*) von E .

25.14. Definition. Sei E eine elliptische Kurve über \mathbb{Q} . Ein *Torsor* unter E ist eine (glatte) projektive (nicht notwendig ebene) über \mathbb{Q} definierte Kurve X zusammen mit einem ebenfalls über \mathbb{Q} definierten Morphismus $\mu: E \times X \rightarrow X$, der einer Operation von E auf X entspricht: **DEF**
Torsor

$$\forall P, Q \in E, x \in X: \mu(O, x) = x \quad \text{und} \quad \mu(P + Q, x) = \mu(P, \mu(Q, x))$$

und sodass die induzierte Operation von $E(\bar{\mathbb{Q}})$ auf $X(\bar{\mathbb{Q}})$ transitiv mit trivialen Stabilisatoren ist (d.h., für $x, y \in X(\bar{\mathbb{Q}})$ gibt es *genau einen* Punkt $P \in E(\bar{\mathbb{Q}})$ mit $\mu(P, x) = y$).

Zwei Torsore (X, μ) und (X', μ') unter E sind *isomorph*, wenn es einen über \mathbb{Q} definierten Isomorphismus $\phi: X \rightarrow X'$ von Kurven gibt, sodass das folgende Diagramm kommutiert:

$$\begin{array}{ccc} E \times X & \xrightarrow{\text{id} \times \phi} & E \times X' \\ \downarrow \mu & & \downarrow \mu' \\ X & \xrightarrow{\phi} & X' \end{array}$$

Ein Torsor (X, μ) heißt *trivial*, wenn $X(\mathbb{Q}) \neq \emptyset$ ist. Er heißt *lokal trivial*, wenn $X(\mathbb{R}) \neq \emptyset$ und $X(\mathbb{Q}_p) \neq \emptyset$ ist für alle Primzahlen p . ◇

Der triviale Torsor ist $(E, +)$ (dabei ist $+$ die Additionsabbildung $E \times E \rightarrow E$). Ein Torsor ist genau dann trivial, wenn er zu $(E, +)$ isomorph ist. („ \Rightarrow “: Sei $x \in X(\mathbb{Q})$. Dann liefert $\phi: E \rightarrow X, P \mapsto \mu(P, x)$, den gewünschten Isomorphismus. „ \Leftarrow “: Sei $\phi: E \rightarrow X$ der Isomorphismus. Dann ist $x = \phi(O) \in X(\mathbb{Q})$.) Da jede Kurve X stets $\bar{\mathbb{Q}}$ -rationale Punkte hat, kann man einen Torsor unter E auch definieren als ein über \mathbb{Q} definiertes Paar (X, μ) , das über \mathbb{Q} zu $(E, +)$ isomorph wird.

Aus zwei Torsoren (X, μ) und (X', μ') kann man einen neuen machen, die Baer-Summe der beiden. Sie ist definiert als der Quotient von $X \times X'$ bezüglich der durch

$$P * (x, x') = (\mu(P, x), \mu'(-P, x'))$$

gegebenen Operation von E . Diese Operation ist kommutativ und assoziativ, verträglich mit Isomorphismen und hat den trivialen Torsor als neutrales Element (bis auf Isomorphie), und jeder Torsor (X, μ) hat bis auf Isomorphie ein Inverses, nämlich $(X, (P, x) \mapsto \mu(-P, x))$. Daher bildet die Menge der Isomorphieklassen von Torsoren unter E eine abelsche Gruppe, die Weil-Châtelet-Gruppe von E .

25.15. Definition. Sei E eine elliptische Kurve über \mathbb{Q} . Die *Shafarevich-Tate-Gruppe* $\text{III}(E)$ von E ist die Untergruppe der Weil-Châtelet-Gruppe von E , die aus den Isomorphieklassen von lokal trivialen Torsoren besteht. **DEF**
Shafarevich-Tate-Gruppe

25.16. **Beispiel.** Torsore sind uns schon begegnet. Sei $\hat{\phi}: E' \rightarrow E$ eine Isogenie vom Grad 2, wobei $E: y^2 = x(x^2 + ax + b)$ ist. Für $d \in \mathbb{Q}^\times$ haben wir dann die Gleichung

BSP
Torsore

$$w^2 = du^4 + au^2v^2 + \frac{b}{d}v^4$$

betrachtet. Diese Gleichung definiert eine glatte Kurve X_d in einer gewichteten projektiven Ebene (man behandelt w so, als ob $\deg(w) = 2$ wäre). Diese Kurve X_d ist ein Torsor unter E' . Um das zu sehen, zeigt man, dass X_d über $\mathbb{Q}(\sqrt{d})$ isomorph zu E' ist (da X_d über $\mathbb{Q}(\sqrt{d})$ zu X_1 isomorph ist, genügt es zu zeigen, dass X_1 über \mathbb{Q} zu E' isomorph ist). Diesen Isomorphismus $\varphi: X_d \rightarrow E'$ kann man so wählen, dass er mit $\hat{\phi}$ und der Abbildung $X_d \rightarrow E$ verträglich ist. Man definiert dann μ so, dass (X_d, μ) über $\mathbb{Q}(\sqrt{d})$ zu $(E', +)$ isomorph ist:

$$\mu(P, x) = \varphi^{-1}(P + \varphi(x));$$

man muss sich dann noch davon überzeugen, dass μ über \mathbb{Q} definiert ist.

Die Isomorphieklasse von (X_d, μ) ist genau dann ein Element von $\text{III}(E')$, wenn X_d überall lokal lösbar ist, also wenn $d \in S_{\hat{\phi}}$ ist. Der Torsor (X_d, μ) ist genau dann trivial, wenn X_d einen rationalen Punkt hat, also wenn $d \in \text{im}(\delta)$ ist. Man bekommt also eine exakte Sequenz

$$E'(\mathbb{Q}) \xrightarrow{\hat{\phi}} E(\mathbb{Q}) \xrightarrow{\delta} S_{\hat{\phi}} \longrightarrow \text{III}(E').$$

Die Isogenie $\hat{\phi}$ induziert einen Homomorphismus $\hat{\phi}_*: \text{III}(E') \rightarrow \text{III}(E)$. Das Bild der letzten Abbildung in der Sequenz oben ist dann genau der Kern $\text{III}(E')[\hat{\phi}_*]$ von $\hat{\phi}_*$. Wir erhalten also die Äquivalenz

$$\text{im}(\delta) = S_{\hat{\phi}} \iff \text{III}(E')[\hat{\phi}_*] = \{0\}.$$

Die analoge Aussage gilt für beliebige Isogenien und die zugehörigen Selmer-Gruppen. ♣

Jetzt haben wir alles beisammen, um die starke Version der Vermutung von Birch und Swinnerton-Dyer formulieren zu können.

25.17. **Vermutung.** Sei E eine elliptische Kurve über \mathbb{Q} , die durch eine minimale Weierstraß-Gleichung gegeben ist. Sei $r = \text{rk}(E(\mathbb{Q}))$.

VERM
starke BSD-
Vermutung

Dann ist die Gruppe $\text{III}(E)$ endlich, es ist

$$\text{ord}_{s=1} L(E, s) = r,$$

und

$$(25.2) \quad \frac{1}{r!} \left(\frac{d}{ds} \right)^r L(E, s) \Big|_{s=1} = \lim_{s \rightarrow 1} \frac{L(E, s)}{(s-1)^r} = \Omega(E)c(E) \frac{R(E)\#\text{III}(E)}{(\#E(\mathbb{Q})_{\text{tors}})^2}.$$

Was ist darüber bekannt? Zunächst einmal gilt, dass die Vermutung für eine elliptische Kurve E genau dann stimmt, wenn sie für eine isogene Kurve E' stimmt. In diesem Fall ist $L(E, s) = L(E', s)$; diese Aussage läuft also darauf hinaus, dass die Ausdrücke auf der rechten Seite der Gleichung (25.2) für E und E' übereinstimmen (dass $\text{rk}(E(\mathbb{Q})) = \text{rk}(E'(\mathbb{Q}))$ ist, ist leicht zu sehen). Das ist nicht offensichtlich, da sich jeder einzelne Term ändern kann!

Das wichtigste Resultat in Richtung der Vermutung wurde 1988 von Kolyvagin bewiesen:

25.18. **Satz.** Sei E eine (modulare) elliptische Kurve über \mathbb{Q} mit

$$\text{ord}_{s=1} L(E, s) \leq 1.$$

Dann gilt

$$\text{ord}_{s=1} L(E, s) = \text{rk}(E(\mathbb{Q})),$$

$\text{III}(E)$ ist endlich, und (25.2) gilt bis auf einen rationalen Faktor, dessen Zähler und Nenner nur durch Primzahlen aus einer explizit bestimmbar endlichen Menge teilbar sind.

Kolyvagin hatte dieses Resultat unter der Voraussetzung bewiesen, dass E modular ist; das war einige Zeit vor dem Beweis des Modularitätssatzes 25.9.

Man erwartet, dass die Bedingung $\text{ord}_{s=1} L(E, s) \leq 1$ für „fast alle“ elliptischen Kurven E erfüllt ist. Genauer ist die Vermutung, dass der Anteil der elliptischen Kurven mit Führer $\leq X$, die das tun, für $X \rightarrow \infty$ gegen 1 strebt. Das beste Ergebnis in dieser Richtung stammt von Bhargava und verschiedenen Coautoren und besagt, dass der Anteil dieser Kurven für großes X größer als 0,66 ist. In diesem Sinn kommt Satz 25.18 schon recht nahe an einen vollständigen Beweis heran. Basierend auf diesem Resultat und weiteren Verbesserungen hinsichtlich der Primzahlen, die in dem „Fehlerfaktor“ vorkommen können, wurde von vielen Mathematikern über diverse Arbeiten verteilt Folgendes gezeigt:

25.19. **Satz.** Sei E eine elliptische Kurve über \mathbb{Q} mit

$$N_E \leq 5000 \quad \text{und} \quad \text{rk}(E(\mathbb{Q})) \leq 1$$

(das schließt lediglich 691 der insgesamt 31 073 Kurven E mit $N_E \leq 5000$ aus). Dann gilt die starke Vermutung von Birch und Swinnerton-Dyer für E .

Auf der anderen Seite ist für den Fall $\text{rk}(E(\mathbb{Q})) \geq 2$ bzw. $\text{ord}_{s=1} L(E, s) \geq 2$ nur wenig bekannt:

- (1) Man kann das Vorzeichen in der Funktionalgleichung von $L(E, s)$ bestimmen. Im Fall $+1$ ist $\text{ord}_{s=1} L(E, s)$ gerade, im Fall -1 ungerade.
- (2) Man kann entscheiden, ob $L(E, 1) = 0$ (im geraden Fall) bzw. $L'(E, 1) = 0$ ist (im ungeraden Fall). Man kann numerisch verifizieren, dass $L^{(n)}(E, 1) \neq 0$ ist. So kann man $\text{ord}_{s=1} L(E, s)$ bestimmen, wenn die Ordnung höchstens 3 ist. Es gibt demgegenüber aber bisher keine Möglichkeit zu beweisen, dass die Ordnung eine gegebene Zahl ≥ 4 ist. Insbesondere kann man nicht einmal die schwache BSD-Vermutung verifizieren, wenn $\text{rk } E(\mathbb{Q}) \geq 4$ ist.
- (3) Es ist keine einzige elliptische Kurve E über \mathbb{Q} mit $\text{rk}(E(\mathbb{Q})) \geq 2$ bekannt, für die man zeigen konnte, dass $\text{III}(E)$ endlich ist.
- (4) Es gibt keinen Kandidaten für ein Gegenbeispiel zur BSD-Vermutung.

Auf den Beweis der schwachen Vermutung von Birch und Swinnerton-Dyer für elliptische Kurven über \mathbb{Q} (es gibt eine allgemeinere Version für abelsche Varietäten über algebraischen Zahlkörpern) hat die Clay Foundation ein Preisgeld von einer Million US-Dollar ausgesetzt: Das ist eines der Clay-Millenniums-Probleme, neben zum Beispiel der Riemannschen Vermutung über die nichttrivialen Nullstellen der Riemannschen Zetafunktion.

SATZ
BSD für
 $\text{ord} \leq 1$



M. Bhargava
* 1974

SATZ
BSD für
 $N_E \leq 5000$

LITERATUR

- [Cas] J.W.S. CASSELS: *Lectures on elliptic curves*, London Mathematical Society Student Texts **24**, Cambridge University Press (1991).
- [Co1] H. COHEN: *A course in computational algebraic number theory*, Springer GTM **138** (1993).
Online-Zugriff (aus dem Uni-Netz)
- [Co2] H. COHEN: *Number Theory. Volume I: Tools and diophantine equations*, Springer GTM **239** (2007).
Online-Zugriff (aus dem Uni-Netz)
- [Cre] J.E. CREMONA: *Algorithms for modular elliptic curves*, second edition. Cambridge University Press, Cambridge, 1997.
Freie online-Version
- [CF+] J.E. CREMONA, T.A. FISHER, C. O'NEIL, D. SIMON und M. STOLL: *Explicit n -descent on elliptic curves. I. Algebra*, J. reine angew. Math. **615**, 121–155 (2008).
- [Hus] D. HUSEMÖLLER: *Elliptic curves*, Springer GTM **111** (1987).
Online-Zugriff (aus dem Uni-Netz)
- [Jae] K. JÄNICH: *Einführung in die Funktionentheorie*, Springer Hochschultext (2. Auflage 1980).
Online-Zugriff (aus dem Uni-Netz)
- [Kna] A.W. KNAPP: *Elliptic curves*, Mathematical Notes **40**, Princeton University Press (1992).
- [Si1] J.H. SILVERMAN: *The arithmetic of elliptic curves*, Springer GTM **106** (1986).
Online-Zugriff (aus dem Uni-Netz)
- [Si2] J.H. SILVERMAN: *Advanced topics in the arithmetic of elliptic curves*, Springer GTM **151** (1994).
Online-Zugriff (aus dem Uni-Netz)
- [ST] J.H. SILVERMAN und J.T. TATE: *Rational points on elliptic curves*, second edition. Springer Undergraduate Texts in Mathematics (2015).
Online-Zugriff (aus dem Uni-Netz)