Arithmetik elliptischer Kurven mit Anwendungen

Sommersemester 2009

Universität Bayreuth

Michael Stoll

Inhaltsverzeichnis

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

$$E : y^2 = x^3 + a\,x + b \tag{1.1}$$

in den Variablen x und y mit Koeffizienten a und b aus einem Körper K (der Charakteristik $\neq 2$), wobei wir noch verlangen, dass $4\,a^3 + 27\,b^2 \neq 0$ 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“ dazu zunehmen. 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 acht geben, 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.

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

wobei

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

und $\mu = \eta_1 - \lambda\xi_1$; $y = \lambda\,x + \mu$ ist 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.)

Als Beispiel betrachten wir die Kurve

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

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

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

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 möchte ich 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.)

Zum Beispiel haben wir 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.)

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 die Verteilung einigermaßen gleichmäßig ist.

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 (beides lässt sich recht leicht nachweisen). Wir wollen einen echten Teiler $d \neq 1$ von N finden. Dazu wählen wir zufällig eine Zahl $a \in \{2, \ldots, N-1\}$. Falls $d = \operatorname{ggT}(a, N) > 1$, 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 $B = \operatorname{kgV}(1, 2, \ldots, L)$. Dann berechnen wir $d = \operatorname{ggT}(a^B - 1, N)$. Dazu berechnet man am besten $b = a^B \bmod N$ durch sukzessives Quadrieren und dann $d = \operatorname{ggT}(b - 1, N)$. Der Aufwand dafür ist etwa $\log B (\log N)^2$, und $\log B \approx L$. Wenn $1 < d < N$, 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, so dass $p - 1$ 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 $a^{p-1} \equiv 1 \bmod p$. Auf der anderen Seite ist $a^B - 1$ sehr wahrscheinlich nicht durch q teilbar — dazu müsste a eine $k$te Potenz mod q sein mit $k = (q-1)/\operatorname{ggT}(B', q-1)$. Wir können also erwarten, dass $d = \operatorname{ggT}(a^B - 1, N)$ durch p, aber nicht durch q teilbar ist. In der Praxis wird man eine Folge von Werten von B verwenden, die man durch sukzessive Multiplikation

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

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

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

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

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

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 $\#\tilde{E}(\mathbb{F}_p) = p + 1 - t$. Analog zu eben multiplizieren wir P mit $B = \mathrm{kgV}(1, 2, \ldots, L)$. 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, so dass $m = \#\tilde{E}(\mathbb{F}_p)$ 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

$$Ce^{(1+o(p))\sqrt{\log p \log\log p}}$$

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 $p \to \infty$ gegen null geht.

Als Baby-Beispiel wollen wir die Zahl $N = 851$ faktorisieren. Wir nehmen als Kurve $E : y^2 = x^3 + 9\,x - 9$ 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.

(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 $23 = \mathrm{ggT}(851, 23)$ 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* $\mathbb{A}^2_K$ über K hat folgende Eigenschaften.

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

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

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

$$f_L : \mathbb{A}^2_K(L) \longrightarrow L\,.$$

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

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

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

(Und $f_{\bar{K}}$ bestimmt wieder f eindeutig.)

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

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

Diese Definition ist operational, d.h. sie sagt nicht so sehr, was $\mathbb{A}^2_K$ „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 $\mathbb{A}^2_K$ als „geringte 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 $\mathbb{A}^n_K$, den *n-dimensionalen affinen Raum über K*.

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

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

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

(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 $g/h \in K(x, y)$, so dass f und h keinen nicht-konstanten gemeinsamen Teiler haben. Dabei sind g_1/h_1 und g_2/h_2 äquivalent, wenn f $g_1h_2 - g_2h_1$ teilt.

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

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

(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 Integritätsring. 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.**

(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 $f = x^2 + y^2 - 1$). Für jedes L haben wir die rationalen Punkte $(0, \pm 1)$ und $(\pm 1, 0)$, aber normalerweise natürlich noch mehr. Man kann zeigen, dass $C(L) = \{(\frac{2t}{1+t^2}, \frac{1-t^2}{1+t^2}) \mid t \in L, t^2 \neq -1\} \cup \{(0, -1)\}$ 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 Punktmengen $\mathbb{C}^2$ 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* $\mathbb{P}^2_K$ über K hat folgende Eigenschaften.

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

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

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 $\mathbb{A}^2_K$ in $\mathbb{P}^2_K$ bekommen. Die übrigen Geraden entsprechen den unendlich fernen Punkten, entsprechend ihrer Richtung in der xy-Ebene. In Formeln haben wir für die Einbettung:

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

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

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

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

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 $\mathbb{P}^2_k(\bar{K})$, in denen g verschwindet.

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

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. (Wir schreiben $C : f(X,Y,Z) = 0$.)

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

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

(2) Eine *rationale Funktion* auf C ist eine Äquivalenzklasse rationaler Funktionen auf $\mathbb{P}^2_K$, deren Nenner mit f keinen nicht-konstanten gemeinsamen Teiler hat. Dabei heißen g_1/h_1 und g_2/h_2 äquivalent, wenn $f \mid g_1h_2 - g_2h_1$.

Eine rationale Funktion ϕ ist *regulär* in $P \in C(L)$, wenn sie einen Repräsentanten g/h hat, so dass 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 *Funktionenkörper* $K(C)$ von C.

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, so dass 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 Gerade.

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 $\mathrm{char}(L) \neq 2$, sonst ist es der eine Punkt $(1 : 1 : 0)$). Allgemeiner gilt, dass alle Kreise $(x - a)^2 + (y - b)^2 = r^2$ die selben zwei Punkte im Unendlichen haben.
(3) Der projektive Abschluss der Kurve $y^2 = x^3+ax+b$ ist $Y^2Z-X^3-aX^2Z-bZ^3 = 0$. Er hat genau den einen (stets rationalen) Punkt $(0 : 1 : 0)$ im Unendlichen.

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 das stimmt, müssen die Schnittpunkte aber mit der richtigen Vielfachheit gezählt werden. Deswegen müssen wir erst einmal diese Vielfachheit definieren.

4.1. **Definition.** Sei $P = (\xi : \eta : \zeta) \in \mathbb{P}^2_K(L)$ 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 L in C enthalten). Wir definieren $i(G, C; P)$, die *Vielfachheit des Schnittpunkts P von G und C* wie folgt.

Wenn $P \notin C(L) \cap G(L)$, dann setzen wir $i(G, C; P) = 0$. Ansonsten lösen wir die Gleichung von G nach einer der Variablen auf, z.B. $Z = -\frac{a}{c}X - \frac{b}{c}Y$ (falls $c \neq 0$), 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 $i(G, C; P)$.

Die Definition hängt natürlich nicht davon ab, welche Variable wir eliminieren. Siehe Übungen.

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.)

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.)

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*

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

Gilt für einen Erweiterungskörper $L \supset K$, dass

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

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:

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

Für jeden Schnittpunkt $P = (\xi : \eta : \zeta) \in C(\bar{K}) \cap G(\bar{K})$ gilt $H(\xi, \eta) = 0$ und $\zeta = a'\xi + b'\eta$ und umgekehrt. 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. □

Dieser Satz ist ein Spezialfall des *Satzes von Bézout*, der sagt, dass sich zwei projektive Kurven der Grade d_1 und d_2 stets in genau $d_1 d_2$ Punkten (mit Vielfachheit gerechnet) 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 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* im Punkt $P = (\xi, \eta) \in C(L)$, wenn nicht beide partielle Ableitungen 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* im Punkt $P = (\xi : \eta : \zeta) \in C(L)$, wenn

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

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

Ein Punkt P, in dem C nicht glatt ist, heißt *singulärer Punkt* oder *Singularität* von C.

Ein Punkt auf einer affinen Kurve ist genau dann glatt, wenn er auf dem projektiven Abschluss glatt ist, siehe Übungen.

5.2. **Beispiele.**

(1) Ist die Kurve $Y^2Z - X^3 - Z^3$ 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 $\mathrm{char}(K) \neq 2, 3$ 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.

(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.

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

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

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

Sonst ist $i(C; P) \geq 2$. Die Zahl $i(C; P)$ heißt auch die *Vielfachheit* von P auf C.

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

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

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.

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

(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 so dass $G(R_1, R_2, R_3)$ durch F teilbar ist. Dabei heißen (R_1, R_2, R_3) und (S_1, S_2, S_3) äquivalent, wenn $F \mid R_iS_j - R_jS_i$ für alle i, j.

(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, so dass 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 $\mathrm{id}_C : C \longrightarrow C$.

(5) C und D heißen *birational äquivalent*, wenn es rationale Abbildungen $\phi : C \to D$ und $\psi : D \to C$ gibt, so dass $\phi \circ \psi = \mathrm{id}_D$ und $\psi \circ \phi = \mathrm{id}_C$. Dann ist ϕ eine *birationale* Abbildung. Sind ϕ und ψ sogar Morphismen, dann heißen C und D *isomorph*, und ϕ ist ein *Isomorphismus*.

Es gilt übrigens, dass eine rationale Abbildung von einer glatten Kurve in eine andere Kurve automatisch ein Morphismus ist.

6.2. **Beispiele.**

(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 $\mathrm{char}(K) \neq 2$. Dann definiert $(X^2 - Y^2, 2XY, Z^2)$ einen Morphismus $\phi : C \longrightarrow C$: 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.

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.

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

$$Y^2Z + a_1\,XYZ + a_3\,YZ^2 = X^3 + a_2\,X^2Z + a_4\,XZ^2 + a_6\,Z^3$$

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:

$$E : y^2 + a_1\,xy + a_3\,y = x^3 + a_2\,x^2 + a_4\,x + a_6\,. \tag{7.1}$$

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

Die etwas komische Nummerierung der Koeffizienten wird später verständlich werden.

7.2. **Lemma.** *Sei E eine (nicht notwendig glatte) Kurve wie oben. Dann hat E genau einen Punkt im Unendlichen, nämlich $O = (0 : 1 : 0)$. Der 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 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. □

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

7.3. **Lemma.** *Sei E eine elliptische Kurve über K. Wenn* $\mathrm{char}(K) \neq 2$*, 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 $\mathrm{char}(K) \neq 3$*, 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

$$(X : Y : Z) \longmapsto (X : Y + \frac{a_1}{2} X + \frac{a_3}{2} Z : Z) ;$$

für die Koeffizienten gilt dann

$$a_2' = a_2 + \tfrac{1}{4} a_1^2 , \quad a_4' = a_4 + \tfrac{1}{2} a_1 a_3 , \quad a_6' = a_6 + \tfrac{1}{4} a_3^2 .$$

Wenn $\mathrm{char}(K) \neq 3$, 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. □

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.

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

Dabei gilt

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

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

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

bzw.

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

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.

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

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 $E : y^2 = x^3 + a\, x + b$ 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

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

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

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

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. □

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

7.5. **Beispiele.**

(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 $\mathrm{char}(K) \neq 2$ 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 $\mathrm{char}(K) \neq 2, 3$ ist. Ihre j-Invariante ist 0.

8. Isomorphismen elliptischer Kurven

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

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

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

8.2. **Lemma.** *Wenn E (E') in der Situation der Definition oben durch eine Weierstraß-Gleichung mit Koeffizienten a_i (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

$$u^4\,c_4 = c_4'\,, \quad u^6\,c_6 = c_6'\,, \quad u^{12}\,\Delta = \Delta' \quad \textit{und} \quad j = j'\,.$$

Beweis. Das rechnet man nach. □

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

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

Beweis. Man prüft nach, dass durch

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

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:

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

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. □

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.

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.*

(1) *Sei* $\operatorname{char}(K) \neq 2, 3$. *Wenn es ein $u \in K^\times$ gibt mit $c_4(E') = u^4\, c_4(E)$ und $c_6(E') = u^6\, c_6(E)$, dann sind E und E' über K isomorph.*
(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 $\operatorname{char}(K) \neq 2, 3$ voraus.

(1) Die gegebenen Kurven sind nach Lemma 7.3 und der Bemerkung vor Lemma 7.4 isomorph zu den Kurven

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

Lemma 8.3 zeigt, 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 $u \in \bar{K}^\times$, so dass $c_4(E') = u^4\, c_4(E)$ und $c_6(E') = u^6\, c_6(E)$. 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

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

abgedeckt werden. In den übrigen Fällen tut es die Kurve

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

(Um drauf zu kommen, mache man in der kurzen Weierstraß-Gleichung $y^2 = x^3 + ax + b$ den Ansatz $a = b$.) □

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. **Proposition.** *Sei* $\mathrm{char}(K) \neq 2, 3$, $j \in K$ *und* $E : y^2 = x^3 + a\,x + b$ *eine elliptische Kurve über* K *mit* $j(E) = j$.

(1) *Wenn* $j \neq 0, 1728$, *dann 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*

$$y^2 = x^3 + d^2\, a\, x + d^3\, b\,.$$

(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*

$$y^2 = x^3 + d\, b\,.$$

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

$$y^2 = x^3 + d\, a\, x\,.$$

Beweis. (1) Die j-Invariante hängt bei einer kurzen Weierstraß-Gleichung nur von a^3/b^2 ab. Daher hat $E' : y^2 = x^3 + a'x + b'$ genau dann dieselbe j-Invariante wie E, wenn $a' = d^2\, a$ und $b' = d^3\, b$ für ein $d \in K^\times$. 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.** *Sei E eine elliptische Kurve über K und $L \supset K$ ein Erweiterungskörper. Durch folgende Festlegungen wird $E(L)$ zu einer abelschen Gruppe.*

(i) *Der Punkt $O \in E(L)$ ist das Nullelement.*
(ii) *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:

- Der Punkt $-P$ ist der dritte Schnittpunkt der Geraden durch O und P mit E.
- 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.

9.2. **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 $\eta + \eta' \neq -a_1\,\xi - a_3$ (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 (z.B. durch implizites Differenzieren), oder man überlegt sich, dass

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

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.

Jetzt müssen wir Satz 9.1 aber auch wirklich beweisen. Der Punkt O ist per Definitionem das Nullelement, und wir haben auch schon gesehen, dass zu jedem Punkt P das Inverse $-P$ existiert (beachte, dass der dritte Schnittpunkt in $E(L)$ liegt, wenn das für die anderen beiden gilt). 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 schließlich 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. Da $Z_1=-((P+Q)+R)$ und $Z_2=-(P+(Q+R))$, genügt es zu zeigen, dass $Z_1=Z=Z_2$ ist.

9.3. **Lemma.** *Seien G_i und G'_j (für $i, j \in \{1, 2, 3\}$) paarweise verschiedene Geraden in der projektiven Ebene, so dass die 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}.*

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 Koeffizienten. 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 $L = L_1L_2L_3$ und $L' = L'_1L'_2L'_3$ 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 Körper K (oder L) 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 $C : F(X, Y, Z) = 0$ die (oder 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 L_1 ein Teiler von F sein: $F = L_1F'$ 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: $F' = L_2F''$. 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 $F = cD + c'D'$ 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 $P_{33} \in C$. □

Wir wollen nun das Lemma anwenden auf unsere Geraden G_i und G'_j. Diese Geraden sind alle verschieden, denn sonst hätten wir mindestens fünf Punkte im Schnitt von E mit einer Geraden; 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. Man kann sich aber vorstellen, dass man zum Beispiel P und Q festhält; dann hat man Morphismen $E \to E$. Man kann sich leicht überlegen, dass $P + (-Q) = 0$ impliziert, dass $P = Q$ ist, also kann man den einen Morphismus

$$E \ni R \longmapsto ((P + Q) + R) + (-(P + (Q + R))) \in E$$

betrachten, der für fast alle R den Wert O hat und deswegen konstant sein muss.

9.4. **Bemerkung.** Man kann die Gruppenstruktur auch wie folgt „intrinsisch“ charakterisieren. Seien P, Q, R Punkte von E. Dann ist $P + Q = R$ genau dann, 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 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 $-R$, so dass 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 $E \to E'$, 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.

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 $\phi : C \to D$ über K und der Menge der K-linearen Homomorphismen $\phi^* : K(D) \to K(C)$ der Funktionenkörper. Dabei ist $K(C)$ eine endliche Körpererweiterung von $\phi^*\big(K(D)\big)$.*

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 $\phi : C \to D$ ist in affinen Koordinaten gegeben durch zwei rationale Funktionen $r(x,y), s(x,y) \in K(x,y) = K(C)$, so dass $\big(r(x,y), s(x,y)\big) \in D\big(K(C)\big)$ (das heißt einfach, dass r und s die affine Gleichung von D' erfüllen):

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

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

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

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 $K(C)/\phi^*\big(K(D)\big)$ ist endlich, weil x und y über $\phi^*\big(K(D)\big) = K\big(r(x,y), s(x,y)\big)$ algebraisch sind.

Ist umgekehrt ein K-linearer Homomorphismus $\psi : K(D) \to K(C)$ 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

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

also ist

$$\phi : C \to D\,, \qquad (x,y) \mapsto \big(r(x,y), s(x,y)\big)$$

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

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

Sei $\phi^*\big(K(D)\big) \subset L \subset K(C)$ der maximale Zwischenkörper, der über $\phi^*\big(K(D)\big)$ separabel ist. Dann heißt der Grad $[L : \phi^*\big(K(D)\big)]$ 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 $\mathrm{char}(K) = 0$ ist), sonst *inseparabel*. ϕ heißt *rein inseparabel*, wenn $L = \phi^*\big(K(D)\big)$ ist.

10.3. **Folgerung.** *Seien C und D zwei irreduzible Kurven über K. Dann sind C und D birational äquivalent über K genau dann, wenn ihre Funktionenkörper $K(C)$ und $K(D)$ K-isomorph sind.*

Beweis. Aus dem Satz folgt, dass zwischen der Menge der birationalen Isomorphismen $C \to D$ und der Menge der K-linearen Isomorphismen $K(D) \to K(C)$ eine Bijektion besteht. □

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

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

Beweis. Die erste Gleichung folgt aus der Multiplikativität der Grade in der Körpererweiterung $K(C_3) \hookrightarrow K(C_2) \hookrightarrow K(C_1)$. Die zweite Gleichung folgt aus der analogen Aussage für die maximal separablen Zwischenkörper. Die dritte Gleichung folgt aus den ersten beiden. □

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 $\phi : E \to E'$, so dass $\phi(O) = O$. Die Kurven E und E' heißen *isogen*, wenn es eine nicht konstante Isogenie $E \to E'$ gibt.

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

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

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

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 $\phi : E \to E'$ 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. □

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

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

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 $\mathrm{char}(K) \neq 2$ 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 $\mathrm{char}(K) = 2$ kann man mit einem Punkt der Ordnung 3 ähnlich verfahren.

10.7. **Definition.** Die Isogenien $E \to E$ (wie zum Beispiel die Multiplikationsabbildungen) heißen dann auch *Endomorphismen*; sie bilden einen Ring $\mathrm{End}_K(E)$ (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: $\phi \cdot \psi = \phi \circ \psi$.

Wie für jede nicht-konstante rationale Abbildung zwischen Kurven haben wir für jede nicht-konstante Isogenie $\phi : E \to E'$ den Grad $\deg \phi$, den separablen Grad $\deg_s \phi$ und den inseparablen Grad $\deg_i \phi$. Der Vollständigkeit halber setzt man noch $\deg 0 = \deg_s 0 = \deg_i 0 = 0$ (wo links 0 die konstante Isogenie $[0]$ bezeichnet). Es gilt dann

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

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

Beweis. Seien $\phi, \psi \in \mathrm{End}_K(E)$ 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 $\mathrm{End}_K(E)$ nullteilerfrei ist.

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

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

Man kann die möglichen Endomorphismenringe ziemlich genau klassifizieren. In Charakteristik 0 ist $\mathrm{End}_K(E) = \mathbb{Z}$ 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.

Eine ganz wichtige Eigenschaft ist auch die folgende.

10.9. **Satz.** *Sei $\phi : E \to E'$ eine nicht-konstante Isogenie. Dann gibt es genau eine Isogenie $\hat{\phi} : E' \to E$, die zu ϕ* duale *Isogenie, so dass $\hat{\phi} \circ \phi = [m]$, wobei $\deg(\phi) = m$. Es gilt dann auch $\phi \circ \hat{\phi} = [m]$. Weitere Eigenschaften sind:*

(i) *Ist $\psi : E' \to E''$ eine weitere Isogenie, dann gilt $\widehat{\psi \circ \phi} = \hat{\phi} \circ \hat{\psi}$.*
(ii) *Ist $\psi : E \to E'$ eine weitere Isogenie, dann gilt $\widehat{\phi + \psi} = \hat{\phi} + \hat{\psi}$.*
(iii) $\hat{\hat{\phi}} = \phi$.
(iv) $\deg(\hat{\phi}) = \deg(\phi)$.
(v) *Für $m \in Z$ gilt $\widehat{[m]} = [m]$ und $\deg([m]) = m^2$.*

Man setzt noch $\hat{0} = 0$; dann gelten (i)-(v) 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 $\psi, \psi' : E' \to E$ Isogenien mit $\psi \circ \phi = \psi' \circ \phi = [m]$. Dann folgt $(\psi - \psi') \circ \phi = \psi \circ \phi - \psi' \circ \phi = 0$; weil $\phi \neq 0$ 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.

(i) 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.

(v) 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. Wegen $[m] \circ [m] = [m^2] = \big[\deg[m]\big]$ folgt daraus $\widehat{[m]} = [m]$.

(iv) Es gilt $m^2 = \deg[m] = (\deg \hat{\phi})(\deg \phi) = (\deg \hat{\phi})m$; es folgt $\deg \hat{\phi} = m = \deg \phi$.

(iii) Es gilt $\phi \circ \hat{\phi} = [m] = [\deg \hat{\phi}]$, also folgt $\phi = \hat{\hat{\phi}}$.
(ii) Das beweisen wir hier nicht.

□

10.10. **Bemerkung.** *Ist $\phi : E \to E'$ 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* $\deg \phi = \max\{\deg p, \deg q\}$.

Beweis. Da y eine quadratische Gleichung über $K(x)$ erfüllt, kann man jede rationale Abbildung $E \to E'$ 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

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

Aus der Multiplikativität der Grade in Körpererweiterungen folgt dann

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

□

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

Nach so viel neuen Begriffen ist ein Beispiel angebracht.

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

$$E : y^2 = x^3 + a\,x^2 + b\,x$$

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

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

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

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

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

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 $E \to E$) anwenden, dann bekommen wir folgendes Resultat.

10.12. **Satz.** *Die Abbildung* $\mathrm{End}_K(E) \ni \phi \mapsto \hat{\phi} \in \mathrm{End}_K(E)$ *ist eine Anti-Involution von* $\mathrm{End}_K(E)$ *(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* $\mathrm{End}_K(E)$ *identifizieren, dann gilt*

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

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

Beweis. Dass das Dualisieren eine Anti-Involution ist, folgt aus Satz 10.9, (i) bis (iii). Der erste Teil dieses Satzes zeigt auch $\phi\hat{\phi} = \deg(\phi)$. Um zu sehen, dass $\phi + \hat{\phi} \in \mathbb{Z}$ ist, betrachten wir

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

Dass deg eine quadratische Form ist, bedeutet (definitionsgemäß), dass

$$(\phi, \psi) \mapsto \deg(\phi+\psi) - \deg(\phi) - \deg(\psi)$$

$\mathbb{Z}$-bilinear in ϕ und ψ ist. Das ist aber klar, da sich die rechte Seite umformen lässt zu $\phi\hat{\psi} + \psi\hat{\phi}$ (und da $\phi \mapsto \hat{\phi}$ natürlich $\mathbb{Z}$-linear ist). deg ist positiv definit, weil $\deg(\phi) \geq 0$ für alle ϕ, und $\deg(\phi) = 0$ 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 $\mathrm{End}_K(E)$ als $\mathbb{Z}$-Modul durch 4 beschränkt).

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

10.13. **Satz.** *Sei* K *algebraisch abgeschlossen und* $\phi : E \to E'$ *eine nicht-konstante Isogenie. Dann gilt für alle* $P \in E'(K)$*, dass*

$$\#\phi^{-1}(P) = \deg_s(\phi)\,.$$

Insbesondere hat der Kern von ϕ *die Ordnung* $\deg_s(\phi)$.

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 $f_1(x^{p^k}) - \xi f_2(x^{p^k}) = 0$, wo p^k der inseparable Grad von ϕ ist und $\deg_s \phi = \max\{\deg f_1, \deg f_2\}$. (Dabei ist p die Charakteristik von K. Im Fall $\mathrm{char}(K) = 0$ ist $p^k = \deg_i \phi = 1$.) Für fast alle ξ gibt es dann genau $\deg_s \phi$ Lösungen in K, die zu genau $\deg_s \phi$ Punkten Q führen mit $\phi(Q) = P$. Da die Mengen $\phi^{-1}(P)$ für verschiedene P durch Translation (Addition eines geeignetes Punktes in $E(K)$) auseinander hervorgehen, müssen dann alle diese Mengen genau $\deg_s \phi$ Elemente haben. □

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 Funktionenkörper) 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 alle Koeffizienten a_j durch ihre 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

$$\phi : E \ni (x, y) \mapsto (x^q, y^q) \in E^{(q)}$$

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

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

(1) *Sei $\phi : E \ni (x, y) \mapsto (x^p, y^p) \in E^{(p)}$. Dann ist ϕ rein inseparabel:* $\deg(\phi) = \deg_i(\phi) = p$.
(2) *Sei $\pi \in \mathrm{End}_K(E)$ der Frobenius-Endomorphismus. Dann ist für $m, n \in \mathbb{Z}$ der Endomorphismus $m + n\pi$ separabel genau dann, 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 Bemerkung 10.10.

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 $\psi : E^{(p)} \ni (x, y) \mapsto (x^{p^{e-1}}, y^{p^{e-1}}) \in E$). 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 separabel ist, wenn $p \nmid m$, 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. □

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

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

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).

11. Torsion und Weil-Paarung

In diesem Abschnitt wollen wir die Struktur der *n-Torsionspunkte* einer elliptischen Kurve genauer untersuchen. Das sind die Punkte P mit $n \cdot P = 0$.

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

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

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

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

$$E[p^e] = \{O\} \qquad \text{für } e = 1, 2, 3, \ldots, \quad \text{oder}$$
$$E[p^e] \cong \mathbb{Z}/p^e\mathbb{Z} \quad \text{für } e = 1, 2, 3, \ldots.$$

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

Beweis. (1) In diesem Fall ist $[m]$ separabel, also gilt $\#E[m] = \deg([m]) = m^2$. 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 $\phi : E \ni (x, y) \mapsto (x^p, y^p) \in E^{(p)}$, und sei $\hat{\phi} : E^{(p)} \to E$ die duale Isogenie. Dann gilt

$$\#E[p^e] = \deg_s[p^e] = (\deg_s[p])^e = (\deg_s \hat{\phi}\phi)^e = (\deg_s \hat{\phi})^e\,;$$

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

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*

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

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

(1) *e_m ist bilinear:*

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

(2) *e_m ist alternierend: $e_m(T, T) = 1$.*
(3) *e_m ist nicht-ausgeartet: Wenn $e_m(S, T) = 1$ für alle $T \in E[m]$, dann ist $S = O$. Insbesondere ist e_m surjektiv.*
(4) *e_m ist verträglich mit der Operation der Galoisgruppe von $\bar{K}$ über K, d.h. für $\sigma \in \operatorname{Gal}(\bar{K}/K)$ 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 $\phi : E \to E'$ 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, dass es rationale Funktionen l_j in $\bar{K}(E)$ gibt, so dass 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 $P_1, \ldots, P_n$ und Polstellen in $Q_1, \ldots, Q_n$ (mit Vielfachheit gerechnet), falls $P_1 + \cdots + P_n = Q_1 + \cdots + Q_n$. 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 die selben 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

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

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

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

(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 $P \in E$

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

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 $f_{S_1+S_2} h^m = c f_{S_1} f_{S_2}$ mit einer Konstanten $c \neq 0$. Es folgt $g_{S_1+S_2}(h \circ [m]) = c' g_{S_1} g_{S_2}$ 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

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

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-Ausgeartetheit 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

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

(5) ist nicht allzu schwer (Übung).

(6) beweisen wie 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.*

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 $d_1 d_2 = \#G$, 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 Galoisgruppe $\mathrm{Gal}(\bar{K}/K)$ 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 $d_1 = \mathrm{ord}(\zeta) \mid \#\mu(K)$. □

Dieser Satz ist eine Analogie 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 $\mu(\mathbb{Q}) = \{\pm 1\}$, 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 $\#T \le 10$ oder $= 12$ gibt. Im anderen Fall muss $d \le 4$ 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.

12.1. **Wiederholung: Endliche Körper.** Zur Erinnerung hier eine Zusammenstellung der wichtigsten Tatsachen über endliche Körper.

(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 Galoisgruppe der Ordnung n. Die Galoisgruppe wird erzeugt vom *Frobeniusautomorphismus* $x \mapsto x^q$.
(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
$$\mathbb{F}_q^\times = \{x \in \bar{\mathbb{F}}_q \mid x^{q-1} = 1\} = \mu_{q-1}\,.$$

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.

Anzahl der rationalen Punkte. Eine heuristische Überlegung lässt einen vermuten, dass eine elliptische Kurve über dem endlichen Körper $\mathbb{F}_q$ ungefähr q rationale Punkte haben sollte. Das stimmt tatsächlich, und man kann die Abweichung sogar sehr genau beschränken.

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

(1) *Sei $t = \phi + \hat{\phi} \in \mathbb{Z}$ die* Spur des Frobenius. *Dann gilt in $\mathrm{End}_{\mathbb{F}_q}(E)$ die Relation*
$$\phi^2 - t\,\phi + q = 0\,,$$
und $|t| \leq 2\sqrt{q}$.
(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) Wir haben in $\mathrm{End}_{\mathbb{F}_q}(E)$
$$0 = (\phi - \phi)(\phi - \hat{\phi}) = \phi^2 - (\phi + \hat{\phi})\phi + \phi\hat{\phi} = \phi^2 - t\,\phi + q\,,$$
denn $\phi\hat{\phi} = \deg(\phi) = q$.

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

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

also hat das Polynom $X^2 - t\,X + q$ 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 (Prop. 10.14), gilt $\#E(\mathbb{F}_q) = \#\ker(\phi - 1) = \deg(\phi - 1)$ (Satz 10.13). Außerdem ist

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

□

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

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

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 äquivalent:*

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

Beweis. Wir werden hier nur die Richtung „(1) $\Longrightarrow$ (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 $\psi : E \longrightarrow E'$. 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 $\mathrm{End}(E)$. Da $\deg(\psi) \neq 0$ und $\mathrm{End}(E)$ ein Integritätsring der Charakteristik 0 ist (Satz 10.8), folgt $t = t'$, also nach Satz 12.2 auch $\#E(\mathbb{F}_q) = \#E'(\mathbb{F}_q)$. □

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 $n = 2, 3, 4, \ldots$. In diesem Abschnitt wollen wir uns damit beschäftigen, wie die Zahlen

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

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.

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

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 $C(\bar{\mathbb{F}}_q) = \bigcup_{n\geq 1} C(\mathbb{F}_q)$. Sie zerfällt in Bahnen unter der Operation des Frobenius-Endomorphismus ϕ. Sei a_d die Anzahl der Bahnen der Länge d. Dann gilt $\#C(\mathbb{F}_n) = \sum_{d|n} d a_d$, und die Zetafunktion schreibt sich als

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

(Siehe Übungen.) Außerdem stellt sich heraus, dass die Zetafunktion in der definierten Form eine besonders einfache Gestalt erhält.

Über diese Zetafunktion gilt nun folgender Satz.

12.5. **Satz.** (Weil-Vermutungen für Kurven) *Sei C eine glatte projektive Kurve über $\mathbb{F}_q$. Dann gilt:*

(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* $P(T)$ *vom Grad* $2g$, *das über* $\mathbb{C}$ *faktorisiert als*

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

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

12.6. **Kommentar.**

(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 $\zeta(C,1-s) = q^{(g-1)(2s-1)}\zeta(C,s)$, was auch 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 $\phi \in \mathrm{End}(E)$ der Frobenius-Endomorphismus. Wir hatten gesehen, dass ϕ die Gleichung $X^2 - t\,X + q = 0$ löst (Satz 12.2), wobei $t = \phi + \hat{\phi}$ die Spur des Frobenius ist. Weiterhin galt $|t| \le 2\sqrt{q}$, woraus folgt, dass

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

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

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

das heißt, dass wir einen Isomorphismus

$$\mathbb{Z}[\alpha] \longrightarrow \mathbb{Z}[\phi] \subset \mathrm{End}(E)\,, \qquad \alpha \mapsto \phi$$

haben. (Im Falle $\alpha = \pm\sqrt{q}$ verwenden wir dabei, dass $\mathrm{End}(E)$ ein Integritätsring 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

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

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

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 Primzahltest: 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:

$$a^{p-1} \equiv 1 \bmod p\,.$$

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 $a^{N-1} \equiv 1 \bmod N$ (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 $a^{N-1} \not\equiv 1 \bmod N$. Hierbei ist wichtig, dass sich der Rest $a^{N-1} \bmod N$ 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 1994[1])
Es gibt unendlich viele Carmichael-Zahlen.

Wenn N eine Carmichael-Zahl ist, dann gilt stets $a^{N-1} \equiv 1 \bmod N$, außer wenn $\mathrm{ggT}(a, N) > 1$, 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.

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:

$$a^q \equiv 1 \bmod N \qquad \text{oder} \qquad a^{2^e q} \equiv -1 \bmod N \text{ für ein } 0 \le e < t.$$

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. **Proposition.**

(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 $1 < a < N$.*

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

(2) Wir betrachten zunächst den Homomorphismus

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

Sei $G \subset (\mathbb{Z}/N\mathbb{Z})^\times$ sein Kern. Dann gilt $\#G \le \#(\mathbb{Z}/N\mathbb{Z})^\times < N$. ($N$ 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

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

$\mathrm{ggT}(N-1,(p_j-1)p_j^{e_j-1})$ hat. Sei $G_j' \subset G_j$ die Untergruppe vom Index 2. Für $a \in \mathbb{Z}$ gilt dann: Ist $a \bmod p_j^{e_j}$ in G_j', dann ist $a^{(N-1)/2} \equiv 1 \bmod p_j^{e_j}$, andernfalls ist $a^{(N-1)/2} \equiv -1 \bmod p_j^{e_j}$. 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

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

ein surjektiver Gruppenhomomorphismus. Es gilt

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

Es folgt

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

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

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

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

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

so dass die Behauptung ebenfalls gilt. □

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

13.5. **Algorithmus.** (Miller-Rabin-Test)
Eingabe: $N > 1$; $m \geq 1$ (Anzahl der Tests)
  Schreibe $N - 1 = 2^t q$ mit q ungerade
  Für $j = 1, \ldots, m$:
    Wähle $1 < a < N$ zufällig und berechne $b := a^q \bmod N$
    Falls $b = \pm 1$: nimm das nächste j
    Für $e = 1, \ldots, t-1$:
      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 Proposition 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. **Proposition.** *Sei $N > 0$ eine ganze Zahl und p ein Primteiler von $N-1$. Sei weiter $a_p \in \mathbb{Z}$ mit*

$$a_p^{N-1} \equiv 1 \bmod N \qquad \text{und} \qquad (a_p^{(N-1)/p} - 1) \perp N\,. \tag{13.1}$$

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

$$d \equiv 1 \bmod p^{e_p}\,.$$

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

Wenn wir über die Faktorisierung von $N-1$ gut genug bescheid wissen, 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.*

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. so dass (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 $d \equiv 1 \bmod F$ 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. □

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, so dass 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 Saxena[2] 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

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

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

$$(X-a)^N \equiv X^N - a \bmod (N, X^r - 1)$$

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.

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 $\mathrm{ord}_r(N) > (\log_2 N)^2$.
Wenn $1 < \mathrm{ggT}(a, N) < N$ für ein $1 \leq a \leq r$, gib aus „zusammengesetzt“; Stop.
Wenn $N \leq r$, gib aus „prim“; Stop.
Für $a = 1, \ldots, \lfloor \sqrt{\phi(r)} \log_2 N \rfloor$:
  Wenn $(X-a)^N \not\equiv X^N - a \bmod (N, X^r - 1)$:
    gib aus „zusammengesetzt“; Stop.
Gib aus „prim“; Stop.

Hier bezeichnet $\mathrm{ord}_r(N)$ die Ordnung von N in der multiplikativen Gruppe $(\mathbb{Z}/r\mathbb{Z})^\times$, und $\phi(r)$ 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. Kommen wir nun zur 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.

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

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.

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, so dass $p-1$ B-potenzglatt ist.

Die Idee ist wie folgt. Wir wählen eine Schranke B und eine ganze Zahl a. Wenn N einen Primteiler p hat, so dass $p-1$ B-potenzglatt ist, dann ist $p-1$ ein Teiler von $L(B) = \mathrm{kgV}(1,2,\ldots,B)$, und nach dem kleinen Satz von Fermat gilt $a^{L(B)} \equiv 1 \bmod p$, also

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

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 $a^{L(1)} \bmod N$, $a^{L(2)} \bmod N$, …, $a^{L(B)} \bmod N$ 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 $a^{59} \equiv 25 \bmod 119$ und $a^{118} \equiv 30 \bmod 119$, so dass N den Test nicht besteht.

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

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

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

Man kann auch diesen Algorithmus modifizieren, so dass er eine Gruppe der Ordnung $p+1$ verwendet; dann findet man Teiler p, so dass $p+1$ 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 faktorisierende Zahl. Wir betrachten eine einfach auszuwertende Funktion $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{Z}/N\mathbb{Z}$, 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 $(x_n \bmod p)$ irgendwann periodisch; es gibt also n und $m \geq 1$ mit $x_{n+m} \equiv x_n \bmod p$. Mit etwas Glück gilt diese Relation nicht für alle Primteiler von N, und wir erhalten einen nichttrivialen Teiler von N mittels $\mathrm{ggT}(x_{n+m} - x_n, N)$.

Um die Anzahl der Vergleiche in einem vernünftigen Rahmen zu halten, kann man parallel zu (x_n) die Folge (x_{2n}) berechnen und dann jeweils $\mathrm{ggT}(x_{2n} - x_n, N)$ 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, wo 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

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

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

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 $x^2 \equiv y^2 \bmod N$, dann ist mit Wahrscheinlichkeit $\geq 1/2$ (unter der Annahme, dass x und y zufällig aus den Restklassen mod N gewählt sind) $\mathrm{ggT}(x - y, N)$ ein nichttrivialer Teiler von N. Der Grund dafür ist, dass für jeden Primteiler p von N $x \equiv \varepsilon_p y \bmod p^{v_p(N)}$ 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 $x^2 \equiv y^2 \bmod N$ zu generieren. Dazu legt man eine Schranke B fest und betrachtet die Primzahlen $q_1, q_2, \ldots, q_k$, die kleiner als B sind. Die Menge $\{-1, q_1, \ldots, q_k\}$ heißt die *Faktorbasis*. Man versucht dann, Relationen der Form

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

zu bekommen. Hat man genügend viele davon gesammelt, kann man (durch lineare Algebra über $\mathbb{F}_2$) Teilmengen dieser Relationen finden, so dass 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 $k = 1, 2, \ldots$ 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, so dass man hoffen kann, dass sich t über der Faktorbasis faktorisieren lässt. Beachte, dass $r^2 \equiv t \bmod N$.

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

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

Wir erhalten die Relationen

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

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

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

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

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 mod p abhängt. Dadurch lassen sich die Primfaktoren aus der Faktorbasis sehr schnell 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 $O(e^{c\sqrt[3]{\log N (\log\log N)^2}})$, wird aber wegen der komplizierteren Rechnungen erst in einem Bereich schneller, der schon an der Grenze des Machbaren liegt.

14. Faktorisierung und Primzahltest 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 (so dass die Diskriminante invertierbar ist).

Wir bemerken noch, dass ein Ringhomomorphismus $\phi : R \to S$ eine Abbildung $\mathbb{P}^2(R) \to \mathbb{P}^2(S)$ 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 $E(R) \to E'(S)$. (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 vorausstzen, dass $N \perp 6$, d.h. dass 6 in $Z/N\mathbb{Z}$ invertierbar ist. In diesem Fall lässt sich eine lange Weierstraß-Gleichung wieder transformieren in eine kurze Weierstraß-Gleichung (affin geschrieben)

$$E : y^2 = x^3 + a\,x + b$$

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

Es ist zwar 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 Prop. 13.6.

14.1. **Proposition.** *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, so dass gilt*

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

Dann ist N prim.

Beweis. Angenommen, N ist nicht prim; dann gibt es einen Primteiler p von N mit $p \le \sqrt{N}$. Der kanonische Homomorphismus $\mathbb{Z}/N\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z} = \mathbb{F}_p$ 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 ζ mod N invertierbar ist, also auch mod p nicht verschwindet). Es folgt, dass q diese Ordnung n teilt. Andererseits gilt aber

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

ein Widerspruch. □

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

14.2. **Proposition.** *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.*

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. $(m/q) \cdot E(\mathbb{Z}/N\mathbb{Z}) = 0$. 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

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

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. □

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

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 $4\,a^3 + 27\,b^2 \in (\mathbb{Z}/N\mathbb{Z})^\times$. Sei E die durch $y^2 = x^3 + a\,x + b$ gegebene elliptische Kurve über $\mathbb{Z}/N\mathbb{Z}$.
2. Wir benutzen den Polynomzeit-Algorithmus von Schoof-Elkies-Atkin, um $m = \#E(\mathbb{Z}/N\mathbb{Z})$ zu berechnen. (Wenn dabei etwas schief geht, dann wissen wir, dass N nicht prim ist.)
3. Durch Probedivision (bis zu einer vernünftigen Schranke) faktorisieren wir $m = u \cdot q$, wo u nur kleine Primteiler hat. 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 Jacobisymbol $\left(\frac{x^3+ax+b}{N}\right)$ den Wert 0 oder 1 hat. Dann finden wir $y \in \mathbb{Z}/N\mathbb{Z}$ mit $y^2 = x^3 + a\,x + b$. (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 Prop. 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).

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.

Das größte praktische Problem ist die Bestimmung von m in Schritt 2. 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, von 1000-stelligen Zahlen zu beweisen, dass sie prim sind. Damit ist er etwa 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 anderen 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 Proposition 14.1 sehr schnell davon überzeugen, dass N tatsächlich prim ist.

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 genau so 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 $E(\mathbb{Z}/N\mathbb{Z}) \to E'(\mathbb{F}_p)$. 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 ζ mod 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 $i = 1, \ldots, m$ wiederhole Schritte 2 bis 5.
2. Wähle $A \in \{1, \ldots, N-1\}$ zufällig
und setze $d_1 = \mathrm{ggT}(A, N)$, $d_2 = \mathrm{ggT}(4A + 27, N)$.
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 $p \in \{\text{Primzahlen} \le B\}$, 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 $\mathrm{ggT}(d, N)$ 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 $\ell(x)^{-1/(2a)}$.*

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 $B \to \infty$ gilt* $\log L(B) \sim B$.

Diese Aussage ist äquivalent zum Primzahlsatz, der sagt, dass für die Anzahl $\pi(x)$ der Primzahlen $\le x$ 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 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 $e^{c \log N} = N^c$

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

(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.

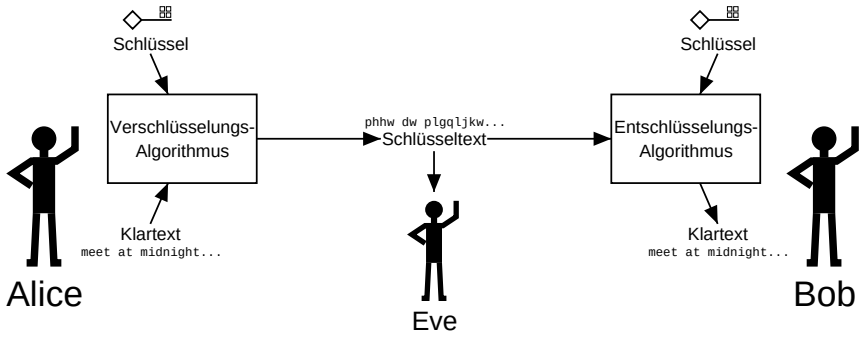
Faktoriserungsalgorithmen, die in Computeralgebrasystemen implementiert sind, verwenden in der Regel verschiedene Methoden nacheinander. Üblicherweise beginnt man mit Probedivision durch Primzahlen aus einer gegebenen Liste. Dann prüft man, ob der verbleibende Faktor prim ist. Wenn nicht, kann man die $p-1$- 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; das wird sich aber wohl bald ändern.

Hier ist ein Vergleich der verschiedenen Komplexitätsklassen.

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

15. Kryptographie: Grundlagen

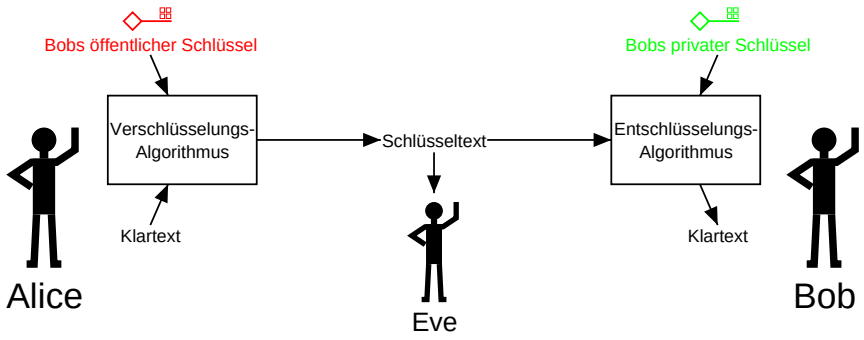
Die Grundaufgabe der Kryptographie besteht darin, eine geheime Nachricht („Klartext“) sicher vom Sender („Alice“) zum Empfänger („Bob“) zu bringen, obwohl der Übertragungskanal (von „Eve“) abgehört werden kann. Die Nachricht muss also so verschlüsselt werden (in einen „Schlüsseltext“), dass sie von möglichen 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 ebenfals 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. **Verfahren.** (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 $(p-1)(q-1)$.
3. Berechne d mit $de \equiv 1 \bmod (p-1)(q-1)$.
4. **Öffentlicher Schlüssel:** (N, e),
   Privater Schlüssel: d.
5. **Verschlüsselung:**
   $\{0, 1, \ldots, N-1\} \ni m \longmapsto c = (m^e \bmod N) \in \{0, 1, \ldots, N-1\}$.
6. **Entschlüsselung:**
   $\{0, 1, \ldots, N-1\} \ni c \longmapsto m = (c^d \bmod N) \in \{0, 1, \ldots, N-1\}$.

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

$$m^{de} = m \cdot (m^{p-1})^{k(q-1)} \equiv m \bmod p$$

und ebenso mod q, wobei wir $de = 1 + k(p-1)(q-1)$ gesetzt haben.

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: Kennt man d, dann ist $de - 1 = k(p-1)(q-1) = 2^t u$ mit u ungerade. Für zufälliges a wird mit Wahrscheinlichkeit wenigstens $1/2$ eine der Zahlen $\mathrm{ggT}(a^u - 1, N)$, $\mathrm{ggT}(a^{2u} - 1, N)$, $\ldots$, $\mathrm{ggT}(a^{2^{t-1}u} - 1, N)$ einen nichttrivialen Teiler liefern.

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. Sei dazu $G = \langle g \rangle$ 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.

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.

15.2. **Verfahren.** (Diffie-Hellman)

1. Man einigt sich auf eine Gruppe G mit Erzeuger g.
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 $s = A^b$.

Wegen $A^b = (g^a)^b = g^{ab} = g^{ba} = (g^b)^a = B^a$ berechnen beide tatsächlich das selbe 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 zu Grunde liegende Idee auch direkt zum Verschlüsseln benutzen.

15.3. **Verfahren.** (El Gamal)

1. Man einigt 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 Gewinn bringend einsetzen kann, möchte ich noch ein wenig auf Algorithmen für diskrete Logarithmen eingehen.

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

15.4. **Algorithmus.** (Durchprobieren)

1. Setze $x := 1_G$.
2. Für $a = 0, 1, \ldots, n-1$ 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 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. Das führt auf folgenden Algorithmus.

15.5. **Algorithmus.** (Baby-Step-Giant-Step)

1. Sei $m := \lceil \sqrt{n} \rceil$ und $\gamma := g^m$.
2. Berechne $\gamma^0, \gamma^1, \ldots, \gamma^{m-1}$ und speichere die Paare (j, γ^j) in einer Tabelle T.
3. Für $r = 0, 1, \ldots, m-1$ 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

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

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

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

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)

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

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.

15.6. **Algorithmus.** (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 $m = 1, 2, \ldots$ 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
$$a(y - y_m) \equiv x_m - x \bmod n$$
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

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

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 $y'_{m'} \not\equiv y_m \bmod n$, 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, so dass 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.

15.7. **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 Hellman[4] 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 $j = 1, \ldots, k$). 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 $a \bmod p_j^{e_j}$. Mit dem Chinesischen Restsatz können wir aus diesen Informationen a berechnen.

Jetzt nehmen wir an, G habe Primzahlpotenzordnung $n = p^e$. Wir bestimmen zunächst $a \bmod p$. 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 $a \bmod p$. Sei etwa $a \equiv a_0 \bmod p$. 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

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

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.

[4] G.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).

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.

15.8. **Algorithmus.** (Index Calculus)

1. Initialisiere eine leere Liste L.
2. Wiederhole Schritte 3 und 4 solange, bis $\#L \geq \#F_B + 10$.
3. Wähle $x \in \{1, \ldots, p-2\}$ 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 $x \in \{0, \ldots, p-2\}$ 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 der selben 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 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 $O(e^{c\sqrt[3]{\log x (\log\log x)^2}})$.

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. **Verfahren.** (Diffie-Hellman-Schlüsselaustausch mit elliptischen Kurven)

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. **Verfahren.** (ElGamal-Verschlüsselung mit elliptischen Kurven)

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, so dass 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 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) 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 $q \equiv 1 \bmod n$, dann ist die Tate-Paarung auch nicht-ausgeartet.

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 Ordnung $n \perp q$. Sei l die kleinste Zahl, so dass $q^l \equiv 1 \bmod n$ ist. Wenn $l > 1$, dann gilt $E[n] \subset E(\mathbb{F}_{q^l})$.*

Beweis. ... □

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 $l > 1$, dann gibt es nach dem Lemma einen Punkt $P' \in E(\mathbb{F}_{q^l})$, der nicht in $\langle P\rangle$ liegt. Unter diesen Umständen gilt

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

Wir haben den Isomorphismus

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

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

$$r = \alpha\big(\langle P, P'\rangle_{\text{Tate}}\big) \quad \text{und} \quad s = \alpha\big(\langle Q, P'\rangle_{\text{Tate}}\big) \,.$$

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 diesm Fall ist $\langle P, P\rangle_{\text{Tate}}$ 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 $q \equiv 1 \bmod n$ 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 $\mathbb{Q}_p$, so dass 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 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

$$\begin{array}{ccccc} E(\mathbb{F}_p) & \longrightarrow & \tilde{E}_1(\mathbb{Q}_p)/\tilde{E}_2(\mathbb{Q}_p) & \longrightarrow & \mathbb{Z}/p\mathbb{Z} \\ Q & \longmapsto & p\tilde{Q} & & \\ & & R & \longmapsto & \frac{x}{py}(R) \bmod p \end{array}$$

falls $p\tilde{P} \notin \tilde{E}_2(\mathbb{Q}_p)$. 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 jeweils 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. Struktur der Gruppe $E(\mathbb{Q})$

(Für den Inhalt der vier von Dr. Elsenhans vertretenen Vorlesungstermine wird hier demnächst der Text nachgeliefert.)

17.1. **Beispiel.** Wir wollen an einem einfachen Beispiel sehen, wie man die Gruppe $E(\mathbb{Q})$ bestimmen kann. Wir betrachten die elliptische Kurve

$$E : y^2 = x^3 - x = (x+1)x(x-1)\,.$$

Wir haben die Abbildung

$$\phi : E(\mathbb{Q}) \longrightarrow H \times H \times H\,, \qquad (\xi, \eta) \longmapsto (\xi + 1, \xi, \xi - 1)\,,$$

wobei $H = \langle -1, 2\rangle \subset \mathbb{Q}^\times/(\mathbb{Q}^\times)^2$ ist und die Werte auf den 2-Torsionspunkten durch

$$\begin{aligned} \phi(O) &= (1,1,1)\,, \quad \phi\big((-1,0)\big) = (2,-1,-2)\,, \\ \phi\big((0,0\big) &= (1,-1,-1)\,, \quad \phi\big((1,0)\big) = (2,1,2) \end{aligned}$$

gegeben sind. Wir wissen, dass ϕ ein Homomorphismus mit Kern $2E(\mathbb{Q})$ ist. Es folgt

$$\mathrm{im}(\phi) \cong E(\mathbb{Q})/2E(\mathbb{Q}) \cong (\mathbb{Z}/2\mathbb{Z})^{r+2}$$

wobei r der Rang von $E(\mathbb{Q})$ ist.

Da das Produkt der drei Komponenten in $\phi(P)$ stets trivial ist, ist die dritte Komponente durch die ersten beiden eindeutig bestimmt. Durch Projektion auf die ersten beiden Komponenten erhalten wir einen Homomorphismus

$$\phi' : E(\mathbb{Q}) \longrightarrow H \times H\,,$$

dessen Kern ebenfalls $2E(\mathbb{Q})$ ist. Da $H \cong (\mathbb{Z}/2\mathbb{Z})^2$, liefert das bereits die Abschätzung $r \leq 2$. Wir betrachten nun den Homomorphismus ϕ' genauer, um sein Bild besser abzuschätzen.

Zuerst betrachten wir die möglichen Vorzeichen der Komponenten von ϕ'. Da $\eta^2 \geq 0$, ist $-1 \leq \xi \leq 0$ oder aber $\xi \geq 1$. In jedem Fall ist $\xi \geq -1$; damit ist die erste Komponente von $\phi'(P)$ stets positiv (beachte, dass $\phi'\big((-1,0)\big) = (2,-1)$). Das zeigt schon einmal, dass

$$\operatorname{im}(\phi') \subset \langle 2\rangle \times \langle -1, 2\rangle$$

und damit $r \leq 1$.

Jetzt betrachten wir die 2-adische Bewertung von ξ. Wenn $v_2(\xi) < 0$ ist, dann haben $\xi+1$, ξ und $\xi-1$ dieselbe Bewertung; es folgt $3v_2(\xi) = 2v_2(\eta)$, also ist $v_2(\xi)$ gerade. Wenn $v_2(\xi) > 0$ ist, dann sind $\xi+1$ und $\xi-1$ ungerade, also haben wir

$$2v_2(\eta) = v_2(\eta^2) = v_2\big((\xi+1)\xi(\xi-1)\big) = v_2(\xi)\,,$$

und $v_2(\xi)$ ist ebenfalls gerade. Wir sehen, dass $v_2(\xi)$ nicht ungerade sein kann. Damit folgt

$$\operatorname{im}(\phi') \subset \langle 2\rangle \times \langle -1\rangle$$

und somit $r = 0$.

Die Gruppe $E(\mathbb{Q})$ ist also endlich, und alle ihre Elemente sind Torsionspunkte. Wir kennen bereits vier davon, nämlich die 2-Torsionspunkte O, $(-1,0)$, $(0,0)$ und $(1,0)$. Gibt es weitere? Dies können wir zum Beispiel mit dem Satz von Nagell-Lutz feststellen. Dieser besagt, dass jeder Torsionspunkt (ξ,η), der kein 2-Torsionspunkt ist, ganzzahlige Koordinaten ξ und η hat, wobei η^2 ein Teiler der Diskriminante von E ist. Diese ist 64, also muss $\pm\eta \in \{1,2,4,8\}$ sein. Das Polynom x^3-x-1 ist irreduzibel, also kommt $\eta = \pm 1$ nicht in Frage. Da $x^3 - x$ für $x = \pm 1$ den Wert 0 annimmt, muss eine ganzzahlige Nullstelle ξ von $x^3 - x - \eta^2$ die Form $\pm 2^t$ haben mit $t \geq 1$. Dann ist aber $\xi^3 - \xi = \xi(\xi^2-1) = \pm 2^t(2^{2t}-1)$, und der zweite Faktor ist eine ungerade Zahl > 1, so dass das Produkt kein Teiler von 64 sein kann. Es gibt also keine weiteren Torsionspunkte, und wir haben bewiesen:

$$E(\mathbb{Q}) = \{O, (-1,0), (0,0), (1,0)\}\,.$$

17.2. **Kanonenkugeln.** Als weiteres Beispiel betrachten wir die folgende Aufgabe:

> $N > 1$ Kanonenkugeln sind zu einer Pyramide gestapelt. Durch eine Unachtsamkeit stürzt die Pyramide ein. Um weitere Unfälle zu vermeiden, werden die Kugeln nicht mehr gestapelt, sondern in Form eines Quadrats ausgelegt. Wie viele Kugeln können es gewesen sein?

Sei X die Seitenlänge des Basisquadrats der Pyramide und Y die Seitenlänge des Quadrats aus allen Kugeln. Dann gilt

$$Y^2 = N = \sum_{k=1}^{X} k^2 = \frac{X(X+1)(2X+1)}{6}\,,$$

und wir suchen ganzzahlige Lösungen (X, Y) mit $X, Y > 1$.

Wir setzen $X = x/12$, $Y = y/72$ und erhalten nach Multiplikation mit dem Hauptnenner $72^2 = 12^2 \cdot 6^2$ die Gleichung einer elliptischen Kurve

$$E : y^2 = x(x+6)(x+12)\,.$$

Wenn wir die ganzzahligen Punkte auf dieser Kurve finden können, dann können wir auch das ursprüngliche Problem lösen; denn jede Lösung (X, Y) der ersten Gleichung führt zu einer Lösung $(x, y) = (12X, 72Y)$ der zweiten Gleichung.

Zuerst müssen wir die Gruppe $E(\mathbb{Q})$ bestimmen. Der Homomorphismus ist nun

$$\phi : E(\mathbb{Q}) \longrightarrow H \times H \times H\,, \qquad (\xi, \eta) \longmapsto (\xi, \xi+6, \xi+12)$$

mit $H = \langle -1, 2, 3\rangle$. Wir projizieren wieder auf die ersten beiden Komponenten und erhalten $\phi' : E(\mathbb{Q}) \to H \times H$. Die anfängliche Schranke für den Rang ist hier $r \le 4$, da $H \times H \cong (\mathbb{Z}/2\mathbb{Z})^6$. Wie oben sehen wir, dass die dritte Komponente $\xi + 12$ positiv ist; weil das Produkt der drei Komponenten ebenfalls positiv ist, müssen die beiden Komponenten von ϕ' dasselbe Vorzeichen haben. Wir erhalten $r \le 3$.

Jetzt betrachten wir die 3-adischen Möglichkeiten für ξ. Wenn $v_3(\xi) \le 0$ ist, dann ist $\eta^2/\xi^3 = (1 + 6/\xi)(1 + 12/\xi) \equiv 1 \bmod 3$ und damit ein Quadrat in $\mathbb{Q}_3$; dann müssen auch ξ und $\xi + 6$ Quadrate in $\mathbb{Q}_3$ sein. Ist $v_3(\xi) \ge 2$, dann ist $(\eta/3)^2/\xi = (2 + \xi/3)(4 + \xi/3) \equiv -1 \bmod 3$; es folgt $\xi = -\square$, $\xi + 6 = -3\square$ in $\mathbb{Q}_3$. Es bleibt der Fall $v_3(\xi) = 1$ zu betrachten. Wir setzen $\xi = 3\xi'$ mit $v_3(\xi') = 0$. Die Gleichung wird zu

$$\eta^2 = 3^3\xi'(\xi'+2)(\xi'+4)\,.$$

Wir können ξ' schreiben als $\xi' = 3\xi'' - 2$ oder $3\xi'' - 4$, je nachdem ob $\xi' \equiv 1$ oder $\xi' \equiv 2 \bmod 3$. Im ersten Fall ergibt sich

$$\eta^2 = 3^4\xi''(3\xi''-2)(3\xi''+2)$$

und damit $\xi = 3(3\xi'' - 2) = 3\square$, $\xi + 6 = 3^2\xi'' = -\square$; im zweiten Fall haben wir

$$\eta^2 = 3^4\xi''(3\xi''-4)(3\xi''-2)$$

und damit $\xi = 3(3\xi''-4) = -3\square$, $\xi+6 = 3(3\xi''-2) = 3\square$. Unter Beachtung, dass $2 = -\square$ in $\mathbb{Q}_3$, bleiben folgende Elemente von $H \times H$ als mögliche Bilder übrig:

$$(1,1), (-1,-3), (2,6), (-2,-2), (3,2), (-3,-6), (6,3), (-6,-1)\,.$$

Die Schranke für den Rang ist also $r \le 1$. Die 2-Torsionspunkte werden auf $(1, 1)$, $(2, 6)$, $(-6, -1)$ und $(-3, -6)$ abgebildet. Die triviale Lösung $(X, Y) = (1, 1)$ entspricht dem Punkt $(x, y) = (12, 72)$ mit dem Bild $(3, 2)$. Da das Bild von ϕ' eine Gruppe ist, folgt, dass alle acht möglichen Bilder tatsächlich auftreten; damit ist $r = 1$.

Mit dem Satz von Nagell-Lutz überzeugt man sich relativ schnell davon, dass es außer den 2-Torsionspunkten O, $(0, 0)$, $(-6, 0)$ und $(-12, 0)$ keine weiteren Torsionspunkte gibt. Als abstrakte abelsche Gruppe hat $E(\mathbb{Q})$ also die Form $(\mathbb{Z}/2\mathbb{Z})^2 \times \mathbb{Z}$. Der Punkt $P = (12, 72)$ hat unendliche Ordnung. Man kann eine Schranke für die Höhe der Punkte eines Erzeugendensystems bestimmen und findet auf diese Weise, dass $E(\mathbb{Q})$ von den 2-Torsionspunkten und P erzeugt wird.

Es bleiben die ganzzahligen Punkte (ξ, η) auf E zu bestimmen mit $\xi > 12$ und durch 12 teilbar. Diese liegen alle auf der Zusammenhangskomponente $E(\mathbb{R})^0$ von $E(\mathbb{R})$, die den Punkt O enthält. Sie haben daher die Form nP oder $nP + T$, wobei $T = (0, 0)$ der auf dieser Komponente liegende nichttriviale 2-Torsionspunkt

ist. (Die auf der anderen Zusammenhangskomponente liegenden ganzzahligen Punkte lassen sich sehr leicht bestimmen, da dort $-12 \le x \le -6$ gilt. Wir finden die Punkte $(-12, 0)$, $(-9, \pm 9)$, $(-8, \pm 8)$ und $(-6, 0)$.)

Für die weitere Argumentation benutzen wir den *elliptischen Logarithmus*. Als Gruppe ist $E(\mathbb{R})^0$ isomorph zur Kreisgruppe. Der elliptische Logarithmus liefert einen expliziten Isomorphismus. Sei dazu $f(x) = x(x+6)(x+12)$ die rechte Seite der Gleichung für E. Wir setzen

$$\omega = \int_0^\infty \frac{dx}{\sqrt{f(x)}}\,,$$

das ist die sogenannte *reelle Periode* von E. Der elliptische Logarithmus ist dann definiert als

$$\psi : E(\mathbb{R})^0 \longrightarrow \mathbb{R}/\omega\mathbb{Z}\,, \qquad (\xi, \eta) \longmapsto \operatorname{sign}(\eta) \int_\infty^\xi \frac{dx}{2\sqrt{f(x)}} \quad \text{für } \eta \neq 0\,;$$

man setzt noch $\psi(T) = \pm\omega/2$. Man kann nachprüfen, dass ψ tatsächlich ein stetiger Gruppenisomorphismus ist; das liegt daran, dass die Differentialform $dx/2y$ auf E translationsinvariant (bezüglich der Gruppenstruktur von E) ist.

Wenn $(\xi, \eta) \in E$ ganzzahlig ist mit ξ groß, dann liegt unser Punkt „nahe bei O“. Das bedeutet, dass $|\psi(\xi, \eta)|$ klein ist. Genauer gilt

$$|\psi(\xi, \eta)|^2 \le \frac{c}{\xi}$$

mit einer berechenbaren Konstanten c. Ist $(\xi, \eta) = nP$ oder $nP + T$, dann gilt außerdem (unter Verwendung der Theorie der Höhen)

$$\frac{1}{\xi} \le c_1 e^{-c_2 n^2}$$

mit berechenbaren Konstanten c_1 und c_2, so dass wir eine Abschätzung

$$-\log|\psi(\xi, \eta)| \ge \frac{c_2}{2} n^2 - c'$$

erhalten. Auf der anderen Seite (und das ist die Stelle, wo die wirklich schwere Mathematik eingeht) kann man eine obere Abschätzung der Form

$$-\log|\psi(\xi, \eta)| \le c_3(\log|n| + c_4)(\log\log|n| + c_5)^2$$

beweisen, mit wiederum expliziten Konstanten c_3, c_4 und c_5. Dabei ist im konkreten Fall $c_3 \approx 10^{19}$. Wenn $|n|$ groß wird, wächst die untere Schranke (wie n^2) schneller als die obere (im wesentlichen wie $\log|n|$); für $|n| \ge 4 \cdot 10^{11}$ etwa bekommt man einen Widerspruch. Es folgt $|n| < 4 \cdot 10^{11}$.

Man beachte nun, dass ψ ein Homomorphismus ist (das wird auch schon bei der oberen Abschätzung oben benutzt). Daraus folgt, dass

$$\psi(\xi, \eta) = n\psi(P) + m\omega$$

mit $m \in \mathbb{Z}$ falls $(\xi, \eta) = nP$ und $m \in \mathbb{Z} + \frac{1}{2}$ falls $(\xi, \eta) = nP + T$. Man kann nun eine untere Schranke für $|n\psi(P) + m\omega|$ bestimmen, die für alle $0 < |n| < 4 \cdot 10^{11}$ und alle $m \in \frac{1}{2}\mathbb{Z}$ gültig ist. In unserem Fall erhält man zum Beispiel als untere Schranke $1,7 \cdot 10^{-13}$. Das liefert eine obere Schranke für $-\log|\psi(\xi, \eta)|$, die zusammen mit der obigen unteren Schranke impliziert, dass $|n| \le 8$ sein muss. Die verbleibenden Möglichkeiten sind schnell durchprobiert. Das Ergebnis ist, dass die einzigen ganzzahligen Punkte auf $E(\mathbb{R})^0$ die Punkte $T = (0, 0)$, $\pm P = (12, \pm 72)$,

$\pm P + T = (6, \pm 36)$ und $\pm 2P + T = (288, \pm 5040)$ sind. Nur das letzte Paar hat $\xi > 12$ und durch 12 teilbar. Dies zeigt, dass die einzige Lösung des Kanonenkugelproblems durch $X = 24$, $Y = 70$, $N = 4900$ gegeben ist.

Bemerkung. Es gibt auch eine „elementare“ Lösung des Kanonenkugelproblems, siehe z.B. [Co2] (wo man auch eine recht ausführliche Beschreibung der Methode zur Bestimmung der ganzzahligen Punkte findet).

17.3. **Die Selmergruppe.** Zum Abschluss möchte ich die in den Beispielen benutzte Methode, den Rang nach oben abzuschätzen, etwas systematischer behandeln.

Wir betrachten wiederum eine elliptische Kurve

$$E : y^2 = (x-a)(x-b)(x-c)$$

mit $a, b, c \in \mathbb{Z}$ und den Homomorphismus

$$\phi : E(\mathbb{Q}) \longrightarrow H_1 = \{(\alpha, \beta, \gamma) \in H \times H \times H \mid \alpha\beta\gamma = 1\}$$

mit $H = \langle -1, 2, p_1, \dots, p_k \rangle \subset \mathbb{Q}^\times/(\mathbb{Q}^\times)^2$, wo $p_1, \dots, p_k$ die ungeraden Primteiler der Differenzen $b-a$, $c-a$, $c-b$ sind. Wir können annehmen, dass $a < b < c$ ist.

Unser Ziel ist es, eine möglichst kleine Untergruppe H_0 von H_1 zu bestimmen, die das Bild von ϕ enthält. Man beachte, dass man die Gruppe $\mathbb{Q}^\times/(\mathbb{Q}^\times)^2$ auch als $\mathbb{F}_2$-Vektorraum auffassen kann; die Untergruppen sind dann Untervektorräume, und wir haben

$$2 + r = \dim_{\mathbb{F}_2} \operatorname{im}(\phi) \le \dim_{\mathbb{F}_2} H_0 \le \dim_{F_2} H_1 = 4 + 2k\,,$$

wobei r der Rang der Gruppe $E(\mathbb{Q})$ ist.

Die Idee ist, den Homomorphismus ϕ statt über $\mathbb{Q}$ auch über $\mathbb{R}$ und über geeigneten $\mathbb{Q}_p$ zu betrachten. Die früher bewiesenen Eigenschaften von ϕ gelten über jedem Körper (wir haben nirgends benutzt, dass wir über $\mathbb{Q}$ arbeiten). Wir bekommen kommutative Diagramme

$$\begin{array}{ccc} E(\mathbb{Q}) & \xrightarrow{\phi} & H_1 \\ \downarrow & & \downarrow{\scriptstyle \rho_{\mathbb{R}}} \\ E(\mathbb{R}) & \xrightarrow{\phi_{\mathbb{R}}} & H_{1,\mathbb{R}} \end{array} \qquad \text{und} \qquad \begin{array}{ccc} E(\mathbb{Q}) & \xrightarrow{\phi} & H_1 \\ \downarrow & & \downarrow{\scriptstyle \rho_{\mathbb{Q}_p}} \\ E(\mathbb{Q}_p) & \xrightarrow{\phi_{\mathbb{Q}_p}} & H_{1,\mathbb{Q}_p} \end{array}$$

Dabei ist $H_{1,\mathbb{R}} = \{(\alpha, \beta, \gamma) \in \big(\mathbb{R}^\times/(\mathbb{R}^\times)^2\big)^3 \mid \alpha\beta\gamma = 1\}$ und entsprechend für $H_{1,\mathbb{Q}_p}$, und die jeweils rechte senkrechte Abbildung wird von der Inklusion $\mathbb{Q} \subset \mathbb{R}$ bzw. $\mathbb{Q} \subset \mathbb{Q}_p$ induziert. Es folgt

$$\operatorname{im}(\phi) \subset \rho_{\mathbb{R}}^{-1}\big(\operatorname{im}(\phi_{\mathbb{R}})\big) \quad \text{und} \quad \operatorname{im}(\phi) \subset \rho_{\mathbb{Q}_p}^{-1}\big(\operatorname{im}(\phi_{\mathbb{Q}_p})\big)\,.$$

Im ersten Fall finden wir, dass $a \le \xi \le b$ oder $c \le \xi$ für alle $(\xi, \eta) \in E(\mathbb{R})$ gilt, so dass das Bild in $H_{1,\mathbb{R}}$ genau aus den Tripeln $(1,1,1)$ und $(1,-1,-1)$ besteht. Diese bilden einen Untervektorraum der Kodimension 1 in $H_{1,\mathbb{R}}$, und da $\rho_{\mathbb{R}}$ surjektiv ist, führt das zu einer Reduktion der Dimension des Bildes von ϕ um 1, wie wir das auch in den Beispielen gesehen haben.

Für Primzahlen p gilt, dass man für $p \notin \{2, p_1, \dots, p_k\}$ keine neue Information erhält (in diesem Fall kann man zeigen, dass das Bild immer $H_{1,\mathbb{Q}_p} \cap \big(\mathbb{Z}_p^\times/(\mathbb{Z}_p^\times)^2\big)^3$ ist). Für $p = 2$ oder $p = p_j$ sieht das anders aus. Ist $p = p_j$ ungerade, dann gilt $\dim_{\mathbb{F}_2} \operatorname{im}(\phi_{\mathbb{Q}_p}) = 2$, und das Bild ist in vielen Fällen bereits durch $\phi_{\mathbb{Q}_p}(E[2])$

gegeben. Sonst findet man zufällige Punkte in $E(\mathbb{Q}_p)$ (unter Benutzung des Henselschen Lemmas) und berechnet ihr Bild, so lange bis der erzeugte Untervektorraum zweidimensional ist. Für $p = 2$ ist $\dim_{\mathbb{F}_2} \mathrm{im}(\phi_{\mathbb{Q}_2}) = 3$; man braucht also in jedem Fall mindestens einen zusätzlichen Erzeuger.

Die Gruppe

$$\mathrm{Sel}^{(2)}(E/\mathbb{Q}) = \rho_{\mathbb{R}}^{-1}\big(\mathrm{im}(\phi_{\mathbb{R}})\big) \cap \bigcap_{p\in\{2,p_1,\dots,p_k\}} \rho_{\mathbb{Q}_p}^{-1}\big(\mathrm{im}(\phi_{\mathbb{Q}_p})\big) \subset H_1$$

ist die gesuchte Untergruppe H_0. Sie heißt die *2-Selmergruppe* von E über $\mathbb{Q}$. Es gilt

$$r \leq \dim_{\mathbb{F}_2} \mathrm{Sel}^{(2)}(E/\mathbb{Q}) - 2\,.$$

Bemerkungen.

(1) Die Abschätzung für r oben ist oft scharf, aber nicht immer. Man bekommt untere Schranken, indem man nach Punkten in $E(\mathbb{Q})$ sucht und die von ihren Bildern unter ϕ erzeugte Untergruppe bestimmt. Ist diese gleich der Selmergruppe, dann hat man den Rang bestimmt, vergleiche die Beispiele oben.

(2) Allgemein gilt

$$\frac{\mathrm{Sel}^{(2)}(E/\mathbb{Q})}{\mathrm{im}(\phi)} \cong Ш(E/\mathbb{Q})[2]\,.$$

Dabei ist $Ш(E/\mathbb{Q})$ die *Shafarevich-Tate-Gruppe* von E, ein recht mysteriöses Objekt. Es wird vermutet, dass diese Gruppe stets endlich ist; das ist aber nur in gewissen Fällen bewiesen. Die Ordnung dieser Gruppe tritt in der verfeinerten Version der Vermutung von Birch und Swinnerton-Dyer auf; die Fälle, in denen ihre Endlichkeit bekannt ist, sind genau die, für die die Vermutung (aber nicht unbedingt ihre Verfeinerung) bewiesen werden konnte.

(3) Ist $Ш(E/\mathbb{Q})$ endlich, dann ist $\dim_{\mathbb{F}_2} Ш(E/\mathbb{Q})[2]$ gerade. Das bedeutet, dass die Differenz zwischen der oberen Schranke und dem tatsächlichen Rang stets gerade sein sollte.

(4) Der Beweis des schwachen Satzes von Mordell-Weil und die Definition der 2-Selmergruppe können auf beliebige elliptische Kurven über beliebigen algebraischen Zahlkörpern verallgemeinert werden. Dafür werden allerdings Methoden der algebraischen Zahlentheorie benötigt.

(5) Wie der Name vermuten lässt, gibt es außer der 2-Selmergruppe auch eine n-Selmergruppe für jedes $n \geq 2$, die ebenfalls eine Schranke für den Rang r liefert. In diesem Fall liegt das Hindernis dagegen, dass diese Schranke scharf ist, in der n-Torsion der Shafarevich-Tate-Gruppe. Die Berechnung von $\mathrm{Sel}^{(n)}(E/\mathbb{Q})$ ist allerdings im allgemeinen wesentlich komplizierter als die der 2-Selmergruppe, siehe etwa [CF+].

Literatur

[BSS] I. Blake, G. Seroussi, N. Smart: *Elliptic curves in cryptography*, LMS Lecture Notes **265**, Cambridge University Press (1999).

[Buc] J. Buchmann: *Einführung in die Kryptographie*, Springer (1999).

[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).

[Co2] H. Cohen: *Number Theory. Volume I: Tools and diophantine equations*, Springer GTM **239** (2007).

[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).

[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).

[Si2] J.H. Silverman: *Advanced topics in the arithmetic of elliptic curves*, Springer GTM **151** (1994).