Elliptische Kurven I
Vorlesung im Sommersemester 2000

Michael Stoll

KAPITEL 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

$$E : y^2 = x^3 + a\,x + b \tag{0.1}$$

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 (0.1). Es gibt gute Gründe (die bald erklärt werden), zu dieser Menge noch einen Punkt O „im Unendlichen“ dazuzunehmen. 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 $\mu = \eta_1 - \lambda\xi_1$. (Die Gerade durch die beiden Punkte hat die Gleichung $y = \lambda\,x + \mu$.)

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 $7 \cdot P = O$. (Tatschlich 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). Im zweiten Teil der Vorlesung (Wintersemester) 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 $|t| \le 2\sqrt{p}$. (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	–	

Hier ist $|t| \le \lfloor 2\sqrt{23} \rfloor = 9$, 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	4	2	2	3	4	1	4	1	2	1

Sei nun N eine große zusammengesetzte Zahl (daß eine Zahl zusammengesetzt ist, läßt sich recht leicht nachweisen). Wir wollen einen Primfaktor p von N finden. Dazu tun wir erst einmal so, als sei N eine Primzahl. Wir wählen 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 $\xi, \eta \in \mathbb{Z}/N\mathbb{Z}$). 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 $\#\tilde{E}(\mathbb{F}_p) = p+1-t$. Nun multiplizieren wir P mit einer geeigneten natürlichen Zahl M. Wenn $p + 1 - t$ ein Teiler von M ist, dann gilt $M \cdot \tilde{P} = \tilde{O}$. Normalerweise wird aber nicht gelten, daß $M \cdot P = O$ 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, M geeignet und nicht zu groß zu wählen, so daß $m = \#\tilde{E}(\mathbb{F}_p)$ ein Teiler von M ist. Man wählt M etwa als das kleinste gemeinsame Vielfache aller Zahlen zwischen 1 und B. Die Bedingung an m ist dann, „B-glatt“ zu sein, d.h. alle Primzahlpotezen, die m teilen, sind hïchstens gleich B. Tatsächlich kann man zeigen, daß man bei optimaler Wahl von B einen Algorithmus erhält, dessen (erwartete) Laufzeit etwa durch

$$e^{\sqrt{\log N \log\log N}}$$

beschränkt ist — der Algorithmus ist *subexponentiell*.

Als Baby-Beispiel wollen wir die Zahl $N = 851$ faktorisieren. Wir nehmen als Kurve $E : y^2 = x^3 + 9\,x - 9$ über $\mathbb{Z}/851\mathbb{Z}$. Sie hat den Punkt $P = (1, 1)$. Um $M \cdot P$ 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. Die Folge der Multiplikatoren

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

kommt dabei aus der Folge der Primzahlpotenzen

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

Auf diese Weise sammelt man gerade die kleinsten gemeinsamen Vielfachen der ersten paar 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 nichttrivialer 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.

KAPITEL 2

Elliptische Kurven

1. 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.

1.1. Affine ebene Kurven. 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.

Definition 2.1. 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
$$f_L : \mathbb{A}^2_K(L) \longrightarrow L\,.$$
(Umgekehrt ist f durch $f_{\bar{K}}$ eindeutig bestimmt.)

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 auch mit $K[\mathbb{A}^2_K]$ bezeichnet.

(3) Eine *rationale Funktion* auf $\mathbb{A}^2_K$ ist gegeben durch ein Element $f = g/h \in K(X, Y)$. ($K(X, Y)$ ist der Quotientenkörper von $K[X, Y]$. Seine Elemente werden üblicherweise „rationale Funktionen (in zwei Variablen)“ genannt, was im vorliegenden Zusammenhang etwas verwirrend ist.)

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

DEFINITION 2.2. Eine *affine ebene Kurve* C über K ist gegeben durch ein Polynom $0 \neq f \in K[X,Y]$. (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). Umgekehrt bestimmt $g_{\bar{K}}$ die Klasse von g eindeutig.

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 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
$$(g/h)_L : \{P \in C(L) \mid g/h \text{ regulär in } P\} \longrightarrow L\,.$$
(Wir bezeichnen eine rationale Funktion der Einfachheit halber durch einen Repräsentanten.)

(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.

Beispiele 2.3.

(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.)

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.

1.2. 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 (jedenfalls an Beispielen), daß die beschriebenen Punktmengen $\mathbb{C}^2$ oder $C(\mathbb{C})$ nicht kompakt sind (in der üblichen Topologie). 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.

DEFINITION 2.4. 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

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

(Das ist deswegen 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.

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.

Definition 2.5. Eine *projektive ebene Kurve C vom Grad d* 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 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
$$(g/h)_L : \{P \in C(L) \mid g/h \text{ regulär in } P\} \longrightarrow L\,.$$
(Wir bezeichnen der Einfachheit halber die Klasse mit einem Repräsentanten. Wir müssen dabei aber im Kopf behalten, daß wir den Repräsentanten evtl. wechseln müssen, um die Regularität zu prüfen und ggfs. den Wert zu berechnen.)

(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. Der besseren Unterscheidbarkeit halber schreiben wir ab jetzt affine Koordinaten(funktionen) mit kleinen Buchstaben x, y und projektive Koordinaten(funktionen) mit großen Buchstaben X, Y, Z.

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 $\bar{C}$; meistens verwischt man diesen Unterschied).

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).

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.

Beispiele 2.6.

(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 in L ein Quadrat ist (und $\mathrm{char}(L) \neq 2$, sonst ist es der eine Punkt $(1 : 1 : 0)$).

(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.

1.3. 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 (das brauchen wir dann 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.

DEFINITION 2.7. 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 eine Komponente von C). 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.

BEISPIEL 2.8. 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 X und bekommen $H = Y^2Z$, 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.

SATZ 2.9. *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 keine Komponente*

von C ist (d.h. so daß $aX + bY + cZ \nmid F$). 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.

1.4. 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.

Definition 2.10.

(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

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

Die Kurve C heißt (schlechthin) *glatt*, wenn sie in allen Punkten $P \in C(\bar{K})$ glatt ist.

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

Beispiele 2.11.

(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.

Bemerkung 2.12. Sei $C : F(X;Y,Z) = 0$ eine projektive Kurve, sei weiter $P = (\xi : \eta : \zeta) \in C(K)$ und K algebraisch abgeschlossen. Es ist nicht allzu schwer, folgendes zu zeigen (siehe z.B. [**Hus**, Ch. 2]).

(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.

1.5. 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.

Definition 2.13. $C : F(X,Y,Z) = 0$ 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 $F \mid R_i S_j - R_j S_i$ 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 $\phi : C \longrightarrow D$ und $\psi : D \longrightarrow C$ gibt, so daß $\phi \circ \psi = \mathrm{id}_D$ und $\psi \circ \phi = \mathrm{id}_C$. Sind ϕ und ψ sogar Morphismen, dann heißen C und D *isomorph* (und ϕ und ψ sind *Isomorphismen*).

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

Beispiele 2.14.

(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 unbeding 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.

Übungsaufgaben 2.1.

(1) Zeigen Sie, daß die Definition von $i(G, C; P)$ nicht davon abhängt, welche Variable eliminiert wird.
(2) Zeigen Sie, daß für ein homogenes Polynom $F(X, Y, Z)$ vom Grad n gilt:
$$n\,F = X\,\frac{\partial F}{\partial X} + Y\,\frac{\partial F}{\partial Y} + Z\,\frac{\partial F}{\partial Z}\,.$$
Folgern Sie daraus, daß für eine affine Kurve C gilt:
$$P = (\xi, \eta) \text{ glatt auf } C \iff \bar{P} = (\xi : \eta : 1) \text{ glatt auf } \bar{C}\,,$$
wobei $\bar{C}$ der projektive Abschluß von C ist.
(3) Zeigen Sie, daß folgende rationale Abbildung ein Isomorphismus ist. (Der Grundkörper habe Charakteristik $\neq 2$.) C sei die Kurve $X^2 - Y^2 + Z^2 = 0$, G sei die Gerade $Y = 0$, die Abbildung sei $\phi : G \longrightarrow C$, gegeben durch $(X : 0 : Z) \mapsto (X^2 - Z^2 : X^2 + Z^2 : 2XZ)$.
(4) Zeigen Sie, daß die Kurven $C : Y^2Z - X^3 = 0$ und $D : Y = 0$ birational äquivalent sind. (Hinweis: Als Abbildung von D nach C kann man $(X : 0 : Z) \mapsto (X^2Z : X^3 : Z^3)$ nehmen.) C und D sind aber nicht isomorph, da C nicht glatt ist (wo?), D aber schon.

2. Elliptische Kurven

In diesem Abschnitt werden wir elliptische Kurven zunächst über einem beliebigen Grundkörper einführen. Im folgenden Kapitel werden wir uns dann speziell dem Fall eines endlichen Grundkörpers widmen.

2.1. Definition und erste Eigenschaften. 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.

DEFINITION 2.15. 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:

$$E : y^2 + a_1\,xy + a_3\,y = x^3 + a_2\,x^2 + a_4\,x + a_6\,. \tag{2.1}$$

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

Die etwas komische Numerierung der Koeffizienten wird später verständlich werden.

LEMMA 2.16. *Sei E eine elliptische 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.

LEMMA 2.17. *Sei E über K eine elliptische Kurve. Wenn* $\mathrm{char}(K) \neq 2$*, dann ist E isomorph zu einer elliptischen Kurve mit einer Gleichung 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 heißt eine kurze Weierstraß-Gleichung.*

BEWEIS: Der Isomorphismus von E auf E' ist (in projektiven Koordinaten) gegeben durch

$$(X : Y : Z) \longmapsto (2\,X : 2\,Y + a_1\,X + a_3\,Z : 2\,Z)\,;$$

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. □

BEMERKUNG 2.18. Die Transformationen im gerade gezeigten Lemma sind Isomorphismen von elliptischen Kurven im Sinne der im nächsten Abschnitt gegebenen Definition.

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

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

bzw.

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

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.

LEMMA 2.19. *Eine Weierstraß-Gleichung der Form (2.1) definiert genau dann eine elliptische (d.h. eine glatte) Kurve, wenn die Disriminante 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 nächsten Abschnitt). Da es sich um einen Isomorphismus handelt, ändert sich auch nichts daran, ob die Kurve glatt ist oder nicht. (Das haben wir hier zwar nicht bewiesen, ist aber nicht schwer. Wer mag, kann es als Übungsaufgabe betrachten.) 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 (d.h. nicht glatter) 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. □

2.2. Isomorphismen. Da der Punkt O für elliptische Kurven eine so große Rolle spielt (das werden wir im nächsten Abschnitt sehen), fassen wir den Isomorphismusbegriff für elliptische Kurven etwas enger.

Definition 2.20. Zwei (durch Weierstraß-Gleichungen gegebene) elliptische Kurven heißen *isomorph*, wenn es zwischen ihnen einen Isomorphismus (als ebene projektive Kurven) gibt, der den Punkt O fest läßt.

(Man kann aber zeigen, daß zwei elliptische Kurven, die als ebene Kurven isomorph sind, auch als elliptische Kurven isomorph sein müssen.)

Wie sehen solche Isomorphismen aus?

Lemma 2.21. *Seien E und E' elliptische Kurven über K, gegeben durch Weierstraß-Gleichungen mit den Koeffizienten a_j bzw. a'_j. Wenn es einen Isomorphismus $E \to E'$ gibt, dann hat er die Form*

$$(x, y) \longmapsto (u^2\,x + r, u^3\,y + su^2\,x + t)$$

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

Für die Koeffizienten gilt dann

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

(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: Aus Resultaten der Algebraischen Geometrie folgt, daß der Isomorphismus (in projektiven Koordinaten) durch lineare Polynome gegeben sein muß:

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

Wenn wir das voraussetzen, dann sehen wir, daß außer dem Punkt O auch die unendlich ferne Gerade als Ganzes fest gehalten werden muß (denn sie ist die einzige Gerade, die die Kurven E und E' in O mit Vielfachheit 3 schneidet). 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 Form der „Leitkoeffizienten“ u^2 und u^3. Diese ergibt sich aber aus der Form der Weierstraß-Gleichung, die die Beziehung $\alpha_1^3 = \beta_2^2$ impliziert. Schließlich kann u nicht verschwinden, weil der Morphismus sonst konstant wäre.

Die Transformationsformeln für die Koeffizienten rechnet man nach. □

BEMERKUNG 2.22. 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 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 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.

SATZ 2.23. *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 2.17 und der Bemerkung vor Lemma 2.19 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 2.21 zeigt, daß diese beiden Kurven ducrh $(x, y) \mapsto (u^2\,x, u^3\,y)$ isomorph sind.

(2) Aus $j(E) = j(E') = j$ folgt entweder $c_4(E) = c_4(E') = 0$ oder $j \neq 0$ und $c_6(E)^2/c_4(E)^3 = c_6(E')^2/c_4(E')^3$. In beiden 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.

PROPOSITION 2.24. *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 Kruven* 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 2.23 sind die beiden Kurven genau dann bereits über K isomorph, wenn d ein Quadrat ist.

(2) und (3) werden analog bewiesen. □

ÜBUNGSAUFGABEN 2.2.

(1) Betrachten Sie die beiden elliptischen Kurven über $\mathbb{Q}$

$$E : y^2 + y = x^3 - x^2 - 10\,x - 20$$
$$E' : y^2 + y = x^3 - 93\,x + 625$$

Berechnen Sie die Größen $b_j(E)$, $b_j(E')$ sowie $c_j(E)$, $c_j(E')$, $j(E)$, $j(E')$. Sind die Kurven isomorph über $\mathbb{Q}$? Über $\bar{\mathbb{Q}}$?

(2) Wie viele Isomorphieklassen von elliptischen Kurven gibt es über dem endlichen Körper $\mathbb{F}_5$? Über $\mathbb{F}_{13}$?
(3) Bestimmen Sie Repräsentanten aller Isomorphieklassen von elliptischen Kurven über $\mathbb{R}$ mit

$$j = 0\,, \qquad j = 1728\,, \qquad j = 1536\,.$$

(4) Sei $E : y^2 = x^3 - 43\,x + 166$ eine elliptische Kurve über $\mathbb{Q}$. Sei weiter $P = (3, 8) \in E(\mathbb{Q})$. Berechnen Sie (mit den Formeln aus dem Einführungskapitel) die Vielfachen $n \cdot P$ für $2 \leq n \leq 7$.

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

SATZ 2.25. *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.

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 den Satz 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 dirtter 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.

LEMMA 2.26. *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 $D_i(X,Y,Z) = 0$ bzw. $D'_j(X,Y,Z) = 0$ mit linearen Polynomen D_i, D'_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 $D = D_1D_2D_3$ und $D' = D'_1D'_2D'_3$ in diesem Raum und sind linear unabhängig. Wir zeigen, daß die Dimension tatsächlich genau 2 ist, d.h. der Raum wird von D und D' 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 D_1 ein Teiler von F sein: $F = D_1F'$ 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 D_2 ein Teiler von F' sein: $F' = D_2F''$. 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 $F = cD$ 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 (da glatt) 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 dswegen 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.

2.4. Isogenien und Endomorphismen. In diesem und dem folgenden Abschnitt werden wir den größeren Teil der Resultate nicht beweisen. Das geschieht hauptsächlich, damit am Ende des Semesters noch Zeit bleibt, ein paar Anwendungen von elliptischen Kurven zu diskutieren. Ein anderer Grund ist, daß in den meisten Fällen weitergehende Vorbereitungen nötig wären (für die wiederum in einer zweistündigen Vorlesung, die keine Kenntnisse in Algebraischer Geometrie voraussetzt, keine Zeit ist).

Wir hatten bereits den Isomorphismusbegriff für elliptische Kurven dahingehend eingeschränkt, daß wir verlangen, daß der Punkt O festgehalten wird. Allgemeiner kann man auch Morphismen mit dieser Eigenschaft betrachten. Solche Morphismen haben einen eigenen Namen.

Definition 2.27. 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.

Satz 2.28. *Sei $\Phi : E \to E'$ eine Isogenie. Dann gilt $\Phi(P + Q) = \Phi(P) + \Phi(Q)$ für alle $P, Q \in E$. Anders gesagt, Φ ist ein Gruppenhomomorphismus.*

Beweis: Siehe zum Beispiel [Si1], Thm. III.4.8. □

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)).

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 jede nicht-konstante rationale Abbildung zwischen Kurven induziert eine nicht konstante Isogenie $\phi : E \to E'$ eine Inklusion der Funktionenkörper durch

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

Die Körpererweiterung $\phi^*(K(E')) \subset K(E)$ ist dann endlich; ihr Grad wird dann auch der *Grad von* ϕ genannt:

$$\deg(\phi) = [K(E) : \phi^*(K(E'))]\,.$$

Man kann dann auch noch analog den *separablen* und den *inseparablen* Grad von ϕ, $\deg_s(\phi)$ bzw. $\deg_i(\phi)$, definieren. Der Vollständigkeit halber setzt man noch $\deg(0) = 0$ (wo links 0 die konstante Isogenie $[0]$ bezeichnet). Es gilt dann (auf Grund der Multiplikativität des Grades in Körpererweiterungen)

$$\deg(\phi\circ\psi) = \deg(\phi)\cdot\deg(\psi)\,;$$

außerdem (auf Grund der Definitionen)

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

Satz 2.29. *Der Endomorphismenring* $\mathrm{End}_K(E)$ *ist ein Integritätsring 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 $\mathrm{End}_K(E)$ ein Integritätsring 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.

Satz 2.30. *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* $\deg([m]) = m^2$.

BEWEIS: Siehe zum Beispiel [Si1], Thms III.6.1 und 6.2. □

Nach so viel neuen Begriffen ist ein Beispiel angebracht.

BEISPIEL 2.31. 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

$$\phi : E \ni (\xi,\eta) \longmapsto \left(\frac{\eta^2}{\xi^2}, \frac{\eta(b-\xi^2)}{\xi^2}\right) \in E'$$

$$\hat{\phi} : E' \ni (\xi',\eta') \longmapsto \left(\frac{\eta'^2}{4\xi'^2}, \frac{\eta'(a^2-4b-\xi'^2)}{8\xi'^2}\right) \in E$$

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.

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.

SATZ 2.32. *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-Involutiom ist, folgt aus Satz 2.30, (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).

2.5. 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 $n \cdot P = 0$. Damit wir diese Punkte auch alle „sehen“ können, nehmen wir an, daß unser Grundkörper K algebraisch abgeschlossen ist.

Zunächst brauchen wir noch eine Vorbemerkung über den Zusammenhang der Größe des Kerns einer Isogenie und ihrem (separablen) Grad.

SATZ 2.33. *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. □

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 *Frobenius-Endomorphismus.*

PROPOSITION 2.34. *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 $\psi \in \mathrm{End}_K(E)$ der Frobenius-Endomorphismus. Dann ist für $m, n \in \mathbb{Z}$ der Endomorphismus $m + n\phi$ separabel genau dann, wenn m nicht durch p teilbar ist.*

BEWEIS: Siehe [Si1], Cor. III.5.5 und Prop. II.2.11. □

Ist $\phi : E \to E'$ eine Isogenie, dann schreiben wir $E[\phi]$ für ihren Kern, d.h. $E[\phi] = \{P \in E(\bar{K}) \mid \phi(P) = 0\}$. 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).

SATZ 2.35. *Sei E eine elliptische Kurve über K und $m \in \mathbb{Z}_{>0}$.*

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

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

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

$$E[p^e] = \{0\} \qquad \text{für } e = 1, 2, 3, \ldots, \quad \text{oder}$$
$$E[p^e] \cong \mathbb{Z}/p^e\mathbb{Z} \quad \text{für } e = 1, 2, 3, \ldots.$$

Im ersten Fall heißt E supersingulär*, im zweiten Fall* gewöhlich *(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 mm 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.

SATZ 2.36. *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 $\sigma \in \mathrm{Gal}(\bar{K}/K)$ gilt*

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

(5) *Die Abbildungen sind miteinander kompatibel: Für $S \in E[mm']$ und $T \in E[m]$ gilt*

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

(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).

BEWEIS: Siehe [Si1], § III.8. □

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

KOROLLAR 2.37. *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 2.35. 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 $\mathbb{Q}$ werden wir im nächsten Semester sehen, daß die Gruppe $E(\mathbb{Q})$ endlich erzeugt ist. 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 $\#T \leq 10$ oder $= 12$ gibt. Im anderen Fall muß $d \leq 4$ sein.

ÜBUNGSAUFGABEN 2.3.

(1) Wir betrachten die elliptische Kurve

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

über $K = \mathbb{Q}(i)$. Zeigen Sie: $[i] : E \ni (x, y) \mapsto (-x, iy) \in E$ definiert einen Automorphismus von E, und $\widehat{[i]} = -[i]$. Folgern Sie, daß

$$\mathbb{Z}[i] \ni a + bi \mapsto [a + bi] = [a] + [b] \cdot [i] \in \mathrm{End}_K(E)$$

eine Einbettung ist. (Tatsächlich sogar ein Isomorphismus.) Was ist der Grad von $[a + bi]$?

(2) Sei E über K eine elliptische Kurve und $\mathrm{char}(K) \neq 2$. Dann können wir E durch eine Gleichung der Form

$$E : y^2 = x^3 + a_2\, x^2 + a_4\, x + a_6$$

definieren. Wir wissen, daß $E[2] \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ ist. Bestimmen Sie die Elemente von $E[2]$. Wie muß die Weil-Paarung

$$e_2 : E[2] \times E[2] \longrightarrow \mu_2 = \{\pm 1\}$$

aussehen?

(3) Sei wieder $\mathrm{char}(K) \neq 2$ und

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

eine elliptische Kurve. Sei $f \in K(E)$ die rationale Funktion x. Zeigen Sie, daß $f \circ [2] = g^2$ ist mit einer rationalen Funktion $g \in K(E)$.

(4) (Fortsetzung) Sei $S = (0,0) \in E[2]$ und $T = (\alpha, 0) \in E[2]$. Sei $P \in E(\bar{K})$ ein Punkt, so daß g in P und in $P + T$ definiert ist. Dann gilt

$$g(P+T)/g(P) = e_2(S,T) \in \{\pm 1\}\,.$$

Bemerkung: Die Weil-Paarung e_m läßt sich in ähnlicher Weise definieren.

KAPITEL 3

Elliptische Kurven über endlichen Körpern

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

1. Wiederholung: Endliche Körper

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

- Die Anzahl der Elemente eines endlichen Körpers ist eine Primzahlpotenz p^f (mit $f \geq 1$).
- Umgekehrt gibt es zu jeder Primzahlpotenz $q = p^f$ bis auf Isomorphie genau einen endlichen Körper $\mathbb{F}_q$.
- 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 *Forbeniusautomorphismus* $x \mapsto x^q$.
- Der algebraische Abschluß von $\mathbb{F}_q$ ist die (aufsteigend filtrierte) Vereinigung $\bar{\mathbb{F}}_q = \bigcup_n \mathbb{F}_{q^n}$. Es gilt
$$\mathbb{F}_q = \{x \in \bar{\mathbb{F}}_q \mid x^q = x\}\,.$$
- Es gilt
$$\mathbb{F}_q^\times = \{x \in \bar{\mathbb{F}}_q \mid x^{q-1} = 1\} = \mu_{q-1}\,.$$

2. Elliptische Kurven über endlichen Körpern

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.

2.1. 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.

SATZ 3.1. *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 $|t| \le 2\sqrt{q}$.
(2) *Es gilt $\#E(\mathbb{F}_q) = \deg(\phi - 1) = q + 1 - t$. Insbesondere haben wir*
$$|\#E(\mathbb{F}_q) - (q+1)| \le 2\sqrt{q}\,.$$

BEWEIS: (1) Wir haben in $\mathrm{End}_{\mathbb{F}_q}(E)$
$$\begin{aligned} 0 &= (\phi - \phi)(\phi - \hat{\phi}) \\ &= \phi^2 - (\phi + \hat{\phi})\phi + \phi\hat{\phi} \\ &= \phi^2 - t\,\phi + q\,, \end{aligned}$$
denn $\phi\hat{\phi} = \deg(\phi) = q$.

Für eine rationale Zahl $r/s \in \mathbb{Q}$ gilt
$$\left(\frac{r}{s}\right)^2 - t\,\frac{r}{s} + q = \frac{1}{s^2}(r^2 - t\,rs + q\,s^2) = \frac{1}{s^2}\deg(r - s\phi) \ge 0\,,$$
also hat das Polynom $X^2 - t\,X + q$ nicht-positive Diskriminante: $t^2 - 4q \le 0$, d.h. $|t| \le 2\sqrt{q}$.

(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. 2.34), gilt $\#E(\mathbb{F}_q) = \#\ker(\phi - 1) = \deg(\phi - 1)$ (Satz 2.33). Außerdem ist
$$\deg(\phi - 1) = (\phi - 1)(\hat{\phi} - 1) = \phi\hat{\phi} - (\phi + \hat{\phi}) + 1 = q - t + 1\,.$$
□

Unter Berücksichtigung von Kor. 2.37 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 $|d^2d' - (q+1)| \le 2\sqrt{q}$.

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

SATZ 3.2. *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 tiefliegende 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 komponieren 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 2.29), folgt $t = t'$, also nach Satz 3.1 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 mod ℓ zu bestimmen und daraus auf den Wert von t (und damit von $\#E(\mathbb{F}_q) = q + 1 - t$) zu schließen.

2.2. 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.

DEFINITION 3.3. 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-Endomorphimus 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.

SATZ 3.4 (Weil-Vermutungen für Kurven). *Sei C eine glatte projektive Kurve über $\mathbb{F}_q$. Dann gilt:*

(1) *$Z(C,T) \in \mathbb{Q}(T)$ ist eine rationale 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*

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

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

Zu diesem Satz ein paar Bemerkungen.

- 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 (Deligne) erledigt.
- 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 $g = 1$.
- 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 3.1), wobei $t = \phi + \hat{\phi}$ die Spur des Frobenius ist. Weiterhin galt $|t| \leq 2\sqrt{q}$, 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 2.29.) 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.

ÜBUNGSAUFGABEN 3.1.

(1) Beweisen Sie die Weil-Vermutungen für die projektive Gerade $\mathbb{P}^1$. (Man kann sich $\mathbb{P}^1$ vorstellen als die ebene projektive Kurve, die durch die Gleichung $Y = 0$ gegeben ist. Das Geschlecht ist $g = 0$ für $\mathbb{P}^1$.)

(2) Sei $(a_d)_{d\geq 1}$ eine Folge rationaler Zahlen, und sei $b_n = \sum_{d|n} d a_d$. Zeigen Sie folgende Identität von Potenzreihen:

$$\exp\left(\sum_{n=1}^{\infty} \frac{b_n}{n} T^n\right) = \prod_{d=1}^{\infty} (1 - T^d)^{-a_d}\,.$$

(3) Zeigen Sie (für die projektive Gerade über $\mathbb{F}_q$):

$$Z(\mathbb{P}^1, q^{-s}) = (1 - q^{-s})^{-1} \prod_f (1 - (\#F_q[T]/(f))^{-s})^{-1}\,,$$

wobei das Produkt über alle irreduziblen normierten Polynome $f \in \mathbb{F}_q[T]$ läuft. (Der erste Faktor gehört zum Punkt im Unendlichen.) Vergleichen Sie mit der Riemannschen Zetafunktion

$$\zeta(s) = \prod_p (1 - \#(\mathbb{Z}/(p))^{-s})^{-1}$$

(dabei durchläuft p alle Primzahlen).

Hinweis: Es gilt

$$\prod_{f\,:\,(\deg f)|n} f(T) = \prod_{x \in \mathbb{F}_{q^n}} (T - x) = T^{q^n} - T\,.$$

Wenn a_d die Anzahl der normierten irreduziblen Polynome vom Grad d ist, dann folgt daraus $\sum_{d|n} d a_d = q^n$.

(4) Sei

$$E : y^2 + y = x^3 - x$$

eine elliptische Kurve über $\mathbb{F}_{11}$. Bestimmen Sie $\#E(\mathbb{F}_{11})$ (durch Abzählen) und daraus die Spur des Frobenius sowie die Zahlen $\#E(\mathbb{F}_{11^n})$ für $n = 2, 3, 4, 5$.

KAPITEL 4

Anwendungen I: Faktorisierung und Primzahltest

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 dieses Kapitel ist [Coh].

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.

Beim ersten Punkt wird heutzutage meistens der *Miller-Rabin-Test* verwendet. Er basiert auf folgendem Resultat.

PROPOSITION 4.1. *Sei N eine ungerade natürliche Zahl. Wir schreiben $N-1 = 2^t q$ mit q ungerade. Sei weiter a eine ganze Zahl. Wir sagen, N sei eine* starke Pseudoprimzahl zur Basis a, *wenn folgendes gilt:*

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

Dann gilt:

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

Teil (1) ist leicht zu beweisen (man benutzt den kleinen Satz von Fermat $a^{N-1} \equiv 1 \bmod N$ und daß ± 1 die beiden einzigen Quadratwurzeln von $1 \bmod N$ sind). Teil (2) ist etwas komplizierter, aber immer noch elementar.

Wenn man wissen möchte, ob eine gegebene Zahl N zusammengesetzt ist, wählt man also zufällig Zahlen $1 < a < N$ und prüft, ob N eine starke Pseudoprimzahl zur Basis a ist. Wenn dies für ein a nicht der Fall ist, ist bewiesen, daß N zusammengesetzt ist. Anderenfalls kann man ziemlich sicher sein (jedenfalls wenn man ausreichend viele Werte von a probiert hat), daß N prim ist, und kann daran gehen, das auch zu beweisen.

1. Faktorisierung und Primzahltest mit der multiplikativen Gruppe

Wir wenden uns zunächst dem Problem zu, von einer Zahl N (von der wir überzeugt sind, daß sie prim ist) zu beweisen, daß sie prim ist. Eine Möglichkeit dafür besteht darin, eine geeignete Umkehrung des (kleinen) Satzes von Fermat ($a^{p-1} \equiv 1 \bmod p$) zu verwenden. Im Folgenden schreiben wir $a \perp b$, wenn a und b teilerfremd sind.

PROPOSITION 4.2. *Sei $N > 0$ eine ganze Zahl und p ein Primteiler von $N-1$. Sei weiter $a_p \in \mathbb{Z}$ mit*

$$a_p^{N-1} \equiv 1 \bmod N \qquad \text{und} \qquad (a_p^{(N-1)/p} - 1) \perp N\,. \tag{1.1}$$

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 (nach dem kleinen Fermat) $a_p^{d-1} \equiv 1 \bmod d$. Andererseits ist $a_p^{(N-1)/p)} \not\equiv 1 \bmod d$, da $(a_p^{(N-1)/p} - 1) \perp N$. 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.

KOROLLAR 4.3. *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 (1.1) 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 (1.1).

Seien nun umgekehrt für alle $p \mid F$ Zahlen a_p mit (1.1) gegeben. Aus Prop. 4.2 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 zusammengestzt wäre, hätte N einen nichttrivialen Teiler $\leq \sqrt{N}$; 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 $2^p - 1$.

Kommen wir nun zur Faktorisierung. Hier haben wir eine Zahl N gegeben, von der wir wissen, daß sie zusammengestzt ist (weil sie den Miller-Rabin-Test nicht bestanden hat beispielsweise). Das Ziel ist, einen nichttrivialen Teiler d von N zu finden. (Um eine vollständige Faktorisierung zu erhalten, macht man dann mit d und N/d rekursiv weiter.)

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 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 $a^{L(1)} \bmod N$, $a^{L(2)} \bmod N$, $\ldots$, $a^{L(B)} \bmod N$ 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.

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.

2. Verwendung von 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); die Menge der R-rationalen Punkte $E(R)$ trägt wiederum eine Gruppenstruktur mit Nullelement $O = (0 : 1 : 0)$.

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 einen Gruppenhomomrphismus $E(R) \to E'(S)$. (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 $4a^3 + 27b^2 \in (\mathbb{Z}/N\mathbb{Z})^\times$.

Die Addition und Berechnung von Vielfachen in $E(\mathbb{Z}/N\mathbb{Z})$ geht dann mit denselben Formeln wie vorher über einem Körper. 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.

2.1. Primzahltest.

Wir betrachten zuerst wieder das Problem, zu beweisen, daß N prim ist. Das folgende Resultat steht in Analogie zu Prop. 4.2.

PROPOSITION 4.4. *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*

$$m \cdot P = O \qquad \text{und} \qquad (m/q) \cdot P = (\xi : \eta : \zeta) \text{ mit } \zeta \in (\mathbb{Z}/N\mathbb{Z})^\times . \tag{2.1}$$

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.

Proposition 4.5. *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* (2.1) *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

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

daher $(\sqrt[4]{N}+1)^4 < q^2 \le m \le (\sqrt{N}+1)^2$, ein Widerspruch. □

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

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 $(\sqrt[4]{N}+1)^2 < q \le m/2$ 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. 4.4, *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 polynomial ist; er ist aber (bisher) nicht praktikabel.

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 der andere schnelle Primzahltest (der mit sogenannten Jacobi-Summen arbeitet und ziemlich viel algebraische Zahlentheorie benutzt).

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 sehr schnell davon überzeugen, daß N tatsächlich prim ist.

2.2. 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 von P selbst auf E 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.

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.

PROPOSITION 4.6 (Canfield, Erdős, Pomerance). *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; die Rechenzeit für jede Kurve ist etwa B, also insgesamt $\ell(M)^{a+1/(2a)}$. Das wird minimal für $a = 1/\sqrt{2}$ bei einer (erwarteten) Rechenzeit von ungefähr $\ell(M)^{\sqrt{2}}$. 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 $\ell(N)$. Insbesondere ist die Rechenzeit *subexponentiell*.

Im Vergleich zu anderen Methoden hat die 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 ($\exp(C\sqrt[3]{\log N(\log\log N)^2})$ beim Zahlkörpersieb).

KAPITEL 5

Anwendungen II: Kryptographie

Wir kommen nun zu einer weiteren Anwendung von elliptischen Kurven über endlichen Körpern, nämlich in der Kryptographie. Etwas genauer handelt es sich dabei um sogenannte *Public-Key-Kryptographie*. Im Unterschied zur klassischen symmetrischen Verschlüsselung, bei der sich die beiden Kommunikationspartner vorher auf sichere Weise auf einen gemeinsam benutzten Schlüssel einigen müssen (was ein großes Problem sein kann, insbesondere dann, wenn vor der Übertragung der zu verschlüsselnden Nachricht noch kein Kontakt bestand — man denke zum Beispiel an die Abwicklung von Geschäften über das Internet), beruht die Public-Key-Kryptographie auf einem Paar von Schlüsseln für jeden Teilnehmer. Ein Teil des Schlüssels ist ein geheimer privater Schlüssel, der nur dem jeweiligen Teilnehmer bekannt ist (und also nicht auf geschütztem Wege ausgetauscht werden muß), der andere Teil ist ein davon abgeleiteter öffentlicher Schlüssel, der allgemein bekannt gemacht wird. Dieser öffentliche Schlüssel dient dazu, Nachrichten an den betreffenden Teilnehmer zu verschlüsseln; zum Entschlüsseln verwendet er seinen geheimen privaten Schlüssel. Die Sicherheit dieser Verfahren beruht darauf, daß es zwar relativ einfach ist, aus dem privaten Schlüssel den zugehörigen öffentlichen zu generieren, aber praktisch unmöglich, umgekehrt aus dem öffentlichen den privaten Schlüssel abzuleiten. Das wird üblicherweise dadurch erreicht, daß man sich auf mathematische Probleme stützt, für deren Lösung keine schnellen Algorithmen bekannt sind (und die andererseits so gut untersucht sind, daß man davon ausgehen kann, daß bessere Algorithmen schnell bekannt werden, so daß die Sicherheit gegebenenfalls angepaßt werden kann).

1. Kryptographie mit zyklischen Gruppen

Ein solches Problem ist das *Diskrete Logarithmus-Problem*, kurz DLP genannt. Dazu sei G eine endliche zyklische Gruppe der Ordnung n mit gegebenem Erzeuger $g \in G$. Dazu gehört auch noch eine konkrete Darstellung der Elemente von G (z.B. als ganze Zahlen in einem gegebenen Intervall oder als Paare von ganzen Zahlen). Wir setzen folgendes voraus. (Wir behandeln G hier als multiplikative Gruppe; gegebenenfalls muß man die Multiplikation durch Addition ersetzen.)

- Man kann in G effizient rechnen. D.h., es gibt schnelle Algorithmen, um aus (durch ihre konkrete Darstellung) gegebenen Elementen $a, b \in G$ ihr Produkt ab und das Inverse a^{-1} zu berechnen (genauer: die konkrete Darstellung von ab bzw. a^{-1}).
- Das DLP in G ist schwierig zu lösen (d.h. es gibt keinen schnellen Algorithmus dafür, oder es ist wenigstens keiner bekannt).

Das DLP besteht dabei darin, zu einem gegebenen Element $h \in G$ (d.h., seiner konkreten Darstellung) die Zahl $x \in \mathbb{Z}/n\mathbb{Z}$ zu finden, so daß $h = g^x$ ist.

Ist zum Beispiel G die additive Gruppe $(\mathbb{Z}/n\mathbb{Z}, +)$, dann läuft das DLP darauf hinaus, eine Gleichung $g \cdot x = h \bmod n$ zu lösen, was mit dem erweiterten euklidischen Algorithmus sehr schnell geht. So eine Gruppe eignet sich also nicht für kryptographische Zwecke. Daran sieht man auch, daß die Wahl der konkreten Darstellung der Elemente von G sehr wesentlich ist. Abstrakt ist G nämlich isomorph zur additiven Gruppe $(\mathbb{Z}/n\mathbb{Z}, +)$; das bedeutet, daß die Schwierigkeit des DLP allein von der gewählten Darstellung von G abhängt.

Wir wollen nun zwei kryptographische Verfahren kennenlernen, die auf der Verwendung einer Gruppe basieren.

1.1. Diffie-Hellman-Schlüsselaustausch.

Hier möchten A(lice) und B(ob) über einen sogenannten unsicheren Kanal (d.h. z.B. eine abhörbare Leitung) Informationen austauschen, die am Ende zu einem gemeinsamen geheimen Schlüssel führen. Dieser kann dann zur Übermittlung einer längeren Nachricht mit einem effizienten symmetrischen Verschlüsselungsverfahren dienen. Das Vorgehen ist dabei wie folgt.

1. A wählt eine zufällige Zahl $a \in \mathbb{Z}/n\mathbb{Z}$ und sendet g^a an B.
2. B wählt eine zufällige Zahl $b \in \mathbb{Z}/n\mathbb{Z}$ und sendet g^b an A.
3. Beide können jetzt den gemeinsamen geheimen Schlüssel $g^{ab} = (g^a)^b = (g^b)^a$ berechnen.

Jemand, der die Leitung abhört und die Verschlüsselung knacken will, steht vor der Aufgabe, aus g, g^a und g^b das Element g^{ab} zu berechnen. Dies ist das sogenannte *Diffie-Hellman-Problem*, kurz DHP. Es ist klar, daß eine Lösung des DLP zu einer Lösung des DHP führt. Umgekehrt ist es sehr plausibel (und in manchen Fällen auch beweisbar (siehe z.B. [BSS, IX.4]), daß ein schneller Algorithmus für das DHP auch zu einem schnellen Algorithmus für das DLP führt. Man kann also beide als von vergleichbarer Schwierigkeit ansehen, so daß das angegebene Verfahren für Gruppen mit schwerem DLP sicher sein sollte.

1.2. ElGamal-Verschlüsselung.

Hier geht es nun darum, daß A eine verschlüsselte Nachricht an B senden möchte. B hat als seinen privaten und geheimen Schlüssel eine zufällige Zahl $b \in \mathbb{Z}/n\mathbb{Z}$ gewählt und den öffentlichen Schlüssel $h = g^b$ bekannt gemacht. Die Nachricht, die A an B schicken möchte, sei durch das Element $m \in G$ repräsentiert.

1. A wählt eine zufällige Zahl $a \in \mathbb{Z}/n\mathbb{Z}$.
2. A sendet das Paar $(r, s) = (g^a, h^a \cdot m)$ an B.
3. B entschlüsselt die Nachricht als $m = r^{-b} \cdot s$.

Dem liegt dieselbe Idee zu Grunde wie beim Diffie-Hellman-Verfahren, nur daß der zeitliche Ablauf etwas anders ist und der geheime gemeinsame Schlüssel sofort zur Übertragung der Nachricht m verwendet wird. Ein Unbefugter, der die Leitung abhört, hat wieder ein DHP zu lösen, um die Nachricht zu entschlüsseln.

2. Verwendung von elliptischen Kurven

Es kommen nun verschiedene Arten von Gruppen G in Frage. Beispiele sind

1. die multiplikative Gruppe $F_q^\times$;
2. die Gruppe $E(\mathbb{F}_q)$ der rationalen Punkte einer elliptischen Kurve über $\mathbb{F}_q$.

Die besprochenen Verfahren wurden ursprünglich für die Gruppen $\mathbb{F}_q^\times$ vorgeschlagen. Der Vorteil einer solchen Gruppe (jedenfalls wenn $q = p$ eine Primzahl ist) ist die Einfachheit und daher Effizienz des Rechnens in der Gruppe. Auf der anderen Seite sind aber inzwischen subexponentielle Algorithmen zur Lösung des DLP in solchen Gruppen bekannt (die Komplexität ist vergleichbar mit Faktorisierungsalgorithmen). Das bedeutet, daß man recht große Werte von q (1000, 2000 oder sogar 4000 Bit) verwenden muß, um ein sicheres System zu haben.

Für elliptische Kurven ist demgegenüber kein schnelleres Verfahren für das DLP bekannt als die, die auf jede Gruppe anwendbar sind. (Ausnahmen sind spezielle elliptische Kurven, die man für kryptographische Zwecke nicht verwenen sollte. Was das genau heißt, kann man in [BSS] nachlesen.) Diese allgemeinen Verfahren haben eine Komplexität von $O(\sqrt{n})$. Das bedeutet, daß (nach gegenwärtigem Kenntnisstand) man q bei der Verwendung von $G = E(\mathbb{F}_q)$ wesentlich kleiner wählen kann (200 bis 400 Bit), um ein vergleichbares Niveau an Sicherheit zu bekommen. Dies bedeutet auch eine nicht unerhebliche Speicherplatzersparnis, was für Implementationen z.B. auf Chipkarten wichtig ist. Auf der anderen Seite ist das Rechnen in der Gruppe komplizierter als in $\mathbb{F}_q^\times$ und braucht daher mehr Zeit und eine aufwendigere Implementation.

3. Das DLP für eine beliebige zyklische Gruppe

Zum Abschluß möchte ich noch etwas genauer auf Algorithmen zur Lösung des DLP eingehen, die auf jede zyklische Gruppe anwenbar sind (sofern man in der Gruppe rechnen kann). Diese Algorithmen sind gleichzeitig auch die besten bekannten, die für eine allgemeine elliptische Kurve funktionieren.

3.1. Durchprobieren.

Der offensichtlichste Algorithmus besteht darin, die Elemente $g^0, g^1, g^2, \dots$ zu berechnen, bis man auf $g^x = h$ stößt. Der Zeitaufwand dafür ist $O(n)$, der Platzbedarf ist $O(1)$ an Gruppenelementen.

3.2. Baby-Step-Giant-Step.

Diese Methode läßt sich in vielen Situationen in ähnlicher Weise verwenden. Sie reduziert die Laufzeit auf $O(\sqrt{n})$, hat aber einen Platzbedarf von ebenfalls $O(\sqrt{n})$ Gruppenelementen. Die Idee ist wie folgt. Sei $h = g^x$ und $m = \lceil\sqrt{n}\rceil$. Dann können wir schreiben $x = r \cdot m + s$ mit $0 \le r, s < m$, und die urpsrüngliche Gleichung $h = g^x$ ist äquivalent zu $hg^{-s} = (g^m)^r$.

Wir berechnen nun zuerst die linke Seite hg^{-s} für alle $s = 0, 1, \dots, m-1$ und speichern die Werte (zusammen mit dem zugehörigen s) in einer Tabelle, die schnellen Zugriff auf die Einträge erlaubt. In einem zweiten Schritt berechnen wir für $r = 0, 1, 2, \dots$ die rechte Seite $(g^m)^r$, bis wir das Ergebnis in der Tabelle finden. Aus r und s können wir dann x bestimmen.

3.3. Pollard-Lambda.

Dieses Verfahren wird auch Methode der zahmen und wilden Känguruhs genannt. Sie ist probabilistisch und hat eine erwartete Laufzeit von ebenfalls $O(\sqrt{n})$, dafür aber beliebig langsam wachsenden Platzbedarf. Bei der vorherigen Methode sucht man nach einer Übereinstimmung („Kollision“) zwischen den Elementen hg^{-s} und $(g^m)^r$.

Eine ähnliche Idee liegt hier zu Grunde. Wir definieren in geeigneter zufälliger Weise eine Funktion $f = (f_1, f_2) : G \to \mathbb{Z} \times \mathbb{Z}$ (es genügt, wenn sie nur einige verschiedene Werte (z.B. etwa 20) annimmt). Damit definieren wir dann eine weitere Abbildung $F : G \ni z \mapsto z \cdot g^{f_1(z)} \cdot h^{f_2(z)} \in G$. Beginnend mit zwei Startwerten iterieren wir nun diese Abbildung, bis die zweite Iterationsfolge auf die erste trifft. (Dazu merken wir uns die Glieder der ersten Folge in gewissen Abständen. Sobald die zweite Folge (wildes Känguruh) auf die erste (zahmes Känguruh) trifft, muß sie ihr folgen und schließlich auf ein Glied treffen, das wir uns gemerkt haben (Loch, das das zahme Känguruh gegraben hat, und in das das wilde Känguruh hineinfällt).) Dabei wählen wir Startwerte der Form $z_0 = g^{x_0} \cdot h^{y_0}$ und $z'_0 = g^{x'_0} \cdot h^{y'_0}$ und führen die Folgenglieder in der Form $g^a \cdot h^b$ mit.

Wenn eine Kollision $g^{x_k} \cdot h^{y_k} = z_k = z'_m = g^{x'_m} \cdot h^{y'_m}$ auftritt, erhalten wir die Relation $g^{x_k - x'_m} = h^{y'_m - y_k}$, aus der wir die Lösung des DLP durch eine Division mod n berechnen können, jedenfalls wenn $y'_m - y_k$ mod n invertierbar ist. Ist dies nicht der Fall, müssen wir es halt noch einmal versuchen (mit anderen Startwerten oder einer anderen Funktion f).

Literaturverzeichnis

[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).

[Coh] H. Cohen: *A course in computational algebraic number theory*, Springer GTM **138** (1993).

[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).