Elliptische Kurven und ihre Weierstraßsche Normalform

Nach vielen Vorbereitungen sind wir nun endlich so weit, den zentralen Begriff, um den es in dieser Vorlesung geht, exakt zu definieren und erste wichtige Eigenschaften zu beweisen.

Definition: Eine *elliptische Kurve* über einem Körper k ist eine irreduzible ebene Kurve vom Grad drei in $\mathbb{P}^2(k)$, die keine singulären Punkte hat und mindestens einen Punkts mit Koordinaten in k enthält.

Falls der Grundkörper k algebraisch abgeschlossen ist, hat natürlich jedes homogene Polynom positiven Grades aus $k[X, Y, Z]$ Nullstellen in $\mathbb{P}^2(k)$; für Körper wie $k = \mathbb{Q}$ oder auch für endliche Körper muß das aber nicht der Fall sein. Betrachten wir etwa über den rationalen Zahlen das kubische Polynom $X^3 + pY^3 + p^2Z^3$ für irgendeine Primzahl p. Wenn es einen Punkt $(x : y : z) \in \mathbb{P}^2(\mathbb{Q})$ gibt, der diese Gleichung erfüllt, können wir wegen der Homogenität der Koordinaten annehmen, daß x, y, z ganze Zahlen sind und keinen gemeinsamen Teiler haben. Aus $x^3 + py^3 + p^2z^3 = 0$ folgt, daß $x^3 = -py^3 - p^2z^3$ durch p teilbar ist; sei etwa $x = px_0$ mit $x_0 \in \mathbb{Z}$. Dann ist $py^3 = -p^3x_0^3 - p^2z^3$, also $y^3 = -p^2x_0^3 - pz^3$, so daß auch $y = py_0$ durch p teilbar ist. Schließlich folgt aus $p^2z^3 = -p^3x_0^3 - p^4y_0^3$, daß auch noch z durch p teilbar sein muß, im Widerspruch zur angenommenen Teilerfremdheit von x, y und z.

Die letzte Forderung ist also für beliebige Körper eine echte Bedingung, und da wir keine Aussagen über die leere Menge beweisen wollen, ist sie auch sinnvoll.

Wir werden allerdings häufiger auch Punkte betrachten müssen, die keine Koordinaten im Grundkörper k haben; deshalb definieren wir

Definition: Ist $E = V(F) \subset \mathbb{P}^2(k)$ eine elliptische Kurve über dem Körper k und ist K irgendein Körper, so bezeichnen wir mit $E(K)$ die Menge aller Punkte $(x : y : z) \in \mathbb{P}^2(K)$ mit $F(x, y, z) = 0$.

Der Vektorraum aller homogener Polynome vom Grad drei über einem Körper k hat die Dimension $\binom{3+2}{2} = 10$; die Gleichung einer elliptischen Kurve hat also zehn Koeffizienten. Wir wollen uns überlegen, daß

wir diese Zahl durch geschickte Wahl des Koordinatensystems deutlich verkleinern können.

Wir nehmen dazu an, daß die Charakteristik des Körpers k von zwei verschieden sei, daß also $1 + 1 \neq 0$ ist.

Elliptische Kurven über Körpern der Charakteristik zwei sind zwar auch interessant und werden in der Kryptographie auch gelegentlich angewandt; die Situation in Charakteristik zwei ist aber deutlich unübersichtlicher, so daß wir sie nicht betrachten wollen.

Zur Erinnerung sei der Begriff der Charakteristik noch einmal kurz erläutert: Für jeden Körper k gibt es genau einen Homomorphismus $\mathbb{Z} \to k$, denn die Eins muß auf die Eins des Körpers abgebildet werden, und $n \in \mathbb{N}$ als Summe von n Einsen auf die Summe von n Körpereinsen. Die Null muß auf die Null des Körpers gehen, und $-n$ auf das additive Inverse des Bilds von n. Falls diese Abbildung (wie beispielsweise bei den rationalen, reellen oder komplexen Zahlen) injektiv ist, sagen wir, der Körper habe die Charakteristik Null, in Zeichen $\operatorname{char} k = 0$. Andernfalls sei p die kleinste natürliche Zahl, die auf die Null des Körpers abgebildet wird; wir schreiben dann $\operatorname{char} k = p$. Wie die Bezeichnung schon andeutet, muß p eine Primzahl sein, denn wäre $p = rs$ mit $r, s > 1$, so würde das Produkt der Bilder von r und s in k verschwinden, während diese Bilder selbst von Null verschieden wären. Das ist in einem Körper nicht möglich. Die einfachsten Körper der Charakteristik $p > 0$ sind die endlichen Körper $\mathbb{F}_p = \mathbb{Z}/p$. Wie man in der Algebra zeigt, gibt es für jede Primzahlpotenz p^n bis auf Isomorphie genau einen Körper mit p^n Elementen; er enthält $\mathbb{F}_p$ und hat somit die Charakteristik p. Natürlich gibt es auch unendliche Körper der Charakteristik p, z.B. den Körper aller rationaler Funktionen (Quotienten zweier Polynome) über $\mathbb{F}_p$.

Vor allem Körper von Zweipotenzordnung werden in der Kryptographie, auch der mit elliptischen Kurven, teilweise benutzt, da sie Vektorräume über $\mathbb{F}_2$ sind, so daß man ihre Elemente als Bitfolgen kodieren kann, und damit kann ein binärer Computer leicht und effizient rechnen. Der Nachteil ist allerdings, daß es für eine gegebene Größenordnung von Zahlen erheblich mehr Primzahlen als Zweierpotenzen gibt, und ein

Gegner könnte sich Vorteile bei Angriffen verschaffen, indem er für die in Frage kommenden Körper Spezialhardware für ihre Arithmetik baut. Die Anwender könnten das natürlich auch, aber erfahrungsgemäß zeigen Angreifer eine deutlich höhere Bereitschaft für teure Investitionen als Anwender. Aus diesem Grund sind Kryptosysteme über Körper von Primzahlordnung häufiger im Einsatz als solche über Körpern mit Zweipotenzordnung.

Sei also $E = V(F) \subset \mathbb{P}^2(k)$ eine elliptische Kurve über einem Körper, mit von zwei verschiedener Charakteristik und

$$\begin{aligned} F =& a_1X^3 + a_2X^2Y + a_3XY^2 + a_4Y^3 + a_5X^2Z \\ &+ a_6XYZ + a_7Y^2Z + a_8XZ^2 + a_9YZ^2 + a_{10}Z^3 \,. \end{aligned}$$

Zur besseren Übersicht schreiben wir dies als

$$F = F_3 + F_2Z + F_3Z^2 + a_{10}Z^3$$

mit

$$\begin{aligned} F_3 &= a_1X^3 + a_2X^2Y + a_3XY^2 + a_4Y^3, \\ F_2 &= a_5X^2 + a_6XY + a_7Y^2 \quad \text{und} \\ F_1 &= a_8X + a_9Y \,. \end{aligned}$$

Nach Definition hat E mindestens einen Punkt P mit Koordinaten in k; wir betrachten zunächst denn Fall, daß E sogar einen *Wendepunkt* P mit Koordinaten in k hat. Dann können wir das Koordinatensystem so wählen, daß die Wendetangente die Gleichung $z = 0$ hat und P die Koordinaten $(0 : 1 : 0)$. Außer P gibt es dann keinen weiteren Punkt auf E mit z-Koordinate Null, denn die Wendetangente schneidet die Kurve mindestens mit Vielfachheit drei, und nach dem Satz von BÉZOUT hat sie, mit Vielfachheiten gezählt, insgesamt höchstens drei Schnittpunkte mit E.

Ein Punkt $(x : y : 0)$ mit z-Koordinate Null liegt genau dann auf E, wenn $F(x, y, 0) = F_3(x, y)$ verschwindet. F_3 ist ein homogenes Polynom vom Grad drei in X und Y; da außer $(0 : 1 : 0)$ kein weiterer Punkt $(x : y : 0)$ auf E liegt, muß F_3 in $(0 : 1)$ eine dreifache Nullstelle haben. Somit ist $F_3 = a_1X^3$ mit $a_1 \neq 0$, denn sonst wäre F kein Polynom vom Grad drei.

Da eine elliptische Kurve nach Definition keine singulären Punkte enthält, kann insbesondere der Punkt $P = (0 : 1 : 0)$ nicht singulär sein, d.h. die drei partiellen Ableitungen nach x, y, z dürfen dort nicht alle drei verschwinden. Nun ist

$$\begin{aligned}
\frac{\partial F}{\partial x}(x,y,z) &= 3a_1x^2 + \frac{\partial F_2}{\partial x}(x,y)z + \frac{\partial F_1}{\partial x}(x,y)z^2 \\
\frac{\partial F}{\partial y}(x,y,z) &= \frac{\partial F_2}{\partial y}(x,y)z + \frac{\partial F_1}{\partial y}(x,y)z^2 \\
\frac{\partial F}{\partial z}(x,y,z) &= F_2(x,y) + 2F_1(x,y)z + 3a_{10}z^2 \,.
\end{aligned}$$

Im Punkt $(0 : 1 : 0)$ verschwinden daher die partiellen Ableitungen nach x und nach y, und

$$\frac{\partial F}{\partial z}(0,1,0) = F_2(0,1,0) = a_7 \,.$$

Damit P kein singulärer Punkt ist, darf also der Koeffizient a_7 von Y^2Z in F nicht verschwinden. Da jedes skalare Vielfache von F die gleiche Nullstellenmenge wie F hat, können wir durch a_7 dividieren und somit annehmen, daß $a_7 = 1$ ist. Damit ist

$$\begin{aligned}
F &= a_1X^3 + Y^2Z + a_5X^2Z + a_6XYZ + a_8XZ^2 + a_9YZ^2 + a_{10}Z^3 \\
&= a_1X^3 + Y^2Z + (a_6XZ + a_9Z^2)Y + (a_5X^2Z + a_8XZ^2 + a_{10}Z^3) \,.
\end{aligned}$$

In der affinen Ebenen $Z \neq 0$ können wir $Z = 1$ setzen und erhalten

$$F(X,Y,1) = a_1X^3 + Y^2 + (a_6X + a_9)Y + (a_5X^2 + a_8X + a_{10}) \,.$$

Ersetzen wir hier Y durch die neue Koordinate $\widetilde{Y} = Y - \frac{1}{2}(a_6X + a_9)$, so wird das Polynom in den neuen Variablen zu

$$a_1X^3 + \widetilde{Y}^2 + b_2X^2 + b_1X + b_0$$

mit geeigneten Elementen $b_0, b_1, b_2 \in k$. Man beachte, daß wir hier (und nur hier) die Voraussetzung benutzen mußten, daß $\operatorname{char} k \neq 2$ ist: In einem Körper der Charakteristik zwei ist $2 = 0$, wir können also nicht durch zwei dividieren und haben somit auch keine quadratische Ergänzung.

Zur weiteren Vereinfachung ersetzen wir noch X durch $X' = -a_1 X$ und $\widetilde{Y}$ durch $Y' = a_1 \widetilde{Y}$; das ist möglich, da a_1 nicht verschwindet. In diesen Koordinaten wird das Polynom zu

$$\frac{-X'^3}{a_1^2} + \frac{Y'^2}{a_1^2} + \frac{b_2}{a_1^2} X'^2 + \frac{b_1}{a_1} X' + b_0 \,.$$

Multiplikation mit a_1^2 führt zur (fast) endgültigen Form

$$Y'^2 - X'^3 + b_2 X'^2 + a_1 b1 X' + a_1^2 b_0 \,.$$

Über einem Körper k, dessen Charakteristik von zwei verschieden ist, können wir also für eine elliptische Kurve, die einen Wendepunkt mit Koordinaten in k hat, immer ein Koordinatensystem finden, in dem die Nullstellenmenge eines Polynoms der Form

$$Y^2 - (X^3 + c_2 X^2 + c_1 X + c_0)$$

ist.

Falls es keinen Wendepunkt mit Koordinaten in k gibt, ist die Situation etwas komplizierter: Jetzt reicht ein linearer Koordinatenwechsel im Allgemeinen nicht mehr aus, aber wie TRYGVE NAGELL 1928 gezeigt hat, ist über einem Körper mit von zwei verschiedener Charakteristik jede elliptische Kurve *birational äquivalent* zu einer Kurve in obiger Form.

Die klassische algebraische Geometrie des neunzehnten und frühen zwanzigsten Jahrhunderts betrachtete vor allem *rationale* Abbildungen. Eine rationale Abbildung von $\mathbb{P}^n$ nach $\mathbb{P}^m$ ist gegeben durch $m + 1$ homogene Polynome $f_0, \ldots, f_m$ desselben Grades und bildet einen Punkt $(x_0 : \ldots : x_n)$ ab auf den Punkt

$$\big(f_0(x_0, \ldots, x_n) : \ldots : f_m(x_0, \ldots, x_m)\big) \,.$$

Letzteres muß natürlich keinen Punkt in $\mathbb{P}^m$ definieren, denn es könnte ja sein, daß die Polynome f_j allesamt im Punkt $(x_0 : \ldots : x_n)$ verschwinden. Die Abbildung ist daher nur auf einer Teilmenge von $\mathbb{P}^n$ wohldefiniert, nämlich auf $\mathbb{P}^n$ ohne den Durchschnitt der $m + 1$ Hyperflächen $V(f_j)$.

Rationale Abbildungen haben ihren Namen daher, daß sie im Affinen durch rationale Funktionen gegeben sind: Beschränken wir uns jeweils

auf den affinen Raum, in dem die nullte Koordinate nicht verschwindet. so wird $(x_1, \ldots, x_n)$ abgebildet auf

$$\left(\frac{f_1(1, x_1, \ldots, x_n)}{f_0(1, x_1, \ldots, x_n)}, \ldots, \frac{f_m(1, x_1, \ldots, x_n)}{f_0(1, x_1, \ldots, x_n)} \right) ,$$

was für die Punkte definiert ist, an denen f_0 nicht verschwindet.

Beispiel einer rationalen Abbildung der projektiven Ebene auf sich selbst ist die sogenannte CREMONA-Transformation

$$(x : y : z) \mapsto (yz : xz : xy) .$$

Falls keine der drei Koordinaten x, y, z verschwindet, ist der Bildpunkt

$$(yz : xz : xy) = \left(\frac{yz}{xyz} : \frac{xz}{xyz} : \frac{xy}{xyz} \right) = \left(\frac{1}{x} : \frac{1}{y} : \frac{1}{z} \right) ;$$

für solche Punkte ist die Abbildung also zu sich selbst invers und damit insbesondere bijektiv. In den drei Punkten $(0 : 0 : 1)$, $(0 : 1 : 0)$ und $(1 : 0 : 0)$ ist die Abbildung nicht definiert, und ansonsten wird jeder Punkt auf der Geraden $x = 0$ abgebildet auf $(1 : 0 : 0)$, jeder auf $y = 0$ auf $(0 : 1 : 0)$, und die Punkte mit $z = 0$ und $xy \neq 0$ gehen auf $(0 : 0 : 1)$. Alle Punkte auf der Geraden durch zwei der drei Ausnahmepunkte $(0 : 0 : 1)$, $(0 : 1 : 0)$ und $(1 : 0 : 0)$ mit Ausnahme dieser Punkte selbst werden also abgebildet auf den noch verbleibenden Ausnahmepunkt.

Aus der Geometrie bekannte Beispiele rationaler Abbildungen sind Zentralprojektionen; beispielsweise läßt sich die Projektion von $\mathbb{P}^3(k)$ auf $\mathbb{P}^2(k)$ mit Zentru $(1 : 0 : 0 : 0)$ beschreiben ddurch die rationale Abbildung $(u : x : y : z) \mapsto (x : y : z)$, die in allen Punkten mit Ausnahme des Projektionszentrums definiert ist.

Definition: Zwei ebene Kurven C, C' in $\mathbb{P}^2(k)$ heißen *birational äquivalent,* wenn es eine birationale Abbildung von $\mathbb{P}^2(k)$ nach $\mathbb{P}^2(k)$ gibt, die C auf C' abbildet, sowie eine weitere, deren Einschränkung auf C' die Umkehrabbildung dazu ist.

Wir wollen also zeigen, daß jede elliptische Kurve $E = V(F)$ in $\mathbb{P}^2(k)$ birational äquivalent ist zu einer, deren Gleichung die Form $y^2 = G(x, z)$ hat mit einem homogenen Polynom G vom Grad drei.

Nach Definition hat E mindestens einen Punkt P mit Koordinaten in k; falls dies ein Wendepunkt ist, wissen wir bereits, daß wir diese Form sogar durch eine lineare Koordinatentransformation erreichen können. Andernfalls schneidet die Tangente an E im Punkt P dort nur mit Vielfachheit zwei; nach dem Satz von BÉZOUT muß es also noch einen weiteren Schnittpunkt Q mit E geben. Dessen Koordinaten müssen *a priori* nicht unbedingt in k liegen, denn der Satz von BÉZOUT gilt ja nur über algebraisch abgeschlossenen Körpern. Hier bekommen wir jedoch für die Koordinaten von Q kubische Gleichungen über k, die die entsprechende Koordinate von P als doppelte Lösung haben, und die liegt in k. Durch Abdividieren erhalten wir eine lineare Gleichung über k, und auch deren Lösung liegt in k. (Alternativ könnte man natÜrlich auch den Wurzelsatz von VIÈTE anwenden.)

Durch eine lineare Koordinatentransformation können wir erreichen, daß die Gleichung der Tangenten an E in P die Gleichung $x = 0$ hat und Q die Koordinaten $(0 : 0 : 1)$. In der affinen (x, y)-Ebene ist Q also der Nullpunkt, und die y-Achse ist Tangente an E in P. Als Punkt der y-Achse hat P die x-Koordinate Null, hat also Koordinaten $(0 : y_0 : 1)$ mit einem noch unbekannten $y_0 \in k$. Zur Bestimmung von y_0 schneiden wir die y-Achse mit E, setzen also in der Kurvengleichung $F(x, y, z) = 0$ die Werte $x = 0$ und $z = 1$ ein. Mit Bezeichnungen wie oben ist

$$F(0, y, 1) = F_3(0, y) + F_2(0, y) + F_1(y + a_{10} = 0\,.$$

Da $Q = (0 : 0 : 1)$ auf der Kurve liegt, ist $a_{10} = 0$, und da F_j homogen vom Grad j ist. folgt $F_j(0.y) = y^j F_j(0, 1)$. Damit wird die Gleichung zu

$$y^3 F_3(0, 1) + y^2 F_2(0, 1) + y F_1(0, 1) = 0\,.$$

Diese kubische Gleichung hat die einfache Lösung $y0 = 0$, die dem Punkt Q entspricht, sowie die doppelte Lösung $y = y_0$ für P, denn dort ist die y-Achse ja Tangente. Die quadratische Gleichung

$$y^2 F_3(0, 1) + y F_2(0, 1) + F_1(0, 1) = 0$$

hat also eine zweifache Lösung, d.h. ihre Diskriminante

$$F_2(0, 1)^2 - 4F_1(0, 1)F_3(0, 1) = 0$$

verschwindet.

Als nächstes schneiden wir E mit der Geraden $y = tx$, wobei t ein neuer Parameter sei.

$$\begin{aligned} F(x, tx, 1) &= F_3(x, tx) + F_2(x, tx) + F_1(x, tx) \\ &= x^3 F_3(1, t) + x^2 F_2(1, t) + x F_1(1, t) = 0 \end{aligned}$$

hat die Lösung $x = 0$, was geometrisch klar war, da Q auf der Geraden liegt. Die beiden anderen Lösungen sind die der quadratischen Gleichung

$$x^2 F_3(1, t) + x F_2(1, t) + F_1(1, t) = 0 \,.$$

Da wir Körper der Charakteristik zwei ausgeschlossen haben, können wir diese Gleichung durch quadratische Ergänzung umformen zu

$$F_3(1, t) \left(\left(x + \frac{F_2(1, t)}{2F_3(1, t)} \right)^2 - \frac{F_2(1, t)^2}{4F_3(1, t)^2} + \frac{F_1(1, t)}{F_3(1, t)} \right) = 0 \,.$$

Multiplikation mit $4F_3(1, t)$ bringt die Nenner weg und liefert die neue Gleichung

$$\big(2F_3(1, t)x + F_2(1, t)\big)^2 = F_2(1, t)^2 - 4F_1(1, t)F_3(1, t) \,.$$

$\widetilde{G} = F_2^2 - F_1 F_3$ ist ein homogenes Polynom vom Grad vier in X und Y. Wie wir oben gesehen haben, verschwindet es für den Punkt $(0, 1)$, ist also durch X teilbar. Wir schreiben $\widetilde{G} = XG$ mit einem homogenen Polynom G vom Grad drei. Dann ist

$$G(1, t) = \widetilde{G}(1, t) = F_2(1, t)^2 - 4F_1(1, t)F_3(1, t)$$

ein Polynom vom Grad drei in t.

Wir führen nun zwei neue Variablen S, T ein durch

$$T = \frac{Y}{X}, \quad S = 2F_3(1, T)X + F_2(1, T) = 2F_3\left(1, \frac{Y}{X}\right) + F_1\left(1, \frac{Y}{X}\right) \,.$$

Die liefert eine rationale Abbildung der affinen Ebene auf sich selbst. (Beim Übergang zum projektiven müssen wir homogenisieren und können dann die X-Potenzen im Nenner eliminieren.) Es gibt auch eine rationale Abbildung in Gegenrichtung, denn

$$X = -\frac{F_2(1, T)}{2F_3(1, T)} \qquad \text{und} \qquad Y = TX = -\frac{TF_2(1, T)}{2F_3(1, T)} \,.$$

Somit ist E birational äquivalent zur Kurve $s^2 = G(1,t)$, wobei $G(1,T)$ ein Polynom vom Grad drei ist.

In der Ebene mit den neuen Koordinaten ist E somit die Nullstellenmenge eines Polynoms

$$S^2 - (aT^3 + bT^2 + cT + d)\,.$$

Um a auf eins zu setzen, machen wir noch einen letzten Koordinatenwechsel mit $U = aS$ und $V = aT$; wir erhalten

$$\frac{U^2}{a^2 =} - \left(a\frac{V^3}{a^3} + b\frac{V^2}{a^2} + c\frac{T}{a} + d\right)\,.$$

Multiplikation mit a^2 macht daraus

$$U^2 - (V^3 + bV^2 + acV + a^2 d)\,;$$

wir bekommen also ein Polynom der gleichen Gestalt wie im ersten Fall, als es einen Wendepunkt in $E(k)$ gab.

Damit haben wir gezeigt

Satz: Jede elliptische Kurve $E \subset \mathbb{P}^2(k)$ über einem Körper mit von zwei verschiedener Charakteristik ist birational äquivalent zu einer Kurve $V(F)$ mit $F = Y^2Z - (X^3 + c_2X^2Z + c_1XZ^2 + c_0Z^3)$. ∎

Wie wissen allerdings noch nicht, ob die Kurven $V(F)$ für beliebige Wahl der Konstanten a, b, c, d elliptisch sind. Tatsächlich sind sie es nicht; beispielsweise ist der Nullpunkt auf der Kurve $V(Y^2Z - X^3)$ offensichtlich singulär. Für das obige Polynom F ist

$$\frac{\partial F}{\partial y}(x,y) = 2y\,;$$

damit sind alle Punkte mit nichtverschwindender y-Koordinate nichtsingulär.

Für Punkte $(x : y : z)$ mit $y = 0$ muß $(x : z)$ eine Nullstelle des kubischen Polynoms $X^3 + aX^2Z + bxZ^2 + cZ^2$ sein. Die partiellen Ableitungen von F nach x und z sind bis aufs Vorzeichen die dieses Polynoms, und die verschwinden genau dann beide, wenn es sich um

eine mehrfache Nullstelle handelt. Wir erhalten also stets eine Gleichung obiger Form, bei der das kubische Polynom in der Klammer keine mehrfachen Nullstellen hat, und umgekehrt definiert eine Gleichung dieser Form dann eine elliptische Kurve.

Falls die Charakteristik des Körpers k auch von drei verschieden ist, können wir die Gleichung noch etwas vereinfachen: Dann können wir analog zur quadratischen Ergänzung eine „kubische Ergänzung“ durchführen, indem wir die Koordinate X ersetzen durch $X' = X + \frac{1}{3}aZ$. Dies bringt den Koeffizienten von X^2Z zum Verschwinden, wir können also sogar eine Gleichung der Form

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

erreichen.

Definition: Eine Gleichung dieser Form heißt WEIERSTRASSsche Normalform.

(WEIERSTRASS untersuchte elliptische Kurven über den komplexen Zahlen; in seiner Originalgleichung hatte x^3 tatsächlich den in der analytischen Theorie sinnvollen Koeffizienten vier.)