Elliptische Kurven in der Kryptographie

Ein Faktorisierungsalgorithmus

Thomas Wunderer

September 2012

Bachelorarbeit im Bachelorstudium Mathematik

Institut für Mathematik und wissenschaftliches Rechnen

Karl-Franzens-Universität Graz

Betreuer: Ao. Prof. Dipl.-Ing. Dr. techn. Günter Lettl

Vorwort

Diese Arbeit wurde von mir ursprünglich im Rahmen des Seminars *Zahlentheorie (Algebra und Kryptographie)* bei Prof. Grabner und Prof. Tichy auf der Technischen Universität Graz verfasst, und dann zu dieser Bachelorarbeit ausgebaut.

Einleitung

Wir alle kennen den Fundamentalsatz der Arithmetik, nach dem sich jede natürliche Zahl auf eindeutige Weise als Produkt von Primzahlen darstellen lässt. Leider lässt sich aus den bisher bekannten Beweisen dieses Satzes aber kein effizientes Verfahren zur Berechnung eben dieser Primfaktoren ableiten. Seit jeher haben Mathematiker versucht ein solches Verfahren zu finden, doch bisher ist noch keines bekannt, das auch bei großen Zahlen schnell funktioniert. Genau diese Tatsache machen sich die meisten modernen Verschlüsselungsverfahren zu Nutze. Somit ist die spannende Frage nach einem effizientem Faktorisierungsalgorithmus in der heutigen Zeit auch mehr denn je von praktischer Natur. Ziel dieser Arbeit ist es nun, einen Algorithmus zur Faktorisierung natürlicher Zahlen, dessen Grundidee auf der Arithmetik elliptischer Kurven über endlichen Körpern beruht, herzuleiten und zu analysieren. Im ersten Teil wird eine Einführung in die Thematik der elliptischen Kurven gegeben. Danach werden speziell Kurven über endlichen Körpern betrachtet, die das Fundament vieler Methoden in der Kryptographie, und auch des Faktorisierungsalgorithmus von Lenstra, der schließlich im letzten Teil behandelt wird, bilden.

Einführung in die elliptische Kurven

Als erstes wollen wir uns mit algebraischen Kurven über einem beliebigen Körper K, insbesondere mit elliptischen Kurven vertraut machen, ohne die Sätze zu beweisen.

Definition: Sei $f(x,y)$ ein Polynom mit Koeffizienten in einem Körper K.

(a) Die Lösungsmenge der Gleichung $C: f(x,y) = 0$ in $K \times K$ vereinigt mit den über K definierten projektiven Fernpunkten[1] heißt *(ebene) Kurve* über K, oder kurz $C(K)$.

(b) Ist $P = (x,y) \in C(K)$ mit $x, y \in K$ oder ist P ein Fernpunkt mit homogenen Koordinaten aus K, so heißt P ein K-rationaler Punkt auf der Kurve C. $\mathbb{Q}$-rationale Punkte werden kurz auch als rationale Punkte bezeichnet.

(c) Ist f vom Grad 3, so heißt $C(K)$ eine *kubische Kurve* oder *Kubik*.

(d) Sei $C: f(x,y) = 0$ eine Kubik über einem Körper K. Verschwinden für keinen Punkt $P \in C(K)$ gleichzeitig die beiden partiellen Ableitungen von f, so heißt C *nicht singulär*.

(e) Eine nicht singuläre Kubik mit einem K-rationalen Punkte heißt *elliptische Kurve* über K.

Satz 1*: Es sei K ein Körper mit $char(K) \neq 2$. Jede elliptische Kurve über K kann durch eine birationale Transformation in eine Gleichung der Form*

$$C: y^2 = f(x) = x^3 + ax^2 + bx + c$$

umgeschrieben werden. Diese Form nennt man Weierstraß Normalform.[2]

Bemerkung: Eine Kubik in Weierstraß Normalform hat genau einen projektiven Fernpunkt. Diesen nennen wir den *Punkt im Unendlichen* und bezeichnen ihn mit $\mathcal{O}$.

Von nun an sei K ein Körper mit $char(K) \neq 2$.

Definition: Seien C und f wie im Satz oben. Dann heißt $D(f) = -4a^3c + a^2b^2 + 18abc - 4b^3 - 27c^2$ die *Diskriminante* von f.

Bemerkung: Seien C und f wie im Satz oben und $F(x,y) \coloneqq y^2 - f(x)$, also ist C gegeben durch $C: F(x,y) = 0$. Ein Polynom $p \in K[X]$ hat genau dann eine mehrfache Nullstelle im Punkt x_0, wenn $p(x_0) = 0$ und $p'(x_0) = 0$ gilt. Nachdem die partielle Ableitung von F nach x mit der negativen Ableitung von f überein stimmt, verschwinden diese in genau den selben Punkten. Die partielle Ableitung von F nach y verschwindet genau im Fall $y = 0$. Somit verschwinden für einen Punkt $(x_0, y_0) \in C(K) \backslash \{\mathcal{O}\}$ beide partiellen Ableitungen von F genau dann, wenn $f(x_0) = y_0^2 = 0$ und $f'(x_0) = 0$, also wenn x_0 eine mehrfache Nullstelle von f ist. Somit ist eine Kubik in Weierstraß Normalform genau dann eine elliptische Kurve, wenn f keine mehrfachen Nullstellen besitzt.

---

[1] Für eine kurze Einführung in die projektive Geometrie siehe nächstes Kapitel.
[2] Siehe [11], Seite 369 ff

Zur Veranschaulichung betrachten wir einmal ein paar Beispiele für reelle singuläre Kubiken und elliptische Kurven.

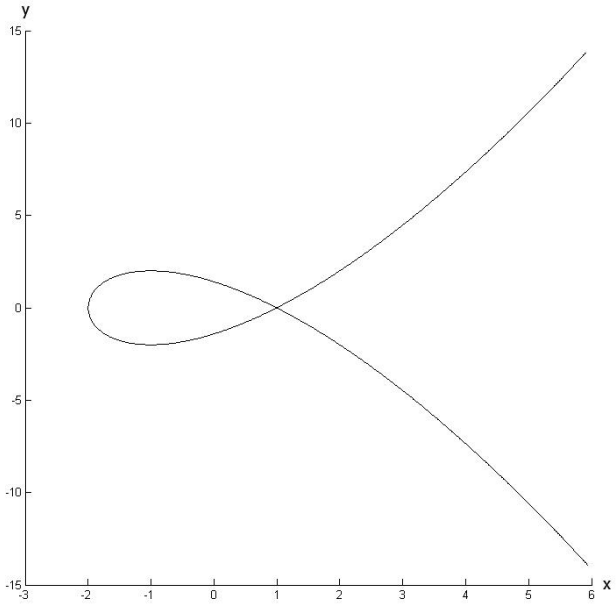

Abbildung 1: $C_1: y^2 = x^3 - 3x + 2$

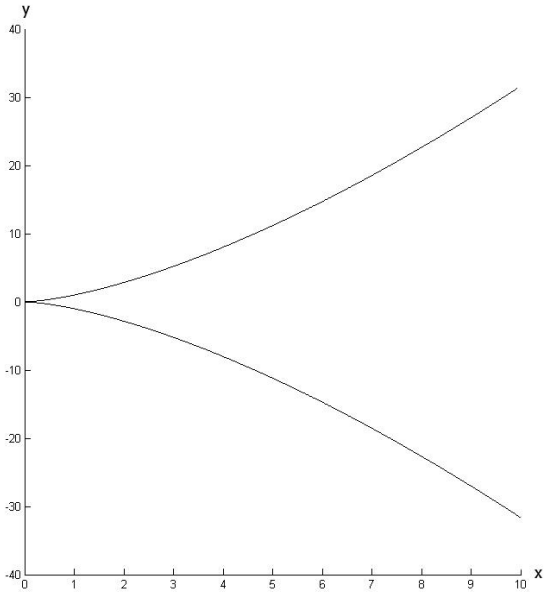

Abbildung 2: $C_2: y^2 = x^3$

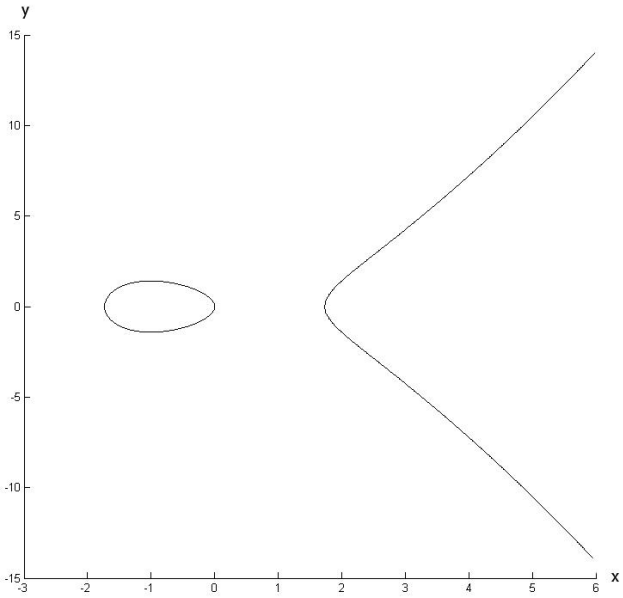

Abbildung 3: $C_3: y^2 = x^3 - 3x$

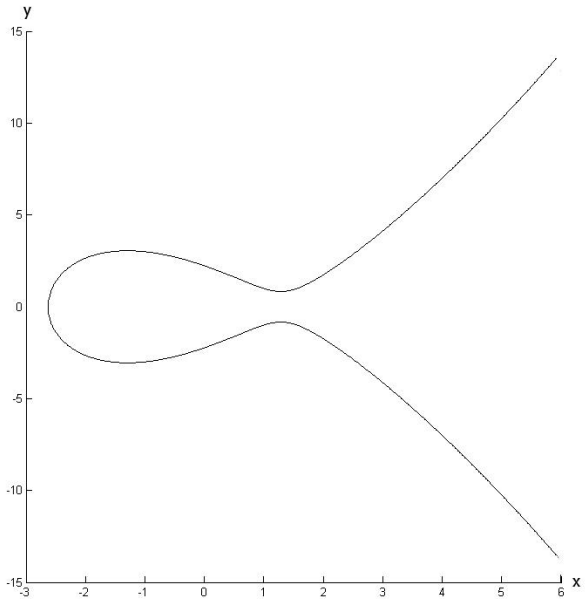

Abbildung 4: $C_4: y^2 = x^3 - 5x + 5$

Die ersten beiden Abbildungen zeigen Kurven mit Singularitäten. Die Kurve C_1 aus Abbildung 1 hat ihren singulären Punkt dort, wo sich die Kurve selbst schneidet. Die Singularität der Kurve C_2 liegt in der Spitze im Punkt $(0,0)$ vor.

Die Kurven C_3 und C_4 aus den Abbildungen 3 und 4 stellen nicht singuläre Kubiken, also elliptische Kurven dar.

Der nächste Satz zeigt, welche algebraische Struktur die Punkte auf einer elliptischen Kurve haben.

Satz 2: *Sei* $C: y^2 = x^3 + ax^2 + bx + c$ *eine elliptische Kurve über* K*. Dann ist* $(C(K), +)$ *eine abelsche Gruppe mit neutralem Element* $\mathcal{O}$*, wobei für* $P_1 = (x_1, y_1)$ *und* $P_2 = (x_2, y_2)$ *die Summe* $P_1 + P_2$ *definiert ist durch:*

$$P_3 = \begin{cases} P_2\ (bzw.\ P_1) & ,falls\ P_1 = \mathcal{O}\ (bzw.\ P_2 = \mathcal{O}) \\ \mathcal{O} & ,falls\ x_1 = x_2\ und\ y_1 = -y_2 \\ (x_3, y_3) = (\lambda^2 - a - x_1 - x_2, -\lambda x_3 - \nu) & ,sonst \end{cases}$$

wobei

$$\lambda = \begin{cases} \frac{y_2 - y_1}{x_2 - x_1} & ,falls\ P_1 \neq P_2 \\ \frac{3{x_1}^2 + 2ax_1 + b}{2y_1} & ,falls\ P_1 = P_2 \end{cases} \quad und\ \nu = y_1 - \lambda x_1 = y_2 - \lambda x_2.$$

Die im obigen Satz definierte Gruppenoperation hat eine anschauliche geometrische Motivation im Falle $K = \mathbb{R}$. Wollen wir zwei verschiedene Punkte P und Q einer elliptischen Kurve $C(\mathbb{R})$ addieren, so ziehen wir zunächst eine Gerade durch die beiden Punkte. Diese Gerade schneidet die Kurve in einem dritten Punkt $P * Q$ (eventuell im Unendlichen). Anschließend spiegeln wir diesen Punkt $P * Q$ an der x-Achse, und erhalten so den Punkt $P + Q$. Die Spiegelung eines Punktes S an der x-Achse entspricht genau dem dritten Schnittpunkt der Geraden durch die Punkte S und $\mathcal{O}$ mit der Kurve C. Wir könnten so auch jeden anderen Punkt der Kurve C zum neutralen Element der Gruppe machen, indem wir im zweiten Schritt anstatt durch $\mathcal{O}$ durch diesen Punkt eine Gerade ziehen. Allerdings vereinfachen sich dadurch, dass sich, wenn wir gerade $\mathcal{O}$ verwenden, einfach eine Spiegelung ergibt, die Formeln für die Addition enorm. Die Addition zweier Punkte einer reellen Kurve sieht also so aus:

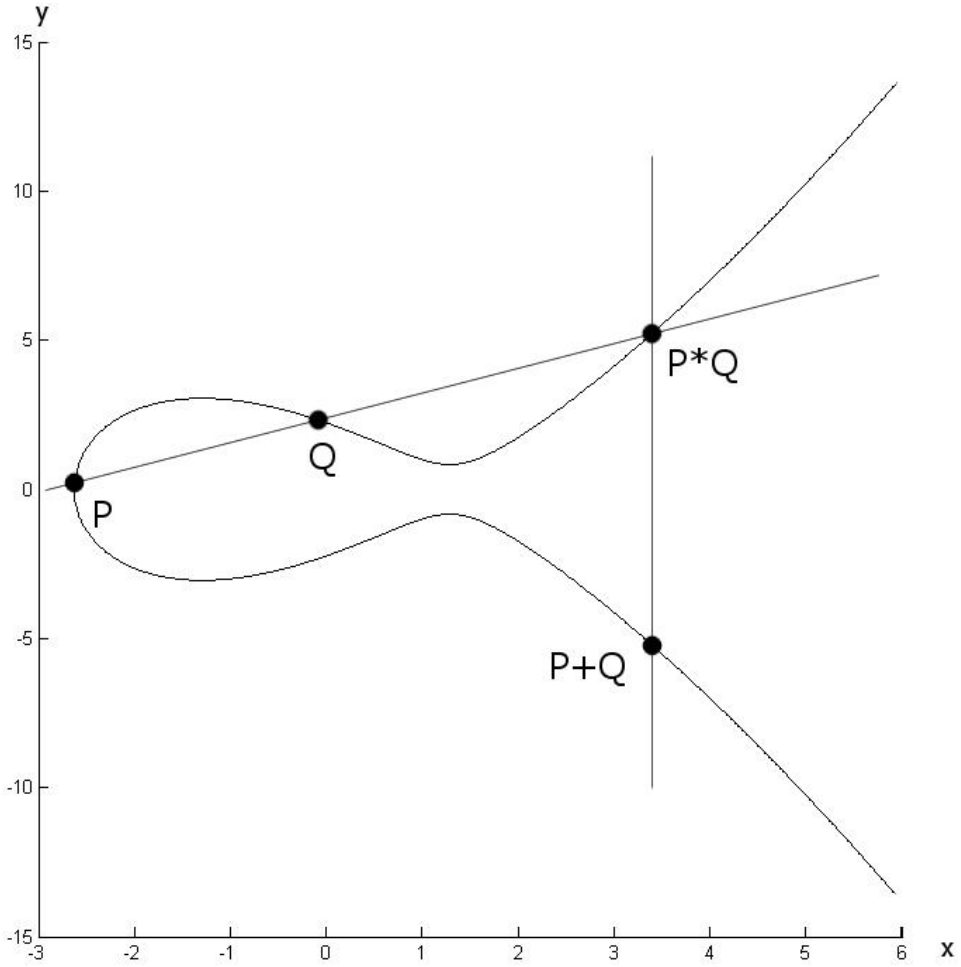

Abbildung 5: Addition von Punkten auf einer elliptischen Kurve

Bemerkung: Die selbe geometrische Definition des Additionsgesetzes über Schnitte von Geraden mit der Kurve liegt auch im Falle eines beliebigen Körpers K mit $char(K) \neq 2$ zugrunde.

Elliptische Kurven über endlichen Körpern

In diesem Abschnitt wollen wir uns ein wenig mit elliptischen Kurven über endlichen Körpern befassen, welche in der Kryptographie ihre Anwendung finden. Da uns später insbesondere eine Reduktion modulo einer Primzahl p interessieren wird, genügt uns sogar die Betrachtung von Kurven über $\mathbb{Z}_p \coloneqq \mathbb{Z}/p\mathbb{Z}$ für eine Primzahl p. Zu diesem Zweck sind alle in diesem Abschnitt folgenden Rechnungen und arithmetischen Ausdrücke im jeweiligen Körper $\mathbb{Z}_p$ zu verstehen. Wir haben im vorherigen Kapitel bereits gesehen, wie Kurven über $\mathbb{Z}_p$ definiert sind. Als rationalen Punkt definieren wir einen Punkt mit Koordinaten in $\mathbb{Z}_p$, der auf einer solchen Kurve liegt.

Wegen Satz 1 genügt es uns im Falle $p \neq 2$, ausschließlich Kurven in Weierstraß Normalform zu betrachten.

Untersuchen wir erst einmal, wann so eine Kurve nicht singulär ist.

Lemma 3: *Sei $C: y^2 = x^3 + ax^2 + bx + c$ eine Kurve über $\mathbb{Z}_p$ mit $p \neq 3$. Dann existiert eine birationale Transformation $X = r_1(x), Y = r_2(y)$, sodass C in der Form $Y^2 = X^3 + BX + C$ mit Koeffizienten in $\mathbb{Z}_p$ geschrieben werden kann.*

Beweis: Setze $X \coloneqq x + \frac{a}{3}$ und $Y \coloneqq y$. Das ist wohldefiniert, nachdem $p \neq 3$ gilt. Mit dieser (offensichtlich birationalen) Transformation gilt:

$$Y^2 = \left(X - \frac{a}{3}\right)^3 + a\left(X - \frac{a}{3}\right)^2 + b\left(X - \frac{a}{3}\right) + c =$$

$$= X^3 + \left(b - \frac{a^2}{3}\right)X + \left(\frac{2a^3 - 9ab}{27} + c\right)$$

Und da $p \neq 3$ sind alle Koeffizienten in $\mathbb{Z}_p$ wohldefiniert. ■

Bemerkung: Eine solche Form heißt kurze Weierstraß Normalform. Ist $p \neq 2{,}3$ kann mit Satz 1 und obigem Lemma jede elliptische Kurve über $\mathbb{Z}_p$ durch birationale Transformation in eine kurze Weierstraß Normalform umgeschrieben werden.

Satz 4: *Sei $p \neq 2{,}3$ eine Primzahl und $C: y^2 = x^3 + bx + c$ eine Kurve über $\mathbb{Z}_p$. Dann ist C genau dann nicht singulär, wenn die Diskriminante der Kurve $D = -4b^3 - 27c^2$ nicht das Nullelement in $\mathbb{Z}_p$ ist.*

Beweis: Sei C singulär. Dann existiert per Definition der Singularität ein Punkt $P = (x, y) \in \mathbb{Z}_p \times \mathbb{Z}_p$, der gleichzeitig die Gleichungen

$$(I)\ y^2 = x^3 + bx + c \quad , \quad (II)\ 2y = 0 \quad , \quad (III)\ 3x^2 + b = 0$$

erfüllt. Aus (II) können wir wegen $p \neq 2$ schließen, dass $y = 0$ gilt. Aus (III) schließen wir, dass $b = -3x^2$. Mit diesen beiden Ergebnissen erhalten wir aus (I) wiederum $c = y^2 -$

$x^3 - bx = 2x^3$. Somit erhalten wir für die Diskriminante $D = -4b^3 - 27c^2 = 108x^6 - 108x^6 = 0$.

Sei andererseits $D = -4b^3 - 27c^2 = 0$, so gilt $c^2 = -4b^3 27^{-1} = -(3^{-1}2b)^2(3^{-1}b)$, oder umgeformt $(3c(2b)^{-1})^2 = -3^{-1}b$. Um nun einen singulären Punkt $P = (x, y) \in \mathbb{Z}_p \times \mathbb{Z}_p$ auf der Kurve zu finden, muss dieser $(I) - (III)$ erfüllen. Wegen (II) setzen wir $y \coloneqq 0$ und wegen (III) und voriger Überlegung setzen wir $x \coloneqq -3c(2b)^{-1}$, wobei wir gleich sehen werden, dass das voran gesetzte Minus zum richtigen Ergebnis führt. P erfüllt also offensichtlich (II) und (III). Bleibt nur noch zu überprüfen, ob P überhaupt auf der Kurve liegt, also (I) erfüllt:

$$x^3 + bx + c = -3c(2b)^{-1}(-3c(2b)^{-1})^2 - b3c(2b)^{-1} + c =$$

$$= 2^{-1}c - 2^{-1}3c + c = 0 = y^2$$

Somit haben wir einen singulären Punkt auf der Kurve gefunden. ■

Wir haben also ein Kriterium für die Singularität einer Kurve gefunden.

Bemerkung: Im Beweis von Lemma 3 und Satz 4 ist lediglich eingegangen, dass die Charakteristik des zugrundeliegenden Körpers ungleich zwei und drei ist. Die Sätze lassen sich also mit den gleichen Beweisen für beliebige Körper K mit $char(K) \neq 2{,}3$ verallgemeinern.

Wie wir bereits wissen, bilden die Punkte auf einer elliptischen Kurve eine abelsche Gruppe:

Satz 5: *Sei $C: y^2 = x^3 + ax^2 + bx + c$ eine elliptische Kurve über $\mathbb{Z}_p$. Dann ist $\big(C(\mathbb{Z}_p), +\big)$ eine abelsche Gruppe mit neutralem Element $\mathcal{O}$ und der Definition für die Addition zweier Punkte aus Satz 2.*

Beweis: Spezialfall von Satz 2 mit $K = \mathbb{Z}_p$. ■

Die nächste Frage, die sich uns stellt, ist natürlich die nach der Größe dieser Gruppe. Da $\mathbb{Z}_p$ ein endlicher Körper ist, und $C(\mathbb{Z}_p) \subseteq (\mathbb{Z}_p \times \mathbb{Z}_p) \cup \{\mathcal{O}\}$, ist $\big|C(\mathbb{Z}_p)\big|$ natürlich durch $\big|\mathbb{Z}_p\big|^2 + 1 = p^2 + 1$ nach oben beschränkt, also insbesondere immer endlich. Sei $C: y^2 = f(x)$ eine elliptische Kurve. Da es im Falle $p \neq 2$ genauso viele quadratische Reste wie nicht-Reste $mod\, p$ gibt, würden wir erwarten, dass in etwa jedes zweite x für $f(x)$ einen quadratischen Rest liefert. Für jedes solche x gibt es dann meist 2 mögliche Werte für y, sodass (x, y) auf C liegt. Rechnen wir noch den Punkt im Unendlichen dazu, so erwarten wir also etwa $p + 1$ Punkte auf der Kurve. Tatsächlich ist $p + 1$ eine gute Schätzung, wie folgender Satz zeigt.

Satz 6 (Hasse-Weil): *Sei $p \neq 2, 3$ und C eine elliptische Kurve über $\mathbb{Z}_p$, so gilt:*

$$\big|C(\mathbb{Z}_p) - p - 1\big| \leq 2\sqrt{p}$$

Beweis: Ohne Beweis.[3] ■

Für unseren Faktorisierungsalgorithmus fehlt uns jetzt nur noch eine Überlegung: was passiert mit einer elliptischen Kurve C über $\mathbb{Q}$ mit ganzzahligen Koeffizienten in kurzer Normalform, wenn wir sie modulo einer Primzahl $p \neq 2{,}3$ reduzieren? Sprich, wenn wir die Kurve

$$C: y^2 = x^3 + bx + c$$

mittels des kanonischen Restklassenhomomorphismus

$\pi: \mathbb{Z} \to \mathbb{Z}_p,\ a \to \pi(a) =: \bar{a}$ auf die Kurve

$$\tilde{C}: y^2 = x^3 + \bar{b}x + \bar{c}$$

reduzieren. Das neutrale Element von $\tilde{C}$ bezeichnen wir mit $\tilde{\mathcal{O}}$. Diese reduzierte Kurve ist wie wir wissen genau dann nicht singulär, wenn für die Diskriminante

$$\widetilde{D} = -4\bar{b}^3 - 27\bar{c}^2 \neq 0_{\mathbb{Z}_p}$$

gilt. Da π ein Homomorphismus ist, ist das der Fall, falls $D = -4b^3 - 27c^2 \not\equiv 0\ (mod\ p)$, also falls p nicht die Diskriminante D teilt. Gilt also $p \nmid D$, so sind $C(\mathbb{Q})$ und $\tilde{C}(\mathbb{Z}_p)$ beides elliptische Kurven und damit abelsche Gruppen. Wie diese Gruppen auf natürliche Weise zusammenhängen, zeigt folgender Satz.

Satz 7: *Sei $p \neq 2{,}3$ eine Primzahl und $C: y^2 = x^3 + bx + c$ eine elliptische Kurve mit ganzzahligen Koeffizienten und $p \nmid 4b^3 + 27c^2$. Seien außerdem $\frac{d}{e}, \frac{f}{g}$ rationale Zahlen in reduzierter Bruchdarstellung. Dann ist $\varphi: C(\mathbb{Q}) \to \tilde{C}(\mathbb{Z}_p)$, mit*

$$\varphi\left(\frac{d}{e}, \frac{f}{g}\right) = \begin{cases} \left(\bar{d}\bar{e}^{-1}, \bar{f}\bar{g}^{-1}\right) & , \quad falls\ p \nmid e, g \\ \tilde{\mathcal{O}} & , \quad sonst \end{cases}$$

$$\varphi(\mathcal{O}) = \tilde{\mathcal{O}}$$

ein Gruppenhomomorphismus. φ heißt Reduktion der Kurve $mod\ p$ und wir schreiben $\varphi(P) = \tilde{P}$.

Für den Beweis des Satzes müssen wir uns zunächst ein wenig mit den Grundzügen der projektiven Geometrie und Kurven in der projektiven Ebene beschäftigen.

Davor betrachten wir aber noch ein nützliches Lemma über die Darstellung rationaler Punkte auf einer elliptischen Kurve.

[3] Die in der Arbeit behandelten Mittel reichen für den Beweis leider nicht aus. Ein Beweis findet sich in [2].

Lemma 8: *Sei $C: y^2 = x^3 + bx + c$ eine elliptische Kurve mit ganzzahligen Koeffizienten und $\mathcal{O} \neq P \in C(\mathbb{Q})$ ein rationaler Punkt auf der Kurve. Dann existiert ein $k \in \mathbb{Z}\backslash\{0\}$, sodass P die reduzierte Bruchdarstellung $P = \left(\frac{d}{k^2}, \frac{f}{k^3}\right)$ besitzt.*

Beweis: Sei $P = \left(\frac{d}{e}, \frac{f}{g}\right)$ ein rationaler Punkt auf C, wobei die Koordinaten in reduzierter Bruchdarstellung vorliegen. Da P auf der Kurve C liegt, ist

$$\frac{f^2}{g^2} = \frac{d^3}{e^3} + b\frac{d}{e} + c = \frac{d^3 + bde^2 + ce^3}{e^3} \Leftrightarrow e^3 f^2 = d^3 g^2 + bde^2 g^2 + ce^3 g^2$$

erfüllt, und somit gilt $g^2 | e^3$, da g^2 die rechte Seite teilt und g und f teilerfremd sind. Auf der anderen Seite bekommen wir aus der Gleichung aber auch $e^2 | d^3 g^2$, und damit wiederum $e^2 | g^2$, beziehungsweise $e | g$ wegen $ggT(e, d) = 1$. Es gilt also auch $e^3 | eg^2$. Dadurch teilt e^3 in der obigen Gleichung sowohl die linke Seite, als auch die letzten beiden Summanden der rechten Seite, also muss es auch $d^3 g^2$ teilen, woraus wir wegen der Teilerfremdheit von e und d schlussendlich $e^3 | g^2$ erhalten.

Somit ist $e^3 = g^2$, also $g = \sqrt{e^3} = e\sqrt{e}$, was wegen $e | g$ bedeutet, dass $\sqrt{e}$ ganzzahlig ist, also $e = k^2$ für ein ganzes k. Dadurch gilt $g = k^3$, und wir sehen dass P die geforderte Gestalt besitzt. ■

Diese Darstellung werden wir später noch verwenden.

Einschub über projektive Geometrie

Definition: Sei K ein Körper. Wir sagen zwei Punkte (a,b,c) und (d,e,f) in K^3 sind äquivalent, kurz $(a,b,c)\sim(d,e,f)$, wenn $(a,b,c)=(td,te,tf)$ für ein $t\in K\setminus\{0\}$ gilt. Die *projektive Ebene* $\mathbb{P}^2(K)$ ist definiert als die Menge aller von $[(0,0,0)]_\sim$ verschiedenen Äquivalenzklassen des K^3 unter $\sim$. Wir schreiben für $[(a,b,c)]_\sim \in \mathbb{P}^2(K)$ auch einfach (a,b,c) und $\mathbb{P}^2$ für $\mathbb{P}^2(\mathbb{R})$.

Bemerkung: Die Schreibweise $\mathbb{P}$ hat sich sowohl für die Primzahlen, als auch im Bezug zur projektiven Geometrie durchgesetzt, und wird im Rahmen dieser Arbeit auch für beides verwendet. Im Kontext sollte dem Leser aber immer klar sein, was gerade gemeint ist.

Ein Punkt (a,b,c) in der projektiven Ebene $\mathbb{P}^2$ entspricht also einer Geraden im $\mathbb{R}^3$, die durch (a,b,c) und den Ursprung geht (eigentlich ohne den Ursprung selbst).

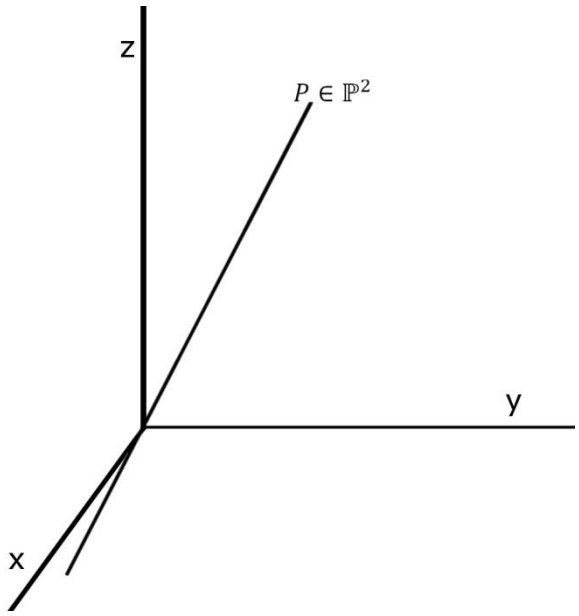

Abbildung 6

Als nächstes identifizieren wir den $\mathbb{R}^2$ im $\mathbb{R}^3$ mit der affinen Ebene $\mathbb{A}^2$, gegeben durch die Gleichung $z=1$, also die um 1 entlang der z-Achse nach oben verschobene x-y-Ebene.

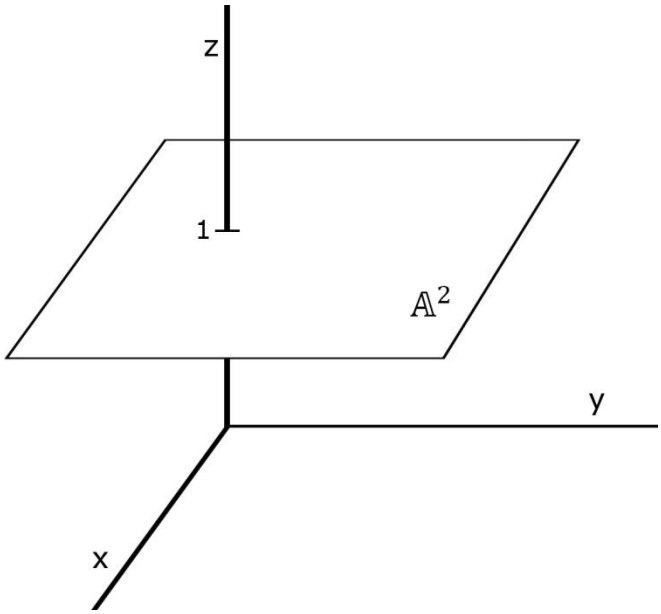

Abbildung 7

Die durch einen Punkt $(a, b, c) \in \mathbb{P}^2$ mit $c \neq 0$ gegebene Gerade im $\mathbb{R}^3$ schneidet die affine Ebene genau in einem Punkt, nämlich $\left(\frac{a}{c}, \frac{b}{c}, 1\right)$. Wir können also jedem Punkt $(x, y) \in \mathbb{R}^2$ eindeutig den Punkt $(x, y, 1)$ in der projektiven Ebene zuordnen, und andererseits jeden Punkt $(x, y, z) \in \mathbb{P}^2$ mit $z \neq 0$ eindeutig mit dem Punkt $\left(\frac{x}{z}, \frac{y}{z}\right) \in \mathbb{R}^2$ identifizieren.

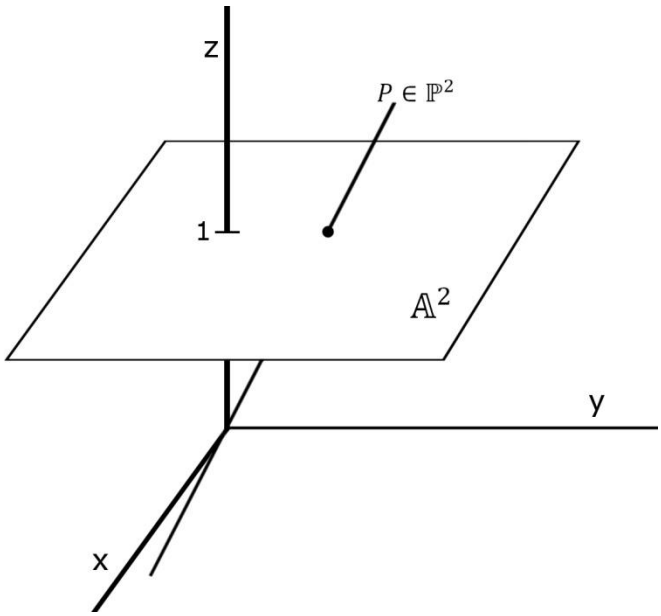

Abbildung 8

Die Punkte der projektiven Ebene, die bei dieser Identifikation ausgelassen werden, also die Punkte $(x, y, 0)$ mit x und y nicht beide Null, heißen *unendlich ferne* Punkte. Die dadurch induzierten Geraden im $\mathbb{R}^3$ schneiden die affine Ebene in keinem Punkt.

Eine *Gerade* in $\mathbb{P}^2(K)$ ist definiert als Lösungsmenge einer Gleichung $G: ax + by + cz = 0$ mit Koeffizienten in K, nicht alle gleich Null. Im anschaulichen Fall von $K = \mathbb{R}$ sieht man, dass eine Gerade in $\mathbb{P}^2$ einer Ebene im $\mathbb{R}^3$, die durch den Ursprung verläuft, entspricht.

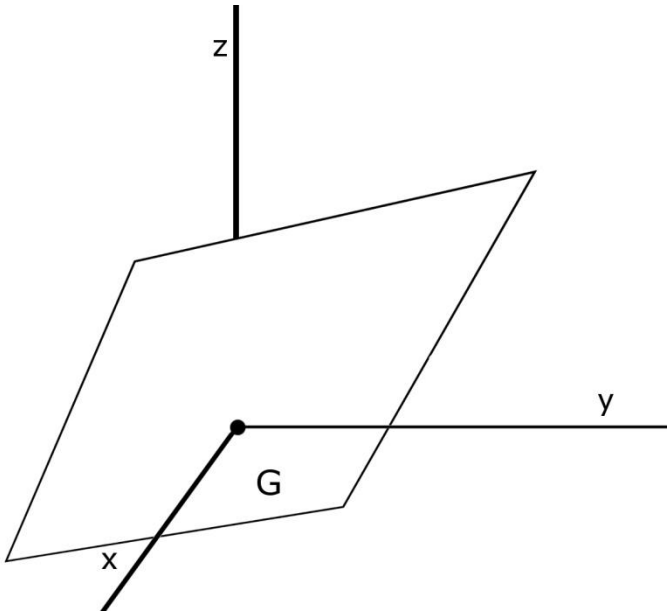

Abbildung 9

Ähnlich wie bei den Punkten in $\mathbb{P}^2$ gehen wir jetzt mit den Geraden vor: jede Gerade in $\mathbb{P}^2$ entspricht genau einer Ursprungsebene im $\mathbb{R}^3$, die, falls die Ebene nicht durch $z = 0$ gegeben ist, eine eindeutige Schnittgerade mit der affinen Ebene $\mathbb{A}^2$ liefert, welche wir wiederum mit einer Geraden im $\mathbb{R}^2$ identifizieren können. Auf der anderen Seite liefert jede Gerade G der affinen Ebene, die eine Gerade im $\mathbb{R}^2$ repräsentiert, genau eine Ebene, die durch den Ursprung verläuft und deren Schnittgerade mit $\mathbb{A}^2$ die Gerade G ist, also eine eindeutige Gerade in der projektiven Ebene.

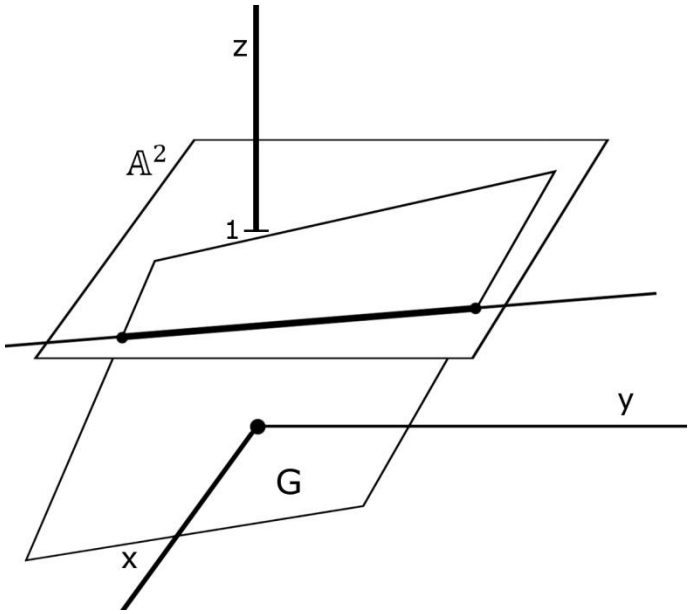

Abbildung 10

Die Gerade in $\mathbb{P}^2$, die wir auf diese Weise nicht übertragen können, nämlich $G: z = 0$, heißt die *unendlich ferne* Gerade. Sie besteht genau aus den unendlich fernen Punkten.

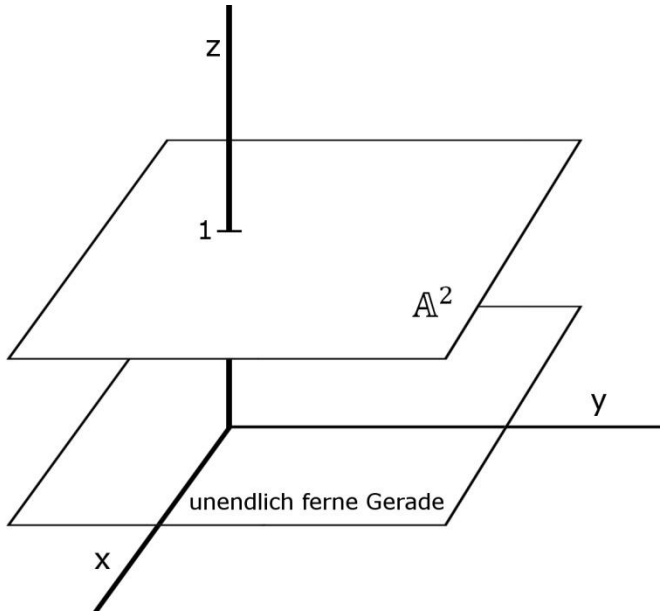

Abbildung 11

Untersuchen wir nun, wie wir dieses Konzept auf Kurven übertragen können. Sei also beispielsweise eine Kurve $C: y^2 = x^3 + bx + c$ gegeben. Greifen wir auf die projektiven Koordinaten zurück, erhalten wir dadurch eine Gleichung $\frac{y^2}{z^2} = \frac{x^3}{z^3} + b\frac{x}{z} + c$,

beziehungsweise $\tilde{C}: y^2z = x^3 + bxz^2 + cz^3$. Die so neu entstandene Gleichung ist dadurch hervorgegangen, dass wir jedes Monom mit einer entsprechenden Potenz von z multipliziert haben, sodass der Grad jedes Monoms in der neuen Gleichung dem höchsten auftretenden Grad in der ursprünglichen Gleichung, in diesem Fall 3, entspricht. Dieses Verfahren nennt sich Homogenisieren und lässt sich analog auf Polynome über beliebigen Körpern übertragen.

Definition: Sei K ein Körper und $f \in K[X,Y]$ ein Polynom vom Grad $n = \deg(f)$. Die Homogenisierung von f ist gegeben durch das Polynom $g \in K[X,Y,Z]$, welches entsteht, indem man in f jedes Monom mit der entsprechenden Potenz von Z multipliziert, sodass jedes Monom von g Grad n hat.

Wir sehen sofort einige wesentlichen Eigenschaften der Homogenisierung g eines Polynoms f. Ist (x,y) eine Nullstelle von f, so ist $(x,y,1)$ wegen $g(x,y,1) = f(x,y)$ eine Nullstelle von g. Und ist (x,y,z) eine Nullstelle von g, so ist es auch (tx,ty,tz). Damit ist aber auch, sofern $z \neq 0$ und (x,y,z) eine Nullstelle von g, $\left(\frac{x}{z}, \frac{y}{z}, 1\right)$ eine Nullstelle von g und dadurch auch $\left(\frac{x}{z}, \frac{y}{z}\right)$ eine Nullstelle von f.

Haben wir ein Polynom $f \in K[X,Y]$ vom Grad n, so erhalten wir für dieses eine eindeutige Homogenisierung $g \in K[X,Y,Z]$. Andererseits liefert uns ein homogenes[4] Polynom $g \in K[X,Y,Z]$ vom Grad n, welches nicht durch Z teilbar ist, durch $g(x,y,1) =: f(x,y)$ genau ein Polynom $f \in K[X,Y]$, welches g als Homogenisierung besitzt. Wir können also eindeutig Polynome in $K[X,Y]$ mit Homogenisierungen in $K[X,Y,Z]$ identifizieren, oder anders gesagt, Kurven in K^2 mit Kurven in $\mathbb{P}^2(K)$.

Bemerkung: Sprechen wir von einer projektiven Kurve $C: f(x,y,z) = 0$, so meinen wir also insbesondere, dass f homogen ist.[5]

Betrachten wir nun den Spezialfall von Geraden. Eine Gerade in K^2 ist gegeben durch die Lösungsmenge einer Gleichung $G: ax + by + c = 0$ mit a und b nicht gleichzeitig Null. Homogenisieren liefert nun die Gleichung $\tilde{G}: ax + by + cz = 0$, also einer Geraden in $\mathbb{P}^2(K)$. Das rechtfertigt auch die geometrische Deutung, die wir schon vorher im Reellen gegeben haben.

Wir haben auch schon gesehen, dass wir für eine Kurve C einem Punkt $(x,y) \in C(K)$ den Punkt $(x,y,1) \in C'(K)$ zuordnen können, und einem Punkt $(x,y,z) \in C'(K)$ für $z \neq 0$ den Punkt $\left(\frac{x}{z}, \frac{y}{z}\right) \in C(K)$. Dabei bezeichne C' die projektive Kurve, die durch Homogenisieren des Polynoms, welches C definiert, hervorgeht. Die übrigen Punkte mit $z = 0$, also gerade die unendlich fernen Punkte, entsprechen genau den projektiven Fernpunkten, die wir bei der Definition von $C(K)$ hinzu genommen haben. Das liefert also eine eins-zu-eins Übertragung der Punkte auf $C(K)$ mit denen auf $C'(K)$. Ist $C(K)$ eine elliptische Kurve,

[4] Ein Polynom heißt homogen vom Grad d, wenn jedes Monom Grad d hat.
[5] Durch die Homogenität impliziert $f(x,y,z) = 0$ für $t \in K$ auch $f(tx,ty,tz) = 0$.

besitzt also insbesondere eine Gruppenstruktur, so bleibt diese bei der Identifikation erhalten. Also liefert in diesem Fall die Identifikation der Punkte sogar einen Isomorphismus zwischen $C(K)$ und $C'(K)$.[6]

An dieser Stelle können wir uns auch noch davon überzeugen, dass eine elliptische Kurve in Normalform genau einen Fernpunkt besitzt. Sei $C: y^2 = x^3 + bx + c$ eine elliptische Kurve in Normalform. Ihre Homogenisierung ist dann gegeben durch $C': y^2z = x^3 + bxz^2 + cz^3$. Bestimmen der Fernpunkte durch Setzen von $z = 0$ liefert $0 = x^3$, also auch $x = 0$. Somit ist $(0,1,0) \in \mathbb{P}^2(K)$ der einzige Fernpunkt.

Definition: Sei K ein Körper und $f \in K[x, y, z]$ ein homogenes Polynom. Jede Änderung der Koordinaten der Form $(x', y', z')^T = M(x, y, z)^T$, mit einer (über K) invertierbaren Matrix $M \in K^{3\times3}$, heißt *projektive (Koordinaten-) Transformation* der Kurve $C: f(x, y, z) = 0$ über $\mathbb{P}^2(K)$ in die projektive Kurve $C': f'(x, y, z) = 0$, wobei $f(x, y, z) = f'(x', y', z')$.

Bemerkung: Da $(x, y, z)^T \mapsto M(x, y, z)^T$ ein Isomorphismus ist, bewahrt man durch eine projektive Transformation die Gruppenstruktur und kann durch die Inverse Transformation ohne Probleme zwischen C und C' hin und her wechseln. Also ist das Problem Punkte auf $C(K)$ zu finden äquivalent mit dem, Punkte auf $C'(K)$ zu finden.

Nun wenden wir uns speziell $\mathbb{P}^2(\mathbb{Q})$ und $\mathbb{P}^2(\mathbb{Z}_p)$ für eine Primzahl $p > 3$ zu. p bezeichne ab nun für den Rest des Kapitels so eine beliebige, aber fest gewählte Primzahl. Außerdem sei für eine ganze Zahl a mit $\bar{a}$ stets die Reduktion von $a\ modulo\ p$ gemeint.

Sei $P = \left(\frac{p}{q}, \frac{r}{s}, \frac{t}{u}\right) \in \mathbb{P}^2(\mathbb{Q})$. Multiplizieren mit qsu ist in $\mathbb{P}^2(\mathbb{Q})$ erlaubt und liefert eine Darstellung $P = (x, y, z)$ mit ganzzahligen Koordinaten für P. Nun dürfen wir weiter durch $ggT(x, y, z)$ dividieren, und erhalten $P = (a, b, c)$ mit teilerfremden, ganzen Koordinaten. Eine solche Darstellung ist bis auf das Vorzeichen eindeutig und heißt *normalisiertes Koordinatentripel*.

Wir betrachten nun die Abbildung

$$p^*: \begin{cases} \mathbb{P}^2(\mathbb{Q}) \to \mathbb{P}^2(\mathbb{Z}_p) \\ (a, b, c) \mapsto (\bar{a}, \bar{b}, \bar{c}) \end{cases}$$

wobei (a, b, c) als normalisiertes Koordinatentripel vorliegt. Wir schreiben kurz $\tilde{P}$ für $p^*(P)$ und nennen $\tilde{P}$ die Reduktion modulo p von P.

Für eine projektive kubische Kurve $C: f(x, y, z) = 0$ mit ganzzahligen Koeffizienten können wir nun genauso vorgehen. Durch Dividieren der Koeffizienten von f durch deren größten gemeinsamen Teiler erhalten wir ein Polynom mit ganzzahligen, teilerfremden Koeffizienten, welches die gleiche Nullstellenmenge besitzt wie das ursprüngliche Polynom. Wir können also ohne Einschränkung annehmen, dass f von dieser Form ist. Wir sagen C ist dann

[6] Dadurch, dass wir bei der Definition der Kurve die Fernpunkte bereits hinzugefügt haben, haben wir die Kurve letztendlich schon „projektiv gemacht“, und betrachten einfach nur die Repräsentanten mit $z = 1$.

normalisiert. Liegt C in normalisierter Form vor, so können wir die Koeffizienten von f modulo p zu einem Polynom $\tilde{f}$ reduzieren, und erhalten dadurch eine reduzierte, projektive Kurve $\tilde{C}$: $\tilde{f}(x,y,z)=0$. Ist (a,b,c) eine Nullstelle von f, so ist, da die Reduktion modulo p ein Homomorphismus ist, $(\bar{a},\bar{b},\bar{c})$ eine Nullstelle von $\tilde{f}$, oder anders gesehen, ist $P \in C(\mathbb{Q})$, so ist $\tilde{P} \in \tilde{C}(\mathbb{Z}_p)$. Damit gilt aber auch für die Schnittpunkte von C mit einer Geraden G, dass $\widetilde{(C(\mathbb{Q}) \cap G(\mathbb{Q}))} \subseteq \tilde{C}(\mathbb{Z}_p) \cap \tilde{G}(\mathbb{Z}_p)$.

Ohne Beweis verwenden wir noch folgendes wichtige Resultat über die Anzahl von Schnittpunkten projektiver Kurven aus der projektiven Geometrie.

Satz 9 (Satz von Bezout für $\mathbb{C}$): *Seien $f,g \in \mathbb{C}[x,y,z]$ nicht konstante, homogene, teilerfremde Polynome. Dann ist die Anzahl der Schnittpunkte, gezählt mit eventuellen Vielfachheiten, der projektiven Kurven C_1: $f(x,y,z)=0$ und C_2: $g(x,y,z)=0$ über $\mathbb{P}^2(\mathbb{C})$ gegeben durch das Produkt der Grade der Polynome $deg\,(f)\cdot deg(g)$.*

Dieser Satz stellt sicher, dass sich eine Gerade und eine elliptische Kurve über $\mathbb{P}^2(\mathbb{C})$, die durch teilerfremde Polynome gegeben sind, in genau drei (nicht notwendigerweise verschiedenen) Punkten schneiden.

Nun sind wir in der Lage, zwei für den Beweis von Satz 7 notwendige Lemma zu beweisen.

Lemma 10: *Sei C: $f(x,y,z)=0$ eine homogene Kubik mit ganzzahligen Koeffizienten in der projektiven Ebene $\mathbb{P}^2$ und G: $g(x,y,z)=z=0$ die unendlich ferne Gerade. Weiterhin seien alle drei (nicht notwendigerweise verschiedenen) Schnittpunkte P_1, P_2 und P_3 von C und G rational (Satz 9 garantiert uns die Existenz der Schnittpunkte). Sei weiterhin $\tilde{G}$ keine Komponente von $\tilde{C}$ ($\tilde{g}$ kein Teiler von $\tilde{f}$). Dann schneiden sich $\tilde{C}$ und $\tilde{G}$ in den drei (nicht notwendigerweise verschiedenen) Punkten$\widetilde{P_1}$, $\widetilde{P_2}$ und $\widetilde{P_3}$.*

Beweis: Seien G und C: $f(x,y,z)=0$ so, dass sie die Voraussetzungen erfüllen. Ohne Einschränkung liege C in normalisierter Form vor, f habe also ganzzahlige, teilerfremde Koeffizienten. Die Ferngerade G ist gegeben durch $z=0$, also ist $C \cap G$ gegeben durch die Nullstellenmenge von $f(x,y,0)$. Schreiben wir $P_i=(x_i,y_i,0)$ mit normalisierten Koordinaten, so zerfällt $f(x,y,0)$ in die Linearfaktoren $f(x,y,0)=c(\mathrm{y_1x}-\mathrm{x_1y})(\mathrm{y_2x}-\mathrm{x_2y})\,(\mathrm{y_3x}-\mathrm{x_3y})$ mit $c\neq 0$. Dabei ist c nicht durch p teilbar, da laut Voraussetzung $\tilde{g}$ kein Teiler von $\tilde{f}$ ist. Es gilt also $\tilde{f}(x,y,0)=\tilde{c}(\tilde{\mathrm{y}}_1\mathrm{x}-\tilde{\mathrm{x}}_1\mathrm{y})(\tilde{\mathrm{y}}_2\mathrm{x}-\tilde{\mathrm{x}}_2\mathrm{y})(\tilde{\mathrm{y}}_3\mathrm{x}-\tilde{\mathrm{x}}_3\mathrm{y})$. Da die Punkte P_i normalisiert waren, teilt p nicht gleichzeitig x_i und y_i, und damit ist $\tilde{f}(x,y,0)\not\equiv 0$, da p nicht c teilt. Das bedeutet aber die $\widetilde{P_i}$ sind die drei Nullstellen von $\tilde{f}(x,y,0)$, also die drei Schnittpunkte von $\tilde{C}$ und $\tilde{L}$. ■

Lemma 11: *Sei G: $ax+by+cz=0$ eine projektive Gerade mit $a,b,c\in\mathbb{Z}$. Dann gibt es eine ganzzahlige Koordinatentransformation (so, dass auch die inverse Transformation ganzzahlig ist), die G auf die unendlich ferne Gerade abbildet.*

Beweis: Sei $G\colon ax + by + cz = 0$ wie im Lemma und $L\colon z = 0$ die Ferngerade. Sei weiterhin $ggT(a, b, c) = 1$. Dann existieren teilerfremde, ganze Zahlen r und s, sodass $rc - sb = g \coloneqq ggT(b, c)$ gilt. Da a und g teilerfremd sind, existieren nun teilerfremde t und u, sodass $tg + ua = 1$ gilt. Letztendlich können wir wegen $ggT(r, s) = 1$ noch v und w so wählen, dass $vs - wr = u$. Betrachten wir nun die Matrix

$$M = \begin{pmatrix} t & v & w \\ 0 & r & s \\ a & b & c \end{pmatrix}$$

Wegen $\det(M) = trc + vsa - arw - bst = t(rc - bs) + a(vs - rw) = tg + au = 1$ ist M über $\mathbb{Z}$ invertierbar. Also stellt M eine zulässige Koordinatentransformation dar. Ferner gilt dann

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = M \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} tx + vy + wz \\ ry + sz \\ ax + by + cz \end{pmatrix}$$

und damit $f(x, y, z) = ax + by + cz = z' = f'(x', y', z')$, also ist $C' = L$. ■

Korollar 12: *Durch Lemma 11 gilt Lemma 10 auch für eine beliebige projektive Gerade G, die nicht die unendlich ferne Gerade ist, sofern ihre Reduktion keine Komponente der reduzierten Kurve ist.*

Beweis: Sei M die Koordinatentransformation nach Lemma 11, welche G auf die Ferngerade transformiert. Reduzieren der Einträge von M modulo p liefert die entsprechende Matrix für die Koordinatentransformation $\overline{M}$ für $\mathbb{P}^2(\mathbb{Z}_p)$. Es gilt $\det(\overline{M}) = \overline{\det(M)} \neq 0_{\mathbb{Z}_p}$, da M über $\mathbb{Z}$ invertierbar ist, also $\det(M) \in \{-1, 1\}$. Also ist $\overline{M}$ über $\mathbb{Z}_p$ invertierbar. Und nachdem die Reduktion ein Homomorphismus ist, ist es egal, ob wir erst die Koordinaten ändern und dann reduzieren, oder erst reduzieren, und dann die Koordinaten ändern. Wir können also die Gerade mittels Änderung der Koordinaten in die Ferngerade transformieren, dann Lemma 10 anwenden um die Schnittpunkte mit der (ebenfalls in neue Koordinaten transformierten) Kurve zu bestimmen, da die Voraussetzung impliziert, dass die Ferngerade keine Komponente der transformierten Kurve ist. Danach können wir die Schnittpunkte mit der inversen, reduzierten Koordinatenänderung $\overline{M}^{-1}$ zurück transformieren, was uns das gewünschte Ergebnis liefert. Das funktioniert, da eine invertierbare Koordinatentransformation als lineare Transformation den Grad der Kurve nicht ändert, also weiterhin eine Kubik mit einer Geraden (jetzt der Ferngeraden) geschnitten wird. ■

Bemerkung: Da eine projektive elliptische Kurve stets irreduzibel ist (anderenfalls gäbe es singuläre Punkte), kann keine Gerade eine Komponente einer solchen Kurve sein. Somit ist Lemma 10 für projektive elliptische Kurven und beliebige projektive Geraden anwendbar.

Damit haben wir nun endlich das nötige Werkzeug, um Satz 7 zu beweisen.

Beweis von Satz 7: Sei C' eine ganzzahlige elliptische Kurve in kurzer Normalform in $\mathbb{P}^2$, sodass $\widetilde{C'}$ nicht singulär ist. $\mathcal{O}$ beziehungsweise $\tilde{\mathcal{O}}$ bezeichnen die jeweiligen Fernpunkte (wir wissen, es gibt jeweils nur einen), die wir auch als neutrale Elemente der Gruppen nehmen. Wir erinnern uns an die geometrische Definition des Additionsgesetzes für Kurven. Zwei Punkte P und Q der Kurve C' werden addiert, indem man die Gerade durch P und Q mit C' schneidet, was einen dritten Schnittpunkt (siehe Satz von Bezout) R mit der Kurve liefert. $P+Q$ ist dann definiert als der dritte Schnittpunkt S der Geraden durch R und $\mathcal{O}$ mit der Kurve C'. Lemma 10 und 11 besagen nun, dass $\tilde{R}$ der dritte Schnittpunkt der Geraden durch $\tilde{P}$ und $\tilde{Q}$ mit $\widetilde{C'}$ ist, und $\tilde{S}$ der dritte Schnittpunkt der Geraden durch $\tilde{\mathcal{O}}$ und $\tilde{R}$ mit $\widetilde{C'}$. In anderen Worten formuliert, $\widetilde{P+Q} = \tilde{S} = \tilde{P} + \tilde{Q}$, also ist π, die Reduktion modulo p, ein Gruppenhomomorphismus zwischen $C'(\mathbb{Q})$ und $\widetilde{C'}(\mathbb{Z}_p)$. Als letzten Schritt vergewissern wir uns jetzt noch, dass φ: $C(\mathbb{Q}) \to \tilde{C}(\mathbb{Z}_p)$ aus Satz 7 gegeben ist durch $\varphi = \tilde{\psi}^{-1} \circ \pi \circ \psi$, wobei ψ beziehungsweise $\tilde{\psi}$ die kanonischen Isomorphismen von $C(\mathbb{Q})$ nach $C'(\mathbb{Q})$, beziehungsweise von $\tilde{C}(\mathbb{Z}_p)$ nach $\widetilde{C'}(\mathbb{Z}_p)$ sind. Ist das gezeigt, so ist φ als Komposition von Homomorphismen selbst ein Homomorphismus.

Sei $P = \mathcal{O}$. Dann gilt:

$$\tilde{\psi}^{-1}\left(\pi\big(\psi(P)\big)\right) = \tilde{\psi}^{-1}\left(\pi\big(\psi(\mathcal{O})\big)\right) = \tilde{\psi}^{-1}\left(\pi\big((0{,}1{,}0)\big)\right) = \tilde{\psi}^{-1}\big((0{,}1{,}0)\big) = \tilde{\mathcal{O}} = \varphi(\mathcal{O})$$

Sei $P \neq \mathcal{O}$, so hat P nach Lemma 8 eine Darstellung $P = \left(\frac{a}{s^2}, \frac{b}{s^3}\right)$ in reduzierter Bruchdarstellung. Sei P nun von dieser Form.

Gilt $p \nmid s$, dann erhalten wir:

$$\begin{aligned}\tilde{\psi}^{-1}\left(\pi\big(\psi(P)\big)\right) &= \tilde{\psi}^{-1}\left(\pi\left(\psi\left(\left(\frac{a}{s^2}, \frac{b}{s^3}\right)\right)\right)\right) = \tilde{\psi}^{-1}\left(\pi\left(\left(\frac{a}{s^2}, \frac{b}{s^3}, 1\right)\right)\right) = \\ &= \tilde{\psi}^{-1}\left(\pi\big((as, b, s^3)\big)\right) = \tilde{\psi}^{-1}\left(\left(\overline{a}\bar{s}, \bar{b}, \bar{s}^3\right)\right) = \tilde{\psi}^{-1}\left(\left(\overline{a}\bar{s}^{-2}, \bar{b}\bar{s}^{-3}, 1\right)\right) = \\ &= \left(\overline{a}\bar{s}^{-2}, \bar{b}\bar{s}^{-3}\right) = \varphi\left(\left(\frac{a}{s^2}, \frac{b}{s^3}\right)\right)\end{aligned}$$

Gilt andererseits $p|s$, so erhalten wir:

$$\begin{aligned}\tilde{\psi}^{-1}\left(\pi\big(\psi(P)\big)\right) &= \tilde{\psi}^{-1}\left(\pi\left(\psi\left(\left(\frac{a}{s^2}, \frac{b}{s^3}\right)\right)\right)\right) = \tilde{\psi}^{-1}\left(\pi\left(\left(\frac{a}{s^2}, \frac{b}{s^3}, 1\right)\right)\right) = \\ &= \tilde{\psi}^{-1}\left(\pi\big((as, b, s^3)\big)\right) = \tilde{\psi}^{-1}\left(\left(\overline{a}\bar{s}, \bar{b}, \bar{s}^3\right)\right) = \tilde{\psi}^{-1}\big((0{,}1{,}0)\big) = \tilde{\mathcal{O}} = \\ &= \varphi\left(\left(\frac{a}{s^2}, \frac{b}{s^3}\right)\right)\end{aligned}$$

Es gilt also wie behauptet $\varphi = \tilde{\psi}^{-1} \circ \pi \circ \psi$. ■

Die Pollard $(p-1)$-Methode

Nehmen wir an, eine natürliche Zahl n hat eine Zerlegung $n = pq$ mit $p \in \mathbb{P}$, die uns allerdings noch unbekannt ist. Dann gilt aufgrund des kleinen Satzes von Fermat, dass $a^{p-1} \equiv 1\ (mod\ p)$ für jede ganze Zahl a, die zu p teilerfremd ist. Folglich gilt auch für jedes Vielfache k von $p-1$, dass $a^k \equiv 1\ (mod\ p)$, also dass $a^k - 1 = mp$ für ein $m \in \mathbb{N}_0$. Somit teilt aber p die Zahl $G \coloneqq ggT(a^k - 1, n) = ggT(mp, pq)$, also ist G ein echter Teiler von n, falls $G \neq n$, sprich wenn m kein Vielfaches von q ist. Diese Idee führt uns zu folgendem Algorithmus zur Faktorisierung natürlicher Zahlen.

Algorithmus von Pollard:

Sei $n \geq 2$ eine zusammengesetzte, natürliche Zahl, für die wir eine Faktorisierung finden sollen.

Schritt 1: Wähle eine Zahl k, welche das Produkt von Primzahlpotenzen mit kleinen Primzahlen und kleinen Exponenten ist, zum Beispiel $k = kgV(1,2,\dots,K)$ für ein natürliches K.

Schritt 2: Wähle eine zufällige, natürliche Zahl a mit $1 < a < n$.

Schritt 3: Berechne $ggT(a,n)$. Ist dieser echt größer als 1, so haben wir einen echten Teiler von n gefunden und sind fertig. Ansonsten gehe zu Schritt 4.

Schritt 4: Berechne $G \coloneqq ggT(a^k - 1, n)$. Gilt $1 < G < n$, so haben wir einen echten Teiler von n gefunden, und sind fertig. Ist $G = 1$, gehe zurück zu Schritt 1 und wähle ein größeres k. Ist $G = n$, gehe zurück zu Schritt 2 und wähle ein anderes a.

Analyse des Pollard Algorithmus

Wieso funktioniert dieser Algorithmus nun?

Haben wir Glück, und das gewählte k ist für einen Primteiler p von n ein Vielfaches von $p-1$, so ist für jedes a, das wir uns im zweiten Schritt wählen können, aufgrund unserer vorherigen Überlegungen G entweder ein echter Teiler von n oder n selbst. Im einen Falle sind wir schon fertig, im anderen variieren wir a, was uns auf Dauer auch ans Ziel führt, wenn Schritt 4 keine Teiler liefern sollte, da a irgendwann einen Teiler von n durchläuft. Gilt allerdings $G = 1$, so kann unser k nicht die gewünschte Form haben, also vergrößern wir es so lange, bis es die Form hat.

Dabei sehen wir schon, wann der Algorithmus in der Regel gut funktioniert, nämlich dann, wenn n einen Primfaktor p besitzt, so dass $p-1$ nur kleine Primfaktoren hat. Denn so erreichen wir in Schritt 1 relativ schnell ein Vielfaches von $p-1$, und erhalten so in Schritt 4, wenn wir nicht gerade viel Pech haben, einen echten Teiler von n. Besitzt n allerdings keinen derartigen Teiler p, so funktioniert der Algorithmus zwar, kommt aber eher einem Probieren gleich, und ist damit nicht gerade effektiv.

Außerdem müssen wir uns noch überlegen, ob die verwendeten Berechnungen schnell durchführbar sind. Den größten gemeinsamen Teiler zweier Zahlen a und b können wir bekanntlich mit Hilfe des Euklidischen Algorithmus in höchstens $2\log_2 \max\{2a, 2b\}$ Operationen, welche jeweils eine Division mit Rest beinhaltet, berechnen.

Die andere Rechenoperation, die wir im Laufe des Algorithmus durchführen müssen, ist die Berechnung von a^k, beziehungsweise noch leichter, da wir $ggT(a^k - 1, n)$ berechnen wollen, genügt es uns $a^k \ (mod\ n)$ zu bestimmen. Natürlich könnten wir dies durch sukzessive, k-fache Multiplikation mit a und anschließender Reduktion $(mod\ n)$ durchführen, aber auch hier gibt es eine effektivere Methode, wie folgendes Lemma zeigt:

Lemma 13: *Seien $a, k, n \in \mathbb{N}$. Dann lässt sich $a^k \ (mod\ n)$ in höchstens $2\ log_2\ k$ Operationen berechnen, wobei jede Operation aus einer Multiplikation und einer Reduktion $(mod\ n)$ besteht.*

Beweis: Sei $k = \sum_{i=0}^{r} k_i 2^i$ mit $k_i \in \{0,1\}$ für $0 \leq i < r$ und $k_r = 1$ die binäre Entwicklung[7] von k. Dann ist $a^k = \prod_{i=0}^{r} a^{k_i 2^i}$. Bei Kenntnis von $A_j \coloneqq a^{2^j} \ (mod\ n)$ kann $A_{j+1} = a^{2^{j+1}} \ (mod\ n) = \left(a^{2^j}\right)^2 \ (mod\ n)$ mit nur einer Multiplikation und anschließender Reduktion $(mod\ n)$ berechnet werden. Wir benötigen so also höchstens r Operationen um alle A_j zu berechnen, und da höchstens r+1 Faktoren A_j in der Entwicklung von a^k vorkommen, noch höchstens r zusätzliche Operationen, also höchstens $2r$ Operationen zur Berechnung von a^k. Und wegen $2^r \leq k$, also $r \leq \log_2 k$ sind das höchstens $2\log_2 k$. ■

Nun wollen wir den Algorithmus an einem kleinen Beispiel ausprobieren.

Beispiel: Gesucht ist eine Faktorisierung für $n = 6887$.

Als erstes stellen wir wegen $2^{n-1} \equiv 800 \ (mod\ n)$ fest, dass n keine Primzahl ist.

Für den Anfang setzen wir zunächst einmal $k \coloneqq kgV(1,2,3,4,5) = 60$ und $a \coloneqq 2$. Da a und n teilerfremd sind, müssen wir nun $ggT(a^k - 1, n)$ berechnen. Dazu bestimmen wir $a^k \ (mod\ n)$ nach unserem eben hergeleiteten Schema:

Entwickeln wir zunächst k binär, so erhalten wir:

$$k = 60 = 2^5 + 2^4 + 2^3 + 2^2$$

und für die A_i und a^k somit

$$A_0 \equiv 2,\ \ A_1 \equiv 4,\ \ A_2 \equiv 16,\ \ A_3 \equiv 256,\ \ A_4 \equiv 3553,\ \ A_5 \equiv 6825 \ \ \ (mod\ n)$$

$$a^k = A_2 * A_3 * A_4 * A_5 \equiv 1962\ (mod\ n)$$

Leider ist $ggT(1962 - 1,6887) = 1$, also war unser k zu klein und wir müssen ein größeres wählen.

[7] Die Binärentwickling einer Zahl kann effektiv bestimmt werden.

Probieren wir $k = kgV(1,2,3,4,5,6,7) = 420$, so erhalten wir:

$$k = 420 = 2^8 + 2^7 + 2^5 + 2^2$$

$$A_6 \equiv 3844, \;\; A_7 \equiv 3721, \;\; A_8 \equiv 2971 \;\; (mod\ n)$$

$$a^k = A_2 * A_5 * A_7 * A_8 \equiv 1918\ (mod\ n)$$

Mit $ggT(1918 - 1,6887) = 71$ finden wir nun die Faktorisierung

$$n = 6887 = 71 * 97$$

Elliptische Kurven Algorithmus von Lenstra

Erinnern wir uns kurz an die Funktionsweise des Algorithmus von Pollard: die Grundidee war es ja, uns den kleinen Satz von Fermat zu Nutze zu machen, um so auf einen Teiler von n zu kommen. Der theoretische Hintergrund dieses Satzes ist aber wiederum die Tatsache, dass in einer endlichen Gruppe G die Ordnung eines Elements $g \in G$ die Gruppenordnung $|G|$ teilt, und somit $g^{|G|} = 1_G$ in einer multiplikativen Gruppe gilt, beziehungsweise $|G| * g = 0_G$ in einer additiven Gruppe. Die Idee für den folgenden Algorithmus ist es, anstelle der Einheitengruppe von $\mathbb{Z}_p$ wie bei Pollard, nun die Gruppe der Punkte auf einer elliptischen Kurve $\tilde{C}(\mathbb{Z}_p)$ über $\mathbb{Z}_p$ zu betrachten.

Algorithmus von Lenstra:

Sei $n \geq 2$ eine zusammengesetzte, natürliche Zahl, für die wir eine Faktorisierung finden sollen.

Schritt 1: Überprüfe, ob $ggT(n, 6) = 1$ und ob $n \neq m^r$ für alle $m \in \mathbb{N}$ und $r \in \mathbb{N}_{\geq 2}$. Ist eines davon nicht der Fall, so haben wir einen echten Teiler gefunden und wir sind fertig.

Schritt 2: Wähle beliebige ganze Zahlen b, x_1, y_1 aus $\{1, 2, \dots, n\}$.

Schritt 3: Setze $c \coloneqq y_1^2 - x_1^3 - bx_1 \ (mod\ n)$ und C die kubische Kurve, geben durch $C: y^2 = x^3 + bx + c$, und sei $P = (x_1, y_1)$.

Schritt 4: Überprüfe, ob $ggT(4b^3 + 27c^2, n) = 1$. Ist das nicht der Fall, so haben wir entweder einen echten Teiler von n gefunden, oder das Ergebnis ist n. Ist letzeres der Fall, gehe zurück zu Schritt 3 und wähle ein anderes b.

Schritt 5: Wähle eine Zahl k, welche das Produkt von Primzahlpotenzen mit kleinen Primzahlen und kleinen Exponenten ist, zum Beispiel $k = kgV(1,2, \dots, K)$ für ein natürliches K.

Schritt 6: Berechne $kP = \left(\frac{a_k}{d_k^2}, \frac{b_k}{d_k^3}\right)$.

Schritt 7: Berechne $G = ggT(d_k, n)$. Gilt $G = 1$, so gehe entweder zu Schritt 5 zurück, und vergrößere k, oder gehe zu Schritt 2 zurück und wähle eine andere Kurve. Gilt $G = n$, so gehe zurück zu Schritt 5 und verkleinere k. Ansonsten haben wir einen echten Teiler von n gefunden und sind fertig.

Analyse des Algorithmus:

Nun wollen wir den Algorithmus analysieren.

Zunächst einmal überlegen wir uns, dass der Algorithmus tatsächlich funktioniert. Im ersten Schritt schließen wir aus, dass 2 oder 3 ein Teiler von n ist, und dass n eine Potenz einer natürlichen Zahl ist, oder finden anderenfalls einen echten Teiler. Dass 2 oder 3 ein Teiler

von n ist müssen wir deshalb ausschließen, da wir bei der Betrachtung elliptischer Kurven über $\mathbb{Z}_p$ stets $p \neq 2,3$ vorausgesetzt haben. Auszuschließen, dass n eine perfekte Potenz ist macht deswegen Sinn, da bei der Berechnung von Summen und Vielfachen von Punkten auf elliptischen Kurven Potenzen im Nenner auftreten. In den Schritten 2 und 3 generieren wir uns einen Punkt P und eine Kurve C in verkürzter Weierstraß Normalform, sodass P offensichtlich für alle Primteiler p von n auf der $(mod\ p)$ reduzierten Kurve $\tilde{C}(\mathbb{Z}_p)$ liegt. Ist nun der in Schritt 4 berechnete größte gemeinsame Teiler von der Diskriminante der Kurve C und n gleich 1, so haben wir, da 2 und 3 keine Teiler von n sind, für jeden Primteiler p von n sichergestellt, dass $\tilde{C}(\mathbb{Z}_p)$ nicht singulär, also eine elliptische Kurve ist. Ist deren ggT allerdings n, so wählen wir eine andere Kurve durch den selben Punkt P, und liegt er zwischen 1 und n, haben wir einen echten Teiler gefunden und sind fertig. In Schritt 5 wählen wir nun k als ein Produkt von „kleinen“ Primzahlpotenzen, genauso wie im Algorithmus von Pollard. Dazu sollte K anfangs relativ klein gewählt werden, und dann nach und nach vergrößert werden. Haben wir nun Glück, so teilt $|\tilde{C}(\mathbb{Z}_p)|$ unser k für einen Primfaktor p von n, und somit ist $\widetilde{kP} = k\tilde{P} = \tilde{\mathcal{O}}$ der Punkt im Unendlichen in $\tilde{C}(\mathbb{Z}_p)$. Das heißt aber wiederum für $kP = \left(\frac{a_k}{d_k^2}, \frac{b_k}{d_k^3}\right)$, dass p den Nenner d_k teilt (siehe Satz 7). In diesem Falle ist das in Schritt 7 berechnete G also entweder ein echter Teiler von n oder gleich n. Ist letzteres der Fall, so kann das daran liegen, dass wir unser k zu groß gewählt haben, um einen echten Teiler zu finden, also müssen wir es verkleinern. Im schlimmsten Fall kann es aber auch daran liegen, dass P bei Reduktion modulo beider Teiler von n die gleiche Ordnung hat, und wir wären besser beraten, eine andere Kurve zu wählen. Ist G aber gleich 1, so kann unser k für unsere Kurve nicht die gewünschte Eigenschaft besessen haben, also vergrößern wir es entweder, oder wählen eine andere Kurve. Der Algorithmus führt irgendwann zum Ziel, selbst wenn wir keine geeignete Kurve mit passendem k finden, da der ggT in Schritt 4 früher oder später einen Teiler von n durchläuft. Beispielsweise gilt bei $n = pq$ für $b = p,\ x_1 = y_1 = n$ dann $ggT(4b^3 + 27c^2, n) \in \{p, p^2, p^3\}$.

Wir sehen also, dass die Grundidee im Prinzip die selbe ist wie die des Algorithmus von Pollard. Gleichzeitig sehen wir aber auch den eindeutigen Vorteil der neuen Methode. War bei Pollard $p - 1$ kein Produkt aus kleinen Primfaktoren, so waren wir machtlos. Jetzt haben wir aber die Möglichkeit, wenn $|\tilde{C}(\mathbb{Z}_p)|$ für keinen Primteiler p von n nur kleine Primfaktoren besitzt, einfach eine andere Kurve zu wählen, die dann möglicherweise diese Eigenschaft hat. Nach dem Satz von Hasse-Weil gilt nun für eine nicht singuläre elliptische Kurve, dass $\left|\tilde{C}(\mathbb{Z}_p)\right| = p + 1 - \varepsilon_p$ mit $\left|\varepsilon_p\right| \leq 2\sqrt{p}$. Weiterhin kann gezeigt werden, dass wenn C alle solche Kurven durchläuft, ε_p recht gut über sein gegebenes Intervall verteilt ist. Also stehen unsere Chancen gut, dass wir relativ bald auf eine Kurve C treffen, für die $\left|\tilde{C}(\mathbb{Z}_p)\right|$ nur kleine Primfaktoren besitzt.[8]

Bleibt uns nur noch zu untersuchen, wie gut Schritt 6, also die Berechnung von kP durchzuführen ist. Die naheliegende Variante, einfach die k-fache Summe zu berechnen, ist

[8] Die Fragen „wie gut?“ und „wie bald?“ bleiben an dieser Stelle leider offen.

leider nicht sehr effektiv, allerdings gibt es einen besseren Weg, ähnlich der Berechnung von a^k, wie folgendes Resultat zeigt.

Lemma 14: *Sei $k \in \mathbb{N}$ und P ein Punkt auf einer Kurve C. Dann lässt sich kP in höchstens $2\,log_2\,k$ Schritten, bestehend aus einer Addition oder Verdoppelung von Punkten, berechnen.*

Beweis: Sei $k = \sum_{i=0}^{r} k_i 2^i$ mit $k_i \in \{0,1\}$ für $0 \leq i < r$ und $k_r = 1$ die binäre Entwicklung von k. Somit ist $kP = \sum_{i=0}^{r} k_i 2^i P = \sum_{i=0}^{r} k_i P_i$ mit $P_i = 2^i P$. Bei Kenntnis von P_j können wir $P_{j+1} = 2^{j+1}P = 2P_j$ mit einer Verdoppelung des Punktes P_j berechnen. Wir müssen r solcher Punkte P_j berechnen ($P_0 = P$ kennen wir ja schon), und höchstens $r+1$ davon aufaddieren um P zu erhalten, also benötigen wir höchstens $2r \leq 2\log_2 k$ Schritte. ■

Bemerkung: Zu beachten ist aber, dass wir kP nicht mit rationalen Koordinaten berechnen wollen (P liegt ja im Allgemeinen nicht einmal auf $C(\mathbb{Q})$, sondern auf der reduzierten Kurve!). Eigentlich würden wir die Berechnungen ja gerne in $\mathbb{Z}_p$ durchführen, wobei p ein zu findender Primteiler von n ist. Da wir diesen aber noch nicht kennen, funktioniert das leider nicht. Deswegen führen wir alle Berechnungen $(mod\ n)$ durch, und bewegen uns damit sozusagen zwischen dem noch unbekannten $\mathbb{Z}_p$ und dem für unsere Zwecke zu großen Körper $\mathbb{Q}$. Das stellt auch nicht weiter ein Problem dar, da eine Reduktion modulo einem Vielfachen von p, in unserem Fall also n, eine spätere Reduktion modulo p nicht beeinflusst. Anders gesprochen könnten wir auch sagen, wir verwenden einfach andere Repräsentanten über $\mathbb{Z}_p$. Satz 7 stellt sicher, dass die Reduktion modulo p ein Homomorphismus ist, und erlaubt uns so das eben beschriebene Vorgehen.

Allerdings ist $\mathbb{Z}_n$ kein Körper, und somit können wir nicht jedes Element invertieren. Wie führen wir also die Berechnungen durch? Seien $P_1 = (x_1, y_1)$ und $P_2 = (x_2, y_2)$ zwei verschiedene Punkte auf der Kurve aus dem Algorithmus mit Koordinaten in $\mathbb{Z}_n$. Nach den Formeln für die Addition ist $P_3 = P_1 + P_2 = (x_3, y_3)$ gegeben durch

$$x_3 = \lambda^2 - x_1 - x_2 \text{ und } y_3 = -\lambda x_3 - (y_1 - \lambda x_1) \text{ mit } \lambda = \frac{y_2 - y_1}{x_2 - x_1}.$$

Ob wir die Berechnung $(mod\ n)$ durchführen können hängt also davon ab, ob $(x_2 - x_1)$ ein Inverses in $\mathbb{Z}_n$ besitzt. Dabei unterscheiden wir nun drei Fälle.

<u>Fall 1</u>: $ggT(x_2 - x_1, n) = 1$

In diesem Fall besitzt $x_2 - x_1$ ein Inverses in $\mathbb{Z}_n$ und wir können die Berechnung durchführen.[9]

<u>Fall 2</u>: $1 < ggT(x_2 - x_1, n) < n$

In diesem Fall haben wir einen echten Teiler von n gefunden und haben unser Ziel erreicht, auch wenn wir P_3 nicht berechnen können.

[9] Der erweiterte Euklidische Algorithmus liefert uns hier schnell das Inverse.

Fall 3: $ggT(x_2 - x_1, n) = n$

Das ist der unglücklichste Fall für uns. Passiert uns das im Laufe von Schritt 7 einmal, so gehen wir am besten zurück zu Schritt 5 und verkleinern k, oder zu Schritt 2 und wählen eine andere Kurve.

Analog müssen wir, um einen Punkt $P = (x, y)$ über $\mathbb{Z}_n$ zu verdoppeln, den Quotienten

$$\lambda = \frac{f'(x)}{2y} = \frac{3x^2 + b}{2y} \ (mod\ n)$$

berechnen und stoßen dabei auf die gleichen drei Fälle: entweder wir können y in $\mathbb{Z}_n$ invertieren, finden einen echten Teiler von n, oder müssen k verkleinern beziehungsweise eine andere Kurve wählen.

Wollen wir uns den Algorithmus nun an einem konkreten Beispiel veranschaulichen, um seine Funktionsweise in der Praxis zu sehen.

Beispiel: Gesucht ist eine Faktorisierung für $n = 10057$.

Als erstes stellen wir wegen $2^{n-1} \equiv 4988\ (mod\ n)$ fest, dass n keine Primzahl ist. Danach verifizieren wir $ggT(6, n) = 1$, und, dass n keine perfekte Potenz ist, indem wir bei 1 beginnend solange natürliche Wurzeln aus n ziehen, bis diese zum ersten Mal kleiner als 5 sind, wobei keine davon ganzzahlig ist.

Nun wählen wir uns willkürlich $x_1 \coloneqq 45, y_1 \coloneqq 863, b \coloneqq 1355$, berechnen daraus $c = 9363$ und erhalten damit unsere Kurve

$$C: y^2 = x^3 + 1355x + 9363$$

und den Punkt $P = (45{,}863)$ auf der Kurve. Wir überprüfen noch, dass $ggT(4 * 1355^2 + 27*93632,\ n) = 1$ gilt, und haben damit sichergestellt, dass die reduzierte Kurve $\mathcal{C}$ für jeden Primteiler von n nicht singulär ist. Für k wählen wir für den Anfang den Wert $kgV(1, 2, \dots, 7) = 420$. Zur Berechnung von kP entwickeln wir wieder zuerst k binär:

$$k = 420 = 2^8 + 2^7 + 2^5 + 2^2$$

Als nächstes berechnen wir die $P_i = 2^i P$, wobei wir gleich $mod\ n$ reduzieren:

$$P_0 = (45, 863), P_1 = (4167, 9035), P_2 = (7386, 7584),$$

$$P_3 = (9189, 7867), P_4 = (9432, 429), P_5 = (2671, 9453),$$

$$P_6 = (2603, 514), P_7 = (9528, 5215), P_8 = (7688, 7151)$$

Jetzt können wir kP berechnen, wiederum gleich $mod\ n$ reduziert:

$$kP = \sum_{i=0}^{8} k_i P_i = (8829, 9923)$$

Wir können also $k\tilde{P}$ tatsächlich berechnen, was uns aber keinen Teiler von n liefert. Die Idee des Algorithmus war es ja gerade, dass wir einen Teiler von n finden, wenn das normale Additionsverfahren zusammenbricht, also wenn wir $k\tilde{P}$ nicht $mod\, n$ berechnen können. Demnach war unser gewähltes k zu klein und wir müssen es vergrößern. Versuchen wir es nun mit:

$$k = kgV(1, 2, \dots, 11) = 27720 = 2^{14} + 2^{13} + 2^{11} + 2^{10} + 2^6 + 2^3$$

Wir müssen also die Liste der P_i bis $i = 14$ erweitern.

$$P_9 = (3933, 2180), P_{10} = (4281, 781), P_{11} = (4264, 1241),$$

$$P_{12} = (8196, 896), P_{13} = (3971, 736), P_{14} = (4697, 9253)$$

Wollen wir nun kP berechnen, so stellen wir fest, dass wir $\sum_{i=0}^{5} k_i P_i = (9189, 767)$ problemlos berechnen können, versuchen wir nun aber $P_6 = (2603, 515)$ dazu zu addieren, so scheitern wie daran, dass wir wegen $ggT(9189 - 2603, n) = 89$ im Ring $\mathbb{Z}_n$ die Zahl $9189 - 2603$ nicht invertieren können, was für die Addition aber notwendig wäre.

Das stört uns aber auch nicht weiter, da wir so die Faktorisierung

$$n = 10057 = 89 * 113$$

gefunden haben.

Literatur

[1] Joseph H. Silverman, John Tate, *Rational Points on Elliptic Curves*, Springer Verlag, 1994

[2] Lawrence C. Washington, *Elliptic Curves – Number Theory and Cryptography*, Chapman & Hall/CRC, 2008

[3] Oleg Bogopolski, Algorithmische *Zahlentheorie mit Anwendungen in der Kryptographie*, Skriptum zur Vorlesung, Sommersemester 2007

[4] Arne Winterhof, *Zahlentheoretische Methoden in der Kryptographie II*, Skriptum zur Vorlesung, Sommersemester 2002

[5] Michael Stoll, *Arithmetik Elliptischer Kurven mit Anwendungen*, Skriptum zur Vorlesung, Sommersemester 2009

[6] Bergit Grußien, *Elliptische Kurven in der Kryptographie*, Skriptum zum Seminarvortrag, Sommersemester 2006

[7] Martin Gubisch, *Elliptische Kurven*, Private Mitschrift zum Seminar Algorithmische Zahlentheorie, Wintersemester 2007

[8] Michael Carter Woodbury, *Finite Groups on Elliptic Curves*, 2003

[9] Prof. Dr. Hans-Georg Rück, *Die projektive Ebene – Was sind unendlich ferne Punkte*

[10] Stephan Klaus, Oliver Labs, Thomas Markwig, *Theorie und Visualisierung algebraischer Kurven und Flächen*, Vortragsausarbeitung, 2009

[11] Henri Cohen, *A Course in Computational Algebraic Number Theory*, Springer Verlag, 1996