Westfälische
Wilhelms-Universität
Münster

Fachbereich 10
Mathematik und
Informatik

Elliptische Kurven und Kryptographie

gelesen von PD Dr. Karin Halupczok

Zusammenfassung von Phil Steinhorst

Sommersemester 2015

Hier kommt bald ein Bild hin!

http://wwwmath.uni-muenster.de/u/karin.halupczok/ellKKSoSe15/

Stand: 8. Juli 2015

Vorwort

Der vorliegende Text ist eine inhaltliche Aufbereitung zur Vorlesung Elliptische Kurven und Kryptographie, gelesen von PD Dr. Karin Halupczok an der WWU Münster im Sommersemester 2015. Der Inhalt entspricht weitestgehend dem handschriftlichen Skript, welches auf der Vorlesungswebsite bereitsgestellt wird. Dieses Werk ist daher keine Eigenleistung des Autors und wird nicht von der Dozentin der Veranstaltung korrekturgelesen. Für die Korrektheit des Inhalts wird keinerlei Garantie übernommen. Bemerkungen, Korrekturen und Ergänzungen kann man folgenderweise loswerden:

- persönlich durch Überreichen von Notizen oder per E-Mail
- durch Abändern der entsprechenden TeX-Dateien und Versand per E-Mail an mich
- direktes Mitarbeiten via GitHub. Dieses Skript befindet sich im `latex-wwu`-Repository von Jannes Bantje:

  https://github.com/JaMeZ-B/latex-wwu

Literatur

- Blake, Seroussi, Smart: Elliptic curves in cryptography
- Menezes, van Oorschot, Vanstone: Handbook of applied cryptography
- Silverman: The arithmetic of elliptic curves
- Silverman: A friendly introduction to number theory, chap. 40-45
- Washington: Elliptic curves, number theory and cryptography
- Werner: Elliptische Kurven in der Kryptographie

Kommentar der Dozentin

In der Vorlesung beschäftigen wir uns mit den arithmetischen und geometrischen Eigenschaften elliptischer Kurven sowie deren Anwendungen in der Kryptographie. Dabei werden wir auch einen Vergleich mit Anwendungen der elementaren Zahlentheorie in der Kryptographie ziehen. Wir verfolgen eine elementare Herangehensweise, d.h. Kenntnisse der algebraischen Geometrie und der Funktionen- oder Zahlentheorie werden nicht benötigt. Es genügen die Vorkenntnisse aus den Grundvorlesungen.

Vorlesungswebsite

Das handgeschriebene Skript sowie weiteres Material findet man unter folgendem Link:

http://wwwmath.uni-muenster.de/u/karin.halupczok/ellKKSoSe15/

Titelbild

Das fehlt noch. Über Ideen und Anregungen freue ich mich sehr!

Phil Steinhorst
p.st@wwu.de

Inhaltsverzeichnis

0 Motivation und Einführung

Kryptologie

Die **Kryptologie** besteht aus den folgenden beiden Gebieten: [1]

Kryptographie: Studium mathematischer Techniken zur Verschlüsselung von Informationen oder geheimen Nachrichten und dem Schutz von Daten.

Kryptoanalyse: Beschreibung der Rückgewinnung von Informationen aus verschlüsselten Texten, der Entschlüsselung.

Oft meint man mit "Kryptographie" die Kryptologie.

Früher wurde die Kryptographie vor allem im militärischen oder diplomatischen Sektor verwendet, heutzutage steht in unserer vernetzten Welt vor allem auch der praktische Nutzen im Alltag im Vordergrund: im Internet einkaufen, Online-Banking, persönliche Daten geheimhalten bzw. Datenschutz, Nachrichten und Dokumente digital unterschreiben etc. Das Internet liefert schnelle Informationswege über öffentliche Kanäle, die leicht abgehört werden können, sodass die Verschlüsselung schützenswerter Daten unumgänglich wird. Auch die Möglichkeit zur Signierung wird nötig, weil sehr leicht Absenderangaben gefälscht werden können. Eventuell nicht abhörsichere Kanäle können außer dem Internet aber auch Briefe, Radio, Boten, etc. sein.

Bei der **symmetrischen Verschlüsselung** von Daten gibt es einen Sender S und einen Empfänger E, die sich beide auf einen gemeinsamen Schlüssel geeinigt haben, der zum Ver- und Entschlüsseln dient. Beim **Caesar-Code** z.B. ist dies die Vereinbarung, jeden Buchstaben durch den dritten nachfolgenden im Alphabet zu ersetzen, also $A \mapsto D, B \mapsto E, C \mapsto F$, usw. Die Entschlüsselung ist klar. Derartige **monoalphabetische Chiffrierungen**, bei der jeder Buchstabe des Alphabets stets durch denselben Geheimtextbuchstaben chiffriert wird, sind durch Häufigkeitsanalysen durch einen Angreifer, der die verschlüsselten Nachrichten abhört, sehr leicht zu entschlüsseln. Übrigens gibt es auch heutzutage PDF-Verschlüsselungsprogramme, die so arbeiten!
In dieser Vorlesung behandeln wir die heutzutage gängigen modernen Methoden, die als sicher gelten. Worauf diese starke Sicherheit beruht, hat mathematische Gründe, die wir besprechen möchten. Vor allem interessiert uns, wie und welche Mathematik in die Kryptologie kommt, sodass wir deren Verfahren verstehen können.

Die Anwendungen erfordern die Lösung folgender Probleme bei symmetrischen Verschlüsselungsverfahren:

- Schlüsselaustausch über öffentliche Kanäle (**öffentliche Schlüssel**)
- Verschlüsselung ohne vorherigen Schlüsselaustausch (mit **geheimen Schlüsseln**, die nicht versendet werden)
- Digitale Signierung und Autentifizierung

Dies können **asymmetrische Verfahren** leisten (auch **Public Key-Kryptographie** genannt) und gehen zurück auf Ideen von Diffie[1] und Hellman[2] aus den 70er Jahren:

Jeder Nutzer eines Kommunikationskanals hat einen privaten Schlüssel, den er geheim hält und niemand sonst kennt, sowie einen öffentlichen Schlüssel, den jeder einsehen kann. Eine Nachricht wird dann unter Ausnutzung einer Funktion $x \mapsto f(x)$ verschlüsselt, die zwar leicht zu berechnen, aber praktisch nur mit Kenntnis des privaten Schlüssels des rechtmäßigen Empfängers entschlüsselt werden kann. Der Sender der Nachricht wird dafür den

[1] Whitfield Diffie, http://de.wikipedia.org/wiki/Whitfield_Diffie
[2] Martin Hellman, http://de.wikipedia.org/wiki/Martin_Hellman

öffentlichen Schlüssel des Empfängers zur Verschlüsselung benutzen. Eine derartige Funktion heißt **Einwegfunktion**.

Beispiele

- **RSA-Verfahren**: $(p, q) \mapsto p \cdot q$ mit p, q prim.
- **ECC-Verfahren**: $x \mapsto mx$ in einer Gruppe auf einer elliptischen Kurve.

In einem ersten Teil der Vorlesung stellen wir gängige Verfahren dar, die leicht mit dem Zahlring $\mathbb{Z}$ und Strukturen darin realisiert werden können. Dabei werden wir nur einige Hilfsmittel der elementaren Zahlentheorie entwickeln und dafür heranziehen. In einem zweiten Teil studieren wir die Eigenschaften elliptischer Kurven als interessante geometrische und arithmetische Objekte, die sich in der Praxis der Kryptographie als nützlich erwiesen haben. Wir besprechen dann auch die Sicherheit und Implementierung dieser Verfahren und vergleichen sie miteinander.

Elliptische Kurven

Was sind elliptische Kurven? Jedenfalls sind elliptische Kurven **keine** Ellipsen. Ellipsen lassen sich durch Gleichungen der Form

$$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1 \text{ mit } a, b \in \mathbb{R} \setminus \{0\}$$

beschreiben. Durch die Parametrisierung $x(t) = a \cdot \cos(t), y(t) = b \cdot \sin(t)$ ergibt sich für die Bogenlänge der Ellipse ein elliptisches Integral zweiter Art, nämlich

$$\int_0^{2\pi} \sqrt{\left(\frac{dx(t)}{dt}\right)^2 + \left(\frac{dy(t)}{dt}\right)^2}\, dt = 4 \int_0^{2\pi} \sqrt{a^2 \cos^2(t) + b^2 \cdot \sin^2(t)} dt$$

Im Allgemeinen lässt sich dies nicht elementar integrieren (außer natürlich, falls $a = b$, d.h. ein Kreis vorliegt). Mit Hilfe von elliptischen Kurven findet man jedoch nicht-elementare Stammfunktionen für diese Integrale ($\Rightarrow$ Funktionentheorie). Aufgrund dieses Zusammenhangs haben elliptische Kurven ihren Namen, sie haben ansonsten nichts mit Ellipsen zutun.

Was sind nun elliptische Kurven? Es sind "abelsche Varietäten der Dimension 1". Elliptische Kurven sind spezielle algebraische Kurven über einem Körper k. Es handelt sich dabei um glatte kubische Kurven, deren definierende algebraische Gleichung sich meist in die Form

$$E\colon y^2 = x^3 + ax + b \text{ mit } a, b \in k$$

bringen lässt. Als Punktmenge haben wir dafür

$$E(k) := \{(x, y) \in k^2 : y^2 = x^3 + ax + b\} \cup \{\mathcal{O}\},$$

die Kurve hängt nur von a, b ab. Die Rolle des zusätzlichen so genannten "unendlich fernen Punkts" $\mathcal{O}$ werden wir dabei noch näher beleuchten.

Zwei typische Beispiele für elliptische Kurven:

1) $E_1\colon y^2 = x^3 + 17$, hier liegen sogar Punkte mit ganzzahligen Koordinaten auf E_1, nämlich $(-2, 3), (-1, 4), (2, 5)$. Die Kurve besteht aus einer Zusammenhangskomponente.
2) $E_2\colon y^2 = x^3 + ax + b$, wenn $f(x) = x^3 + ax + b$ drei verschiedene Nullstellen hat, z.B. $a = -3, b = -1$. Die Kurve besteht dann aus zwei Zusammenhangskomponenten.

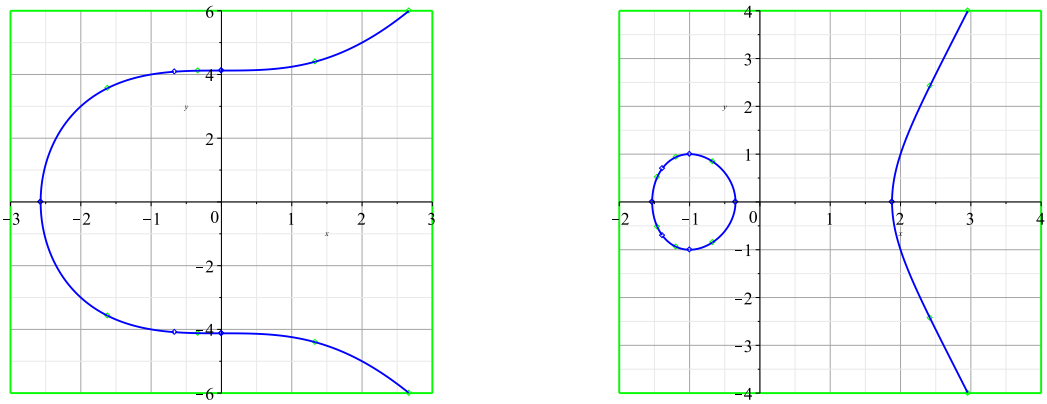


Abbildung 1: Die Kurven E_1 (links) und E_2 (rechts).

Bemerkung

Die kubischen Kurven $C_1 \colon y^2 = x^3 - 3x + 2$ und $C_2 \colon y^2 = x^3$ z. B. sind jedoch keine elliptischen Kurven, weil diese nicht glatt sind.

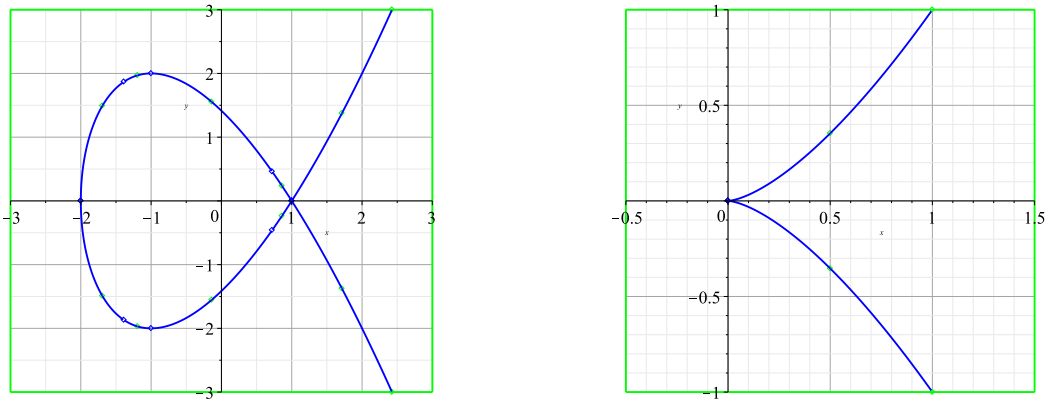


Abbildung 2: Die Kurven C_1 (links) und C_2 (rechts). C_1 ist nicht glatt im Punkt $(1, 1)$, C_2 nicht im Punkt $(0, 0)$.

Für die Kryptographie sind elliptische Kurven interessant, weil sich eine Verknüpfung auf ihrer Punktemenge definieren lässt, mit der diese zu einer Gruppe wird. Dabei gerade auch endliche Körper k zuzulassen, macht diese Verknüpfung auf Rechenmaschinen realisierbar. Die Sicherheit der darauf beruhenden elliptic curve cryptography (ECC) beruht darauf, dass das Problem des diskreten Logarithmus auf einer elliptischen Kurve E, nämlich die Umkehrung der Funktion $P \mapsto mP$ für $m \in \mathbb{N}$ fest, nach heutigem Wissensstand rechnerisch im Allgemeinen extrem schwer realisierbar ist.

1 Allgemeines über Kryptographieverfahren

1.1 Grundlagen aus der elementaren Zahlentheorie und Gruppentheorie

1.1.1 Zahlen, Darstellung von Zahlen

[2] Die Zahlbereiche $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$ sind aus den Grundvorlesungen bekannt. Bezüglich den Verknüpfungen $+$ und $\cdot$ sind verschiedene Axiome erfüllt, die diese Zahlbereiche zu interessante algebraische Strukturen machen:

Halbgruppe	Gruppe	Ring	Körper
$(\mathbb{N}, +), (\mathbb{N}, \cdot)$			
$(\mathbb{Z}, +), (\mathbb{Z}, \cdot)$	$(\mathbb{Z}, +, 0)$	$(\mathbb{Z}, +, \cdot)$	
$(\mathbb{Q}, +), (\mathbb{Q}, \cdot)$	$(\mathbb{Q}, +, 0), (\mathbb{Q} \setminus \{0\}, \cdot, 1)$	$(\mathbb{Q}, +, \cdot)$	$(\mathbb{Q}, +, \cdot)$
$(\mathbb{R}, +), (\mathbb{R}, \cdot)$	$(\mathbb{R}, +, 0), (\mathbb{R} \setminus \{0\}, \cdot, 1)$	$(\mathbb{R}, +, \cdot)$	$(\mathbb{R}, +, \cdot)$
$(\mathbb{C}, +), (\mathbb{C}, \cdot)$	$(\mathbb{C}, +, 0), (\mathbb{C} \setminus \{0\}, \cdot, 1)$	$(\mathbb{C}, +, \cdot)$	$(\mathbb{C}, +, \cdot)$

Weiter sind $\mathbb{Q}$ und $\mathbb{R}$ angeordnete Körper, d.h. es gibt eine Anordnungsrelation $\leq$, die sich mit $+$ und $\cdot$ verträgt. Für $\mathbb{C}$ ist eine solche Anordnung nicht mehr möglich.

Definition 2.1 (Halbgruppe)
Eine Menge $H \neq \emptyset$ mit Verknüpfung $*\colon H \times H \to H$ heißt **Halbgruppe**, falls $*$ assoziativ ist, d.h. für alle $a, b, c \in H$ gilt $a * (b * c) = (a * b) * c$.

Definition 2.2 (Gruppe)
Eine Halbgruppe $(G, *)$ heißt **Gruppe**, falls es ein neutrales Element $e \in G$ gibt mit $e * g = g * e = g$ für alle $g \in G$, und falls zu jedem $g \in G$ ein inverses Element $h \in G$ existiert mit $h * g = g * h = e$. Wir schreiben auch g^{-1}, $\frac{1}{g}$ oder $-g$ für h.

Definition 2.3 (abelsche Gruppe)
Eine Gruppe $(G, *, e)$ heißt **abelsch** bzw. **kommutativ**, falls für alle $a, b \in G$ gilt: $a * b = b * a$.

Definition 2.4 (Ring)
Ring mit Eins
Ein **Ring** $(R, +, \cdot)$ ist eine Menge $R \neq \emptyset$ und zwei Verknüpfungen $+$ und $\cdot$ so, dass $(R, +, 0)$ eine Gruppe ist, $(R, \cdot, 1)$ eine Halbgruppe mit neutralem Element 1, und so, dass die Distributivgesetze gelten, d.h. $(a + b) \cdot c = a \cdot c + b \cdot c$ und $c \cdot (a + b) = c \cdot a + c \cdot b$.

Bemerkung 2.5
Die Addition $+$ ist in einem Ring stets kommutativ. Ein Ring heißt kommutativ, wenn die Multiplikation $\cdot$ kommutativ ist. Soll der Nullring $R = \{0\}$ mit $1 = 0$ ausgeschlossen werden, fordert man zusätzlich noch $1 \neq 0$ in den Ringaxiomen.

Definition 2.6 (Einheit, Einheitengruppe)
Die in einem Ring $(R, +, \cdot)$ bezüglich $\cdot$ invertierbaren Elemente heißen **Einheiten**. Die Menge der Einheiten in R wird mit R^* bezeichnet, d.h. also $R^* := \{a \in R : \exists b \in R \text{ mit } a \cdot b = b \cdot a = 1\}$. Damit ist $(R^*, \cdot, 1)$ also eine Gruppe.

Definition 2.7 (Körper)
Ein **Körper** $(K, +, \cdot)$ ist ein kommutativer Ring mit $1 \neq 0$, für den $K^* = K \setminus \{0\}$ gilt.

Algebraische Strukturen dieser Art können wir auch in Teilmengen von $\mathbb{Z}$ auffinden und diese für kryptographische Anwendungen ausnutzen. Darum geht es in §1 dieser Vorlesung. Dabei wird klar, dass die Anwendungen auch – teilweise – in beliebigen Gruppen, Ringen und Körpern möglich sind. Die Gruppen, die durch elliptische Kurven gegeben sind, haben sich in der Praxis dann als vorteilhaft herausgestellt.

Wenn wir Teilmengen von $\mathbb{Z}$ auch praktisch untersuchen möchten, wird die Frage wichtig, wie man ganze Zahlen auf geschickte und kompakte Art darstellen kann. Dafür benutzen wir im Alltag das Dezimalsystem, für Rechenmaschinen ist auch das Binär- und das Hexadezimalsystem nützlich. Dabei werden die Ziffern $0, 1, \dots 9$ bzw. $0, 1$ bzw. $0, 1, \dots, 9, A, \dots, F$ verwendet. Allgemein erhalten wir die g-adische Darstellung von $n \in \mathbb{N}$ so:

Satz 2.8
Sei $g \in \mathbb{N}, g \geq 2$ und $n \in \mathbb{N}$. Dann gibt es ein $k \in \mathbb{N}_0$ und $c_k, c_{k-1}, \dots, c_0 \in \{0, \dots, g-1\}$ (genannt "Ziffern"), sodass $n = c_k g^k + c_{k-1} g^{k-1} + \dots + c_0 = \sum_{i=0}^{k} c_i g^i$. Fordern wir $c_k \neq 0$, ist k und die Folge $c_k, \dots, c_1, c_0$ eindeutig bestimmt.

Beweis

Existenz: Sei $k \in \mathbb{N}_0$ so, dass $g^k \leq n < g^{k+1}$ gilt, das heißt wir setzen $k := \left\lfloor \frac{\log(n)}{\log(g)} \right\rfloor$. Zeige durch Induktion nach k die Existenz:
$k = 0$: Setze $c_0 := n$.
$k \rightsquigarrow k+1$: Sei $g^{k+1} \leq n < g^{k+2}$. Setze $n' = n - \left\lfloor \frac{n}{g^{k+1}} \right\rfloor \cdot g^{k+1}$. Es folgt $0 \leq n' < g^{k+1}$, d.h. auf n' ist die Induktionsvoraussetzung anwendbar. Nach dieser hat n' eine g-adische Zifferndarstellung $n' = \sum_{i=0}^{k} c_i g^i$.
Wegen $1 \leq \frac{n}{g^{k+1}} < g$ ist $1 \leq \left\lfloor \frac{n}{g^{k+1}} \right\rfloor < g$, also setze $c_{k+1} := \left\lfloor \frac{n}{g^{k+1}} \right\rfloor$.
$\Rightarrow n = c_{k+1} g^{k+1} + n' = \sum_{i=0}^{k+1} c_i g^i$.

Eindeutigkeit: Sind $\sum_{i=0}^{k} a_i g^i = m = \sum_{i=0}^{r} b_i g^i$ zwei verschiedene Darstellungen von $m \in \mathbb{N}$. Ist $r > k$, so sei $a_{k+1} = \dots = a_r := 0$, sonst sei $b_{r+1} = \dots = b_k := 0$, falls $r < k$. Dann sei $l := \max\{i \in \mathbb{N}_0 : i \leq \max\{k, r\}, a_i \neq b_i\}$ die größte Stelle, an der sich die Darstellungen unterscheiden.
$\Rightarrow 0 = \sum_{i=0}^{l} \underbrace{(a_i - b_i)}_{=0 \text{ für } i>l} g^i \Rightarrow \underbrace{|b_l - a_l|}_{\geq 1} g^l = \left| \sum_{i=0}^{l-1} (a_i - b_i) g^i \right|$
$\Rightarrow g^l \leq \sum_{i=0}^{l-1} |a_i - b_i| g^i \leq \sum_{i=0}^{l-1} (g-1) g^i = (g-1) \frac{g^l - 1}{g-1} = g^l - 1 \quad \lightning$ □

Definition 2.9 (g-adische Darstellung)
Die Ziffernfolge $c_k, c_{k-1}, \dots c_0$ aus Satz 2.8 heißt g-**adische Darstellung** von n. Die Zahl c_k heißt **Leitziffer**, die Zahl c_0 die **Endziffer**. Die Zahl $k+1$ heißt **Stellenzahl** bzw. **Länge** der g-adischen Darstellung. Die Zahl g heißt auch **Basis** der Darstellung. Eine m-**Bit-Zahl** ist eine Zahl $n \in \mathbb{N}$ der Länge $\leq m$ zur Basis 2.

Bemerkung 2.10
Wir können jede natürliche (und dann auch jede ganze) Zahl n also eindeutig schreiben als Linearkombination endlich vieler Potenzen von g.

Beispiel 2.11

$$\begin{aligned} 163_{(10)} &= 1 \cdot 10^2 + 6 \cdot 10^1 + 3 \cdot 10^0 \\ 43_{(10)} &= 1 \cdot 2^5 + 0 \cdot 2^4 + 1 \cdot 2^3 + 0 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0 = 101011_{(2)} \\ &= 2 \cdot 16^1 + 11 \cdot 16^0 = 2B_{(16)} \end{aligned}$$

Die bekannten schriftlichen Additions- und Multiplikationsrechnungen, die unter Beachtung von Überträgen ziffernweise geschehen, können in jeder Basis ausgeführt werden. Es gibt weiter für die Multiplikation großer Zahlen (d.h. mit großer Stellenzahl bis etwa $2 \cdot 10^{10}$) schnelle Algorithmen, die wir hier aber nicht näher behandeln möchten; etwa mit der schnellen Fouriertransformation (FFT) nach Schönhage/Strassen[3].
Der Beweis von Satz 2.8 zeigt, dass die Länge von n gleich $\left\lfloor \frac{\log(n)}{\log(g)} \right\rfloor + 1$ ist, so viele Ziffern müssen zum Hinschreiben bzw. Eintippen von n angegeben werden. Bei verschiedenen Basen ändert sich hier nur der Faktor $\frac{1}{\log(g)}$. Deswegen sagt man, die Länge sei $\mathcal{O}(\log(n))$ und meint damit die Aussage: Es existiert eine Konstante $C > 0$, sodass $k + 1 \leq C \cdot \log(n)$. (Landau-Symbolik[4], "Groß-O-Notation")
Entscheidend für das Studium von $\mathbb{Z}$ ist der Grundbegriff der Teilbarkeit.

Definition 2.12 (Teilbarkeit)
Für $a, b \in \mathbb{Z}$ heißt a **Teiler** von b bzw. a **teilt** b, in Zeichen $a \mid b$, falls ein $c \in \mathbb{Z}$ existiert mit $ac = b$. Ist a kein Teiler von b, schreibt man $a \nmid b$.

Beispiel 2.13
$3 \mid 12$, $4 \mid 0$, $0 \mid 0$, $7 \nmid 12$, $0 \nmid 4$. Es kann 0 nur die 0 teilen.

Definition 2.14 (Primzahl)
Eine natürliche Zahl $p \in \mathbb{N}$ heißt **Primzahl** bzw. **prim**, wenn sie genau zwei Teiler in $\mathbb{N}$ besitzt (nämlich 1 und p, $1 \neq p$). Eine natürliche Zahl $n > 1$ heißt **zusammengesetzt**, falls n keine Primzahl ist.

Primzahlen sind die "Bausteine" der natürlichen Zahlen:

Satz 2.15 (Satz von der eindeutigen Primfaktorzerlegung, Hauptsatz der Arithmetik)
Jede natürliche Zahl $n > 1$ besitzt genau eine Darstellung

$$n = p_1^{e_1} \cdot p_r^{e_r} = \prod_{i=1}^{r} p_i^{e_i}$$

mit $r \in \mathbb{N}$, Primzahlen $p_1, \ldots, p_r$ mit $e_1, \ldots, e_r \in \mathbb{N}$ und $p_1 < p_2 < \ldots < p_r$. Diese heißt die **Primfaktorzerlegung** (PFZ) von n.

Bemerkung 2.16
Lässt man die letzte Bedingung weg, ist die Darstellung eindeutig bis auf die Reihenfolge der Primpotenzen. Die Zahl e_i ist dabei die Vielfachheit (auch **Exponent** genannt), mit der p_i als Faktor in n auftritt, d.h. $p_i^{e_i} \mid n$, aber $p^{e_i+1} \nmid n$. Dafür gibt es das Symbol $p^{e_i} \parallel n$, und die Primfaktorzerlegung lässt sich kompakt auch schreiben als $n = \prod_p p^{e(p)}$, wobei $e(p) := e$ mit $p^e \parallel n$, falls $p \mid n$, und $e(p) := 0$, falls $p \nmid n$. Weiter ist $\omega(n) := r$ die Anzahl der verschiedenen Primteiler von n.

[3] siehe http://de.wikipedia.org/wiki/Sch%C3%B6nhage-Strassen-Algorithmus
[4] siehe http://de.wikipedia.org/wiki/Landau-Symbole

Beweis

Existenz: Ist n prim, ist nichts zu zeigen, und ist n nicht prim, gibt es $k, l \in \mathbb{N} \setminus \{1\}$ mit $n = kl$. Da $\min\{k, l\} > 1$, folgt $\max\{k, l\} < n$. Nach Induktionsvoraussetzung sind also k, l Produkte von Potenzen von Primzahlen, also auch $n = kl$.

Eindeutigkeit: Sei $n > 1$ minimal mit zwei verschiedenen Zerlegungen $n = \prod_{i=1}^{r} p_i^{e_i} = \prod_{i=1}^{s} q_i^{f_i}$, die p_i, q_i prim und angeordnet. Da $p_1 \neq q_i$ für alle i gilt (sonst hätte $\frac{n}{p_1} < n$ zwei verschiedene Zerlegungen), ist $\mathrm{ggT}(p_1, q_i) = 1$, und mit den Zerlegungen folgt $p_1 \mid q_1^{f_1 - 1}$ aus Lemma 2.21. Die Fortsetzung des Verfahrens zeigt schließlich $p_1 \mid q_s$, was wegen $\mathrm{ggT}(p_1, q_s) = 1$ ein Widerspruch ist. (Beachte: Im Beweis von Lemma 2.21 wird nie die Eindeutigkeit der Primfaktorzerlegung benutzt.) □

Die Eindeutigkeit der Primfaktorzerlegung zeigt, dass auch diese eine Möglichkeit zur Darstellung natürlicher Zahlen ist. Diese ist jedoch unpraktisch, weil das folgende Problem im Allgemeinen schwer zu lösen ist, worauf einige kryptographische Verfahren (insb. RSA) beruhen.

Definition 2.17 (Faktorisierungsproblem)
Zu einer natürlichen zusammengesetzten Zahl $n > 1$ bestimme man einen nichttrivialen Teiler t mit $1 < t < n$.

Klar: Ist das Faktorisierungsproblem rechnerisch leicht zu machen, kann auch (durch Iteration) die Primfaktorzerlegung von n leicht bestimmt werden. In der Praxis, wenn n nicht gerade schon von einer speziellen Form ist, können Teiler großer Zahlen n jedoch nur sehr schwer aufgefunden werden.

- Das derzeit schnellste algorithmische Verfahren zur Faktorisierung (auf einem klassischen Computer) ist das **Zahlkörpersieb** mit einer Laufzeit von nur $\mathcal{O}(\exp(C(\log n)^{1/3}(\log\log n)^{2/3}))$, d.h. es handelt sich um so genanntes **subexponential schnelles Verfahren**, weil $(\log n)^B \ll \exp(C(\log n)^{1/3}(\log\log n)^{2/3}) \ll \exp(D \log n) = n^D$.

- Peter Shor[5] entdeckte um 1994, dass das Faktorisierungsproblem auf einem Quantencomputer mit einer Laufzeit von (meist) nur $\mathcal{O}((\log n)^3)$ sehr (d.h. polynomiell) schnell gelöst werden kann, was die Sicherheit gängiger Kryptoverfahren wie RSA untergräbt. Allerdings ist die Konstruktion solcher Quantencomputer (physikalisch) extrem schwierig, diverse Forschergruppen arbeiten daran. Am 2.1.2014 meldete die Washington Post unter Berufung auf Dokumente von Edward Snowden[6], dass die NSA an der Entwicklung eines kryptographisch mützlichen Quantencomputers arbeitet[7]. Zum Begriff Quantencomputer siehe Wikipedia.

Im Folgenden besprechen wir noch den ggT zweier natürlicher Zahlen, der sich in vielerlei Hinsicht als wichtig und nützlich erweist:

Definition 2.18
Seien $a, b \in \mathbb{Z}$. Der **größte gemeinsame Teiler** (ggT) von a und b in $\mathbb{N}$ ist die Zahl $d := \max\{t \in \mathbb{N} : t \mid a \wedge t \mid b\}$. Notation: $\mathrm{ggT}(a, b) := d$. Ist $\mathrm{ggT}(a, b) = 1$, heißen a und b **teilerfremd**.

Haben wir für a und b die Primfaktorzerlegungen $a = \prod_p p^{e(p)}$ und $b = \prod_p p^{f(p)}$ vorliegen, kann ihr ggT leicht bestimmt werden als $\mathrm{ggT}(a, b) = \prod_p p^{\min(e(p), f(p))}$, z.B. $\mathrm{ggT}(2^3 \cdot 3^6 \cdot 5^4, 2^4 \cdot 3^5) = 2^3 \cdot 3^5$. Wegen des Faktorisierungsproblems kann dies aber so nicht praktisch umgesetzt werden. Stattdessen benutzt man den (polynomiell) schnellen euklidischen Algorithmus, vgl. Übungsaufgabe.

[5] http://de.wikipedia.org/wiki/Peter_Shor
[6] http://de.wikipedia.org/wiki/Edward_Snowden
[7] Link zum Artikel

Satz 2.19 (Teilen mit Rest)
Zu $a \in \mathbb{Z}, b \in \mathbb{N}$ existieren eindeutigen $q, r \in \mathbb{Z}, 0 \leq r < b$ mit $a = qb + r$, nämlich $q = \left\lfloor \frac{a}{b} \right\rfloor = \max\{m \in \mathbb{Z} : m \leq \frac{a}{b}\}$ und $r = a - qb$. Dabei heißt r der **kleinste nichtnegative Rest**. Statt $0 \leq r < b$ kann auch $r \in \mathbb{Z}$, $|r| < \frac{b}{2}$, erfüllt werden; r heißt dann der **absolut kleinste Rest** (bei Division durch b).

Satz 2.20 (Euklidischer Algorithmus)
Seien $a, b \in \mathbb{N}$. Durch fortgesetztes Teilen mit Rest erhalten wir als letzten Rest $\neq 0$ den $\mathrm{ggT}(a, b)$, sowie $x, y \in \mathbb{Z}$ mit $\mathrm{ggT}(a, b) = xa + yb$ (siehe Schema).

Beschreibung des Rechenverfahrens
Rechnen sukzessive mit $r_{-1} := a, r_0 := b$:

$$\begin{aligned} r_{-1} &= q_0 r_0 + r_1 \\ r_0 &= q_1 r_1 + r_2 \\ r_1 &= q_2 r_2 + r_3 \\ &\vdots \end{aligned}$$

Das Verfahren wird fortgeführt, bis erstmals ein Rest $r_{m+1} = 0$ auftritt, was wegen $r_0 > r_1 > r_2 > \ldots$ nach höchstens $b + 1$ vielen Schritten der Fall sein wird. Sind die Quotienten $q_0, \ldots, q_m$ bekannt, können mit den Rekursionen

$$\begin{aligned} c_{-2} = 0, c_{-1} = 1 \text{ und } c_k = q_k c_{k-1} + c_{k-2}, k = 0, 1, 2, \ldots, n \\ d_{-2} = 1, d_{-1} = 0 \text{ und } d_k = q_k d_{k-1} + d_{k-2}, k = 0, 1, 2, \ldots, n \end{aligned}$$

die **Bézout-Elemente** als $x = (-1)^{n-1} d_{n-1}, y = (-1)^n c_{n-1}$ berechnet werden.
Wir behaupten also:

(1) Es ist $\mathrm{ggT}(a, b) = r_n$.

(2) $\mathrm{ggT}(a, b) = \underbrace{(-1)^{n-1} d_{n-1}}_{x} a + \underbrace{(-1)^n c_{n-1}}_{y} b$

Beweis
zu (1) : Da $r_n \mid r_{n-1}, r_n \mid r_{n-2}, \ldots, r_n \mid r_0 = b, r_n \mid r_{-1} = a$, ist r_n ein Teiler von a und b (Teilen mit Rest von unten nach oben). Ist d irgendein Teiler ≥ 1 von a und b, folgt $d \mid r_1 = a - q_0 b \Rightarrow d \mid r_2 = r_0 - q_1 r_1 \Rightarrow d \mid r_3 = \ldots$, also auch r_n, sodass $d \leq r_n$ folgt (Teilen mit Rest von oben nach unten). Somit ist $r_n = \mathrm{ggT}(a, b)$.

zu (2) : Induktiv kann $c_{k-1} d_k - c_k d_{k-1} = (-1)^k$ gezeigt werden. Daher genügt zu zeigen: $c_n = \frac{a}{\mathrm{ggT}(a,b)}, d_n = \frac{b}{\mathrm{ggT}(a,b)}$.

Details siehe EZT-Skript Lorenz

Mit den $\frac{c_k}{d_k}$ wird die Kettenbruchentwicklung von $\frac{a}{b}$ berechnet und diese bricht bei $\frac{c_n}{d_n} = \frac{a}{b}$ ab. Da bei der Kettenbruchentwicklung alle Brüche $\frac{c_k}{d_k}$ gekürzt sind wegen $c_{k-1} d_k - c_k d_{k-1} = (-1)^k$, folgt dies.

Der Satz vom Euklidischen Algorithmus sichert uns konstruktiv also die Existenz ganzer Zahlen $x, y \in \mathbb{Z}$ mit $\mathrm{ggT}(a, b) = xa + yb$. Die Zahlen x und y heißen auch **Bézout-Elemente** von a und b. Deren Existenz ist auch in der Theorie immer wieder wichtig, z.B. hierfür:

Lemma 2.21
Seien $a, b, c \in \mathbb{Z}$ und $b, c \neq 0$. Gilt $c \mid ab$ und $\mathrm{ggT}(b, c) = 1$, dann ist $c \mid a$.

Beweis
Aus den Voraussetzungen und $c \mid ac$ folgt, dass $c \mid \operatorname{ggT}(ab, ac) = |a| \cdot \operatorname{ggT}(b, c) = |a|$, also $c \mid a$. Zur ersten Gleichheit: Nach Satz 2.20 existieren $x, y \in \mathbb{Z}$ mit $\operatorname{ggT}(b, c) = xb + yc$.
$|a| \cdot \operatorname{ggT}(b, c)$ teilt $|a| \cdot b$ und $|a| \cdot c$, also auch ba und ca, d.h. die rechte Seite ist ein gemeinsamer Teiler von ba und ca. Ist t ein solcher, so teilt t auch $\operatorname{sgn}(a) \cdot (xba + yca) = xb \cdot |a| + yc \cdot |a| = |a| \cdot (xb + yc) = |a| \operatorname{ggT}(b, c)$. □

1.1.2 Kongruenzenrechnen und die modulare Brille

Wir behandeln nun, wie man mit Teilmengen von $\mathbb{Z}$ und neuen Definitionen von "$+$" und "$\cdot$" zu neuen algebraischen Strukturen (Gruppe, Ringe, Körper) kommt. Dazu ist das Kongruenzenrechnen modulo m wesentlich. [3]

Definition 3.1 (Kongruenz, Modul)
Sei $m \in \mathbb{N}$. Dann heißen $a \in \mathbb{Z}$ und $b \in \mathbb{Z}$ **kongruent modulo** m, wenn $m \mid (b - a)$. Wir schreiben dann $a \equiv b \bmod m$ oder $a \equiv b\ (m)$. Die Zahl m heißt der **Modul** der Kongruenz.

Folgerung 3.2

(1) $a \equiv b \bmod m$ bedeutet, dass a und b bei Division durch m denselben kleinsten nichtnegativen (absolut kleinsten) Rest lassen.

(2) $a \equiv b \bmod m, b \equiv c \bmod m \Rightarrow a \equiv c \bmod m$

(3) $a_1 \equiv b_1 \bmod m, a_2 \equiv b_2 \bmod m \Rightarrow a_1 + a_2 \equiv b_1 + b_2 \bmod m, a_1 \cdot a_2 \equiv b_1 \cdot b_2 \bmod m.$

(4) $ca \equiv cb \bmod m \Rightarrow a \equiv b \bmod \left(\frac{m}{\operatorname{ggT}(c,m)}\right)$, insbesondere $a \equiv b \bmod m$, falls $\operatorname{ggT}(c, m) = 1$.

(5) $a \equiv b \bmod m_i$ für $i = 1, \ldots, k \Rightarrow a \equiv b \bmod \operatorname{kgV}(m_1, \ldots, m_k)$

Dies zeigt, dass $\equiv$ für festes m eine Äquivalenzrelation ist und $\mathbb{Z}$ in m paarweise disjunkte Äquivalenzklassen zerlegt.

Definition 3.3 (Restklasse)
Die Äquivalenzklassen von $\equiv$ modulo m heißen **Restklassen** modulo m. (auch: Kongruenzklassen modulo m).

Folgerung 3.4
Die Restklassen modulo m sind Teilmengen von $\mathbb{Z}$ der Gestalt $x + m\mathbb{Z} := \{x + ma : a \in \mathbb{Z}\}$. Die Restklasse $x + m\mathbb{Z}$ heißt auch die Restklasse von x modulo m. Davon gibt es m Stück; wird in jeder Restklasse ein Element x_i, $i = 1, \ldots, m$ ausgewählt, können die m Restklassen mit $x_1 + m\mathbb{Z}, x_2 + m\mathbb{Z}, \ldots, x_m + m\mathbb{Z}$ angegeben werden; die Menge $\{x_1, \ldots, x_m\}$ heißt dann **vollständiges Restsystem** modulo m. Sind $y_1, \ldots, y_m \in \mathbb{Z}$ so, dass $y_i \not\equiv y_j \bmod m$ für alle $i \neq j$, $1 \leq i, j \leq m$, gilt (d.h. die y_i sind paarweise inkongruent modulo m), dann ist $\{y_1, \ldots, y_m\}$ ein vollständiges Restsystem modulo m. Die Zahl x heißt **Repräsentant** der Restklasse $x + m\mathbb{Z}$, und $x + m\mathbb{Z} = z + m\mathbb{Z} \Leftrightarrow x \equiv z \bmod m$, weil laut Definition in der Restklasse von $x \mod m$ genau alle zu x kongruenten Zahlen liegen.

Beispiel 3.5
$\{0, 1, 2\}$ ist vollständiges Restsystem modulo 3, und vollständige Restsysteme modulo 8 sind etwa $\{1, \ldots, 8\}$ und $\{3, 6, 9, 12, 15, 18, 21, 24\} = \{3a : 1 \leq a \leq 8\}$, da $12 \equiv 4 \bmod 8, 15 \equiv 7 \bmod 8, 18 \equiv 2 \bmod 8$, $21 \equiv 5 \bmod 8, 24 \equiv 0 \bmod 8$. Die Menge $\{2a : 1 \leq a \leq 8\}$ ist kein vollständiges Restsystem modulo 8. Die Reste $0, 1, \ldots, m - 1$ könnte man auch als "Standardrepräsentanten" modulo m bezeichnen, da sie immer ein vollständiges Restsystem modulo m bilden.

Folgerung 3.6
Ist $\{x_1, \ldots, x_m\}$ ein vollständiges Restsystem modulo m und $a \in \mathbb{Z}, c \in \mathbb{Z}$ mit $\mathrm{ggT}(c, m) = 1$, so sind auch $\{x_1 + a, x_m + a\}$ und $\{x_1 \cdot c, \ldots, x_m \cdot c\}$ vollständige Restsysteme modulo m (vgl. (4) aus Folgerung 3.2).

Das nützliche an den Restklassen modulo m ist, dass wir nun durch folgende naheliegende Definitionen von $\oplus$ und $\odot$ mit ihnen neue algebraische Strukturen gewinnen können:

Definition 3.7 (Addition und Multiplikation auf $\mathbb{Z}_m$)
Ist der Modul m klar, schreiben wir auch $\underline{x} := x + m\mathbb{Z}$ für die Restklasse von x modulo m. Wir definieren für $x, y \in \mathbb{Z}$ dann

$$\underline{x} \oplus \underline{y} := \underline{x + y}$$
$$\underline{x} \odot \underline{y} := \underline{x \cdot y}$$

Weiter sei $\mathbb{Z}_m = \mathbb{Z}/m\mathbb{Z} = \mathbb{Z}/m := \{x + m\mathbb{Z} : x \in \mathbb{Z}\}$ die Menge der m vielen Restklassen modulo m.

Folgerung 3.8
Wir addieren bzw. multiplizieren zwei Restklassen, indem wir Repräsentanten x, y auswählen und diese addieren bzw. multiplizieren. Das ist nur sinnvoll, wenn bei unterschiedlicher Repräsentantenwahl dieselbe Restklasse als Ergebnis herauskommt. Man sagt, die Definition von $\oplus$ und $\odot$ ist wohldefiniert, da repräsentantenunabhängig. Dies ist klar: $\underline{x_1} = \underline{x_2}$ und $\underline{y_1} = \underline{y_2} \Rightarrow x_1 \equiv x_2 \bmod m$ und $y_1 \equiv y_2 \bmod m \Rightarrow x_1 + y_1 \equiv x_2 + y_2 \bmod m \Rightarrow \underline{x_1 + y_1} = \underline{x_2 + y_2}$, also erhalten wir so dieselbe Restklasse für $\underline{x_1} \oplus \underline{y_1}$ und $\underline{x_2} \oplus \underline{y_2}$, wenn $\underline{x_1} = \underline{x_2}$ und $\underline{y_1} = \underline{y_2}$ (analog für die Multiplikation). Damit kann $(\mathbb{Z}_m, \oplus)$ oder $(\mathbb{Z}/m, \odot)$ auf algebraische Strukturen hin untersucht werden. Wir schreiben ab jetzt auch $+$ für $\oplus$ und $\cdot$ für $\odot$.

Folgerung 3.9
$(\mathbb{Z}_m, +)$ ist eine abelsche Gruppe mit neutralem Element $\underline{0} = 0 + m\mathbb{Z}$, denn Kommutativität und Assoziativität gelten wie in $\mathbb{Z}$, und $\underline{0} + \underline{x} = \underline{0 + x} = \underline{x}$ gilt für alle $x \in \mathbb{Z}$, sowie $\underline{x} + \underline{-x} = \underline{x - x} = \underline{0}$, sodass $-\underline{x} = \underline{-x} = \underline{m - x}$ für alle $x \in \mathbb{Z}$ gilt. Ebenso gilt, dass $(\mathbb{Z}_m, +, \cdot)$ ein kommutativer Ring mit 1 ist.

Das Beispiel $\underline{2} \cdot \underline{0} = \underline{0}, \underline{2} \cdot \underline{1} = \underline{2}, \underline{2} \cdot \underline{2} = \underline{0}$ modulo 4 zeigt, dass es Restklassen ohne Inversen bezüglich $\cdot$ geben kann. Der folgende Satz gibt an, welche Restklassen invertierbar sind, d.h. im Ring $\mathbb{Z}_m$ eine Einheit sind:

Satz 3.10 (Einheiten in $\mathbb{Z}_m$)
Zu $\underline{x} \in \mathbb{Z}_m$ existiert genau dann ein multiplikatives Inverses, d.h. ein $\underline{y} \in \mathbb{Z}_m$ mit $\underline{x} \cdot \underline{y} = \underline{1} \Leftrightarrow x \cdot y \equiv 1 \bmod m$, falls $\mathrm{ggT}(x, m) = 1$. Wir schreiben dann $\underline{x}^{-1}$ oder $\underline{x}^*$ für $\underline{y}$, die Bezeichnungen $\frac{1}{\underline{x}}$ oder $1/\underline{x}$ sind didaktisch ungeschickt.

Beweis
"⇒": Sei $\underline{y} \in \mathbb{Z}_m$ mit $\underline{x} \cdot \underline{y} = 1$, d.h. $xy \equiv 1 \bmod m$, also existiert $k \in \mathbb{Z}$ mit $1 - xy = km \Rightarrow xy + km = 1$. Wäre $d = \mathrm{ggT}(x, m) > 1$, so folgt $d \mid xy + km = 1 \lightning$.

"⇐": Sei $\mathrm{ggT}(x, m) = 1$. Nach Satz 2.20 existiert $y, k \in \mathbb{Z}$ mit $1 = yx + km$, also folgt $\underline{x} \cdot \underline{y} = 1$. □

Fazit: Mit dem euklidischen Algorithmus können wir also Inverse schnell explizit berechnen.

Definition 3.11 (Prime Reste, Eulersche φ-Funktion)
$\underline{x} = x + m\mathbb{Z}$ heißt **prime** oder **reduzierte Restklasse** modulo m, falls $\mathrm{ggT}(x, m) = 1$ gilt. Diese sind genau die Einheiten in $(\mathbb{Z}_m, +, \cdot)$, d.h.

$$\mathbb{Z}_m^* = \{\underline{x} \in \mathbb{Z}_m : \mathrm{ggT}(x, m) = 1\}$$

Die Anzahl der Einheiten sei $\varphi(m) := \#\mathbb{Z}_m^* = \#\{a \in \mathbb{N} : a \leq m, \mathrm{ggT}(a,m) = 1\}$, die so erklärte Funktion $\varphi\colon \mathbb{N} \to \mathbb{N}$ heißt **Eulersche φ-Funktion**. Jedes Repräsentantensystem $\{x_1, \ldots, x_{\varphi(m)}\}$ von $\mathbb{Z}_m^*$ heißt **reduziertes** oder **primes Restsystem** modulo m.´

Satz 3.12 (Multiplikativität von φ)
Es ist $\varphi(p^k) = p^k - p^{k-1}$ für alle p prim, alle $k \in \mathbb{N}$, und $\varphi(mn) = \varphi(m) \cdot \varphi(n)$, falls $\mathrm{ggT}(m,n) = 1$.

Beweis
Unter den Zahlen $1, 2, \ldots, p^k$ sind genau die Vielfachen von p zu p^k nicht teilerfremd, d.h. $p, 2p, \ldots, p^{k-1} \cdot p$, was p^{k-1}-viele Zahlen sind. Zur Multiplikativität siehe Zusatz 3.18.
Ist $n = \prod_{p|n} p^{e(p)}$ die Primfaktorzerlegung von n, folgt aus Satz 3.12:

$$\varphi(n) = \prod_{p|n}\left(p^{e(p)} - p^{e(p)-1}\right) = \prod_{p|n} p^{e(p)} \cdot \left(1 - \frac{1}{p}\right) = n \cdot \prod_{p|n}\left(1 - \frac{1}{p}\right)$$

Folgerung 3.13
$(\mathbb{Z}_m^*, \cdot)$ ist eine Gruppe, die multiplikative Gruppe von $\mathbb{Z}_m$, und die Gruppe $(\mathbb{Z}_m, +)$ heißt additive Gruppe von $\mathbb{Z}_m$. Im Fall $\mathbb{Z}_m^* = \mathbb{Z}_m \setminus \{0\}$ ist $(\mathbb{Z}_m, +, \cdot)$ ein Körper; dies ist genau dann richtig, wenn $m = p$ Primzahl ist, weil genau dann alle $1, 2, \ldots, m-1$ zu m teilerfremd sind. Wir bezeichnen für p prim diesen Körper mit p Elementen mit $\mathbb{F}_p$. Der Körper $\mathbb{F}_p$ hat die Eigenschaft, dass $p \cdot a := \sum_{i=1}^{p} a = 0$ in $\mathbb{F}_p$ für alle $a \in \mathbb{F}_p$ gilt. Wir sagen, er hat die Charakteristik p.

Weitere endliche Körper später

Definition 3.14 (Charakteristik)
Sei k ein Körper. Er hat die **Charakteristik** 0, falls für alle $m \in \mathbb{N}$ gilt: $m \cdot 1 := \underbrace{1 + \ldots + 1}_{m\text{-mal}} \neq 0$.
Falls es ein $m \in \mathbb{N}$ mit $m \cdot 1 = 0$ gibt, so heißt das kleinste solche $m \in \mathbb{N}$ die Charakteristik von k. Wir schreiben kurz $\mathrm{char}(k) = 0$ bzw. $\mathrm{char}(k) = m$.
Zum Beispiel ist $\mathrm{char}(\mathbb{Q}) = \mathrm{char}(\mathbb{R}) = \mathrm{char}(\mathbb{C}) = 0$ und $\mathrm{char}(\mathbb{F}_p) = p$.

Bemerkung
Die Charakteristik eines Körpers k ist entweder 0 oder eine Primzahl, denn sonst wäre $0 = (m \cdot n) \cdot 1 = m \cdot (n \cdot 1) = (m \cdot 1) \cdot (n \cdot 1) \Rightarrow m \cdot 1 = 0$ oder $n \cdot 1 = 0$, da $k^* = k \setminus \{0\}$. Widerspruch zu $m \cdot n$ minimal.
Die Struktur der Zahlringe $(\mathbb{Z}_m, +, \cdot)$ versteht man besser, indem man sie auf "kleinere" Zahlringe zurückführt:

Satz 3.15 (Chinesischer Restsatz für Zahlringe)
Sei $m > 1$ eine natürliche Zahl und $m = m_1 \cdot m_2 \cdot \ldots \cdot m_r$ eine Zerlegung von m in paarweise teilerfremde Zahlen $m_i > 1$. Dann ist die Abbildung

$$\begin{aligned} F\colon \mathbb{Z}/m\mathbb{Z} &\longrightarrow (\mathbb{Z}/m_1\mathbb{Z}) \times \ldots \times (\mathbb{Z}/m_r\mathbb{Z}) \\ x + m\mathbb{Z} &\longmapsto (x + m_1\mathbb{Z}, \ldots x + m_r\mathbb{Z}) \end{aligned}$$

ein Isomorphismus von Ringen.

Satz 3.16 (Chinesischer Restsatz für simultane Kongruenzen)
Seien $m_1, \ldots, m_r > 1$ paarweise teilerfremde Zahlen und sei $a_1, \ldots, a_r \in \mathbb{Z}$. Dann ist das simultane Kongruen-

zensystem

$$x \equiv a_1 \bmod m_1$$
$$x \equiv a_2 \bmod m_2$$
$$\vdots$$
$$x \equiv a_r \bmod m_r$$

in x lösbar, die Lösungen sind alle kongruent modulo $m_1 \cdots m_r$.

Bemerkung

Satz 3.16 folgt aus 3.15 wegen der Bijektivität von F, denn $(a_1 + m_1\mathbb{Z}, \ldots, a_r + m_r\mathbb{Z})$ hat dann genau ein Urbild $x + m\mathbb{Z}$.

Zusatz 3.17 (zu Satz 3.16)

Genau alle $x \equiv x_0 \bmod m_1 \cdots m_r$ lösen das oben angegebene System, wobei $x_0 = a_1 M_1^* M_1 + \ldots + a_r M_r^* M_r$ mit $M_i := \frac{m_1 \cdots m_r}{m_i}$ und $M_i^* \in \mathbb{Z}$ ein multiplikatives Inverses von $M_i \mod m_i$ repräsentiert, d.h. es gilt $M_i^* \cdot M_i \equiv 1 \bmod m_i$, wobei die M_i^* mit dem euklidischen Algorithmus (schnell) berechnet werden können.

Zusatz 3.18 (zu Satz 3.15)

Die Gruppe $\mathbb{Z}_m^*$ ist isomorph zu $\mathbb{Z}_{m_1}^* \times \ldots \times \mathbb{Z}_{m_r}^*$, beide Gruppen haben dann gleich viele Elemente, es folgt

$$\varphi(m) = \varphi(m_1) \cdot \varphi(m_2) \cdots \varphi(m_r),$$

d.h. die Multiplikativität von φ ist ein Korollar des chinesischen Restsatzes.

Beweis von Satz 3.16

Existenz: Ist $x \equiv x_0 \bmod m_1 \cdots m_r$, wie in Zusatz 3.17 angegeben, so folgt für alle $1 \leq i \leq r$:

$$x \equiv x_0 = \underbrace{a_1 M_1^* M_1}_{\equiv 0 \bmod m_i} + \ldots + \underbrace{a_i M_i^* M_i}_{\equiv a_i \cdot 1 \bmod m_i} + \ldots + \underbrace{a_r M_r^* M_r}_{\equiv 0 \bmod m_i} \equiv a_i \bmod m_i$$

Eindeutigkeit modulo $m_1 \cdots m_r$: Ist $y \in \mathbb{Z}$ eine weitere Lösung des Kongruenzensystems, so gilt für alle $j \neq i$: $y \equiv a_j \bmod m_j$, also $\underbrace{M_j^* \cdot M_j}_{\equiv 1 \bmod m_j} \cdot y \equiv a_j \bmod m_j$ und $M_i M_i^* a_i \equiv 0 \bmod m_j$ (Division durch m_j), und somit

$$y \equiv a_j \bmod m_j \equiv \sum_{j=1}^{k} M_j M_j^* a_j \bmod m_j \equiv x_0 \bmod m_j \text{ für alle } j = 1, \ldots, r.$$

Damit folgt, dass m_j Teiler von $y - x_0$ ist. Da die $m_1, \ldots, m_r$ alle paarweise teilerfremd sind, folgt daraus $y \equiv x_0 \bmod m_1 \cdots m_r$, vgl. 3.2 (5). □

Beispiel zum chinesischen Restsatz

Das System

$$x \equiv 2 \bmod 7$$
$$x \equiv 3 \bmod 8$$

hat die Lösung $x \equiv 2 \cdot 1 \cdot 8 + 3 \cdot (-1) \cdot 7 = -5 \equiv 51 \bmod 56$, denn $1 \equiv 8^{-1} \bmod 7$ und $-1 \equiv 7^{-1} \bmod 8$.
Beispiel-Textaufgabe dazu: Gegeben seien zwei Tüten mit gleich vielen Bonbons. Beim gleichmäßigen Aufteilen der einen Tüte an sieben Kinder bleiben zwei Bonbons übrig. Beim Aufteilen der anderen auf acht Kinder bleiben

drei Bonbons übrig. Wie viele Bonbons waren in einer Tüte?
Lösung: Möglich sind 51, 107, 163, ...Stück.

Beispiele zum Rechnen mit Kongruenzen

- Es ist $5x \equiv 4 \bmod 12 \Leftrightarrow 5^{-1} \cdot 5x \equiv 4 \cdot 5^{-1} \bmod 12 \Leftrightarrow x \equiv 4 \cdot 5^{-1} \equiv 4 \cdot 5 = 20 \equiv 8 \bmod 12$.
  Analog rechnet man in der Restklasse modulo 12:
  $5x \cdot 4 \Leftrightarrow x = 5^{-1} \cdot 4 = 4 \cdot 5 = 20 = 8$.

- Es ist
$$\begin{aligned} & 8x^2 - 2x + 3 \equiv -1 \bmod 7 \\ \Leftrightarrow \quad & (x-1)^2 - 2 + 3 \equiv -1 \bmod 7 \\ \Leftrightarrow \quad & (x-1)^2 \equiv -2 \equiv 5 \bmod 7 \end{aligned}$$
  Da nun wegen $0^2 \equiv 0 \bmod 7, 1^2 \equiv 1 \bmod 7, 2^2 \equiv 4 \bmod 7, 3^2 \equiv 2 \bmod 7$ die Zahl 5 kein Quadrat modulo 7 ist, hat die Kongruenz keine Lösung.
  Die Kongruenz $(x-1)^2 \equiv 4 \bmod 7$ hat die beiden Lösungen $x \equiv 3 \bmod 7$ und $x \equiv -1 \bmod 7$.

  man spricht auch von quadratischen Resten modulo 7

- Die Kongruenz $(x-3) \cdot 4 \equiv 1 \bmod 33$ ist schreibbar als System:
$$\begin{aligned} (x-3) \cdot 4 &\equiv 1 \bmod 3 \\ (x-3) \cdot 4 &\equiv 1 \bmod 11 \end{aligned}$$
  Die beiden einzelnen Kongruenzen haben die Lösungen $x \equiv 1 \bmod 3$ sowie $x \equiv 6 \bmod 11$. Mittels chinesischem Restsatz erhält man eine Lösung der Ausgangskongruenz modulo 33:
  $x \equiv 1 \cdot 2 \cdot 11 + 6 \cdot 4 \cdot 3 = 22 + 6 \cdot 12 = 94 \equiv -5 \equiv 28 \bmod 33$

- Bei manchen zahlentheoretischen Aufgaben, wie z.B. die Frage, ob es ganzzahlige Lösungen zu bestimmten Gleichungen geben kann, ist die "modulare Brille" ein nützliches Hilfsmittel. Hier ein Beispiel, wo wir die modulare Brille modulo 8 aufsetzen, um mehr zu sehen:
  Betrachte die Gleichung $8x + 7 = u^2 + v^2 + w^2$ in $u, v, w, x \in \mathbb{N}_0$. Sie ist unlösbar, denn modulo 8 erhalten wir $7 \equiv u^2 + v^2 + w^2 \bmod 8$. Alle quadratischen Reste modulo 8 sind $0, 1$ und 4:

z	0	± 1	± 2	± 3	4
z^2	0	1	4	1	0

  Daher ist $v^2 + w^2 \equiv 0, 1, 4, 2, 5 \bmod 8$, also $u^2 + v^2 + w^2 \equiv 0, 1, 4, 2, 5, \quad 1, 2, 5, 3, 6, \quad 4, 5, 0, 6, 1 \bmod 8$, aber nie $\equiv 7 \bmod 8$. Es kann keine Lösungen modulo 8 geben, also auch keine in $\mathbb{Z}$.

1.1.3 Gruppen

Die Gruppen $(\mathbb{Z}_m, +, 0)$ und $(\mathbb{Z}_m^*, \cdot)$ sind endliche abelsche Gruppen. Wir untersuchen ein paar ihrer allgemeinen Eigenschaften und führen dabei ein paar Grundbegriffe ein. [4]

Definition 4.1 (Gruppenordnung)
Die **Ordnung** einer endlichen Gruppe G ist die Anzahl ihrer Elemente, kurz $\mathrm{ord}(G) := \#G$.

Definition 4.2 (Untergruppe)
Eine Teilmenge H einer Gruppe G mit Verknüpfung $*$ heißt **Untergruppe**, falls auch $(H, *)$ eine Gruppe ist.

Satz 4.3 (Satz von Lagrange)
Ist $(G, *)$ eine endliche Gruppe, so ist die Ordnung einer Untergruppe H stets ein Teiler von $\mathrm{ord}(G)$.

Beweis

Die **Linksnebenklassen** $a * H := \{a * h : h \in H\}$ für $a \in G$ sind paarweise disjunkt, das heißt es gilt stets $a * H = b * H$ oder $a * H \cap b * H = \emptyset$.
(Denn: ist $c \in a * H \cap b * H$, so ist $c = a * g = b * h$ für $g, h \in H$, also $a = b * (h * g^{-1})$, somit $a * H = \{a * m : m \in H\} = \{b * h * g^{-1} * m : m \in H\} = \{b * n : n \in H\} = b * H$.)
Also ist G die disjunkte Vereinigung endlich vieler Linksnebenklassen $a_1 * H, \ldots, a_r * H$. Da $\#(a * H) = \#H$ für alle $a \in G$ gilt, folgt mit $\mathrm{ord}(G) = r \cdot \mathrm{ord}(H)$ die Behauptung. □

Definition 4.4 (Erzeugnis, zyklisch)

Sei $(G, +)$ eine abelsche Gruppe und $a \in G$. Für $k \in \mathbb{Z}$ definieren wir $ka := \underbrace{a + \ldots + a}_{k\text{-mal}}$, falls $k > 0$, $k \cdot 0 := 0$ und $k \cdot a := -(-k) \cdot a$, falls $k < 0$. Dann ist $\langle a \rangle := \{ka : k \in \mathbb{Z}\}$ eine Untergruppe von G. Wir nennen $\langle a \rangle$ die von a **erzeugte Untergruppe** bzw. das **Erzeugnis** von a und a einen **Erzeuger**. Ist $\langle a \rangle$ eine endliche Untergruppe, heißt ihre Ordnung die **Ordnung** von a, kurz $\mathrm{ord}(a) := \#\langle a \rangle$. Eine Gruppe G mit Erzeuger a, das heißt $G = \langle a \rangle$, heißt **zyklisch**.

klar!

Schreibt man die Gruppe multiplikativ mit Verknüpfung "$\cdot$", so setzt man $a^k := \underbrace{a \cdots a}_{k\text{-mal}}$, falls $k > 0$, $a^0 := 1$, $a^k := \left(a^{(-k)}\right)^{-1}$, falls $k < 0$, und $\langle a \rangle := \{a^k : k \in \mathbb{Z}\}$. Ansonsten ist bis auf Schreibweise die Begrifflichkeit und Theorie zu Erzeugern und Ordnungen dieselbe.
Nach dem Satz von Lagrange gilt für jede endliche Gruppe G und $a \in G$ stets $\mathrm{ord}(a) \mid \mathrm{ord}(G)$.

Beispiel 4.5

$\mathbb{Z}/m\mathbb{Z} = \{0, 1, \ldots, m-1\}$ ist "die" zyklische Gruppe mit $\mathrm{ord}(G) = m$. Ist $m = p$ prim, können außer $\{0\}$ und $\mathbb{Z}/p\mathbb{Z}$ keine weiteren Untergruppen existieren.

Lemma 4.6

Sei $(G, +)$ eine Gruppe, $a \in G$. Es ist $\mathrm{ord}(a)$ die kleinste natürliche Zahl m mit $ma = 0$.
Es gilt: $ka = 0 \Leftrightarrow \mathrm{ord}(a) \mid k$.
(Bei multiplikativer Schreibweise: $\mathrm{ord}(a) = \min\{m \in \mathbb{N} : a^m = 1\}$ und $a^k = 1 \Leftrightarrow \mathrm{ord}(a) \mid k$.)

Beweis

Der erste Teil ist klar. Zum zweiten Teil:

"⇒": Falls $k \in \mathbb{N}$ mit $ka = 0$ ist, nehme Division von k durch $\mathrm{ord}(a)$ vor: $k = q \cdot \mathrm{ord}(a) + r$ mit $0 \leq r < \mathrm{ord}(a)$. Wegen $0 = ka = q \cdot \underbrace{\mathrm{ord}(a) \cdot a}_{=0} + ra$ folgt $ra = 0$, wegen der Minimalität von $\mathrm{ord}(a)$ also $r = 0$, also $\mathrm{ord}(a) \mid k$.

"⇐": Für $k = m \cdot \mathrm{ord}(a)$ folgt $ka = m \cdot (\mathrm{ord}(a) \cdot a) = 0$. □

Folgerung 4.7

$\mathrm{ord}(G) \cdot a = 0$ bzw. multiplikativ: $a^{\mathrm{ord}(G)} = 1$, da $\mathrm{ord}(a) \mid \mathrm{ord}(G)$ nach Lemma 4.6

Folgerung 4.8 (Kleiner Satz von Fermat)

Da $\mathrm{ord}((\mathbb{Z}/m\mathbb{Z})^*) = \varphi(m)$, ist $a^{\varphi(m)} \equiv 1 \bmod m$, falls $\mathrm{ggT}(a, m) = 1$. Für p prim: $a^{p-1} \equiv 1 \bmod p$ für $p \nmid a$.

Bemerkung 4.9 (Satz von Euler-Fermat)
Die Kongruenz $a^{\varphi(m)} \equiv 1 \bmod m$, falls $\mathrm{ggT}(a,m) = 1$, heißt auch **Satz von Euler-Fermat**. Als Ordnung eines $a \in \mathbb{Z}_m^*$ (Notation: $\mathrm{ord}_m(a)$) kommt also nur ein Teiler von $\varphi(m)$ in Frage.

Beispiel 4.10
Wir haben $\varphi(15) = \varphi(3) \cdot \varphi(5) = 2 \cdot 4 = 8$. Die möglichen Ordnungen von Zahlen $a \mod 15$ mit $\mathrm{ggT}(a,15) = 1$ sind also $0, 1, 2, 4, 8$.
Wegen $4^2 = 16 \equiv 1 \bmod 15$ ist zum Beispiel $\mathrm{ord}_{15}(4) = 2$. Bei anderen Zahlen muss man unter Umständen Potenzen mit größeren Exponenten ausrechnen, um die Ordnung zu bestimmen.

Generell stellt sich in Anwendungen die Frage, wie man leicht und schnell (modulare) Potenzen $a^k \bmod m$ mit großem k berechnen kann. Der Satz von Euler-Fermat erlaubt bereits eine Reduktion von $k \bmod \varphi(m)$: Ist $k = q \cdot \varphi(m) + r$ mit $0 \leq r < \varphi(m)$, folgt

$$a^k = a^{\varphi(m)\cdot q + r} = \left(a^{\varphi(m)}\right)^q \cdot a^r \equiv 1^q \cdot a^r = a^r \bmod m.$$

Ist aber auch $\varphi(m)$ bzw. r groß, hilft man sich mit folgender Methode des schnellen Potenzierens weiter:

Lemma 4.11 (Methode des schnellen Potenzierens)
Gegeben sei eine Gruppe $(G, \cdot)$, zu berechnen ist für $r \in \mathbb{N}, a \in G$ die Potenz a^r in der Gruppe G.

1. Schritt: Mit höchstens $d := \left\lfloor \frac{\log r}{\log 2} \right\rfloor$ vielen Verknüpfungen in G berechne durch sukzessives Quadrieren a^2, $\left(a^2\right)^2 = a^4, \ldots, a^{2^d}$.

2. Schritt: Schreibe r als Binärzahl: $r = \sum_{i=0}^{d} c_i \cdot 2^i$ mit $c_i \in \{0,1\}$.

3. Schritt: Berechne $a^r = a^{c_0} \cdot a^{2c_1} \cdot a^{2^2 c_2} \cdots a^{2^d c_d} = (a^{c_0}) \cdot \left(a^2\right)^{c_1} \cdot \left(a^{2^2}\right)^{c_2} \cdots \left(a^{2^d}\right)^{c_d}$ mit maximal d weiteren Verknüpfungen in G.

Somit reichen höchstens $2d = \mathcal{O}(\log r)$ viele Anwendungen der Gruppenverknüpfung "$\cdot$". Bei additiver Schreibweise einer Gruppe $(G, +)$ geht das Verfahren zur Berechnung von $r \cdot a$ analog. Man nennt es dann auch das **dual and add**-Verfahren.

Beispiel 4.12
$5^{12} = 5^{2^2+2^3} = 5^{2^2} \cdot 5^{2^3}$. Modulo 11 rechnen wir:

$$5^2 \equiv \bmod 11, 5^{2^2} \equiv 3^2 \equiv -2 \bmod 11, 5^{2^3} \equiv (-2)^2 \equiv 4 \bmod 11,$$

also $5^{12} \equiv (-2) \cdot 4 \equiv 3 \bmod 11$. Das geht schneller als 5^{12} von Hand auszurechnen und durch 11 zu teilen.

Anwendung 4.13 (Lösen quadratischer Kongruenzen)
Im Fall $p \equiv 3 \bmod 4$ prim können wir Lösungen quadratischer Kongruenzen modulo p bestimmen: Sei $p = 4k + 3$ prim und a mit $p \nmid a$ ein quadratischer Rest modulo p, d.h. es existiert ein $b \in \mathbb{Z}$ mit $a \equiv b^2 \bmod p$, und wir möchten $\pm b \bmod p$ ausrechnen können. Nach dem kleinen Fermat folgt $b^{4k+2} = b^{p-1} \equiv 1 \bmod p$. Es folgt $\left(a^{k+1}\right)^2 \equiv \left(b^2\right)^{2(k+1)} = b^{(4k+2)+2} \equiv 1 \cdot b^2 \equiv a \bmod p$, d.h. die Lösungen von $b^2 \equiv a \bmod p$ sind $b = \pm a^{k+1} \bmod p$. Da $a^{k+1} \not\equiv -a^{k+1} \bmod p \Leftrightarrow 2a^{k+1} \not\equiv 0 \bmod p$, gibt es genau zwei Lösungen modulo p, die wir etwa im Restsystem $\{0, 1, \ldots, p-1\}$ angeben können und mit $\pm a^{k+1} \bmod p$ berechnen können, zum Beispiel mit dem schnellen Potenzieren.

Anwendung 4.14
Sei nun n eine zusammengesetze Zahl, etwa $n = pq$ mit $p \equiv q \equiv 3 \bmod 4$ prim, etwa $p = 4k + 3$, $q = 4l + 3$ mit $k, l \in \mathbb{N}_0$, und sei $p \neq q$. Sei $a \bmod n$ ein quadratischer Rest modulo n. Gesucht seien die Lösungen der Kongruenz $a \equiv x^2 \bmod n$. Nach dem chinesischen Restsatz gilt $x^2 \equiv a \bmod n \Leftrightarrow x^2 \equiv a \bmod p$ und $x^2 \equiv a \bmod q$, und die jeweiligen Lösungen $\pm a^{k+1} \bmod p$ und $\pm a^{l+1} \bmod q$ kann man zusammensetzen zu (maximal) vier Lösungen modulo n. Es sind genau vier Lösungen, die explizit wie folgt bestimmt werden können:
Sind $r, s \in \mathbb{Z}$ gegeben mit $rp + sq = 1$, d.h. die Bézout-Elemente von p und q, und ist $\pm b$ Lösung von $x^2 \equiv a \bmod p$ sowie $\pm c$ Lösung von $x^2 \equiv a \bmod q$, so liefert die Formel des chinesischen Restsatzes

$$x = \pm b \cdot s \cdot q \pm c \cdot r \cdot p$$

genau vier Lösungen von $x^2 \equiv a \bmod pq$. Diese müssen paarweise inkongruent modulo pq sein, da wir laut chinesischem Restsatz den Ringisomorphismus $\mathbb{Z}_{pq} \simeq \mathbb{Z}_p \times \mathbb{Z}_q$ haben und die vier verschiedenen Lösungspaare $(b, c), (-b, c), (b, -c), (-b, -c)$ deswegen genau vier Restklassen in $\mathbb{Z}_{pq}$ entsprechen.

Beispiel 4.15
Betrachte $p = 11$, $q = 19$, d.h. $k = 2, l = 4$. Wähle $a = 47$.
Die Lösungen von $x^2 \equiv 47 \equiv 3 \bmod 11$ sind $\pm 3^3 \equiv \pm 5 \bmod 11$, die Lösungen von $x^2 \equiv 47 \equiv 9 \bmod 19$ sind $\pm 3 \bmod 19$.
Bézout-Elemente bestimmen: Das Inverse von $19 \equiv 8 \bmod 11$ ist 7, das von $11 \bmod 19$ ist 7.
$\Rightarrow s = r = 7$ und $x \equiv \mp 5 \cdot 7 \cdot 19 \pm 3 \cdot 7 \cdot 11 \bmod (11 \cdot 19)$ ergibt $x \in \{\pm 16, \pm 60\}$.
Probe: $16^2 \equiv 47 \bmod (11 \cdot 19), 60^2 \equiv 47 \bmod (11 \cdot 19)$. ✓

Man beachte, dass wir hier benötigen, dass a ein quadratischer Rest modulo 11 und modulo 19 sein muss. Würde man a zufällig wählen, wäre das nicht unbedingt der Fall. Dann ist $x^2 \equiv a \bmod n$ ohnehin unlösbar, falls a kein quadratischer Rest modulo $11 \cdot 19$ ist.

Anwendung 4.16 (Faires Münzwurfknobeln)
Zwei Spieler, Alice (A) und Bob (B), möchten etwas ausknobeln (zum Beispiel, wer beim Fernschach beginnen soll und dann einen Vorteil hat, etc.), allerdings sprechen sie sich am Telefon oder mailen sich, und können sich daher nicht sehen. A wirft eine Münze, und B denkt vorher ”Kopf” oder ”Zahl”, verrät das aber nicht (Würde A die Wahl von B vorher kennen, so würde B das mitgeteilte Ergebnis des Münzwurfs unter Umständen anzweifeln). A teilt B das Ergebnis mit, und B verkündet, wer gewonnen hat: A, wenn ihr Münzwurfergebnis mit der Wahl von B übereinstimmt, ansonsten gewinnt B. Sei B’s geheime Wahl ”Zahl”.
Teilt A mit, dass sie ”Zahl” geworfen hat, akzeptieren A und B den Spielausgang, weil dann A gewinnt und B ihr dies verkündet. Falls A jedoch mitteilt, dass sie ”Kopf” geworfen hat, teilt B mit, dass A verloren habe, was A natürlich nicht akzeptieren würde.
Problem: Wie kann bei Ergebnis ”Kopf” Spieler B seine Mitspielerin A überzeugen, dass er vor dem Münzwurf die Wahl ”Zahl” getroffen hat?
Unsere Antwort: Wenn B dann eine Zahl $n = pq$ faktorisieren könnte, deren Primteiler p, q ansonsten nur A kennt.

Erläuterung 4.17
Das Verfahren funktioniert wie folgt:

Schritt 1: A wählt Primzahlen $p, q \equiv 3 \bmod 4$, $p \neq q$, berechnet $n = pq$ und schickt n an B.

Schritt 2: B wählt $1 \leq b \leq n - 1$ zufällig und behält b geheim, er berechnet $a \equiv b^2 \bmod n$ und schickt a an A.

Schritt 3: A berechnet die vier Lösungen von $x^2 \equiv a \bmod n$ (vgl. 4.14), die vier Lösungen seien $\pm b, \pm c \in \mathbb{Z}$, (mit b von B), die Lösungen $\pm c$ sind andere, die B nicht kennt.

Soweit die Vorbereitung, dann der eigentliche Münzwurf:

Schritt 4: A wählt eine der vier Lösungen zufällig aus (etwa durch Münzwurf!), das heißt entweder $\pm b$ oder $\pm c$, und schickt B das Ergebnis. A kann nicht wissen, dass B die Zahl b gewählt hat. Die Vereinbarung ist nun: Schickt A eine der Zahlen $\pm b$, gewinnt A. Schickt A eine der Zahlen $\pm c$, gewinnt B, und das verkündet B.

Schritt 5: Es erfolgt die Verifikation, dass A wirklich verloren hat im 2. Fall, dazu muss A sich davon überzeugen, dass B vorher wirklich $\pm b$ gewählt hat. Das kann B nun beweisen, indem er ihr die Primfaktoren von n nennt: Er berechnet $b + c \bmod n$ und $d := \mathrm{ggT}(b + c, n)$ mit dem euklidischen Algorithmus. Dann ist $d = p$ oder $d = q$. (Denn aus $b^2 \equiv a \equiv c^2 \bmod pq$ folgt: $pq \mid (b - c)(b + c) = b^2 - c^2$, und da $b \not\equiv c \bmod p, b \not\equiv c \bmod q$ folgt $q \mid b + c$ oder $p \mid b + c$, und $d \neq n$, weil sonst $b \equiv -c \bmod n$ wäre $\lightning$.)
Also kann B, weil er c kennt, die von A gewählten Primfaktoren bestimmen und A mitteilen und auf diese Art A überzeugen. Das konnte B nur, weil er vorher auch wirklich nicht die von A genannte Lösung $\pm b$ hatte. Damit ist das Spiel fair.

In der praktischen Umsetzung wird noch ein Verfahren zur Erzeugung großer, möglichst zufälliger Primzahlen p, q gebraucht. Man kennt in der Praxis schnelle Tests (den Miller-Rabin-Test), um zu entscheiden, ob eine große Zahl n (mit evtl. hunderten von Stellen in Dezimaldarstellung) zusammengesetzt ist oder (sehr wahrscheinlich) prim. Daher erzeugt man solange Zufallszahlen, bis der Primzahltest ”anschlägt”.

1.2 Public-Key-Kryptographie

[5] **Public-Key-Kryptographie** bezeichnet man auch als asymmetrische Kryptographie. Bei diesem Kommunikationsverfahren hat jeder Nutzer einen **öffentlichen Schlüssel**, den jeder einsehen kann, und einen **privaten Schlüssel**, den jeder Nuter geheim hält. Möchte Nutzer B eine Nachricht an Nutzer A senden, benutzt er zur Verschlüsselung den öffentlichen Schlüssel von A, die Entschlüsselung gelingt aber nur A mit dem privaten Schlüssel.
Ein solches Szenario (auch **Protokoll** genannt) ist das RSA-Verfahren, das wir in Abschnitt 1.2.1 behandeln. Die Verfahren in 1.2.2 und 1.2.3 sind Kryptographie-Verfahren, die mit allgemeinen Gruppen machbar sind (RSA arbeitet in $(\mathbb{Z}_n^*, \cdot)$).

1.2.1 RSA-Verfahren

Das **RSA-Verfahren** ist benannt nach einer Arbeit von Rivest[8], Shamir[9] und Adleman[10] aus dem Jahr 1978. Seine Sicherheit beruht auf der Schwierigkeit des Faktorisierungsproblems und wird bis heute zur sicheren Kommunikation benutzt.
Die Methode verlangt auch die Möglichkeit, große Primzahlen zu erzeugen, die möglichst zufällig gewählt sein sollen, ähnlich wie beim Münzwurfproblem. $n = pq$ muss so groß sein, dass alle bekannten Faktorisierungsverfahren zu langsam wären.

Anwendung 5.1 (Durchführung des RSA-Verfahrens)
Die beiden Protagonisten heißen wieder Nutzer Alice (A) und Bob (B). Sie kommunizieren über einen unsicheren Kanal miteinander.

Schritt 1: Jeder Nutzer, z.B. (A), wählt zwei große Primzahlen $p \neq q$, etwa gleich groß mit ähnlicher Stellenanzahl, und berechnet $n = pq$ sowie $\varphi(n) = (p-1)(q-1)$. Dann wählt (A) eine Zahl e mit $1 < e < \varphi(n)$ und berechnet $1 < d < \varphi(n)$ als Inverses von $e \mod \varphi(n)$, das heißt $de \equiv 1 \bmod \varphi(n)$, unter Zuhilfenahme des euklidischen Algorithmus.

Schritt 2: Bob möchte Alice seinen Geheimtext (als eine Zahl x kodiert) schicken. Er besorgt sich die Daten n, e vom Server und verschlüsselt x zu $x^e \bmod n$. Dann schickt er ihr das Ergebnis $v \equiv x^e \bmod(n)$ zwischen 1 und n.

Schritt 3: Alice entschlüsselt den geheimen Text v durch Berechnen von $v^d \bmod n$, sie erhält x, weil für ein $k \in \mathbb{Z}$ gilt: $ed = 1 + k \cdot \varphi(n)$, also folgt mit Euler-Fermat, falls $\mathrm{ggT}(x, n) = 1$:

$$v^d \equiv (x^e)^d \equiv x^{1+k\cdot\varphi(n)} \equiv x \cdot \underbrace{\left(x^{\varphi(n)}\right)^k}_{\equiv 1 \bmod n} \equiv x \bmod n$$

Bemerkung 5.2
Die nötigen Berechnungen sind: schnelles modulares Potenzieren modulo n, d.h. Berechnungen in der multiplikativen Gruppe $(\mathbb{Z}_n^*, \cdot)$, Berechnen von d mit dem euklidischen Algorithmus und Erzeugen großer Primzahlen p, q.

Bemerkung 5.3
Ein Unbefugter, der die Daten n, e, v dieser Kommunikation abfängt, ist nicht in der Lage, x ohne Kenntnis von $d, p, q, \varphi(n)$ zu berechnen. Dazu müsste man n faktorisieren.

[8] http://de.wikipedia.org/wiki/Ronald_L._Rivest
[9] http://de.wikipedia.org/wiki/Adi_Shamir
[10] http://de.wikipedia.org/wiki/Leonard_Adleman

Bemerkung 5.4
Wie sicher das Verfahren ist, hängt davon ab, wie groß die verwendeten Schlüssel sind. Aktuell ist eine Verschlüsselung, bei der p, q eine Bitlänge von mindestens 512 haben sollten, besonders sicher: 2048 Bit. Empfehlung der Bundesnetzargentur bis Ende 2020: mindestens 1976 Bit. Gegen einen Angriff mit einem Quantencomputer hat man allerdings keine Chance.

Bemerkung 5.5
Auch in den seltenen Fällen $p \mid x$ oder $q \mid x$, das heißt $\mathrm{ggT}(x, n) > 1$, arbeitet das Verfahren korrekt (ohne Beweis).

Bemerkung 5.6
Das Verfahren kann auch ohne Schlüsselserver benutzt werden. B kann A erst mitteilen, dass er ihr eine Nachricht schicken will. Dann erst erledigt A Schritt 1 und teilt ihm die Daten n, e mit. Der Rest geht dann wie oben.

Bemerkung 5.7 (Zur Geschichte von RSA)
RSA wurde 1983 als Patent angemeldet, welches 2000 erlosch. Bis Ende der 90er Jahre verbot die US-Regierung Firmen, Software mit starker Verschlüsselung zu exportieren (z.B. T-Shirts mit aufgedruckter RSA-Anleitung...). Weiter sollten per Gesetzesvorlage Anbieter elektronischer Kommunikationsdienste dazu verpflichtet werden, Behörden die Möglichkeit zum Zugriff zu verschaffen; das Gesetz scheiterte am Widerstand von Industrie und Bürgerrechtlern. Es motivierte Phil Zimmermann[11] dazu, den Standard **PGP** (pretty good privacy) zu entwickeln, mit dem bis heute E-Mails und anderes für Jedermann sicher verschlüsselt werden können (speziell mit RSA; öffentliche Schlüsselserver dafür gibt es im Internet, z.B. auf pgp.mit.edu). Zimmermannn stellte sein Programm 1991 kostenlos zur Verfügung. Es wurde ein Verfahren gegen ihn eröffnet, das sich über drei Jahre lang hinzog (Vorwurf: er exportierte Verschlüsselungstechnologie, die wie Waffentechnologie einzustufen sei). Der Fall wurde fallengelassen, heute ist die Benutzung und Export in den USA straffrei. Bis heute zählt PGP als sicherste und empfehlenswerteste Verschlüsselung privater Kommunikation.

Anwendung 5.8 (Kodierung von Textnachrichten)
Wir beschreiben hier ein Verfahren, das die Machbarkeit der Kodierung "Text $\to$ Zahl" demonstrieren soll. Wenn man es so anwenden möchte, sind aber größere Blöcke erforderlich, damit nicht durch Häufigkeitsanalysen der Blöcke Rückschlüsse auf die Geheimnachricht möglich werden.
Die Buchstaben A, ..., Z des Alphabets werden mit $0, \ldots, 25$ identifiziert, das Leerzeichen mit 26. Klartexte werden zu Blöcken aus je drei Zahlen zusammengefasst, also z.B.

$$\texttt{KLARTEXT\_} \quad \Rightarrow \quad 10, 11, 0 | 17, 19, 4 | 23, 19, 26$$

Jedem Block x_1, x_2, x_3 ordnen wir die Zahl $x = x_1 \cdot 27^2 + x_2 \cdot 27 + x_3$ (im 27er System) zu, also

$$\texttt{KLARTEXT\_} \quad \Rightarrow 7587 | 12910 | 17306,$$

welche beim RSA-Verfahren gemäß $x^e \equiv v \bmod n$ verschlüsselt wird. Jeder Wert v wird im 29er-System umgewandelt gemäß $v = v_1 \cdot 29^2 + v_2 \cdot 29 + v_3$ zu einem Block $v_1, v_2, v_3 \in \{0, \cdot, 28\}$, der wieder als Text geschrieben werden kann (mit zusätzlichen Zeichen für 27 und 28, z.B. "."=27, ","=28).
Ist n zwischen 27^3 und 29^3, werden Ver- und Entschlüsselung eindeutig (ohne Beweis) $\rightsquigarrow$ für größere n werden größere Blöcke nötig!

1.2.2 Diffie-Hellman-Verfahren

Definition 5.9 (Das Problem des diskreten Logarithmus (DL-Problem))
Gegeben sei eine abelsche Gruppe. Wir beschreiben das Problem multiplikativ und additiv:

[11] http://de.wikipedia.org/wiki/Phil_Zimmermann

In $(G, \cdot, 1)$: Sei $x \in G$, $n = \mathrm{ord}(x), y \in \langle x \rangle = \{x^l : l \in \mathbb{Z}\}$. Bestimme $k \bmod n$ mit $y = x^k$.

In $(G, +, 0)$: Sei $x \in G, n\,\mathrm{ord}(x), y \in \langle x \rangle = \{lx : l \in \mathbb{Z}\}$. Bestimme $k \bmod n$ mit $y = kx$.

Bemerkung 5.10
Ist eine Gruppe G gegeben, in der das DL-Problem schwer ist, kann dies für ein Kryptoverfahren genutzt werden.

- Im Fall $G = (\mathbb{Z}_m^*, \cdot, 1)$ ist das DL-Problem ähnlich schwer wie das Faktorisierungsproblem. Auch dafür konnte Short 1994 zeigen, dass es auf einem Quantencomputer schnell lösbar ist.
- Im Fall, dass $G = (E(k), +, \mathcal{O})$ die Gruppe einer (kryptographisch) geeigneten elliptischen Kurve ist, ist das DL-Verfahren quasi unlösbar. Die besten bekannten Algorithmen sind langsamer als die für das DL-Problem für $\mathbb{Z}_m^*$. Darauf beruht die als höher angesehene Sicherheit bei der Kryptographie mit elliptischen Kurven. Algorithmen auf Quantencomputern, die das DL-Problem für elliptische Kurven schnell lösen könnten, sind derzeit unbekannt.

Anwendung 5.11 (Diffie-Hellman-Schlüsselaustausch)
Hier vereinbaren Alice (A) und Bob (B) durch einen öffentlichen Kanal einen gemeinsamen geheimen Schlüssel, die sie dann für ein symmetrisches Kryptoverfahren nutzen können. Gegeben sei eine Gruppe G und $x \in G$, sowie $n \in \mathbb{N}$. Diese Daten seien öffentlich bekannt.
Das Verfahren in $(G, \cdot, 1)$:

Schritt 1: Alice denkt sich eine Zahl $a \in \{1, \ldots, n-1\}$ und schickt $x^a \in G$ an Bob.
Bob denkt sich eine Zahl $b \in \{1, \ldots, n-1\}$ und schickt $x^b \in G$ an Alice.

Schritt 2: Alice berechnet mit a das Gruppenelement $(x^b)^a$.
Bob berechnet mit b das Gruppenelement $(x^a)^b$.
Danach besitzen beide den gemeinsamen geheimen Schlüssel $(x^b)^a = x^{ab} = x^{ba}$.

Das Verfahren in $(G, +, 0)$:

Schritt 1: Alice denkt sich eine Zahl $a \in \{1, \ldots, n-1\}$ und schickt $ax \in G$ an Bob.
Bob denkt sich eine Zahl $b \in \{1, \ldots, n-1\}$ und schickt $bx \in G$ an Alice.

Schritt 2: Alice berechnet mit a das Gruppenelement $a \cdot (bx)$.
Bob berechnet mit b das Gruppenelement $b \cdot (ax)$.
Danach besitzen beide den gemeinsamen geheimen Schlüssel $a \cdot (bx) = abx = b \cdot (ax)$.

Bemerkung 5.12 (Diffie-Hellman-Problem, DH-Problem)
Ein Unbefugter, der die Daten x^a, x^b bzw. ax, bx abhört, kann die geheimen Schlüssel berechnen, wenn er das DL-Problem lösen kann. Es genügt aber schon, dafür das folgende, eventuell leichtere Problem zu lösen:
Berechne zu $x^a, x^b \in \langle x \rangle \subseteq G$ in $(G, \cdot, 1)$ das Element $x^{ab} \in \langle x \rangle$.
Es ist aber davon auszugehen, dass auch DH ein schweres Problem ist.
(Bemerkung: DL lösbar $\Rightarrow$ DH lösbar ist klar, "$\Leftarrow$" ist unbekannt.)

Bemerkung 5.13
Weiter ist beim Schlüsselaustausch entscheidend, dass sich Alice und Bob sicher sein können, wirklich mit dem angegebenen Teilnehmer zu kommunizieren: Ein Unbefugter könnte versuchen, sich erst als Alice auszugeben, und so mit Bob einen Schlüssel x^{eb} auszutauschen und dies Ebenso mit Alice tun (x^{ea}). Gelingt dies, braucht der Unbefugte nur die verschlüsselten Nachrichten zwischen Alice und Bob abzufangen:
Die Nachrichten von Alice an Bob dekodiert er mit dem Alice-Schlüssel x^{ea}, schickt sie mit dem Bob-Schlüssel x^{eb} kodiert an Bob weiter, und umgekehrt. Er kann so die gesamte geheime Kommunikation abhören. Man nennt dies eine **man-in-the-middle-Attacke**.

1.2.3 ElGamal-Verschlüsselung

Allen Teilnehmern bekannt sei eine abelsche Gruppe $(G, +)$ und ein Gruppenelement $x \in G$ von (großer) Ordnung $n = \mathrm{ord}(x)$. Jeder Nutzer wählt eine Zufallszahl $d \in \{1, \ldots, n-1\}$ als privaten Schlüssel und erzeugt einen öffentlichen Schlüssel dx. [6]

Anwendung 6.1 (ElGamal-Verschlüsselung)
Alice möchte eine geheime Botschaft $m \in G$ an Bob schicken. Die **ElGamal-Verschlüsselung**[12] geht wie folgt:

Schritt 1: Alice wählt eine Zufallszahl $\tilde{a} \in \{1, \ldots, n-1\}$ und berechnet $\tilde{a} \cdot x$. Alice besorgt sich Bobs öffentlichen Schlüssel bx und berechnet $R = \tilde{a} \cdot (bx) + m$.

Schritt 2: Alice schickt $\tilde{a}x$ und R an Bob.

Schritt 3: Bob berechnet $b \cdot (\tilde{a}x) = \tilde{a} \cdot (bx)$ und die Nachricht durch $R - b \cdot (\tilde{a}x) = m$.

Bemerkung 6.2
Ein Unbefugter, der die Daten $G, x, n, bx, \tilde{a}x$ kennt und R abgehört hat, kann m genau dann berechnen, wenn er ein Diffie-Hellman-Problem lösen kann (d.h. das Element $\tilde{a}b \cdot x \in G$ berechnen kann).

Bemerkung 6.3
Alice könnte $\tilde{a} = a$ wählen. Für die Sicherheit dieses Verfahrens ist es aber wichtig, dass sie bei jeder ihrer Nachrichten ein neues $\tilde{a}$ wählt: Sonst könnte ein Unbefugter, der die Übertragungen $\tilde{a}x, R_1 = \tilde{a}(bx) + m_1$ und $\tilde{a}x, R_2 = \tilde{a}(bx) + m_2$ abhört und schon die Nachricht m_1 kennt, über $R_2 - R_1 + m_1 = (m_2 - m_1) + m_1 = m_2$ auch m_2 berechnen.

[12] http://de.wikipedia.org/wiki/Taher_Elgamal

1.3 Digitale Unterschriften

1.3.1 DSA-Signatur

Gegeben sei wieder eine abelsche Gruppe $(G,+), x \in G$ mit $n = \operatorname{ord}(x)$ groß. Alice will eine Nachricht m an Bob digital unterschreiben. Wieder hat sie einen geheimen Schlüssel $a \in \{1,\dots,n-1\}$ und einen öffentlichen Schlüssel $ax \in G$.

Definition 6.4 (Hashfunktion)
Sei $\mathcal{M}$ die Menge aller möglichen Nachrichten (etwa beliebig lange Folgen von 0 und 1), und gegeben sei eine Funktion $h\colon \mathcal{M} \to \{0,1,\dots,n-1\}$, deren Werte $h(m)$ für $m \in \mathcal{M}$ leicht zu berechnen sind und die die folgenden beiden Eigenschaften hat:

(i) Es ist praktisch unmöglich, Urbilder unter h zu berechnen, d.h. zu $d \in \{0,1,\dots,n-1\}$ ein $m \in \mathcal{M}$ zu finden mit $h(m) = d$.

(ii) h ist **kollisionsresistent**, das bedeutet, dass es praktisch unmöglich ist, zwei verschiedene Elemente $m, m' \in \mathcal{M}$ mit $h(m) = h(m')$ zu finden. Eine solche Funktion heißt **Hashfunktion**.

Beispiel 6.5
Sei p prim mit $2^{1023} < p \leq 2^{1024} - 1$ und g ein Erzeuger der multiplikativen Gruppe $\mathbb{Z}_p^*$, d.h. $\langle g \rangle = \mathbb{Z}_p^*$. Dann ist nach heutigem Wissen die Funktion

$$\begin{aligned} h\colon \mathbb{Z}_p^* &\longrightarrow \mathbb{Z}_p^* \\ z &\longmapsto g^z \bmod p \end{aligned}$$

eine Hashfunktion. Das ab 6.7 beschriebene Verfahren kann dann mit $G = \mathbb{Z}_p^*, x = g$ durchgeführt werden (in der Praxis nimmt man für p eine **Sophie-Germain-Primzahl**[13], d.h. p prim mit $\frac{p-1}{2}$ auch prim, denn dann ist etwa jedes zweite Element ein Erzeuger. Daher ist leicht ein Erzeuger findbar.

Bemerkung 6.6
Öffentlich zugänglich seien die Daten $(G,+), x \in G, n = \operatorname{ord}(x), h$ und $ax \in G$ sowie eine Bijektion $\Psi\colon \langle x \rangle \to \{0,1,\dots,n-1\}$, deren Werte effektiv berechenbar seien (in der Praxis reicht eine Funktion, deren Urbildmenge $\Psi^{-1}(k)$ von jedem $k \in \{0,\dots,n-1\}$ klein ist).

Anwendung 6.7 (DSA-Verfahren)
Beschreibung für eine additive Gruppe G!

Nun das **DSA-Verfahren** zur Signatur, wie Alice ihre Nachricht m unterschreiben kann:

Schritt 1: Alice wählt eine Zufallszahl $\tilde{a} \in \{1,\dots,n-1\}$ mit $\operatorname{ggT}(\tilde{a}, n) = 1$ und berechnet das Gruppenelement $\tilde{a}x \in G$.

Schritt 2: Alice berechnet das Inverse $\tilde{a}^{-1}$ in $\mathbb{Z}_n$ (euklidischer Algorithmus) sowie $s := \tilde{a}^{-1}(h(m) - \Psi(\tilde{a}x) \cdot a)$ in $\mathbb{Z}_n$.

Schritt 3: Alice schickt die Nachricht m und ihre Unterschrift $\tilde{a}x, s$ an Bob.

Schritt 4: Bob berechnet $\Psi(\tilde{a}x) \cdot ax + s\tilde{a}x$ sowie den Hashwert $h(m)$. Bob akzeptiert die Unterschrift als echt, wenn $\Psi(\tilde{a}x)ax + s\tilde{a}x = h(m) \cdot x$ in G ist, was nur stimmt, wenn $\Psi(\tilde{a}x)a + s\tilde{a} \equiv h(m) \bmod m$ gewählt ist, da ja $n = \operatorname{ord}(x)$ in G gilt.

[13] http://de.wikipedia.org/wiki/Sophie_Germain

Bemerkung 6.8
Kann hier ein Unbefugter die Unterschrift von Alice fälschen? Dazu müsste er s, k, x finden mit $\Psi(kx)ax + skx = h(m)x$ für ein beliebiges k anstelle $\tilde{a}$. Er würde kx berechnen und s passend wählen, wofür ein DL-Problem in $\langle x \rangle \subseteq G$ zu lösen wäre, denn a kennt er nicht.

Bemerkung 6.9
Auch hier ist für die Sicherheit des Verfahrens nötig, dass Alice für jede Unterschrift ein neues $\tilde{a}$ wählt: erzeugt Alice zwei Unterschriften $(\tilde{a}x, s_1)$ für m_1 und $(\tilde{a}x, s_2)$ für m_2, ist $s_2 - s_1 \equiv \tilde{a}^{-1}(h(m_2) - h(m_1)) \bmod n$. Wenn $h(m_2) - h(m_1)$ invertierbar in $\mathbb{Z}_n$ ist, kann der Unbefugte $\tilde{a} \bmod n$ berechnen. Wegen $\Psi(\tilde{a}x)a \equiv h(m_1) - s_1\tilde{a} \bmod n$ ist dann auch a berechenbar, falls $\Psi(r_1)$ invertierbar in $\mathbb{Z}_n$ ist.

Bemerkung 6.10
Wozu eine Hashfunktion h?

- Könnte man leicht Urbilder unter h berechnen, ist das Unterschriftenfälschen einfach: Der Unbefugte wählt $j \in \mathbb{Z}$ beliebig und berechnet $r = jx - ax, s = \Psi(r)$ und bestimmt m (nicht von Alice!) mit $h(m) \equiv \Psi(r)j \bmod n$. Dann ist r, s eine für Bob verifizierbare Unterschrift der falschen Nachricht m, denn es gilt

$$\Psi(r)ax + \underbrace{\Psi(r)}_{s}\underbrace{(jx - ax)}_{r} = \Psi(r)jx = h(m)x$$

- Wäre h nicht kollisionsresistent und ein Auffinden von $m' \in \mathcal{M}$ mit $h(m) = h(m')$ leicht, kann man Alice' Unterschrift unter m' fälschen, wenn man eine gültige Unterschrift $\tilde{a}x, s$ für m hat, da

$$\Psi(\tilde{a}x)ax + s\tilde{a}x = h(m) \cdot x = h(m')x$$

Bemerkung 6.11
Bob muss sicher sein, dass Alice' öffentlicher Schlüssel ax auch wirklich von Alice stammt und nicht von einem Unbefugten gefälscht wurde. Man löst das Problem, indem sich jeder Nutzer bei einer Certification Authority (CA) registrieren lässt. Bob würde von dieser eine "beglaubigte Kopie" von Alice' öffentlichen Schlüssel erhalten; Einzelheiten vgl. Fachliteratur.

Motivation 6.12
Eine auf Koblitz und Miller zurückgehende Idee ist nun, dass für die ElGamal-Verfahren eine beliebige zyklische Gruppe $\langle x \rangle$ verwendbar ist, wie etwa die, die von Punkten auf elliptischen Kurven erzeugt werden. Da für (geeignete) elliptische Kurven das DL-Problem bzw. DH-Problem schwieriger für $\mathbb{Z}_m^*$ ist, gilt diese Art von Verschlüsselungstechnik heute als besonders sicher und wird vielfältig industriell angewendet. Wir werden die Mathematik elliptischer Kurven im folgenden Abschnitt der Vorlesung näher kennenlernen.

2 Elliptische Kurven

2.1 Grundlagen aus der Algebra

2.1.1 Polynome

[7] Sei k ein beliebiger Körper.

Definition 7.1 (Polynom)
Ein **Polynom** über k in den n Variablen $x_1, \ldots, x_n$ ist ein Ausdruck der Form

$$f(x_1, \ldots, x_n) = \sum_{\nu_1, \ldots, \nu_n \geq 0} \alpha_{\nu_1, \ldots, \nu_n} x_1^{\nu_1} \cdots x_n^{\nu_n}$$

mit Koeffizienten $\alpha_{\nu_1,\ldots,\nu_n} \in k$, von denen nur endlich viele $\neq 0$ sind. Hat man es mit mehreren Variablen ($n \geq 2$) zu tun, kann man auch kurz

$$f(\underline{x}) = \sum_{\underline{\nu} \in \mathbb{N}_0^n} \alpha_{\underline{\nu}} x_1^{\nu_1} \cdots x_n^{\nu_n}$$

schreiben, wenn man die Tupelschreibweise $\underline{\nu} \in \mathbb{N}_0^n$ bzw. $\underline{x} = (x_1, \ldots x_n)$ einführt, wobei man für das Monom $x_1^{\nu_1} \cdots x_n^{\nu_n}$ auch kurz $\underline{x}^{\underline{\nu}}$ schreiben kann, wenn klar ist, dass $n \geq 2$ viele Variablen vorliegen.
Die Menge aller Polynome über k in n Variablen wird kurz mit $k[x_1, \ldots, x_n]$ oder noch kürzer mit $k[\underline{x}]$ bezeichnet. Wir schreiben dann auch kurz $f \in k[\underline{x}]$, wenn $f(\underline{x})$ ein Polynom ist.

Bemerkung 7.2
Durch eine Addition und Multiplikation definiert durch

$$\sum_{\underline{\nu}} \alpha_{\underline{\nu}} \underline{x}^{\underline{\nu}} + \sum_{\underline{\nu}} \beta_{\underline{\nu}} \underline{x}^{\underline{\nu}} := \sum_{\underline{\nu}} (\alpha_{\underline{\nu}} + \beta_{\underline{\nu}}) \underline{x}^{\underline{\nu}}$$

$$\left(\sum_{\underline{\nu}} \alpha_{\underline{\nu}} \underline{x}^{\underline{\nu}} \right) \cdot \left(\sum_{\underline{\mu}} \beta_{\underline{\mu}} \underline{x}^{\underline{\mu}} \right) := \sum_{\underline{\nu}, \underline{\mu}} \alpha_{\underline{\nu}} \beta_{\underline{\mu}} \underline{x}^{\underline{\nu} + \underline{\mu}}$$

wird $k[\underline{x}]$ zu einem kommutativen Ring mit Eins; das Nullpolynom $0 := \sum_{\underline{\nu}} 0 \underline{x}^{\underline{\nu}}$ ist dabei das Nullelement, das
"Einspolynom" Polynom $1 := 1 \cdot \underline{x}^{\underline{0}} + \sum_{\underline{\nu} \neq 0} 0 \underline{x}^{\underline{\nu}}$ ist das Einselement.
Der Ring $(k[\underline{x}], +, \cdot)$ heißt **Polynomring** über k.

Definition 7.3 (Formale Ableitung)
Für $f(\underline{x}) = \sum_{\underline{\nu}} \alpha_{\underline{\nu}} \underline{x}^{\underline{\nu}} \in k[\underline{x}]$ und $1 \leq j \leq n$ heißt

$$\frac{\partial f}{\partial x_j}(\underline{x}) := \sum_{\underline{\nu}, \nu_j > 0} \alpha_{\underline{\nu}} \nu_j x_1^{\nu_1} \cdots x_j^{\nu_j - 1} \cdots x_n^{\nu_n} \in k[\underline{x}]$$

die **formale Ableitung** von f nach x_j.

Satz 7.4 (Produktregel, Kettenregel)
Für alle $f, g \in k[\underline{x}]$ und $\gamma \in k$ gelten die Ableitungsregeln

$$\frac{\partial(\gamma f)}{\partial x_j} = \gamma \frac{\partial f}{\partial x_j} \qquad \frac{\partial(f+g)}{\partial x_j} = \frac{\partial f}{\partial x_j} + \frac{\partial g}{\partial x_j} \qquad \frac{\partial(fg)}{\partial x_j} = f \frac{\partial g}{\partial x_j} + g \frac{\partial f}{\partial x_j}$$

und für $f \in k[x_1, \ldots, x_m], g_1, \ldots, g_m \in k[x_1, \ldots, x_n]$

$$\frac{\partial f(g_1, \ldots, g_m)}{\partial x_j}(\underline{x}) = \frac{\partial f}{\partial x_1}(g_1, \ldots, g_m) \frac{dg_1}{dx_j}(\underline{x}) + \ldots + \frac{\partial f}{\partial x_m}(g_1, \ldots, g_m) \frac{\partial g_m}{\partial x_j}(\underline{x}).$$

Polynome in einer Variablen $f \in k[x]$ der Form $f(x) = \sum_{\nu=0}^{k} \alpha_\nu x^\nu$ sind aus den Grundvorlesungen bekannt.

Definition 7.5 (Grad)
Ist $f \neq 0$, so heißt $\deg(f) := \min\{j \in \mathbb{N}_0 : a_j \neq 0\}$ der Grad von f. Für $f \in k[\underline{x}]$ in n Variablen ist $\deg(f) := \min\{\nu_1 + \ldots + \nu_n : a_{\underline{\nu}} \neq 0\}$ der **Grad** von f. Neu ist bei uns, dass wir uns hier vor allem mit $n = 2$ oder $n = 3$ Variablen beschäftigen werden, wo wir dann auch $f(x, y)$ oder $f(x, y, z)$ schreiben möchten, zum Beispiel $f(x, y) = \alpha_{(2,0)}x^2 + \alpha_{(1,1)}xy + \alpha_{(0,1)}y$. Wir werden dann für die Koeffizienten einfachere Notationen wählen.

Bemerkung 7.6
Bleiben wir zunächst beim Polynomring $k[x]$ in einer Variablen x. Sei $f \in k[x]$. Wie im Ring $\mathbb{Z}$ können wir Teilbarkeit in $k[x]$ studieren und Divisionen mit Rest durchführen (Polynomdivision), daher kann man wie in $\mathbb{Z}$ zum Beispiel den ggT von Polynomen mit dem euklidischen Algorithmus ausrechnen. Dies ist aus den Grundvorlesungen bekannt, wir erinnern hier nur an folgendes:

Definition 7.7 (Nullstelle)
Gegeben sei die Einsetzabbildung

$$\begin{aligned} k &\longrightarrow k \\ c &\longmapsto f(c) := \sum_{\nu=0}^{k} \alpha_\nu c^\nu \end{aligned}$$

Ein Element $c \in k$ heißt **Nullstelle** von f, falls $f(c) = 0$ in k ist.

Bemerkung 7.8
$c \in k$ ist genau dann Nullstelle, wenn $(x - c)$ ein Teiler von f im Polynomring $k[x]$ ist, d.h. falls ein $g \in k[x]$ existiert mit $(x - c) \cdot g = f$.

Definition 7.9 (Ordnung einer Nullstelle)
Ist c eine Nullstelle von $f \neq 0$, so gibt es ein maximales $k \geq 1$, sodass $(x - c)^k$ ein Teiler von f ist. Die Zahl k heißt **Ordnung der Nullstelle** c. Ist $f(c) \neq 0$, definiert man diese "Nullstellen"ordnung als 0.

Definition 7.10 (irreduzibel, prim)
Ein Polynom $f \in k[x]$ vom Grad ≥ 1 heißt **irreduzibel** (oder **prim**), falls gilt: Ist $f = u \cdot v$ mit $u, v \in k[x]$, dann ist $\deg(u) = 0$ oder $\deg(v) = 0$, das heißt f kann nicht als Produkt zweier Polynome vom Grad ≥ 1 geschrieben werden. (vgl. den Begriff "Primzahl" bei $\mathbb{Z}$; der Satz von der eindeutigen Zerlegung in irreduzible Polynome heißt der **Satz von Gauß**.)

Wenn wir $\mathbb{Z}$ als Vorbild für den Polynomring $k[x]$ nehmen, möchten wir auch das "Modulorechnen" auf $k[x]$ übertragen, um neue Strukturen zu erhalten. Unsere Moduln sind dann Polynome:

Definition 7.11 (Kongruenz, Restklassenring (Polynome))
Sei $f \in k[x]$. Dann heißen $a \in k[x]$ und $b \in k[x]$ **kongruent modulo** f, wenn $f \mid (b - a)$, das heißt falls ein $g \in k[x]$ existiert mit $b = a + fg$. Die Restklassen modulo f sind Teilmengen von $k[x]$ der Gestalt $a + f \cdot k[x] := \{a + fg : g \in k[x]\}$ mit $a \in k[x]$. Das Polynom $a \in k[x]$ heißt ein **Repräsentant** der Restklasse. Ist der Modul $f \in k[x]$ klar, möchten wir dafür auch kurz wieder $\underline{\underline{a}}$ schreiben.
"≡" nur für $\mathbb{Z}$

doppelt unterstreichen!

Die Menge der Restklassen modulo f bezeichnen wir mit

$$k[x]/(f) := \{a + f \cdot k[x] : a \in k[x]\} = \{\underline{\underline{a}} : a \in k[x]\}$$

und nennen diese den **Restklassenring modulo** f, weil diese bezüglich der Definition $\underline{\underline{a}} + \underline{\underline{b}} := \underline{\underline{a + b}}$ (analog für Multiplikation) für Polynome $a, b \in k[x]$ wieder zu einem kommutativen Ring mit $\underline{\underline{1}}$ als Eins wird.

Doch die einfache Frage, wie viele Elemente der Restklassenring hat, hängt unter anderem vom Körper k ab. Im Fall $k = \mathbb{F}_p$ beantworten wir diese. Klar ist wegen der Teilbarkeit mit Rest im Ring $k[x]$ (d.h. sind $b, f \in k[x]$ und $f \neq 0$, so existieren eindeutige $g, r \in k[x]$ mit $r = 0$ oder $\deg(r) < \deg(f)$, sodass $b = f \cdot g + r$ gilt):

Polynomdivision

Bemerkung 7.12
Für jede Restklasse $\underline{\underline{a}} = a + f \cdot k[x] \in k[x]/(f)$ gibt es genau einen Vertreter $b \in \underline{\underline{a}} = a + f \cdot k[x]$, das heißt $\underline{\underline{b}} = \underline{\underline{a}}$ bzw. $b + f \cdot k[x] = a + f \cdot k[x]$, mit $b = 0$ oder $\deg(b) < \deg(f)$.

2.1.2 Endliche Körper

Sei nun $k = \mathbb{F}_p$ mit p prim.

Satz 7.13
Sei $f \in \mathbb{F}_p[x]$ irreduzibel mit $r := \deg(f)$. Dann ist $\mathbb{F}_p[x]/(f)$ ein Körper mit p^r Elementen.

Beweis
Dass $\mathbb{F}_p[x]/(f)$ ein Körper ist, ist klar (Inverse findet man mit dem euklidischen Algorithmus). $\mathbb{F}_p[x]$ hat p^r Elemente, denn jede Restklasse hat genau einen Vertreter

$$b = \underbrace{\alpha_0 + \ldots + a_{r-1}x^{r-1}}_{p \text{ Möglichkeiten für jedes } \alpha_j}$$

□

Bemerkung 7.14
Für jedes $r \in \mathbb{N}$ gibt es (mindestens) ein irreduzibles Polynom $f \in \mathbb{F}_p[x]$ mit $\deg(f) = r$.

Bemerkung 7.15
Es gibt im Wesentlichen (das heißt bis auf Isomorphie) genau einen endlichen Körper mit p^r Elementen, das heißt welches irreduzible f mit $\deg(f) = r$ wir als Modul nehmen, ist für seine Konstruktion (bis auf Isomorphie!) egal. Wir bezeichnen diesen Körper mit $\mathbb{F}_{p^r}$.

Bemerkung 7.16
Jeder Körper mit endlich vielen Elementen ist einer dieser Körper $\mathbb{F}_{p^r}$ mit p prim und $r \geq 1$. (ohne Beweis, vgl. Vorlesung "Einführung in die Algebra")

Bemerkung 7.17
Wegen Bemerkung 7.12 ist nach Wahl eines irreduziblen Polynom $f \in \mathbb{F}_p[x], \deg(f) = r$ also

$$\mathbb{F}_{p^r} = \{(\alpha_{r-1}x^{r-1} + \ldots + \alpha_1 x + \alpha_0) + f \cdot \mathbb{F}_p[x] : \alpha_i \in \mathbb{F}_p\},$$

die Restklassenvertreter $\alpha_{r-1}x^{r-1}+\ldots+\alpha_1 x+\alpha_0$ lassen sich auch durch Koeffizienten-r-Tupel $(\alpha_{r-1}, \alpha_{r-2}, \ldots, \alpha_1, \alpha_0) \in \mathbb{F}_{p^r}$ darstellen. Will man mit ihnen stellvertretend für die Polynomrestklassen in $\mathbb{F}_{p^r}$ rechnen, muss man also erst mit den zugehörigen Polynomen über $\mathbb{F}_p$ rechnen und modulo f reduzieren.

Beispiel 7.18
Unterstreichungen weggelassen!
Sei $p = 2, r = 3$, wir möchten $\mathbb{F}_8$ konstruieren. Das Polynom $f(x) = x^3 + x + 1$ ist irreduzibel über $\mathbb{F}_2 = \{0, 1\}$, also ist

$$\mathbb{F}_8 = \mathbb{F}_2[x]/(f) = \{(0,0,0), (0,0,1), (0,1,0), (0,1,1), (1,0,0), (1,0,1), (1,1,0), (1,1,1)\},$$

und man rechnet zum Beispiel $(0,1,0)\cdot(1,1,1)=(1,0,1)$, weil

$$(0x^2+1x+0)\cdot(x^2+x+1)=x^3+x^2+x=1\cdot(x^3+x+1)+(x^2+1)$$

in $\mathbb{F}_2[x]$ gilt (Division mit Rest durch f).

- Bei Wahl des irreduziblen Polynoms $f(x)=x^3+x^2+1$ ergeben sich zwar andere Rechenregeln für die Vektorenmultiplikation, man erhält aber die selbe "Struktur" bei $+,\cdot$ mit entsprechenden Elementen. Stellen Sie als Übung mal die Multiplikations- und Additionstabellen auf, der Einfachheit halber auch erst mal von $\mathbb{F}_4$.
- Streng genommen müsste man zum Beispiel $\underline{\underline{(\underline{1},\underline{0},\underline{1})}}=\underline{\underline{x^2+\underline{1}}}$ für die Elemente von $\mathbb{F}_8$ schreiben, um die Reduktion modulo f zu verdeutlichen.

Beispiel 7.19

Rechnen in $\mathbb{F}_{5^3}=\mathbb{F}_{125}$: Haben wir diesen Körper mit dem irreduziblen Polynom $f=x^3+x+1\in\mathbb{F}_5[x]$ vom Grad 3 konstruiert, so rechnen wir in $\mathbb{F}_{5^3}$ zum Beispiel

irreduzibel, da keine Nullstelle und Grad 3!

$$\begin{aligned}(1,2,4)\cdot(-1,3,0)&=(x^2+2x-1)(-x^2+3x)=-x^4+3x^3-2x^3+6x+x^2-3x\\&=-x^4+x^3+x^2+3x=(x^3+x+1)\cdot(-x+1)+2x^2+3x+1=(2,3,1)\bmod f\end{aligned}$$

Bemerkung 7.20

Es ist $\mathrm{char}(\mathbb{F}_{p^r})=p$, denn es gilt $\underline{1}+\underline{1}+\ldots+\underline{1}=\underline{p\cdot 1}=\underline{0}$, und p ist minimal mit dieser Eigenschaft, da prim.

Definition 7.21 (algebraisch abgeschlossen)

Ein Körper k ist **algebraisch abgeschlossen**, wenn sich jedes Polynom $f\in k[x],\deg(f)>0$, als Produkt von linearen Polynomen schreiben lässt, das heißt wenn $f(x)=d(x-c_1)\cdots(x-c_m)$ mit $d,c_i\in k$ gilt.

Bemerkung 7.22

Man kann jeden Körper k in einen algebraisch abgeschlossenen Körper einbetten. Ein bezüglich "$\subseteq$" minimaler heißt algebraischer Abschluss von k, dieser ist eindeutig und wird mit $\overline{k}$ bezeichnet. So ist etwa $\overline{\mathbb{R}}=\mathbb{C}$. Der algebraische Abschluss $\overline{\mathbb{F}_p}$ enthält jeden der Körper $\mathbb{F}_{p^r},r\geq 1$, und umgekehrt ist jedes Element von $\overline{\mathbb{F}_p}$ schon in einem dieser Körper $\mathbb{F}_{p^r},r\geq 1$, enthalten. (ohne Beweis)

2.2 Der affine Raum, affine Kurven und der projektive Raum

[8] Wir stellen den zweidimensionalen affinen und projektiven Raum vor, das heißt die wohlbekannte affine Ebene $k^2 = k \times k$ und ihre Ergänzung zur projektiven Ebene $\mathbb{P}^2(k)$ durch "unendlich ferne Punkte". Kurven im Affinen, wie zum Beispiel elliptische Kurven werden dann in der projektiven Ebene intergriert, weil es rechentechnisch einfacher und mathematisch natürlicher ist.

2.2.1 Der affine und projektive Raum

Sei k ein beliebiger Körper. Wir stellen uns meistens $\mathbb{R}$ vor, weil wir über geometrische Objekte nachdenken möchten; k ist in den Anwendungen aber meist ein endlicher Körper.

Definition 8.1 (zweidimensionaler affiner Raum)
Den zweidimensionalen k-Vektorraum $k^2 = k \times k$ schreiben wir auch als $\mathbb{A}^2(k) := \{(x_1, x_2) : x_1, x_2 \in k\}$ und nennen ihn den **zweidimensionalen affinen Raum** über k bzw. **affine Ebene** über k.

Definition 8.2 (Gerade)
Eine **Gerade** in $\mathbb{A}^2(k)$ ist eine Teilmenge der Form

$$g(a, b, c) := \{(x, y) \in \mathbb{A}^2(k) : ax + by + c = 0\} \subseteq \mathbb{A}^2(k)$$

für ein Tripel $(a, b, c) \in k^3$ mit $a, b \neq 0$.

Bemerkung 8.3
Zwei verschiedene Geraden in $\mathbb{A}^2(k)$ schneiden sich in genau einem Punkt, es sei denn, sie sind parallel, das heißt dann haben sie keinen gemeinsamen Punkt in $\mathbb{A}^2(k)$. Soweit nichts Neues.

Bemerkung 8.4
Die Ausnahme, dass in der "Ebene" $k \times k$ Geraden parallel sein können, möchten wir uns beim Rechnen gerne ersparen. Wir ergänzen die Ebene um "unendlich ferne Punkte" und erklären, dass sich zwei parallele Geraden in genau so einem Punkt schneiden. Durch diese Ergänzung wird die affine Ebene zur **projektiven Ebene**. Wie kann das sinnvoll so umgesetzt werden, dass alle Punkte Koordinaten bekommen, mit denen man wie üblich rechnen kann, sodass bei der Schnittpunktberechnung auch die unendlich fernen Punkte erhalten werden können?

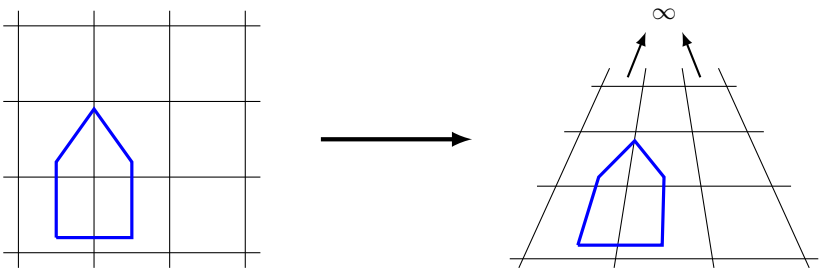


Abbildung 3: Ergänzung des zweidimensionalen affinen Raums zur projektiven Ebene.

Zwei Parallelen $g(a, 1, c)$ und $g(a, 1, d)$ sollen sich dann schneiden, auch rechnerisch. Wir lösen das so, dass in unserer neuen "Ebene" eine dritte "Koordinate" z hinzukommt, welche bei diesen Parallelen also $= 0$ sein müsste, wie folgt:

Definition 8.5 (projektive Ebene)
Die **projektive Ebene** über k ist die Menge

$$\mathbb{P}^2(k) = \{[y_1 : y_2 : y_3] : y_i \in k \text{ nicht alle } 0\}$$

mit der Vereinbarung, dass $[y_1 : y_2 : y_3] = [z_1 : z_2 : z_3]$ genau dann gilt, wenn es ein $\lambda \in k \setminus \{0\}$ gibt mit $y_1 = \lambda z_1, y_2 = \lambda z_2$ und $y_3 = \lambda z_3$.

Definition 8.6 (projektive Ebene (formal))
$\mathbb{P}^2(k)$ ist die Menge der Äquivalenzklassen in k^3 bezüglich der Äquivalenzrelation

$$(y_1, y_2, y_3) \sim (z_1, z_2, z_3) \quad :\Leftrightarrow \quad \exists \lambda \in k \setminus \{0\} : y_i = \lambda z_i, i = 1, 2, 3$$

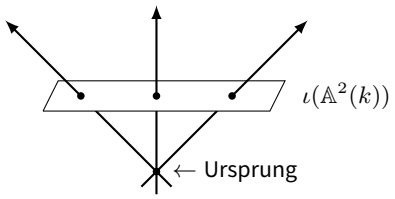
das heißt $\mathbb{P}^2(k) := (k^3 \setminus \{(0,0,0)\}/\sim$.
Wir schreiben $[y_1 : y_2 : y_3]$ für die Äquivalenzklasse, die von (y_1, y_2, y_3) repräsentiert wird und nennen sie einen **projektiven Punkt**. y_1, y_2, y_3 nennen wir **projektive Koordinaten** von $[y_1 : y_2 : y_3]$.

Bemerkung 8.7
Ist $y_3 \neq 0$, gilt $[y_1 : y_2 : y_3] = \left[\frac{y_1}{y_3} : \frac{y_2}{y_3} : 1\right]$, das heißt die dritte (oder jede andere Koordinate $\neq 0$) kann dann auf 1 gebracht ("normiert") werden.

Einem projektiven Punkt $[x : y : z]$ entspricht in unserem Modell in k^3 die Ursprungsgerade $\{(\lambda x, \lambda y, \lambda z) : \lambda \in k\}$.
Diese Punkte sind entweder $[x : y : 1]$ oder $[x : y : 0]$ mit $x, y \in k$ (nicht $[0 : 0 : 0]$!).
Zum Beispiel durch die Abbildung

$$\begin{aligned} \iota \colon \mathbb{A}^2(k) &\longrightarrow \mathbb{P}^2(k) \\ (x, y) &\longmapsto [x : y : 1] \end{aligned}$$

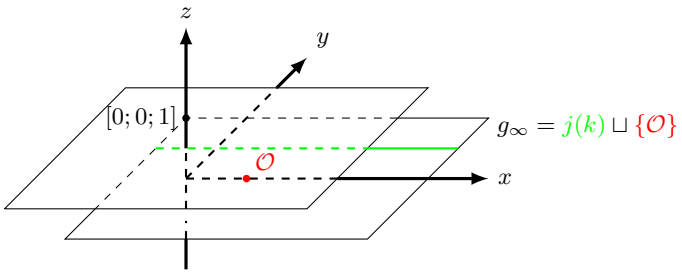
kann die affine Ebene in die projektive eingebettet werden (d.h. ι ist injektiv).



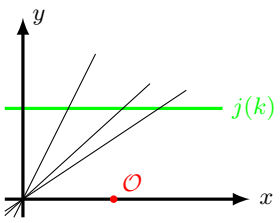
Bemerkung 8.8
Aber $\mathbb{P}^2(k)$ enthält zusätzlich noch die projektiven Punkte $[x : y : 0]$ mit $x, y \in k$ (nicht $x = y = 0$). Offenbar ist $\{[x : y : 0] : x, y \in k, \text{ nicht } x = y = 0\}$ eine Gerade in $\mathbb{P}^2(k)$, die wir **unendlich ferne Gerade** g_∞ nennen möchten, denn mit $j \colon k \to g_\infty, x \mapsto [x : 1 : 0]$ lässt sich k darin einbetten (das heißt j ist injektiv), wobei auffällt, dass $g_\infty \setminus \mathrm{im}(j)$ aus genau den weiteren Punkt $\mathcal{O} := [1 : 0 : 0]$ besteht, das heißt $g_\infty \setminus \mathrm{im}(j) = \{\mathcal{O}\}$.

Bemerkung 8.9
Somit: $\mathbb{P}^2(k) = \iota(\mathbb{A}^2(k)) \sqcup \underbrace{j(k) \sqcup \{\mathcal{O}\}}_{=g_\infty}$ disjunkte Vereinigung

Die in $\mathbb{A}^2(k)$ parallelen Geraden $g(a, 1, c) = \{(x, y) \in k^2 : ax + y + c = 0\} = \{(x, -ax - c) : x \in k\}$ und $g(a, 1, d)$ müssen die projektiven Punkte $[x : -ax - c : 1]$ und $[x : -ax - d : 1]$ enthalten.
Das klappt, wenn die Gleichung $ax + y\{c, d\} = 0$ zu $ax + y + \{c, d\}z = 0$ ergänzt wird. Sie schneiden sich dann im unendlich fernen Punkt $[1 : -a : 0] = \left[-\frac{1}{a} : 1 : 0\right]$, welcher die gemeinsame Steigung $-a$ angibt



Ansicht auf x-y-Ebene



bzw. die gemeinsame "Richtung" $(1, -a)$. Die gemeinsame Richtung $(1, -a)$ wird zum gemeinsamen Schnittpunkt $\left[-\frac{1}{a} : 1 : 0\right]$ erklärt.

Definition 8.10 (projektive Gerade)
Eine **projektive Gerade** ist eine Teilmenge von $\mathbb{P}^2(k)$ der Form

$$g(a, b, c) = \{[x : y : z] : ax + by + cz = 0\} \text{ für } (a, b, c) \in k^3 \setminus \{0\}$$

Man sagt, die "projektive" Gleichung $ax+by+cz = 0$ ist "durch Homogenisierung" aus $ax+by+c = 0$ entstanden: Durch die Ergänzung mit z haben nun alle Summanden ax, by, cz denselben Grad 1 als Polynom aus $k[x, y, z]$. Dieses Prinzip werden wir für allgemeinere Kurven für den Übergang vom Affinen ins Projektive übernehmen. Projektive Geraden werden uns in der Form von Tangenten dann wiederbegegnen.

Beispiel und Bemerkung
Die projektiven Geraden $g(a, 1, c), g(a, 1, d)$ schneiden sich in $\left\{-\frac{1}{a} : 1 : 0\right\} \in g_\infty$. Durch je zwei verschiedene Punkte des $\mathbb{P}^2(k)$ führt genau eine projektive Gerade.

2.2.2 Affine Kurven

Doch zunächst möchten wir im affinen Raum allgemeinere Kurven untersuchen. Dazu benutzen wir Polynome zu ihrer Beschreibung.

Definition 8.11 (affine Kurve)
Sei $f \in k[x, y]$ ein Polynom über k in zwei Variablen x und y. Wir bezeichnen die Menge der Nullstellen von f in $k \times k = \mathbb{A}^2(k)$ als

$$\mathcal{C}_f(k) := \{(u, v) \in \mathbb{A}^2(k) : f(u, v) = 0\}$$

Jede solche Nullstellenmenge $\mathcal{C}_f(k)$ nennen wir eine **affine Kurve**. Ist klar, welches Polynom f vorliegt, schreiben wir auch kurz $\mathcal{C}(k)$ für $\mathcal{C}_f(k)$. Geraden sind spezielle affine Kurven (zu linearen Polynomen $f(x, y, z) = ax+by+c$).

Bemerkung 8.12
Für uns ist interessant, Kurven über verschiedenen Körpern k zu studieren. Der Fall eines endlichen Körpers ist für Anwendungen interessant, weil dann alle Kurven aus nur endlich vielen Punkten bestehen können.

Beispiel 8.13
Sei $k = \mathbb{R}$ und $f(x, y) = y - x^3 - x$. Die Nullstellenmenge $\mathcal{C}_f(k)$ besteht dann aus allen Punkten $(x, y) \in k^2$, welche die Gleichung $y = x^3 + x$ erfüllen. Das reelle Schaubild sieht so aus:

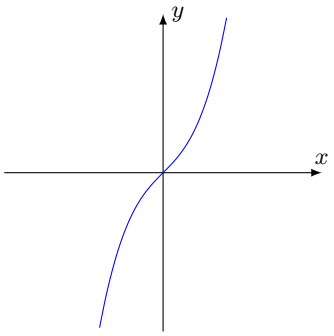


Abbildung 4: Die Menge $\{(x, y) \in \mathbb{R}^2 : y = x^3 + x\}$.

Für $k = \mathbb{F}_5$ können nur wenige Punkte auf der "Kurve" liegen: Die Tabelle

a	0	1	2	3	4
a^3	0	1	3	-3	-1
$a^3 + a$	0	2	0	0	3

zeigt, dass $\mathcal{C}_f(\mathbb{F}_5) = \{(0,0), (1,2), (2,0), (3,0), (4,3)\}$ ist, und mit $f_0(x, y) = y^2 - x^3 - x$ haben wir $\mathcal{C}_{f_0}(\mathbb{F}_5) = \{(0,0), (2,0), (3,0)\}$.
Ist $\tilde{k} \subseteq k$ ein Teilkörper von k (wie zum Beispiell $\mathbb{Q} \subseteq \mathbb{R}$), so folgt auch stets $\mathcal{C}_f(\tilde{k}) \subseteq \mathcal{C}_f(k)$. Unsere Kurvenpunkte in $\mathbb{A}(\mathbb{F}_5)$ finden wir deswegen zum Beispiel in $\mathbb{A}(\mathbb{F}_{25})$ wieder.

Definition 8.14 (Tangente)
Eine (affine) **Tangente** an eine affine Kurve $\mathcal{C}_f(k)$ im Punkt $(a, b) \in \mathcal{C}_f(k)$ ist die Gerade

$$t_f(a, b) = \left\{(x, y) : \frac{\partial f}{\partial x}(a, b)x + \frac{\partial f}{\partial y}(a, b)y + d = 0\right\},$$

falls diese existiert (wir brauchen, dass $\frac{\partial f}{\partial x}(a, b)$, $\frac{\partial f}{\partial y}(a, b)$ nicht beide $= 0$). Dabei ist $d \in k$ so gewählt, dass $(a, b) \in t_f(a, b)$ gilt.

Bemerkung 8.15
Es ist nicht klar, ob Tangenten stets eindeutig existieren. Affine Kurven können sich selbst schneiden oder scharfe "Spitzen" haben. Siehe z.B. Abbildung 2 in Abschnitt 0 auf Seite 7.

Definition 8.16 (singulärer Punkt)
Die affine Kurve $\mathcal{C}_f(k)$ heißt **singulär im Punkt** $(a, b) \in \mathcal{C}_f(k)$, falls $\frac{\partial f}{\partial x}(a, b) = \frac{\partial f}{\partial y}(a, b) = 0$ gilt.

Bemerkung 8.17
Affine Kurven, die in keinem Punkt singulär sind, haben überall eine wohldefinierte Tangente.

Bemerkung 8.18
Es kann vorkommen, dass $\mathcal{C}_f(k)$ gar keine singulären Punkte enthält, wohl aber über einem Erweiterungskörper von k, wie etwa $\overline{k}$, dem algebraischen Abschluss über k.

Beispiel 8.19
Für $f(x, y) = y^2 - x^4 - 2x^2 - 1$ hat $\mathcal{C}_f(\mathbb{R})$ keine singulären Punkte: Es ist $\frac{\partial f}{\partial x}(a, b) = -4a(a^2 + 1)$, $\frac{\partial f}{\partial y}(a, b) = 2b$. Allerdings sind $(i, 0), (-i, 0) \in \mathbb{C}$ singuläre Punkte in $\mathcal{C}_f(\mathbb{C})$, wo $\mathbb{C} = \overline{\mathbb{R}}$.

Beispiel 8.20
Sei $f(x,y) = y^2 - x^3 - x$ und $k = \mathbb{F}_p$. Die Ableitungen sind $\frac{\partial f}{\partial x}(x,y) = -3x^2 - 1$, $\frac{\partial f}{\partial y}(x,y) = 2y$, das heißt die singulären Punkte (a,b) sind die mit $b^2 = a^3 + a$, $-3a^2 = 1, 2b = 0$.

- Für $p \neq 2$ ist $2b = 0$ nur für $b = 0$ richtig, dann ist $0 = a(a^2+1)$ und $3a^2 = -1$. Es folgt $0 = a(3a^2+3) = 2a$ und wegen $p \neq 2$ folgt $a = 0$ im Widerspruch zu $3^2 = -1$. Also existieren keine singulären Punkte für $p \neq 2$.
- Für $p = 2$ ist $\mathcal{C}_f(\mathbb{F}_2) = \{(0,0),(1,0)\}$. Es ist $\frac{\partial f}{\partial x}(1,0) = 0 = \frac{\partial f}{\partial y}(1,0)$, das heißt $(1,0)$ ist singulärer Punkt.

2.3 Projektive Kurven

2.3.1 Homogene Polynome und projektive Kurven

Durch Homogenisierung können wir affine Kurven zu projektiven Kurven machen. [9]

Definition 9.1 (homogenes Polynom, Homogenisierung)
Sei $F \in k[X,Y,Z]$ ein Polynom über k in drei Variablen und $F \neq 0$. Dann heißt F **homogen** vom Grad d, falls gilt:

$$F(X,Y,Z) = \sum_{v_1,v_2,v_3 \geq 0} \alpha_{v_1,v_2,v_3} X^{v_1} Y^{v_2} Z^{v_3}$$

und $\alpha_{v_1,v_2,v_3} \neq 0 \Rightarrow v_1 + v_2 + v_3 = d$, das heißt wenn alle Monome in g den Grad d haben.

Beispiel 9.2
$F(X,Y,Z) = aX + bY + cZ$ ($d = 1$) oder $F(X,Y,Z) = Y^2Z - X^3 - XZ^2$ ($d = 3$).

Bemerkung 9.3
Klar ist, dass ein $f \in k[X,Y]$ durch Ergänzung von Z-Potenzen zu einem homogenen Polynom $F_f \in k[X,Y,Z]$ gemacht werden kann: Ist $f(x,y) = \sum_{v_1,v_2 \geq 0} \alpha_{v_1,v_2} x^{v_1} y^{v_2}$ vom Grad d, so setze

$$F_f(x,y,z) := \sum_{v_1,v_2 \geq 0} \alpha_{v_1,v_2} x^{v_1} y^{v_2} z^{d-v_1-v_2}.$$

Man nennt F_f dann die **Homogenisierung** von f. Für diese gilt $F_f(x,y,1) = f(x,y)$.

Lemma 9.4
Ist $F \in k[X,Y,Z]$ homogen vom Grad d, so gilt für alle $\alpha, \beta, \gamma \in k$ und $\lambda \in k \setminus \{0\}$:

$$F(\alpha,\beta,\gamma) = 0 \Leftrightarrow F(\lambda\alpha, \lambda\beta, \lambda\gamma) = 0$$

Beweis
Nachrechnen zeigt $F(\lambda\alpha, \lambda\beta, \lambda\gamma) = \lambda^d F(\alpha,\beta,\gamma)$, woraus die Behauptung folgt. □

Somit können wir projektive Kurven definieren:

Definition 9.5 (projektive ebene Kurve)
Sei $F \in k[X,Y,Z]$ homogen. Dann bezeichnen wir die Nullstellenmenge mit

$$\mathcal{C}_F(k) := \{[u:v:w] \in \mathbb{P}^2(k) : g(u,v,w) = 0\}.$$

Ist F klar, schreiben wir auch einfach $\mathcal{C}(k)$ für $\mathcal{C}_F(k)$. Jede solche Nullstellenmenge heißt eine **projektive ebene Kurve**.

Beispiel 9.6
Die affine Kurve $\mathcal{C}_f(x,y)$ zu $f(x,y) = y^2 - x^3 - x$ kann durch Homogenisieren zu $C_{F_f}(x,y,z)$ mit $F_f(x,y,z) = y^2z - x^3 - xz^2$ gemacht werden. Die injektive Abbildung $\iota\colon \mathbb{A}^2(k) \to \mathbb{P}^2(k), (x,y) \mapsto [x:y:1]$ bildet $\mathcal{C}_f(k)$ nach $\mathcal{C}_{F_f}(k)$ ab. Die projektive Kurve $\mathcal{C}_{F_f}(k)$ hat aber noch genau einen weiteren Punkt (auf g_∞), nämlich $[0:1:0]$, das heißt $\mathcal{C}_{F_f}(k) = \iota(\mathcal{C}_f(k)) \cup \{[0:1:0]\}$.

Lemma 9.7
$\mathcal{C}_{F_f}(k) \cap \iota(\mathbb{A}^2(k)) = \iota(\mathcal{C}_f(k))$ für jede affine Kurve $\mathcal{C}_f$ und ihre projektive Kurve $\mathbb{C}_{F_f}$.

Beweis
$[x : y : 1] \in \mathcal{C}_{F_f}(k) \cap \iota(\mathbb{A}^2(k)) \Leftrightarrow 0 = F_f(x, y, 1) = f(x, y) \Leftrightarrow [x : y : 1] \in \iota(\mathcal{C}_f(k))$. □

Bemerkung 9.8

- Wir werden hier ι auch weglassen; es ist klar, was gemeint ist.
- Anstelle von ι können auch die Einbettungen $\iota_2(x, y) = [1 : x : y], \iota_3(x, y) = [x : 1 : y]$ betrachtet werden, das Lemma gilt dann entsprechend.
- Geht man für eine projektive Kurve $\mathcal{C}_F(k)$ zu einer dieser Schnitte mit $\mathbb{A}^2(k)$ über, so sagt man, man "geht zu affinen Koordinaten" über.

Definition 9.9 (singulärer Punkt, nicht-singulär)
Sei $F \in k[X, Y, Z]$ homogen vom Grad d. Die projektive ebene Kurve $\mathcal{C}_F(k)$ heißt **singulär im Punkt** $P = [a : b : c] \in \mathcal{C}_F(k)$, falls alle Ableitungen von F in P verschwinden, das heißt

$$\frac{\partial F}{\partial X}(a, b, c) = \frac{\partial F}{\partial Y}(a, b, c) = \frac{\partial F}{\partial Z}(a, b, c) = 0.$$

Die Kurve $\mathcal{C}_f(k)$ heißt **nicht-singulär**, falls $\mathcal{C}_F(\overline{k})$ keinen singulären Punkt enthält, wobei $\overline{k}$ einen algebraischen Abschluss von k bezeichnet.

Bemerkung 9.10
Diese Definition hängt nicht davon ab, welche projektive Koordinaten a, b, c eines Punktes $P = [a : b : c]$ betrachtet werden. Sie passt auch mit der alten Definition von "singulären Punkt" für affine Kurven zusammen, wie folgendes Lemma zeigt. Nach dem Lemma genügt es dann, singuläre Punkte, die im Affinen liegen, auf Singularität im Affinen zu testen.

Lemma 9.11
Sei $F(X, Y, Z) = \sum_{\underline{v} \geq 0} \alpha_{\underline{v}} X^{v_1} Y^{v_2} Z^{v_3}$ homogen vom Grad d und $f(x, y) = \sum_{\substack{v_1, v_2 \\ v_1 + v_2 \leq d}} \alpha_{v_1, v_2, d - v_1 - v_2} x^{v_1} y^{v_2} = F(x, y, 1)$, das heißt $F = F_f$. Weiter sei $P \in \mathcal{C}_F(k)$ mit $P = i(Q) \in \iota(\mathbb{A}^2(k))$. Dann gilt: $\mathcal{C}_F(k)$ singulär in $P \Leftrightarrow \mathcal{C}_F(k)$ singulär in Q.

Beweis
Haben $Q \in \mathcal{C}_f(k)$, etwa $Q = (a, b)$, dann ist $P = \iota(Q) = [a : b : 1]$. Es ist

$$\frac{\partial F}{\partial X}(X, Y, Z) = \sum_{\substack{v_1 > 0 \\ v_2, v_3 \geq 0}} \alpha_{\underline{v}} v_1 X^{v_1 - 1} Y^{v_2} Z^{v_3},$$

also gilt $\frac{\partial F}{\partial X}(a, b, 1) = \frac{\partial f}{\partial x}(a, b)$ und entsprechend $\frac{\partial F}{\partial Y}(a, b, 1) = \frac{\partial f}{\partial y}(a, b)$, sowie

$$\frac{\partial F}{\partial Z}(a, b, 1) = \sum_{v_i \geq 0} \alpha_{\underline{v}} v_3 a^{v_1} b^{v_2} = \sum_{v_i \geq 0} \alpha_{v_1, v_2, d - v_1 - v_2} (d - v_1 - v_2) a^{v_1} b^{v_2} = d \cdot f(a, b) - a \frac{\partial f}{\partial x}(a, b) - b \frac{\partial f}{\partial y}(a, b).$$

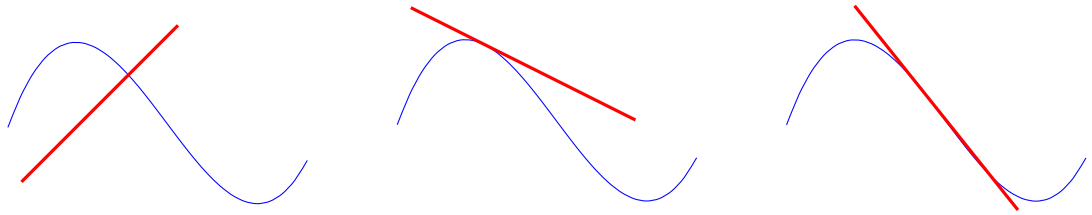
Durch Vergleich der Ableitungen folgt die Behauptung in beide Richtungen. □

Definition 9.12 (Tangente)
Sei $\mathcal{C}_F(k)$ eine projektive ebene Kurve und $P = [a : b : c]$ ein nicht-singulärer Punkt auf $\mathcal{C}_F(k)$. Die projektive Gerade $\mathcal{C}_T(k)$ mit $T(X, Y, Z) := \frac{\partial F}{\partial X}(a, b, c)X + \frac{\partial F}{\partial Y}(a, b, c)Y + \frac{\partial F}{\partial Z}(a, b, c)Z$ heißt **Tangente** in P an $\mathcal{C}_F(k)$. Wir schreiben $T_P(\mathcal{C}_F) := \mathcal{C}_T(k)$ dafür.

Bemerkung 9.13
In nicht-singulären Punkten haben projektive ebene Kurven also eine ”schöne” Tangente. Die Voraussetzung ”nicht-singulär” braucht man, damit nicht alle drei Ableitungen gleichzeitig verschwinden und so eine projektive Gerade definiert werden kann. Bei Übergang zu affinen Koordinaten erhält man wieder die üblichen (affinen) Tangenten, weil wir dann $Z = 1$ setzen.

Beispiel 9.14
Sei $\mathrm{char}(k) \neq 2, f(x,y) := y^2 - 2x^2 - 2, F_f(x,y,z) = y^2 - 2x^2 - 2z^2$. Dann ist $(1,2) \in \mathcal{C}_f(k)$, $\frac{\partial f}{\partial x}(1,2) = -4, \frac{\partial f}{\partial y}(1,2) = 4$, das heißt $(1,2)$ ist nicht-singulär. Die affine Tangende von $\mathcal{C}_f$ in $Q = (1,2)$ ist $t_Q(\mathcal{C}_f) = \{(x,y) \in k^2 : -4x + 4y - 4 = 0\}$, die projektive Tangente von $\mathcal{C}_F$ in $P = [1:2:1] = \iota(Q)$ ist $T_P(\mathcal{C}_F) = \{[X:Y:Z] \in \mathbb{P}^2(k) : -4X + 4Y - 4Z = 0\}$.

Motivation 9.15
Wir möchten studieren, wie sich ebene Kurven mit Geraden schneiden und die folgenden Fälle unterscheiden können:



transversaler Schnitt | einfache Tangente | Wendepunkt-Tangente

Abbildung 5: Wendepunkt-Tangenten liefern gute Approximationen an eine Kurve.

Definition 9.16 (Schnittmultiplizität, Vielfachheit)
Sei $\mathcal{C}_F(k)$ eine projektive Kurve zum homogenen Polynom $F \in k[X,Y,Z]$, sei $G(\alpha,\beta,\gamma)$ eine projektive Gerade und $P \in G(\alpha,\beta,\gamma)$ ein Punkt. Ist P kein Schnittpunkt von $\mathcal{C}_F(k)$ und G, setzen wir $m(P;G,\mathcal{C}_F) := 0$. Ansonsten hat das Polynom $\Psi(t) := F(a + ta', b + tb', c + tc') \in k[t]$ eine Nullstelle in $t = 0$, wobei $P = (a,b,c)$ und $P' = (a',b',c') \in G$ sind. Dann sei $m(P;G,\mathcal{C}_F)$ die Ordnung der Nullstelle $t = 0$ von $\Psi \in k[t]$, falls $\Psi \neq 0$. Die Zahl $m(P;G,\mathcal{C}_F)$ heißt **Schnittmultiplizität** bzw. **Vielfachheit**, mit der sich G und $\mathcal{C}_F$ im Punkt P schneiden.

Bemerkung 9.17
Es ist $m(P;G,\mathcal{C}_F)$ unabhängig von der Wahl von P'.

Beispiel 9.18
Sei $f(x,y) = x(x-1)(x-2)-y \in \mathbb{R}[x,y]$, das heißt $f(x,y) = x^3-3x^2+2x-y$ und $F(X,Y,Z) = F_f(X,Y,Z) = X^3 - 3X^2Z + 2XZ^2 - YZ^2$.
Da $\frac{\partial f}{\partial x} = 3x^2-6x+2, \frac{\partial f}{\partial y} = -1$, hat $\mathcal{C}_f$ in $(0,0) \in \mathcal{C}_f$ die affine Tangente $t_{(0,0)}(\mathcal{C}_f) = \{(x,y) \in \mathbb{R}^2 : 2x-y = 0\}$, projektiv aufgefasst lautet die Tangente $T_{(0,0)}(\mathcal{C}_F) = \{[x:y:z] \in \mathbb{P}^2(\mathbb{R}) : 2X - Y + 0Z = 0\} = G(2,-1,0)$. Die Gerade $G(2,-1,0)$ schneidet $\mathcal{C}_F$ in $[3:6:1]$ und in $[0:0:1]$. Dann haben wir $m([3:6:1];G,\mathcal{C}_F) = 1$, weil

$$\begin{aligned}\Psi(t) &= F(3 + t\cdot 0, 6 + t\cdot 0, 1 + t\cdot 1)\\ &= 3^3 - 3\cdot 3^2(1+t) + 2\cdot 3\cdot (1+t)^2 - 6\cdot(1+t)^2\\ &= 0\cdot t^2 + (-3^3 + 6\cdot 2 - 12)t + (3^3 - 3^3 + 6 - 6) = -3^3t^1\end{aligned}$$

eine einfache Nullstelle in $t = 0$ hat, sowie $m([0:0:1]; G, \mathcal{C}_F) = 2$, weil

$$\widetilde{\Psi}(t) = F(0+3t, 0+6t, 1+t) = (3t)^3 - 3(3t)^2(1+t) + 2(3t)(1+t)^2 - (6t)(1+t)^2 = -3^3t^2.$$

Erläuterung 9.19
Ist $m(P; G, \mathcal{C}_F) = 1$, liegt ein transversaler Schnitt der Geraden G mit der Kurve $\mathcal{C}_f$ vor. Ist $m(P; G, \mathcal{C}_F) = 2$, so ist G eine "einfache" Tangente an $\mathcal{C}_F$. Falls $m(P; G, \mathcal{C}_F) \geq 3$, ist die Tangente eine sehr gute Approximation an $\mathcal{C}_F$ von "Ordnung ≥ 3" (da die Schnittmultiplizität genau die Nullstellenordnung von $\Psi(t)$ in $t = 0$ ist).

Bemerkung 9.20
Ist der Körper k algebraisch abgeschlossen, zerfällt Ψ fast vollständig in Linearfaktoren. Es folgt, dass dann die Summe der Schnittmultiplizitäten aller Schnittpunkte von G mit $\mathcal{C}_F$ genau $\deg(\Psi) = \deg(F)$ ist, das heißt $\sum_{P \in G \cap \mathcal{C}_F} m(P; G, \mathcal{C}_F) = \deg(F)$. Ist k ein beliebiger Körper, so folgt

$$\sum_{P \in G \cap \mathcal{C}_F} m(P; G, \mathcal{C}_F) \leq \deg(F).$$

Bemerkung 9.21
Alle diese Ergebnisse gelten nicht, wenn das lineare Polynom, welches G erklärt, ein Teiler des Polynoms F ist, denn dann lassen sich keine Schnittmultiplizitäten erklären: Ist $G = G(\alpha, \beta, \gamma)$ durch $\alpha X + \beta Y + \gamma Z = 0$ erklärt und $F(X, Y, Z) = (\alpha X + \beta Y + \gamma Z) \cdot H(X, Y, Z)$ für ein $H \in k[X, Y, Z]$, so folgt $G \subseteq \mathcal{C}_F$ und für $[a : b : c], [a' : b' : c'] \in G$ ist dann

$$\Psi(t) = F(a + ta', b + tb', c + tc') = (a(a + ta') + \beta(b + tb') + \gamma(c + tc')) \cdot H(\ldots) = 0 \cdot H(\ldots) = 0$$

das Nullpolynom, also die Nullstellenordnung von $t = 0$ nicht definiert.

Wir zeigen nun, dass wir bei Tangenten in einem Kurvenpunkt immer die Schnittmultiplizität ≥ 2 haben, sofern der Grad der Kurve auch ≥ 2 ist.

Satz 9.22
Sei $P \in \mathbb{P}^2(k)$ ein nicht-singulärer Punkt auf $\mathcal{C}_F$, wobei $\deg(F) \geq 2$ sei, und $T = T_P(\mathcal{C}_F)$ die Tangente an $\mathcal{C}_F$ im Punkt P. Dann ist $m(P; T, \mathcal{C}_F) \geq 2$.

Beweis
Sei $T = F(\alpha, \beta, \gamma) = \{[X : Y : Z] : \alpha X + \beta Y + \gamma Z = 0\}$ die Tangente in $P = [a : b : c] \in G \cap \mathcal{C}_F$, also $\alpha = \frac{\partial F}{\partial X}(a, b, c), \beta = \frac{\partial F}{\partial Y}(a, b, c), \gamma = \frac{\partial F}{\partial Z}(a, b, c)$. Sei $Q = [a', b', c'] \in G$ ein beliebiger weiterer Punkt auf G, und $\Psi(t) = F(a + ta', b + tb', c + tc')$. Dann ist $\Psi(0) = 0$, da $P \in \mathcal{C}_F$, und laut Kettenregel (vgl. Satz 7.4) ist

$$\Psi'(0) = \frac{\partial F}{\partial X}(a, b, c) \cdot a' + \frac{\partial F}{\partial Y}(a, b, c) \cdot b' + \frac{\partial F}{\partial Z}(a, b, c) \cdot c' = \alpha a' + \beta b' + \gamma c' = 0,$$

weil $Q \in G$. Mit $\Psi(0) = 0, \Psi'(0) = 0$ folgt $m(P; T; \mathcal{C}_F) \geq 2$. □

2.3.2 Der Satz von Bézout

[10] Wir zeigen in diesem Abschnitt, dass Kurven im Allgemeinen nicht allzu viele Schnittpunkte haben:

Satz 10.1 (Satz von Bézout)
Zwei Kurven $\mathcal{C}_{F_1}, \mathcal{C}_{F_2}$ in $\mathbb{P}^2(k)$ können sich in nicht mehr als $\deg(F_1) \cdot \deg(F_2)$ vielen Schnittpunkten treffen, es sei denn, F_1 und F_2 haben einen gemeinsamen Teiler vom Grad ≥ 1. Das heißt:

$$\mathrm{ggT}(F_1, F_2) = 1 \Rightarrow \#(\mathcal{C}_{F_1} \cap \mathcal{C}_{F_2}) \leq \deg(F_1) \cdot \deg(F_2)$$

Bemerkung 10.2
Dieser Satz ist eine sehr schwache Form des Satzes von Bézout, welcher besagt:
Sei k ein algebraisch abgeschlossener Körper und seien $F_1, F_2 \in k[X,Y,Z]$ zwei homogene Polynome mit $\mathrm{ggT}(F_1, F_2) = 1$, die zwei ebene projektive Kurven $\mathcal{C}_{F_1}$ und $\mathcal{C}_{F_2}$ definieren. Dann ist

$$\sum_{P \in \mathcal{C}_{F_1} \cap \mathcal{C}_{F_2}} m(P; \mathcal{C}_{F_1}, \mathcal{C}_{F_2}) = \deg(F_1) \cdot \deg(F_2).$$

Ist k ein beliebiger Körper, gilt dies mit "$\leq$" statt "$=$".

Bemerkung 10.3

- Zum Beweis dieses allgemeinen Bézout-Satzes werden mehr Mittel aus der algebraischen Geometrie benötigt, als wir hier zeigen können. Für unsere Zwecke, das Studium elliptischer Kurven, reicht die schwache Version aus Satz 10.1, die wir hier beweisen, und insbesondere die spezielle Verschärfung aus Satz 10.15.
- Die Kurven können singuläre Punkte enthalten.
- Den Fall $\deg(F_1) = 1$, das heißt wenn F_1 eine Gerade $\mathcal{C}_{F_1}$ erklärt, haben wir bereits in Bemerkung 9.20 gezeigt.
- Den Begriff der Schnittmultiplizität müsste man für Schnittpunkte zweier beliebiger ebener Kurven verallgemeinern. Wir verzichten hier darauf.
- Aus diesem (allgemeinen) Satz von Bézout folgt bereits die schwache Version aus Satz 10.1, denn für Schnittpunkte ist $m(P; \mathcal{C}_{F_1}, \mathcal{C}_{F_2}) \geq 1$, also ist

$$\#(\mathcal{C}_{F_1} \cap \mathcal{C}_{F_2}) = \sum_{P \in \mathcal{C}_{F_1} \cap \mathcal{C}_{F_2}} 1 = \sum_{P \in \mathcal{C}_{F_1} \cap \mathcal{C}_{F_2}} m(P; \mathcal{C}_{F_1}, \mathcal{C}_{F_2}) \leq \deg(F_1) \cdot \deg(F_2).$$

Beispiel 10.4
Gegeben seien die Parabeln $F_1(X,Y,Z) = X^2 - 3XZ + Z^2 - YZ$ und $F_2(X,Y,Z) = -X^2 + 3XZ - 3Z^2 - YZ$ mit den beiden affinen reellen Schnittpunkten $[1:-1:1]$ und $[2:-1:1]$. Laut Bézout-Satz haben die Parabeln noch zwei weitere Schnittpunkte über $\mathbb{C}$. Diese sind nicht im Affinen, weil die Gleichung $F_1(X,Y,1) = F_2(X,Y,1)$ genau die Lösungen $(1,-1),(2,-1)$ hat. Mit der Gleichung $F_1(X,Y,0) = F_2(X,Y,0) \Leftrightarrow X^2 = -X^2$ erhält man $X = 0$, also den (unendlich fernen) Punkt $[0:1:0] =: \mathcal{O}$ als einzigen projektiven Schnittpunkt. Eine genaue Analyse würde zeigen, dass $\mathcal{O}$ die Schnittmultiplizität 2 hat.

Definition 10.5 (Resultante)
Seien $f, g \in k[X]$ Polynome vom Grad $m = \deg(f), n = \deg(g)$, etwa gegeben durch

$$f = a_m X^m + a_{m-1} X^{m-1} + \ldots + a_1 X + a_0$$
$$g = b_n X^n + b_{n-1} X^{n-1} + \ldots + b_1 X + b_0$$

Sei

$$M(f,g) := \begin{pmatrix} a_0 & & & & b_0 & & & \\ a_1 & a_0 & & & b_1 & b_0 & & \\ \vdots & a_1 & \ddots & & \vdots & b_1 & \ddots & \\ a_m & \vdots & \ddots & a_0 & b_n & \vdots & \ddots & b_0 \\ & a_m & & a_1 & & b_n & & b_1 \\ & & \ddots & \vdots & & & \ddots & \vdots \\ & & & a_m & & & & b_n \end{pmatrix} \in k^{(m+n)\times(m+n)},$$

das heißt $M(f,g)$ besteht aus n Spalten mit den Koeffizienten von f und m Spalten mit den Koeffizienten von g. Dann heißt $\mathrm{Res}(f,g) = \det(M(f,g)) \in k$ die **Resultante** von f und g.

Bemerkung 10.6

- Anstelle von k können auch beliebige kommutative Ringe mit 1 in der Definition stehen.
- $\mathrm{Res}(f,g)$ kann als Polynom in den Unbestimmten $a_0,\dots,a_m,b_0,\dots,b_n$ angesehen werden. Für einen darin vorkommenden Term $\prod_{i,j} a_i^{\nu_i} b_j^{\mu_j}$ gilt $\sum_{i=0}^{m} \nu_i(m-i) + \sum_{j=0}^{n} \mu_j(n-j) = mn$. (ohne Beweis)

Beispiel 10.7
Sei $k = \mathbb{R}$, $f(x) = x^2 + 2x - 1$, $g(x) = 4x^3 - 3x + 5$. Dann ist

$$M(f,g) = \begin{pmatrix} -1 & 0 & 0 & 5 & 0 \\ 2 & -1 & 0 & -3 & 5 \\ 1 & 2 & -1 & 0 & -3 \\ 0 & 1 & 2 & 4 & 0 \\ 0 & 0 & 1 & 0 & 4 \end{pmatrix}$$

Satz 10.8
Sei S ein faktorieller Ring (z.B. Polynomring oder ein Körper), $f,g \in S[X]$ Polynome mit $\deg(f) = m, \deg(g) = n$. Dann sind äquivalent:

(i) $f,g, \in S[X]$ haben einen gemeinsamen nichtkonstanten Teiler in $S[X]$

(ii) Es gibt $f_0, g_0 \in S[X] \setminus \{0\}$ mit $\deg(f_0) \le m-1, \deg(g_0) \le n-1$ und $f_0 g = g_0 f$

(iii) $\mathrm{Res}(f,g) = 0$

Beweis
(i)⇒(ii): Sei h ein gemeinsamer Teiler, $\deg(h) \ge 1$. Dann setze $f_0 = \frac{f}{h}, g_0 = \frac{g}{h}$.

(i)⇐(ii): Sind f_0, g_0 wie in (ii) und $h = \mathrm{ggT}(f,g)$, folgt $\mathrm{ggT}\left(\frac{f}{h}, \frac{g}{h}\right) = 1$. Nach Voraussetzung ist $\frac{f}{h} \cdot g_0 = f_0 \cdot \frac{g}{h}$, also ist $\frac{f}{h} \mid f_0$, das heißt $\deg\left(\frac{f}{h}\right) \le \deg(f_0) \le m-1$, also $\deg(h) \ge 1$.

(ii)⇔(iii): f_0, g_0 entsprechen den nichttrivialen Lösungen des linearen Gleichungssystems

$$\sum_{k=1}^{n} c_k T^{k-1} f + \sum_{k=1}^{m} c_{n+k} T^{k-1} g = 0.$$

Bezüglich der Basis $T^0, T^1, \dots, T^{n+m-1}$ über S wird das LGS gerade durch $M(f,g)$ beschrieben. □

Beweis 10.9 (von Satz 10.1)
Wir nehmen zum Beweis ohne Einschränkung an, dass k ein unendlicher Körper ist, andernfalls können wir zum Beispiel zum algebraischen Abschluss $\overline{k}$ übergehen, der jedenfalls unendlich ist, vergleiche dazu Bemerkung 7.22; denn für eine Körpererweiterung könnte es mehr Schnittpunkte geben. Sei $d_1 = \deg(F_1)$ und $d_2 = \deg(F_2)$. Angenommen, $\mathcal{C}_{F_1}$ und $\mathcal{C}_{F_2}$ hätten mindestens $d_1 d_2 + 1$ viele Punkte gemeinsam (wir zeigen, dass dann $\deg(\mathrm{ggT}(F_1,F_2)) \ge 1$ sein müsste). Seien $P_0, P_1, \dots, P_{d_1 d_2}$ Schnittpunkte von $\mathcal{C}_{F_1}$ und $\mathcal{C}_{F_2}$.

Beweis 10.10 (Fortsetzung)
Wir können ohne Einschränkung annehmen, dass die Punkte $P_i = (x_i, y_i), i = 0, \dots, d_1 d_2$, verschiedene x-Koordinaten und verschiedene y-Koordinaten haben (sonst erreicht man dies wieder durch eine Verschiebung bzw. lineare Transformation, da k unendlich ist).

Beweis 10.11 (Fortsetzung)
Wir können eine Gerade $G(\alpha,\beta,\gamma) = \{[x:y:z] \in \mathbb{P}^2(k) : \alpha x + \beta y + \gamma z = 0\}$ finden, die durch keine dieser Punkte $P_0, \ldots, P_{d_1 d_2}$ geht, weil k unendlich ist. Diese Gerade sei ohne Einschränkung g_∞, die unendlich ferne Gerade (durch eine Verschiebung bzw. lineare Transformation lässt sich dies erreichen).

Beweis 10.12 (Fortsetzung)
Somit ist das Problem auf ein affines Problem zurückgeführt worden. Die zugehörigen affinen Kurven seien durch $f_1, f_2 \in k[x,y]$ gegeben, das heißt $f_1(x,y) := F_1(X,Y,1), f_2(x,y) := F_2(X,Y,1)$, mit $\deg(f_1) \leq d_1$, $\deg(f_2) \leq d_2$. Wir können ohne Einschränkung sogar $\deg(f_1) = d_1$, $\deg(f_2) = d_2$ annehmen (nach geeigneter Transformation der Koordinaten der Art $X \to X + \varepsilon Y, Y \to Y$ ergeben sich für $F_1(X,Y,0) = \sum_{i+j=d_1} c_{ij} X^i Y^j$, $F_2(X,Y,0) = \sum_{i+j=d_2} d_{ij} X^i Y^j$ die Terme $(\sum_{i+j=d_1} c_{ij}\varepsilon^i)Y^{d_1}$ in $\widetilde{F_1}(X,Y,0)$ und $(\sum_{i+j=d_2} d_{ij}\varepsilon^i)Y^{d_2}$ in $\widetilde{F_2}(X,Y,0)$).

Beweis 10.13 (Fortsetzung)
Wir betrachten $f_1, f_2 \in (k[x])[y]$ als Polynome in y mit Koeffizienten in $k[x]$ und berechnen die Resultante $R(f_1,f_2) \in k[x]$, diese hat den Grad $d_1 d_2$ in x nach Bemerkung 10.6. Sei $R(x) := R(f_1,f_2) \in k[x]$.

Beweis 10.14 (Fortsetzung)
Für jedes x_i haben die Polynome $f_1(x_i,y), f_2(x_i,y) \in k[y]$ eine Faktor $y - y_i \in k[y]$ gemeinsam. Für die $x = x_i$ muss $R(x)$ also verschwinden: $R(x_i) = 0, i = 0, \ldots, d_1 d_2$. Also hat $R(x)$ mehr Nullstellen ($d_1 d_2 + 1$ viele) als sein Grad $d_1 d_2$, $R(x)$ muss also das Nullpolynom (in x) sein. Aber dann haben $f_1, f_2 \in (k[x])[y]$ einen gemeinsamen Teiler vom Grad ≥ 1 wegen Satz 10.8, (iii) $\Rightarrow$ (i). □

Satz 10.15
Sei k ein beliebiger Körper, $F_1, F_2 \in k[X,Y,Z]$ homogene Polynome mit $d_1 = \deg(F_1), d_2 = \deg(F_2)$ und $\mathrm{ggT}(F_1,F_2) = 1$, und es seien $d_1 d_2 - 1$ viele Schnittpunkte von $\mathcal{C}_{F_1}$ und $\mathcal{C}_{F_2}$ gegeben. Dann haben sie einen weiteren Schnittpunkt in $\mathbb{P}^2(k)$ gemeinsam.

Beweis
Wir im Beweis von Satz 10.1 von Bézout erhalten wir ein Polynom $R(x) \in k[x]$ vom Grad $= d_1 d_2$. Es hat $d_1 d_2 - 1$ viele Nullstellen $x_1, \ldots, x_{d_1 d_2 - 1}$ laut Voraussetzung, ist also durch $(x - x_1) \cdots (x - x_{d_1 d_2 - 1})$ teilbar, der Quotient ist vom Grad 1, also $= r \cdot (x - a) \in k[x]$ mit einer (weiteren) Nullstelle $a \in k$.
Somit haben $f_1(a,y), f_2(a,y) \in k[y]$ einen gemeinsamen Faktor vom Grad ≥ 1. Dieser Grad ist 1 (denn wäre er ≥ 2, würde er über $\overline{k}$ in mindestens zwei Linearfaktoren zerfallen, die dann zu zwei weiteren Schnittpunkten mit gleicher x-Koordinate a führen würden, sodass es $\geq (d_1 d_2 - 1) + 2 > d_1 d_2$ viele Schnittpunkte geben müsste – im Widerspruch zu Satz 10.1). Also gibt es nur noch genau einen weiteren Schnittpunkt (a,y) von $\mathcal{C}_{F_1}$ und $\mathcal{C}_{F_2}$. □

Beispiel 10.16
Sei k ein beliebiger Körper, $F_1, F_2 \in k[X,Y,Z]$ homogen, $\deg(\mathrm{ggT}(F_1,F_2)) = 0$, und sei $\deg(F_1) = 1, \deg(F_2) = 3$. Dann ist $\sum_{P \in \mathcal{C}_{F_1} \cap \mathcal{C}_{F_2}} m(P; \mathcal{C}_{F_1}, \mathcal{C}_{F_2}) \in \{0,1,3\}$.

2.4 Elliptische Kurven

2.4.1 Definition elliptischer Kurven und vereinfachte Weierstraßgleichungen

[11] Wir geben nun die Definition einer elliptischen Kurve. Sei k ein Körper.

Definition 11.1 (elliptische Kurve)
Eine **elliptische Kurve** $E(k)$ ist eine nicht-singuläre, irreduzible projektive Kurve vom Grad 3, die einen (rationalen) Wendepunkt enthält.

Bemerkung 11.2

- Es reicht, die Wendepunktbedingung durch $E(k) \cap \mathbb{P}^2(\mathbb{Q}) \neq \emptyset$ zu ersetzen (ist aber aufwendig zu zeigen, lassen dies deswegen sein).
- Eine Kurve $\mathcal{C}$ heißt **irreduzibel**, wenn sie nicht die Vereinigung zweier Kurven $\neq \mathcal{C}$ ist, z.B. ist $\mathcal{C}_F(k)$ mit $F(X, Y, Z) = XY$ reduzibel.

Bemerkung 11.3
Durch eine so genannte **birationale Transformation** kann angenommen werden, dass der Wendepunkt ohne Einschränkung $\mathcal{O} := [0 : 1 : 0]$ ist. Eine Übungsaufgabe zeigt, dass dann die Kurvengleichung die folgende vereinfachte Form hat:

Definition 11.4 (elliptische Kurve (lange Weierstraßform))
Eine **elliptische Kurve** $E_F(k)$ ist eine nicht-singuläre, projektive ebene Kurve $\mathcal{C}_F(k) \subseteq \mathbb{P}^2(k)$, wobei F ein homogenes Polynom vom Grad 3 der Form

$$F(X, Y, Z) = Y^2Z + a_1XYZ + a_3YZ^2 - X^3 - a_2X^2Z - a_4XZ^2 - a_6Z^3 \tag{2.1}$$

ist mit Koeffizienten $a_1, a_2, a_3, a_4, a_6 \in k$. Ist F klar, schreiben wir $E(k)$.

Bemerkung 11.4

- Die Monome X^2Y, Y^3, XY^2 brauchen also nicht vorzukommen.
- Die Nummerierung der Koeffizienten ist historisch bedingt.
- Die affine Version lautet also:

  $$Y^2 + a_1XY + a_3Y = X^3 + a_2X^2 + a_4X + a_6$$

  Die Form (2.1) nennen wir auch die **lange Weierstraßform**, das Polynom heißt **langes Weierstraßpolynom**.
- Wir werden sehen, dass man dies auf eine noch einfachere Form bringen kann.

Bemerkung 11.5
Welche Punkte liegen auf $E(k)$, die nicht affin sind? Ist $P = [r : s : 0] \in \mathbb{P}^2(k) \setminus \mathbb{A}^2(k)$ ein solcher Punkt, dann ergibt Einsetzen in (2.1) dann $r^3 = 0$, dann muss $s \neq 0$ sein, das heißt $P = [0 : s : 0] = [0 : 1 : 0]$. Diesen
lies: "Oh" unendlich fernen Punkt, der allen elliptischen Kurven gemeinsam ist, nennen wir $\mathcal{O} := [0 : 1 : 0]$. Dieser Punkt ist nie singulär, da $\frac{\partial F}{\partial Z}(0, 1, 0) = 1 \neq 0$. Somit genügt es, für ein Polynom F der Form (2.1) die Nichtsingularität auf $\mathcal{C}_F(k) \cap \iota(\mathbb{A}^2(k))$, also im Affinen, zu testen.

Beispiel 11.6
Sei $F(X, Y, Z) = Y^2Z - X^3 - XZ$, für dieses gilt $a_1 = a_2 = a_3 = a_6 = 0$. Dann ist $\mathcal{C}_F(\mathbb{F}_p) \cap \mathbb{A}^2(\mathbb{F}_p)$ für $p \geq 3$ nicht-singulär, also eine elliptische Kurve.

Bemerkung 11.7
Veranschaulichung, dass zum Beispiel alle elliptischen Kurven $E_s(\mathbb{R})$ zur Gleichung $y^2 = x^3 - 3x + s, s \in \mathbb{R}$, den unendlich fernen Punkt $\mathcal{O} = [0 : 1 : 0]$ gemeinsam haben:

Wer mir das folgende Bild text, bekommt Eis oder Bier!

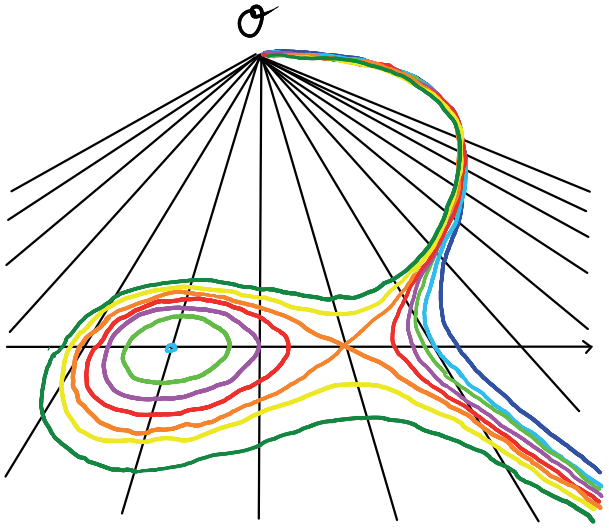


Abbildung 6: $y^2 = x^3 - 3x + s$ für $s = 5$, $s = 3$, $s = 2$, $s = 1$, $s = 0$, $s = -1$, $s = -1.999$, $s = -5$

Das Bild ist perspektivisch so verzerrt, dass der unendlich ferne Punkt $\mathcal{O}$, der für die Richtung der y-Achse steht, am Horizont erscheint.

Satz 11.8 (Vereinfachte Weierstraßgleichungen)
Sei $E_F(k)$ eine elliptische Kurve mit F in der langen Weierstraßform (2.1).

(i) Falls $\mathrm{char}(k) \neq 2$, ist die Abbildung

$$\begin{aligned} \Phi\colon \mathbb{P}^2(k) &\longrightarrow 2\mathbb{P}^2(k) \\ [r : s : t] &\longmapsto \left[r : s + \frac{a_1}{2}r + \frac{a_3}{2}t : t\right] \end{aligned}$$

bijektiv und es ist $\Phi(E_f(k)) = E_{H_1}(k)$ ebenfalls eine elliptische Kurve mit $H_1(X, Y, Z) = Y^2Z - X^3 - \frac{1}{4}b_2X^2Z - \frac{1}{2}b_4XZ^2 - \frac{1}{4}b_6Z^3$, wobei $b_2 = a_1^2 + 4a_2, b_4 = 2a_4 + a_1a_3, b_6 = a_3^2 + 4a_6$.

(ii) Falls $\mathrm{char}(k) \neq 2$ und $\mathrm{char}(k) \neq 3$, ist die Abbildung

$$\begin{aligned} \Psi\colon \mathbb{P}^2(k) &\longrightarrow \mathbb{P}^2(k) \\ [r : s : t] &\longmapsto [36r + 3b_2t : 216s : t] \end{aligned}$$

bijektiv und es ist $\Psi(E_{H_1}(k)) = E_{H_2}(k)$ ebenfalls eine elliptische Kurve mit $H_2(X, Y, Z) = Y^2Z - X^3 + 27c_4XZ^2 + 54c_6Z^3$, wobei $c_4 = b_2^2 - 24b_4, c_6 = -b_2^3 + 36b_2b_4 - 216b_6$.

Bemerkung 11.9
Wir können die lange Weierstraßgleichung im Fall $\mathrm{char}(k) \neq 2$ also stets zur affinen Gleichung $y_2 = x^3 + a_2x^2 + a_4x + a_6$ vereinfachen; falls $\mathrm{char}(k) \neq 2$ und $\mathrm{char}(k) =\neq 3$ gilt, sogar zu $y^2 = x^3 + a_4x + a_6$. Wir nennen diese Gleichung die **kurze Weierstraßform**, das entsprechende Polynom dann das **kurze Weierstraßpolynom**.

Bemerkung 11.10
Auch im Fall $\mathrm{char}(k) = 2$ lässt sich die lange Weierstraßgleichung vereinfachen, das ist nicht schwierig, wenn $a_1 \neq 0$, aber auch für $a_1 = 0$ möglich. Wir behandeln dies hier nicht näher.

Beweis 11.11
Zunächst zu (i):

- Φ ergibt als Abbildung nur Sinn, wenn 2 invertierbar ist in k, das heißt, falls $\mathrm{char}(k) \neq 2$ ist. Φ ist dann bijektiv, da Φ die Umkehrabbildung $\Phi^{-1}([r:s:t]) = \left[r : s - \frac{a_1}{2}r - \frac{a_3}{2}t : t\right]$ hat.
- Weiter bezeichnen wir mit Φ, Φ^{-1} auch die zugehörigen (affinen) Abbildungen $\Phi, \Phi^{-1}\colon k^3 \longrightarrow k$, $\Phi(r,s,t) = \left(r, s + \frac{a_1}{2}r + \frac{a_3}{2}t, t\right)$ bzw. $\Phi^{-1}(r,s,t) = \left(r, s - \frac{a_1}{2}r - \frac{a_3}{2}t, t\right)$. Nun können wir mit den im Satz angegebenen Zahlen b_2, b_4, b_6 nachrechnen, dass $H_1(X,Y,Z) = F\left(X, Y - \frac{a_1}{2}X - \frac{a_3}{2}Z, Z\right)$:

$$
\begin{aligned}
&F\Big(X, Y - \frac{a_1}{2}X - \frac{a_3}{2}Z, Z\Big) \\
&= \Big(Y - \frac{a_1}{2}X - \frac{a_3}{2}Z\Big)^2 Z + a_1 X\Big(Y - \frac{a_1}{2}X - \frac{a_3}{2}Z\Big)Z + a_3\Big(Y - \frac{a_1}{2}X - \frac{a_3}{2}Z\Big)Z^2 \\
&\quad - X^3 - a_2X^2Z - a_4XZ^2 - a_6Z^3 \\
&= Z \cdot \left(Y^2 - 2Y\Big(\frac{a_1}{2}X + \frac{a_3}{2}Z\Big) + \left(\frac{a_1^2}{4}X^2 + 2 \cdot \frac{a_1a_3}{4}XZ + \frac{a_3^2}{4}Z^2\right)\right) \\
&\quad + a_1XYZ - \frac{a_1^2}{2}X^2Z - \frac{a_1a_3}{2}XZ^2 + a_3YZ^2 - \frac{a_1a_3}{2}XZ^2 - \frac{a_3^2}{2}Z^3 - X^3 - a_2X^2Z - a_4XZ^2 - a_6Z^3 \\
&= Y^2Z - X^3 + \Big(-\frac{a_1^2}{4} - a_2\Big)X^2Z + \Big(-\frac{a_1a_3}{2} - a_4\Big)XZ^2 + \Big(-\frac{a_3^2}{4} - a_6\Big)Z^3 \\
&= Y^2Z - X^3 - \frac{1}{4}b_2X^2Z - \frac{1}{2}b_4XZ^2 - \frac{1}{4}b_6Z^3 = H_1(X,Y,Z)
\end{aligned}
$$

- Es folgt $H_1(r,s,t) = F(\Phi^{-1}(r,s,t))$, also gilt $F(r,s,t) = 0$ genau dann, wenn $H_1(\Phi(r,s,t)) = 0$, sodass $\Phi(E_F(k)) = \mathcal{C}_{H_1}(k)$ folgt. Es bleibt zu zeigen, dass $\mathcal{C}_{H_1}(k)$ nicht-singulär ist: Mit der Kettenregel rechnen wir nach:

$$
\begin{aligned}
\frac{\partial H_1}{\partial X}(r,s,t) &= \frac{\partial F}{\partial X}(\Phi^{-1}(r,s,t)) - \frac{a_1}{2}\frac{\partial F}{\partial Y}(\Phi^{-1}(r,s,t)) \\
\frac{\partial H_1}{\partial Y}(r,s,t) &= \frac{\partial F}{\partial Y}(\Phi^{-1}(r,s,t)) \\
\frac{\partial H_1}{\partial Z}(r,s,t) &= -\frac{a_3}{2}\frac{\partial F}{\partial Y}\Phi^{-1}(r,s,t) + \frac{\partial F}{\partial Z}(\Phi^{-1}(r,s,t))
\end{aligned}
$$

- Ist $P = [r:s:t] \in \mathcal{C}_{H_1}(\bar{k})$, dann ist $\Phi^{-1}(P) = \Phi^{-1}([r:s:t])$ als Punkt der Kurve $\mathcal{C}_F(\bar{k})$ nicht-singulär, da F elliptische Kurve ist. Die drei Ableitungen von F in $\Phi^{-1}(P)$ sind also nicht alle $= 0$, also sind auch die drei Ableitungen von H_1 in (r,s,t) nicht alle $= 0$. Also ist P auf $\mathcal{C}_{H_1}(\bar{k})$ nicht-singulär.

Zu (ii): Ψ hat die Inverse $[r:s:t] \mapsto \left[\frac{1}{36}r - \frac{b_2}{12}t : \frac{1}{216}s : t\right]$, da wegen $\mathrm{char}(k) \neq 2, \neq 3$ die Zahlen $\frac{1}{36}, \frac{1}{12}, \frac{1}{216} = \frac{1}{2^3 \cdot 3^3}$ in k existieren, und leicht zu bestätigen ist, dass $\Psi(\Psi^{-1}([r:s:t])) = [r:s:t]$ gilt. Durch geduldiges Nachrechnen zeigt man $H_2(X,Y,Z) = 2^6 3^6 \cdot H_1\left(\frac{1}{36}X - \frac{b_2}{12}Z, \frac{1}{216}Y, Z\right)$, daraus folgt $H_1(r,s,t) = 0$ genau dann, wenn $H_2(\Psi(r,s,t)) = 0$, das heißt $\Psi(E_{H_1}(k)) = \mathcal{C}_{H_2}(k)$. Wieder mit der Kettenregel kann auch die Nicht-Singularität von $\mathcal{C}_{H_2}$ gezeigt werden. □

Wir definieren zwei wichtige Kennzahlen projektiver Kurven wie folgt:

Definition 11.12 (Diskriminante, j-Invariante)
Sei $\mathcal{C}_F(k)$ die projektive ebene Kurve zum langen Weierstraßpolynom (2.1). Dann heißt die Zahl

$$\Delta = \Delta(\mathcal{C}_F(k)) = -b_2^2b_8 - 8b_4^3 - 27b_6^2 + 9b_2b_4b_6$$

mit $b_2 = a_1^2 + 4a_2, b_4 = 2a_4 + a_1a_3, b_6 = a_3^2 + 4a_6$ und $b_8 = a_1^2a_6 + 4a_2a_6 - a_1a_3a_4 + a_2a_6^2 - a_4^2$ die **Diskriminante** der Kurve $\mathcal{C}_F(k)$. Die Zahl

$$j = j(\mathcal{C}_F(k)) := \frac{(b_2^2 - 24b_4)^3}{\Delta} = \frac{c_4^3}{\Delta}$$

heißt die j-**Invariante** der Kurve $\mathcal{C}_F(k)$.

Bemerkung 11.13

- Die j-Invariante legt die Isomorphieklasse der elliptischen Kurve über $\overline{k}$ fest: Zwei elliptische Kurven sind isomorph über $\overline{k}$ genau dann, wenn sie dieselbe j-Invariante besitzen (ohne Beweis).
- j ist unabhängig von der Wahl der speziellen Kurvengleichung.

Bemerkung 11.14
Die Diskriminante einer Kurve $\mathcal{C}_F(k)$ ist ein nützliches Hilfsmittel, um zu testen, ob eine Kurve, die durch eine lange Weierstraßgleichung gegeben ist, nicht-singulär (und damit elliptisch) ist:

Satz 11.15
Sei die Kurve $\mathcal{C}_F(k)$ gegeben durch das lange Weierstraßpolynom F. Dann ist $\mathcal{C}_F(k)$ nicht-singulär genau dann, wenn $\Delta(\mathcal{C}_F(k)) \neq 0$ ist.

Mit der angegebenen Formel für Δ ist dies auch rechnerisch leicht zu testen – wichtig, um elliptische Kurven für die Anwendungen zu konstruieren. Dieses Diskriminantenkriterium zeigen wir im nächsten Abschnitt.

2.4.2 Das Diskriminantenkriterium

Definition 12.1 (Diskriminante)
Sei $\mathcal{C}_F(k)$ die projektive ebene Kurve zum langen Weierstraßpolynom [12]

$$F(X,Y,Z) = Y^2Z + a_1XYZ + a_3YZ^2 - X^3 - a_2X^2Z - a_4XZ^2 - a_6Z^3.$$

Dann heißt die Zahl

$$\Delta = \Delta(\mathcal{C}_F(k)) = -b_2^2b_8 - 8b_4^3 - 27b_6^2 + 9b_2b_4b_6$$

mit $b_2 = a_1^2 + 4a_2, b_4 = 2a_4 + a_1a_3, b_6 = a_3^2 + 4a_6$ und $b_8 = a_1^2a_6 + 4a_2a_6 - a_1a_3a_4 + a_2a_3^2 - a_4^2$ die **Diskriminante** der Kurve $\mathcal{C}_F(k)$.

Bemerkung 12.2
Im Fall einer Kurve $\mathcal{C}_F(k)$ in einer kurzen Weierstraßform $f(x,y) = y^2 - x^3 - ax - b$ haben wir $\Delta(\mathcal{C}_F(k)) = -8 \cdot (2a)^3 - 27 \cdot (4b)^2 = -16 \cdot (4a^3 + 27b^2)$, da $a_1 = 0, a_3 = 0, a_2 = 0, a_4 = a, a_6 = b$, also $b_2 = 0, b_4 = 2a, b_6 = 4b, b_8 = -a^2$. (vgl. Übungsaufgabe 3 auf Blatt 4).

Wir zeigen das Diskriminantenkriterium:

Satz 12.3 (Diskriminantenkriterium)
Sei die Kurve $\mathcal{C}_F(k)$ gegeben durch das lange Weierstraßpolynom F. Dann ist $\mathcal{C}_F(k)$ nicht-singulär genau dann, wenn $\Delta(\mathcal{C}_F(k)) \neq 0$ ist.

Bemerkung 12.4

- Bei diesem Kriterium, wenn $\Delta \neq 0$, erhalten wir, dass $\mathcal{C}_F(\overline{k})$ über dem algebraischen Abschluss $\overline{k}$ keine singulären Punkte enthält (insbesondere auch über k, aber über $\overline{k}$ ist eben noch stärker). Deswegen haben wir uns bei unserer Definition von "nicht-singuläre Kurve" auf $\overline{k}$ bezogen, was wegen Satz 12.3 also mathematisch leichter wird. Für $\mathrm{char}(k) \neq 2$ kann es aber nicht sein, dass $\mathcal{C}_F$ über k keine singulären Punkte hat und über $\overline{k}$ hingegen schon.

- Das Kriterium ist in der Praxis nützlich, da eine Kurve in Weierstraßform (die vielleicht per Zufallsgenerator für die Koeffizienten erzeugt worden ist), damit leicht auf Nicht-Singularität durch Berechnung der einfachen Formel für Δ getestet/überprüft werden kann.

- Der Beweis unterscheidet wesentlich die Fälle $\mathrm{char}(k) = 2, \mathrm{char}(k) = 3$ und sonst.

Beweis 12.5
Die Kurve $\mathcal{C}_F(k)$ ist nicht-singulär genau dann, wenn ihre affine Kurve $\mathcal{C}_F(k)$ mit $f(x,y) = y^2 + a_1xy + a_3y - x^3 - a_2x^2 - a_4x - a_6$ nicht-singulär ist. Wir zeigen den Satz deswegen im Affinen (der einzige nicht-affine Punkt $\mathcal{O} = [0:1:0]$ der Kurve ist immer regulär, vgl. Bemerkung 11.5). Nun enthält $\mathcal{C}_F(\overline{k})$ einen singulären Punkt (r,s) genau dann, wenn

$$f(r,s) = 0 \qquad \underbrace{\frac{\partial f}{\partial x}(r,s)}_{a_1s - 3r^2 - 2a_2r - a_4} = 0 \qquad \underbrace{\frac{\partial f}{\partial y}(r,s)}_{= 2s + a_1r + a_3} = 0$$

gilt. Wir unterscheiden weiter mehrere Fälle nach dem Wert der Charakteristik von k.

Beweis 12.6 (1. Fall: $\mathrm{char}(k) = 2$ und $a_1 = 0$)
"⇐": Dann ist hier $b_2 = b_4 = 0, b_6 = a_3^2, \Delta = -27a_3^4 = a_3^4$. Weiter gilt $\frac{\partial f}{\partial y} = a_3$, sodass, falls ein singulärer Punkt existiert, dann $a_3 = 0$ und $\Delta = 0$ folgt.

"⇒": Ist $\Delta = 0$, folgt $\frac{\partial f}{\partial y} = 0$. Nun existieren $r, s \in \overline{k}$ mit $r^2 + a_4 = 0, s^2 + a_3s = r^3 + a_2r^2 + a_4r + a_6$, also ist $(r,s) \in \mathbb{A}^2(\overline{k})$ singulärer Punkt auf $\mathcal{C}_f(\overline{k})$.

Beweis 12.7 (2. Fall: $\mathrm{char}(k) = 2$ und $a_1 \neq 0$)
In Charakteristik 2 gilt:

$$\begin{aligned}\Delta &= -a_4^4(a_1^2a_6 - a_1a_3a_4 + a_2a_3^2 - a_4^2) - 27a_3^4 + a_1^3a_3^3 \\ &= a_1^6a_6 + a_1^5a_3a_4 + a_1^4a_2a_3^2 + a_1^4a_4^2 + a_1^3a_3^3 + a_3^4.\end{aligned}$$

"⇐": Hat $\mathcal{C}_f(\overline{k})$ einen singulären Punkt (r,s), so ist $f(r,s) = 0$, das heißt $a_1s + r^2 + a_4 = 0$ und $a_1r + a_3 = 0$. Da $a_1 \neq 0$, folgt $r = \frac{a_3}{a_1}$ und $s = \frac{a_3^2 + a_1^2a_4}{a_1^3}$. Durch einsetzen in $f(r,s)$ folgt $0 = f(r,s) = \Delta a_1^{-6}$, also $\Delta = 0$.

"⇒": Ist $\Delta = 0$, definieren wir r, s wie oben, dann ist $f(r,s) = \Delta a_1^{-6}$, also $f(r,s) = 0$, womit ein singulärer Punkt konstruiert ist.

Beweis 12.8 (3. Fall: $\mathrm{char}(k) = 3$)
Via Rechnen in Charakteristik 3 folgt $\Delta = -b_2^2b_8 - 8b_4^3$. Betrachte die Abbildung $\Phi\colon \mathcal{C}_F(k) \to \mathcal{C}_{H_1}(k)$ aus Satz 11.8(i), die die lange Weierstraßform F auf die kurze Form H_1 bringt. Es ist $\Delta(\mathcal{C}_{H_1}(k)) = \Delta(\mathcal{C}_F(k))$ durch Nachrechnen, somit genügt es ohne Einschränkung, das Diskriminantenkriterium für die kurze Form H_1 zu zeigen.

Beweis 12.9 (Fortsetzung 3. Fall)
Die Kurve $\mathcal{C}_{H_1}(\overline{k})$ enthält genau dann einen singulären Punkt, wenn es $r, s \in \overline{k}$ gibt mit

$$s^2 - r^3 - \frac{1}{4}b_2r^2 - \frac{1}{2}b_4r - \frac{1}{4}b_6 = 0, \qquad 3r^2 + \frac{1}{2}b_2r + \frac{1}{2}b_4 = 0, \qquad 2s = 0,$$

das heißt falls r eine doppelte Nullstelle des Polynoms $\sigma(x) := x^3 + \frac{1}{4}b_2x^2 + \frac{1}{2}b_4x\frac{1}{4}b_6$ ist, sprich $\sigma(r) = 0 = \sigma'(r)$ ist.
Über $\overline{k}$ zerfällt σ in drei Linearfaktoren $\sigma(x) = (x-\alpha_1)(x-\alpha_2)(x-\alpha_3)$ mit $\alpha_i \in \overline{k}$. Nun hat ein Polynom σ genau dann eine doppelte Nullstelle, falls seine Diskriminante $\mathrm{disc}(\sigma) := (\alpha_1-\alpha_2)^2(\alpha_1-\alpha_3)^2(\alpha_2-\alpha_3)^2$ verschwindet, vgl. Bemerkung 12.14.

Beweis 12.10 (Fortsetzung)
Somit ist im dritten Fall zu zeigen: $\Delta = 0 \Leftrightarrow \operatorname{disc}(\sigma) = 0$.
Wegen $\operatorname{disc}(x^3 + ux^2 + vx + w) = u^2v^2 - 4u^3w - 4v^3 - 27w^2 + 18uvw$, vgl. Korollar 12.16, erhalten wir mit $u = \frac{b_2}{4}, v = \frac{b_4}{2}, w = \frac{b_6}{4}$ dann in Charakteristik 3, dass $\operatorname{disc}(\sigma) = \frac{1}{64}b_2^2b_4^2 - \frac{1}{64}b_2^3b_6 - \frac{1}{2}b_4^3$. Wegen $4b_8 = b_2b_6 - b_4^2$ erhalten wir $\operatorname{disc}(\sigma) = \frac{1}{16}(-b_2^2b_8 - 8b_4^3) = \frac{1}{16}\Delta$. Aus dieser Formel folgt die Behauptung im dritten Fall.

Beweis 12.11 (4. Fall: $\mathbf{char(k) > 3}$ oder $\mathbf{char(k) = 0}$)
Mit der Bijektion $\Psi \circ \Phi \colon \mathcal{C}_F(k) \to \mathcal{C}_{H_2}(k)$ zum kurzen Weierstraßpolynom H_2 (Satz 11.8(ii)) bzw. $h_2(x, y) = y^2 - x^3 + 27c_4x + 54c_6$ mit $c_4 = b_2^2 - 24b_4, c_6 = -b_2^3 + 36b_2b_4 - 216b_6$, folgt durch Untersuchung der Ableitungen wieder $\mathcal{C}_F(k)$ nicht-singulär genau dann, wenn $\mathcal{C}_{H_2}(k)$ nicht-singulär. Wir berechnen $\Delta(\mathcal{C}_{H_2}(k)) = 2^6 3^9 \cdot (c_4^3 - c_6^2) = \ldots = 2^{12}3^{12}\Delta(\mathcal{C}_F(k))$, also genügt es, die Behauptung für $\mathcal{C}_{H_2}(k)$ zu zeigen. Wie im dritten Fall haben wir:

$$\mathcal{C}_{H_2}(\overline{k}) \text{ enthält singulären Punkt} \quad \Leftrightarrow \quad \sigma(x) = x^3 \underbrace{-27c_4}_{v} x \underbrace{-54c_6}_{w} \text{ hat doppelte Nullstelle} \quad \Leftrightarrow \quad \operatorname{disc}(\sigma) = 0$$

Wegen der Formel in Korollar 12.16 für $\operatorname{disc}(\sigma)$ folgt mit $u = 0, v = -27c_4, w = -54c_6$:

$$\operatorname{disc}(\sigma) = 0 \quad \Leftrightarrow \quad 4 \cdot 27^3 c_4^3 - 27 \cdot 54^2 c_6^2 = 0 \quad \Leftrightarrow \quad c_4^3 - c_6^2 = 0.$$

Daraus folgt die Behauptung. □

Theoretische Ergänzungen zu unserer Definition von $\Delta(\mathcal{C}_F(k))$:

Definition 12.12 (Diskriminante eines Polynoms)
Die **Diskriminante eines Polynoms** $\sigma \in k[x], n := \deg(\sigma) \geq 1$, ist

$$\operatorname{disc}(\sigma) := \prod_{1 \leq i < j \leq n} (\alpha_i - \alpha_j)^2 = (-1)^{\frac{n(n-1)}{2}} \prod_{i \neq j} (\alpha_i - \alpha_j) \in \overline{k},$$

falls $\alpha_1, \ldots, \alpha_n \in \overline{k}$ die Nullstellen von $\sigma \in \overline{k}$ bezeichnen.

Bemerkung 12.13
Man vergleiche dies mit der Diskriminante $\operatorname{disc}(\sigma) = p^2 - 4q$ eines quadratischen Polynoms $\sigma(x) = x^2 + px + q \in k[x]$, wir haben $\alpha_{1,2} = -\frac{p}{2} \pm \sqrt{\frac{p^2}{4} - q} = -\frac{p}{2} \pm \frac{1}{2}\sqrt{p^2 - 4q}$, also genau $(\alpha_1 - \alpha_2)^2 = \operatorname{disc}(\sigma)$. Ist dies $= 0$, ist $\alpha_1 = \alpha_2$ eine doppelte Nullstelle von σ.

Bemerkung 12.14
$\operatorname{disc}(\sigma)$ verschwindet genau dann, wenn σ über $\overline{k}$ eine doppelte Nullstelle hat. Dies folgt unmittelbar aus der Definition von σ.

Bemerkung 12.15

- Ist $\sigma \in k[x], n = \deg(\sigma) \geq 1$, ein normiertes Polynom, kann die Beziehung $\operatorname{disc}(\sigma) = (-1)^{n(n-1)/2} \cdot \operatorname{Res}(\sigma, \sigma')$ mit der in Definition 10.5 behandelten Resultante gezeigt werden.

- Aus dieser wichtigen Formel folgt wegen unserer Definition für die Resultante, dass stets $\operatorname{disc}(\sigma) = k$ gilt.

Korollar 12.16
Es gilt $\operatorname{disc}(x^3 + ux^2 + vx + w) = u^2v^2 - 4wu^3 - 4v^3 - 27w^2 + 18uvw$.

Beweis

Genaue Rechnung im Skript!

Für $\sigma(x) = x^3 + ux^2 + vx + w$ und $\sigma'(x) = 3x^2 + 2ux + v$ ist nach Defintion 10.5:

$$M(\sigma, \sigma') = \begin{pmatrix} w & 0 & v & 0 & 0 \\ v & w & 2u & v & 0 \\ u & v & 3 & 2u & v \\ 1 & u & 0 & 3 & 2u \\ 0 & 1 & 0 & 0 & 3 \end{pmatrix}$$

Durch Berechnung von $\mathrm{Res}(\sigma) = \det(M(\sigma, \sigma'))$ und $(-1)^{3\cdot(3-1)/2} = -1$ folgt die Behauptung. □

2.4.3 Die Gruppenstruktur elliptischer Kurven

[13] Sei $E(k)$ eine elliptische Kurve über einem Körper k.

Satz 13.1

(a) Seien $P, Q \in E(k), P \neq Q$, und $G = G(P, Q) \subseteq \mathbb{P}^2(k)$ die projektive Gerade, die P und Q verbindet. Dann hat G noch einen dritten Schnittpunkt mit $E(k)$ gemäß Vielfachheiten gezählt (das heißt, eventuell P bzw. Q selbst, falls $m(P; G, E(k)) = 2$ bzw. $m(Q; G, E(k)) = 2$).

(b) Sei G die Tangente an $E(k)$ im Punkt $P \in E(k)$. Dann hat G noch einen dritten Schnittpunkt mit $E(k)$ gemäß der Vielfachheiten gezählt (das heißt, eventuell P selbst, falls $m(P; G, E(k)) = 3$).

Beweis 13.2

Als Ergänzung zum Satz von Bézout haben wir Satz 10.15 kennen gelernt, der im Spezialfall $\deg(F_1) = 1$, $\deg(F_2) = 3$ dann $\sum_{P \in \mathcal{C}_{F_1} \cap \mathcal{C}_{F_2}} m(P; \mathcal{C}_{F_1}, \mathcal{C}_{F_2}) \in \{0, 1, 3\}$ liefert (siehe Beispiel 10.16). Also gilt auch hier: $\sum_{R \in G \cap E(k)} m(R; G, E(k)) \in \{0, 1, 3\}$.

zu (a): Ist $G = G(P, Q)$, folgt $2 \leq \#(G \cap E(k)) \leq \sum_{R \in G \cap E(k)} m(R; G, E(k)) \in \{0, 1, 3\}$. Das geht nur, wenn die Vielfachensumme 3 ist, also existiert ein $R \in G \cap E(k)$

- mit $R \notin \{P, Q\}$, falls $m(P; G, E(k)) = 1 = m(Q; G, E(k))$,
- oder mit $R = P$, falls $m(P; G, E(k)) = 2$
- oder mit $R = Q$, falls $m(Q; G, E(k)) = 2$.

zu (b): Ist G die Tangente an $E(k)$ in $P \in E(k)$, ist $m(P; G, E(k)) \geq 2$ nach Satz 9.22. Es folgt wie im Beweis zu (a) wieder, dass die Vielfachensumme 3 ist, also die Existenz eines $R \in G \cap E(k)$ mit $R \neq P$, falls $m(P; G, E(k)) = 2$ und $R = P$, falls $m(P; G, E(k)) = 3$ gilt. □

Beispiel 13.3

Betrachte die elliptische Kurve $E(k)$ zur (kurzen) Weierstraßgleichung $y^2 = x^3 - 3x + 3$. Jede Gerade, die $E(k)$ in zwei Punkten schneidet, schneidet $E(k)$ in einem dritten Punkt, gemäß Vielfachheit gezählt. Der dritte Schnittpunkt kann auch $\mathcal{O} \in g_\infty$ sein.

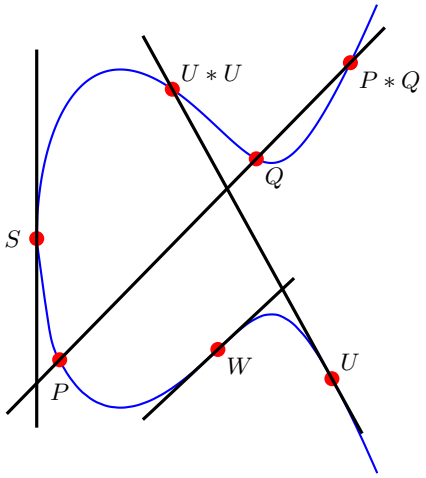


Abbildung 7: Veranschaulichung der Definition 13.4 an dem Beispiel $y^2 = x^3 - 3x + 3$: Für zwei verschiedene Punkte P, Q definieren wir $P * Q$ als den dritten Schnittpunkt der Gerade durch P und Q. Das kann wieder P oder Q sein, falls die Gerade eine Tangente durch P oder Q ist. Für einen Punkt U setzen wir $U * U$ als den weiteren Schnittpunkt der Tangente durch U; dieser kann auch $\mathcal{O}$ sein (z.B. im Fall von S). Ein Wendepunkt W ist bereits dreifacher Schnittpunkt, daher definieren wir hier $W * W = W$.

Wir möchten auf $E(k)$ eine Verknüpfung "$+$" erklären, also eine Punkteaddition $P + Q$, bei der wiederum ein Punkt auf der elliptischen Kurve herauskommt. Den dritten Schnittpunkt, den die Gerade $G(P, Q)$ durch zwei Punkte P und Q auf $E(k)$ mit $E(k)$ hat laut Satz 13.1, bezeichnen wir mit $P * Q$.

Definition 13.4 (dritter Schnittpunkt)
Für $P, Q \in E(k)$, $P \neq Q$ definieren wir also:

$$P * Q := \begin{cases} R \in (G(P,Q) \cap E(k)) \setminus \{P, Q\}, & \text{falls } m(P; G, E(k)) = 1 = m(Q; G, E(k)) \\ P, & \text{falls } m(P; G, E(k)) = 2 \\ Q, & \text{falls } m(Q; G, E(k)) = 2, \end{cases}$$

sowie

$$P * P := \begin{cases} R \in (T_P(E(k)) \cap E(k)) \setminus \{P\}, & \text{falls } m(P; T_p(E(k)), E(k)) = 2 \\ P, & \text{falls } m(P; T_P(E(k)), E(k)) = 3 \text{ (d.h. falls } P \text{ Wendepunkt)} \end{cases}$$

Bemerkung 13.5

- Der unendlich ferne Punkt $\mathcal{O} \in E(k)$ erfüllt $\mathcal{O} * \mathcal{O} = \mathcal{O}$, da er ein Wendepunkt ist.
- Weiter ist offensichtlich $P * Q = Q * P$ aufgrund der Definition.
- Es gilt: $R = P * Q \Rightarrow P = Q * R \Rightarrow Q = R * P$, das heißt $P * (P * Q) = Q$ für alle $P, Q \in E(k)$. [#]
- Damit gilt auch $P * Q = P * R \Leftrightarrow Q = R$, denn $P * Q = P * R \Rightarrow P * (P * Q) = P * (P * R) \Rightarrow Q = R$.

Bemerkung 13.6
Man beachte, dass für $P = [a : b : 1] \in \mathbb{P}^2(k) \cap \iota(\mathbb{A}^2(k))$ die Gerade $G(P, Q) = G(c, 0, -a) = \{[x : y : z] \in \mathbb{P}^2(k) : x - az = 0\}$ im Affinen eine Parallele zur y-Achse darstellt (Gleichung $x = a$). Für eine elliptische Kurve, die durch eine kurze Weierstraßform gegeben und (für $\mathrm{char}(k) \neq 2$) symmetrisch zur x-Achse ist, wird typischerweise

$P * \mathcal{O} \neq P$ sein. Wegen $(\mathcal{O} * \mathcal{O}) * P = \mathcal{O} * P$ einerseits, da $\mathcal{O} * \mathcal{O} = \mathcal{O}$, und $\mathcal{O} * (\mathcal{O} * P) = P$ andererseits wegen [#], kann die Verknüpfung $*$ also nicht assoziativ sein. Stattdessen setzen wir unsere Verknüpfung "$+$" wie folgt:

Definition 13.7 (Punkteaddition auf elliptischen Kurven)
Für $P, Q \in E(k)$ definieren wir

$$P + Q := \mathcal{O} * (P * Q).$$

Ist $E(k)$ in kurzer Weierstraßform und (für $\mathrm{char}(k) \neq 2$) symmetrisch zur x-Achse, erhält man $P + Q$, indem man den dritten Schnittpunkt $P * Q$ von $G(P, Q)$ mit $E(k)$ dann noch an der x-Achse spiegelt, das heißt das Negative des y-Wertes nimmt.

Bemerkung 13.8

- Es gilt $\mathcal{O} + P = \mathcal{O} * (\mathcal{O} * P) = P$ nach [#], das heißt $\mathcal{O}$ ist neutrales Element bezüglich $+$.
- Es gilt $-P = \mathcal{O} * P$, da $P + (\mathcal{O} * P) = \mathcal{O} * (P * (\mathcal{O} * P)) = (P * (\mathcal{O} * P)) * \mathcal{O} = (P * (P * \mathcal{O})) * \mathcal{O} = \mathcal{O}$ nach [#].

Es ist nicht auf Anhieb zu sehen, dass hier mit "$+$" eine elliptische Kurve $E(k)$ zu einer Gruppe $(E(k), +)$ wird, sprich, ob das Assoziativgesetz gilt.

Lemma 13.9
Liegen drei Punkte $P, Q, R \in E(k)$ einer elliptischen Kurve auf einer projektiven Gerade G, so gilt $(P+Q)+R = \mathcal{O}$ und umgekehrt. Dabei müssen P, Q, R nicht notwendig verschieden sein.

Beweis
Da $R = P * Q$ nach Voraussetzung, folgt $-R = \mathcal{O} * R = \mathcal{O} * (P * Q) = P + Q$. Umgekehrt gilt das ebenso. □

Satz 13.10 (Elliptische Kurve mit Punktaddition ist abelsche Gruppe, Poincaré, 1901)
Sei k ein beliebiger Körper und $E(k)$ eine elliptische Kurve. Die Verknüpfung $+\colon (P, Q) \mapsto P + Q$ macht $E(k)$ zu einer abelschen Gruppe $(E(k), +)$ mit neutralem Element $\mathcal{O}$, das heißt:

(i) $P + \mathcal{O} = P$ für alle $P \in E(k)$.

(ii) Für alle $P \in E(k)$ existiert ein $Q \in E(k)$ mit $P + Q = \mathcal{O}$; setze $-P := Q$.

(iii) $P + Q = Q + P$ für alle $P, Q \in E(k)$.

(iv) $(P + Q) + R = P + (Q + R)$ für alle $P, Q, R \in E(k)$.

Beweis

(i) Siehe Bemerkung 13.8.

(ii) Für $P \in E(k)$ setze $Q := \mathcal{O} * P$, siehe Bemerkung 13.8.

(iii) Klar aufgrund der Definition von $P + Q$.

(iv) Beweis folgt im nächsten Vorlesungsteil. □

Bemerkung 13.11

- Anstelle von $\mathcal{O}$ könnte man prinzipiell jeden Punkt $Q \in E(k)$ zum neutralen Element von $+$ machen, indem man $U \oplus V := U + V - Q$ setzt: Dann ist $U \oplus Q = U + Q - Q = U, U \oplus (-U + 2Q) = U - U + 2Q - Q = Q$

und $(U \oplus V) \oplus W = U + V + W - 2Q = U \oplus (V \oplus W)$. Die Eigenschaft von Lemma 13.9 gilt immer noch, wenn für Q ein Wendepunkt genommen wird. Nun kann eine elliptische Kurve bis zu neun Wendepunkte haben, vergleiche Übungsaufgabe 4 (a) auf Blatt 5.

- Die Wahl von $\mathcal{O} := [0 : 1 : 0]$ als Wendepunkt von $E(k)$, das heißt mit $\mathcal{O} * \mathcal{O} = \mathcal{O}$, hat den Vorteil, dass die explizite Formel für $+$ rechnerisch einfacher wird, weil dann die Gleichung von $E(k)$ in einfacher (langer oder kurzer) Weierstraßform vorliegt. Diese Formel wird in Satz 13.13 bzw. Satz 13.14 angegeben.
- Invertieren, das heißt Berechnen von $-P = P * \mathcal{O}$, ist dann bei kurzer Weierstraßform in $\mathrm{char}(k) \neq 2$ und $\mathrm{char}(k) \neq 3$ besonders leicht: Man spiegelt P an der x-Achse und erhält $-P$, das heißt ist $P = [a : b : c]$, gilt $-P = [a : -b : c]$. Schnittpunkte von $E(k)$ mit der x-Achse sind dann selbstinvers, das heißt für $P = [a : 0 : c] \in E(k)$ gilt $-P = P$.

Beispiel 13.12
Betrachte $E(\mathbb{R})$ mit der Gleichung $y^2 = x^3 + 17$ bzw. $Y^2Z = X^3 + 17Z^3$. Dann liegen $P = [-1 : 4 : 1]$ und $Q = [-2 : 3 : 1]$ auf der Kurve. Ihre Verbindungsgerade ist $G(P, Q) = \{[x : y : z] \in \mathbb{P}^2(\mathbb{R}) : x - y + 5z = 0\}$. Um $P + Q$ zu berechnen, bestimmen wir den dritten Schnittpunkt $P * Q$ von $G(P, Q)$ und $E(\mathbb{R})$ wie folgt: Setze $y = x + 5$ ein und erhalte $(x + 5)^2 = x^3 + 17 \Leftrightarrow x^3 - x^2 - 10x - 8 = 0$. Da $x = -1, x = -2$ Nullstelle der linken Seite sind, führt Polynomdivision durch $(x + 1)(x + 2)$ zu $x^3 - x^2 - 10x - 8 = (x + 1)(x + 2)(x - 4)$. Der Punkt $P * Q$ hat also die x-Koordinate 4; da er auf G liegt, folgt $P * Q = [4 : 9 : 1]$. Es folgt $P + Q = [4 : -9 : 1]$.

Ist $E(k)$ gegeben durch das lange Weierstraßpolynom

$$F(X, Y, Z) = Y^2Z + a_1XYZ + a_3YZ^2 - X^3 - a_2X^2Z - a_4XZ^2 - a_6Z^3$$

bzw. $f(x, y) = y^2 + a_1xy + a_3y - x^3 - a_2x^2 - a_4x - a_6$, so möchten wir die Addition "+" für die Krypto-Anwendungen in einer expliziten Formel beschreiben:

Satz 13.13 (Punkteaddition bei langer Weierstraßform)
Sei $E(k)$ gegeben durch das lange Weierstraßpolynom F. Dann gilt:

(a) $P = (u, v) \in E(k) \cap \mathbb{A}^2(k) \Rightarrow -P = (u, -v - a_1u - a_3)$

(b) Sei $P = (u, v), Q = (r, s) \in E(k) \cap \mathbb{A}^2(k)$.

- Ist $u = r$ und $v + s + a_1u + a_3 = 0$, gilt $P + Q = \mathcal{O}$.
- Sonst ist

$$P + Q = (\underbrace{\lambda^2 + a_1\lambda - a_2 - u - r}_{=:x_3}, -(\lambda + a_1)x_3 - \mu - a_3),$$

Beweis wäre langweiliges Nachrechnen…

wobei

$$\lambda = \frac{s - v}{r - u}, \quad \mu = \frac{vr - su}{r - u}, \quad \text{falls } r \neq u,$$

und

$$\lambda = \frac{3u^2 + 2a_2u + a_4 - a_1v}{2v + a_1u + a_3}, \quad \mu = \frac{-u^3 + a_4u + 2a_6 - a_3v}{2v + a_1u + a_3}, \quad \text{falls } r = u.$$

Wir geben den Beweis für die Formel bei kurzer Weierstraßform, wo er leicht zu machen ist:

Satz 13.14 (Punkteaddition bei kurzer Weierstraßform)
Ist $E(k)$ gegeben durch $f(x, y) = y^2 - x^3 - ax - b$, so gilt:

(a) Für $P = (u, v) \in E(k)$ gilt $-P = (u, -v)$.

(b) Für $P = (u, v), Q = (r, s)$ mit $P \neq -Q$ ist

$$P + Q = (\underbrace{\lambda^2 - u - r}_{=:x}, \lambda(u - x) - v),$$

wobei

$$\lambda = \begin{cases} \frac{s-v}{r-u}, & \text{falls } P \neq Q \\ \frac{3u^2+a}{2v}, & \text{falls } P = Q. \end{cases}$$

Beweis
(a) ist klar. Zu (b): Sei $P \neq \pm Q$. Haben $g(P, Q) = \{(u+t, v+\lambda t) : t \in \mathbb{R}\}$ mit $\lambda := \frac{s-v}{r-u}$, sowie $f(u+t, v+\lambda t) = (v + \lambda t)^2 - (u + t)^3 - a(u + t) - b$ mit den Nullstellen $t = 0, t = r - u$. Eine weitere Nullstelle ist $t = x - u$. Die Nullstellensumme $0 + r - u + x - u = r + x - 2u$ ergibt den Koeffizienten vor t^2 des Polynoms, nämlich $\lambda^2 - 3u$; es folgt $x = \lambda^2 - u - r$. Die y-Koordinate des Punkts auf $g(P, Q)$ ist dann $v + \lambda(x - u)$, für $P + Q$ dann das Negative.
Ist $Q = P, P \neq -P$ (das heißt $v \neq 0$), nimmt man die Tangente an $E(k)$ im Punkt P, also

$$\begin{aligned} t_P(E(k)) &= \{(x, y) : (-3u^2 - a)x + 2vy + v^2 - 2au - 3b = 0\} \\ &= \{(u + t, v + \lambda t) : t \in \mathbb{R}\} \end{aligned}$$

mit der Steigung $\lambda := \frac{3u^2+a}{2v}$. Der Rest der Rechnung geht wie eben; es folgt $x = \lambda^2$ (da $u = r$) und der angegebene y-Wert für $P + P$. □

2.4.4 Das Assoziativgesetz

[14] Sei $E(k)$ eine elliptische Kurve über einem Körper k. Für $P, Q \in E(k)$ hatten wir die Verknüpfung "$+$" definiert als $P + Q := \mathcal{O} * (P * Q)$, dafür ist $\mathcal{O}$ das neutrale Element sowie $-P = \mathcal{O} * P$ das Inverse von P. Weiter haben wir die Relation $P * (P * Q) = Q$ für alle $P, Q \in E(k)$.

Bemerkung 14.1
Wir zeigen für eine elliptische Kurve $E(k)$ das Assoziativgesetz: Für alle $P, Q, R \in E(k)$ gilt

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

Lemma 14.2
Das Assoziativgesetz auf $E(k)$ gilt genau dann, wenn für alle $P, Q, R, S \in E(k)$ gilt:

$$(P * Q) * (R * S) = (P * R) * (Q * S)$$

Beweis
Zunächst gilt $P + (Q + R) = \mathcal{O} * (P * (Q + R))$ und $(P + Q) + R = \mathcal{O} * ((P + Q) * R)$. Wegen $\mathcal{O} * (\mathcal{O} * P)$ folgt

$$\mathcal{O} * U = \mathcal{O} * V \Leftrightarrow U = V$$

Also ist die Assoziativität äquivalent zur Gleichung

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

"⇒": Wenn das Assoziativgesetz gilt, folgt mit $P * Q = -(P + Q)$ dann

$$\begin{aligned} (P * Q) * (R * S) &= -((P * Q) + (R * S)) = -(-(P + Q) - (R + S)) = (P + Q) + (R + S) \\ &= (P + R) + (Q + S) \overset{\text{analog}}{=} (P * R) * (Q * S) \end{aligned}$$

"⇐: Gilt die Relation, folgt mit $\tilde{P} = \mathcal{O}, \tilde{Q} = P * Q, \tilde{R} = Q * R, \tilde{S} = Q$ daraus $(\tilde{P} * \tilde{Q}) * (\tilde{R} * \tilde{S}) = (\tilde{P} * \tilde{R}) * (\tilde{Q} * \tilde{S})$, also

$$\underbrace{(\mathcal{O} * (P * Q))}_{P+Q} * \underbrace{((Q * R) * Q)}_{=R} = \underbrace{(\mathcal{O} * (Q * R))}_{Q+R} * \underbrace{((P * Q) * Q)}_{=P},$$

was nach oben Gesagtem die Assoziativität impliziert. □

Als Beweishilfsmitten benötigen wir den **Neunpunktesatz**.

Satz 14.3 (Neunpunktesatz)
Sei k ein algebraisch abgeschlossener Körper. Seien $\mathcal{C}_F, \mathcal{C}_{F_1}$ und $\mathcal{C}_{F_2}$ drei kubische Kurven in $\mathbb{P}^2(k)$ zu paarweise teilerfremden homogenen Polynomen F, F_1, F_2 vom Grad 3, und $\mathcal{C}_F$ enthalte acht der neun Schnittpunkte von $\mathcal{C}_{F_1} \cap \mathcal{C}_{F_2}$. Dann liegt auch der neunte Schnittpunkt auf $\mathcal{C}_F$.

Beweis 14.4 (nicht vollständig, nur Beweisidee)
Dass $\mathcal{C}_{F_1} \cap \mathcal{C}_{F_2}$ (mit Vielfachheiten gezählt) aus genau $9 = 3 \cdot 3$ Punkten besteht, besagt der Satz 10.2 von Bézout. Damit bestehen auch $\mathcal{C}_F \cap \mathcal{C}_{F_1}$ und $\mathcal{C}_F \cap \mathcal{C}_{F_2}$ aus neun Punkten, von denen laut Voraussetzung acht Punkte identisch sind. Eine allgemeine projektive kubische Kurve $\mathbb{C}$ (Singularitäten egal) ist definiert über 10 Koeffizienten:

$$\mathcal{C}\colon a_1X^3 + a_xX^2Y + a_3X^2Z + a_4XY^2 + a_5XYZ + a_6XZ^2 + a_7Y^3 + a_8Y^2Z + a_9YZ^2 + a_{10}Z^3 = 0$$

- Die Kurve verändert sich nicht, wenn das kubische Polynom mit einem Skalar $\rho \neq 0$ multipliziert wird. Deswegen sehen wir eine kubische Kurve als einen projektiven Punkt $\underline{a} = [a_1 : a_2 : \ldots : a_{10}] \in \mathbb{P}^9(k)$ an. Betrachte nun alle kubischen Kurven, die durch die acht vorgegebenen Punkte verlaufen. Sei $M \in k^{8\times 10}$ die Matrix, für die die nichttrivialen Lösungen des linearen Gleichungssystems $M \cdot \underline{a} = \underline{0} \in k^8$ genau die kubischen Kurven $\underline{a} \in \mathbb{P}^9(k)$ ergeben, die durch die acht Punkte verlaufen. Dann ist $\mathrm{Rang}(M) \leq 8$, also folgt wegen der Dimensionsformel $\dim(\ker(M)) = 10 - \mathrm{Rang}(M) \geq 2$.

- Im Allgemeinen ist der Lösungsraum zweidimensional, das heißt $\dim(\ker(M)) = 2$, sofern die acht Punkte in allgemeiner Lage liegen, was wir hier zur Vereinfachung annehmen möchten; liegen die Punkte in spezieller Lage, kann der Lösungsraum mindestens dreidimensional werden. Zur Beweisführung sind dann umständliche Fallunterscheidungen nötig, die wir hier nicht weiter ausführen möchten.

- Ist $\dim(\ker(M)) = 2$, wird der Lösungsraum von den Koeffizienten von F_1 und F_2 aufgespannt, das heißt es existieren $r, s \in k$ mit $F = rF_1 + sF_2$. Für den neunten Schnittpunkt $[X : Y : Z]$ gilt $F_1(X, Y, Z) = 0 = F_2(X, Y, Z)$, und somit auch $F(X, Y, Z) = 0$, das heißt $[X : Y : Z] \in \mathcal{C}_F$. □

Beweis 14.5 (Assoziativität von "+")
Ohne Einschränkung sei k ein algebraisch abgeschlossener Körper (die Assoziativität folgt dann für Teilkörper). Weiter genügt es, die Relation des Lemmas 14.2 nachzuweisen. Seien $P, Q, R, S \in E(k)$ und $E(k)$ durch das kubische homogene Polynom $F \in k[X, Y, Z]$ definiert.

- Wir betrachten dazu die folgenden acht Punkte:

$$P, \quad Q, \quad P * Q, \quad R, \quad S, \quad R * S, \quad P * R, \quad Q * S \tag{\#}$$

Diese definieren sechs Geraden G_1, G_2, G_3 und H_1, H_2, H_3 so, dass die Schnittpunkte der Geraden durch folgende Tabelle gegeben sind:

	G_1	G_2	G_3
H_1	P	Q	$P * Q$
H_2	R	S	$R * S$
H_3	$P * R$	$Q * S$	$H_3 \cap G_3$

Nun ist zu zeigen, dass der Schnittpunkt $H_3 \cap G_3$ ebenfalls wieder auf $E(k)$ liegt, denn daraus folgt dann $\{(P * R) * (Q * S)\} = H_3 \cap G_3 = \{(P * Q) * (R * S)\}$.

Seien $G_1, G_2, G_3, H_1, H_2, H_3$ auch die linearen homogenen Polynome in $k[X, Y, Z]$ bezeichnet, da diese Geraden definieren. Wir betrachten dann die beiden kubischen Polynome $G_1G_2G_3$ und $H_1H_2H_3$. Diese enthalten jeweils alle neun angegebenen Punkte der Tabelle. Die elliptische Kurve tritt acht dieser Punkte des Schnittes $\mathcal{C}_{G_1G_2G_3} \cap \mathcal{C}_{H_1H_2H_3}$, nämlich die Punkte der Liste (#).

Nach dem Neunpunktesatz liegt dann auch der neunte Punkt des Schnittes, nämlich $H_3 \cap G_3$, auf der elliptischen Kurve, wie zu zeigen war. □

Bemerkung 14.6

- Man kann die Assoziativität auch mit Hilfe der expliziten Formeln aus Satz 13.13 nachrechnen, was mühsam ist.
- Algebraiker betrachten zu beliebigen algebraischen Varietäten die so genannte Divisorklassengruppe, welche eine abelsche Gruppe laut Definition ist, sowie eine Abbildung von $E(k)$ auf ihre Divisorklassengruppe. Diese ist ein Isomorphismus nach dem tiefen Satz von Riemann-Roch, das heißt die Gruppenstruktur überträgt sich, insbesondere also die Assoziativität der Verknüpfung "$+$".

2.4.5 Schnelle Arithmetik auf elliptischen Kurven

[15] Die Gruppenoperation "$+$" auf einer elliptischen Kurven $E(k)$ soll rechnerisch praktisch mit den expliziten Formeln aus Satz 13.13 und 13.14 auf dem Computer umgesetzt werden.

Bemerkung 15.1
Ein Blick auf die expliziten Formeln zeigt:

- Bei kurzer Weierstraßform $y^2 = x^3 + ax + b$ spielt der Koeffizient b keine Rolle. Also ist es rechnerisch günstig, Kurven mit kleinem a und großem b zu benutzen.
- Für Kurven mit $\mathrm{char}(k) = 2$ ergeben sich mit Satz 13.13 Formeln für "$+$", die sich besonders gut für Hardwareimplementierungen eignen.
- Sowohl bei der Addition "$+$" als auch bei der Punkteverdopplung $P \mapsto 2P$ (affin) wir eine Division in k benötigt. Das ist zum Beispiel für $k = \mathbb{R}$ unpässlich bzw. zu ungenau. Das Problem lässt sich mit projektiven Koordinaten beheben: Ist $P = \left(\frac{x}{r}, \frac{y}{s}\right) \in \mathbb{P}^2(k)$ mit $r, s \in k \setminus \{0\}$, ist $P = [xs : yr : rs]$ ohne Division berechenbar, aber es müssen mehr Multiplikationen durchgeführt werden. Durch die Einführung einer Variante von projektiven Koordinaten – den so genannten **Jakobinischen Koordinaten** – kann man im Vergleich dazu Multiplikationen einsparen, was den Rechenaufwand vermindert. Wir besprechen dies in diesem Abschnitt zur schnellen Punkteaddition.

Definition 15.2 (Verallgemeinerte projektive Koordinaten)
Sei k ein Körper und $c, d \in \mathbb{N}$, dann definieren wir die Relation $\sim$ auf $k^3 \setminus \{0\}$ durch

$$(x, y, z) \sim (x', y', z') \quad :\Leftrightarrow \quad \exists \sigma \in k \setminus \{0\} : x = \sigma^c z', y = \sigma^d z', z = \sigma z'$$

Bemerkung 15.3

- $\sim$ ist eine Äquivalenzrelation auf $k^3 \setminus \{0\}$, ihre Äquivalenzklassen bezeichnen wir mit

$$(x : y : z) := \{(x', y', z') \in k^3 \setminus \{0\} : (x', y', z') \sim (x, y, z)\}$$

- Im Fall $c = d = 1$ erhalten wir unsere bisherige Definition für einen projektiven Punkt $[x : y : z]$ zurück. Auch hier nennen wir $(x : y : z)$ einen projektiven Punkt.

Bemerkung 15.4

- Wenn $z \neq 0$, gilt durch Normierung $\left(\frac{x}{z^c}, \frac{y}{z^d}, 1\right) \sim (x, y, z)$, vermöge $\sigma = \frac{1}{z}$, damit kann man in der Menge

$$\mathbb{P}^2_{(c,d)}(k) := \{(x : y : z) : (x, y, z) \in k^3 \setminus \{0\}\}$$

  die Punkte mit $z \neq 0$ wieder mit $\mathbb{A}^2(k)$ identifizieren.

- Die Punkte mit $z = 0$ bilden wieder die unendlich ferne Gerade.

- Die projektive Form der Weierstraßgleichung erhält man durch Einsetzen von $\frac{x}{z^c}$ und $\frac{y}{z^d}$ in die Gleichung und Entfernung der Nenner durch Multiplikation:

$$y^2 - x^3 - ax - b = 0 \Rightarrow \left(\frac{y}{z^d}\right)^2 - \left(\frac{x}{z^c}\right)^3 - a \cdot \frac{x}{z^c} - b = 0 \Rightarrow y^2 - x^{3-3c+2d} - axz^{2d-c} - bz^{2d} = 0,$$

  falls etwa $2d = 3c$, was offenbar im Allgemeinen nicht mehr homogen sein muss.

- In der Krpytographie verwendet man folgende projektive Darstellungen:
  - Standard-projektive Koordinaten: $c = d = 1$
  - **Jakobinische Koordinaten**: $c = 2, d = 3$
  - **Chudnovski-Koordinaten**: ein jakobinischer Punkt wird als $(x : y : z : z^2 : z^3)$ dargestellt.

Anwendung 15.5 (Schnelle Punkteaddition mit Jakobinischen Koordinaten)
Sei $E(k)\colon y^2 = x^3 + ax + b$ gegeben mit $4a^3 + 27b^2 \neq 0$. Anhand der affinen Version für die explizite Punkteverdopplung zeigen wir nun, dass man Rechenaufwand sparen kann, wenn man mit Jakobinischen Koordinaten $c = 2, d = 3$ arbeitet. Ist $P = (u, v)$, $P \neq -P$, so ist

$$2P = P + P = (\underbrace{\lambda^2 - 2u}_{=:x}, \underbrace{\lambda(u - x) - v}_{=:y}),$$

wobei $\lambda := \frac{3u^2+a}{2v}$, die affine Version der expliziten Formel in Standard-Darstellung. Für die Koordinaten x, y von $2P = (x : y : 1) = [x : y : 1]$ bei $P = [u : v : z]$ erhält man durch Einsetzen von $\frac{u}{z^2}$ für u und $\frac{v}{z^3}$ für v dann

$$x = \left(\frac{3 \cdot \left(\frac{u}{z^2}\right)^2 + a}{2 \cdot \frac{v}{z^3}}\right)^2 - 2 \cdot \frac{u}{z^2} = \frac{\left(3 \cdot \frac{u^2}{z^4} + a\right)^2 z^6}{4v^2} - 2 \cdot \frac{u}{z^2} = \frac{(3u^2 + az^4)^2 - 8uv^2}{4v^2z^2}$$

und

$$y = \frac{3 \cdot \left(\frac{u}{z^2}\right)^2 + a}{2 \cdot \frac{v}{z^3}} \cdot \left(\frac{u}{z^2} - x\right) - \frac{v}{z^3} = \frac{3u^2 + az^4}{2vz} \cdot \left(\frac{u}{z^2} - x\right) - \frac{v}{z^3}.$$

Setzen wir nun $\sigma := 2vz$, wird damit $x_0 = \sigma^2 x, y_0 = \sigma^3 y, z_0 = \sigma$ und somit

$$x_0 = (3u^2 + az^4)^2 - 8uv^2, z_0 = 2vz$$

und

$$\begin{aligned} y_0 &= \frac{3u^2 + az^4}{2vz} \cdot 8v^3z^3 \cdot \left(\frac{u}{z^2} - x\right) - \frac{v}{z^3} \cdot 8v^3z^3 \\ &= (3u^2 + az^4) \cdot 4v^2 \cdot (u - z^2x) - 8v^4 \\ &= (3u^2 + az^4) \cdot (4uv^2 - x_0) - 8v^4 \end{aligned}$$

Bemerkung 15.6
Eine Umsetzung der Berechnung von $(x_0 : y_0 : z_0) = 2P = (x : y : 1)$ ist somit wie folgt möglich:

$$A := v^2 \qquad B := 4u \cdot A \qquad C := 8A^2 \qquad D := 3u^2 + a \cdot (z^2)^2$$

$$x_0 := D^2 - 2B \qquad y_0 := D \cdot (B - x_0) - C \qquad z_0 := 2v \cdot z$$

Das sind insgesamt sechs Quadrierungen und vier Multiplikationen im Basiskörper k, es sind keine Divisionen nötig! (Die Skalaren Vielfachen mit $2, 3, 4, 8$ zählen wie Additionen: $2 \cdot s = s + s$ usw.)
Analog gewinnt man die folgenden effizienten, expliziten Formeln zur Punkteaddition $P + Q$ mit $P = (u, v), Q = (r, s)$ in jakobinischen Koordinaten:

$$\begin{aligned} x_0 &= (sz^3 - v)^2 - (rz^2 - u)^2(u + rz^2) \\ y_0 &= (sz^3 - v)(u(rz^2 - u)^2 - x_0) - v(rz^2 - u)^3 \\ z_0 &= (rz^2 - u)z \end{aligned}$$

Als Rechenverfahren dient dann:

$$A := z^2 \qquad B := z \cdot A \qquad C := r \cdot A \qquad D := s \cdot B \qquad E := C - u$$

$$F := D - v \qquad G := E^2 \qquad H := G \cdot E \qquad T := u \cdot G$$

$$x_0 := F^2 - (H + 2I) \qquad y_0 := F \cdot (I - x_0) - v \cdot H \qquad z_0 := z \cdot E$$

Bemerkung 15.7
Das sind insgesamt drei Quadrierungen und acht Produkte in k, keine Divisionen! Aufstellung des Rechenaufwands für eine elliptische Kurve $y^2 = x^3 - 3x + b$:

	Punkteverdopplung $2P = P + P$	Punkteaddition $P + Q$
affin	1 Div., 2 Mul., 2 Quad.	1 Div., 2 Mul., 1 Quad.
standard-projektiv	7 Mul., 3 Quad.	12 Mul., 2 Quad.
jakobinische Koordinaten	4 Mul., 4 Quad.	12 Mul., 2 Quad.
Chudnovski-Koordinaten	5 Mul., 4 Quad.	11 Mul., 3 Quad.

Bemerkung 15.8
Fazit: Arbeitet man mit jakobinischen Koordinaten, können bei der Punkteverdopplung Multiplikationen eingespart werden, was dann am Computer zu einem schnelleren Verfahren bei der Berechnung von $2 \cdot P$ bzw. $m \cdot P$ führt.

Bemerkung 15.9
Zur Erinnerung: Wie bei der schnellen Potenzierung zur Berechnung von x^m kann bei additiver Schreibweise einer Gruppe die Berechnung von $m \cdot P$ analog durchgeführt werden, was man "schnelle Vervielfachung nennen könnte, engl. **dual and add**-algorithm. Schritte des Verfahrens:

1) Sei $d = \left\lfloor \frac{\log m}{\log 2} \right\rfloor$, berechne durch sukzessives Verdoppeln $P, 2P, 4P, 8P, \ldots, 2^d P$.
2) Schreibe m als Binärzahl: $m = \sum_{i=0}^{d} c_i 2^i, c_i \in \{0, 1\}$.
3) Berechne $mP = (c_0 P) + (c_1 \cdot 2P) + (c_2 \cdot 4P) + \ldots + (c_d \cdot 2^d P)$ mit maximal d weiteren Additionen von Punkten auf $E(k)$.

3 Elliptische Kurven über verschiedenen Körpern

3.1 Elliptische Kurven über $\mathbb{Q}$

Motivation 16.1

Betrachte die elliptische Kurve $E(k)$ jetzt über $k = \mathbb{Q}$. Wie findet man möglichst viele rationale Punkte (d.h. $P = (x, y) \in \mathbb{A}^2(\mathbb{Q})$) auf der Kurve $E(\mathbb{Q})$? Die expliziten Formeln zeigen: [16]

- Ist P ein rationaler Punkt auf $E(k)$, so auch $2P, 3P, 4P, \ldots$
- Sind P, Q zwei rationale Punkte, so auch $P + Q, P + (P + Q) = 2P + Q, 2P + 2Q$, usw.

Können so unendlich viele rationale Punkte auf $E(k)$ konstruiert werden? Das hängt von der Ordnung des Punktes P in der Gruppe $(E(k), +)$ ab, d.h. von $\operatorname{ord}(P) = \min\{m \in \mathbb{N} : mP = \mathcal{O}\}$, falls diese existiert. Das ist ziemlich unklar, wie auch Beispiele zeigen:

Beispiel 16.2

Sei $E(\mathbb{Q})\colon y^2 = x^3 + 17$. Dann ist $\Delta(E) = -16 \cdot 27 \cdot 17^2 \neq 0$. Zwei Punkte sind $P = (-2, 3)$ und $Q = (-1, 4)$. Es ist

$$\begin{aligned} P + Q &= (4, -9) \\ 2P + Q &= (2, 5) \\ 3P + Q &= \left(\frac{1}{4}, -\frac{33}{8}\right) \\ 4P + Q &= \left(\frac{106}{9}, \frac{1097}{27}\right) \\ 5P + Q &= \left(-\frac{2228}{961}, -\frac{63465}{29791}\right) \\ 6P + Q &= \left(\frac{76271}{289}, -\frac{21063928}{4913}\right) \\ &\vdots \end{aligned}$$

Offenbar werden die Ergebnisse immer komplizierter; unendlich viele rationale Punkte können auf $E(\mathbb{Q})$ wohl derart konstruiert werden, d.h. vermutlich hat P keine (endliche) Ordnung.

Beispiel 16.3

Sei $E(\mathbb{Q})\colon y^2 = x^3 + x$. Der einzige affine rationale Punkt auf $(0, 0)$. Dies kann direkt gezeigt werden unter Verwendung, dass die Gleichung $u^4 + v^4 = w^2$ nur ganzzahlige Lösungen mit $u = 0$ oder $v = 0$ hat (was auch schon nicht so schnell zu zeigen ist). Es kann dennoch gesagt werden, dass durch $P, 2P = \mathcal{O}, 3P = P, 4P = \mathcal{O}, \ldots$ alle rationalen Punkte auf $E(\mathbb{Q})$ konstruiert werden können.

Beispiel 16.4

Sei $E(\mathbb{Q})\colon y^2 = x^3 - 4x^2 + 16$. Dann ist $\Delta(E) = -16 \cdot (4 \cdot (-4)^3 + 27 \cdot 16^2) \neq 0$. Eine kurze Suche liefert die vier rationalen Punkte

$$P_1 = (0, 4) \qquad P_2 = (4, 4) \qquad P_3 = (0, -4) = -P_1 \qquad P_4 = (4, -4) = -P_2$$

Können hier wie in Beispiel 16.2 beliebig viele rationale Punkte konstruiert werden? Hier ist die Gerade durch P_1 und P_2 die Tangente an $E(k)$ in P_1, weil $4^2 = x^3 - 4x^2 + 16 \Leftrightarrow 0 = x^2(x - 4)$ ist und $x = 0$ doppelte Nullstelle. Damit ist $-P_1 = P_1 + P_2 = P_3$, also kann so kein weiterer rationaler Punkt konstruiert werden. Auch mit anderen Paaren P_i und P_j der vier Punkte passiert dies. Vermutlich gibt es außer den vier angegebenen rationalen Punkten

keine weiteren auf $E(\mathbb{Q})$. Wir haben:

$$\begin{aligned} P_1 &= (0,4) \\ 2P_1 &= -P_2 = P_4 \\ 3P_1 &= P_1 + P_4 = P_1 - P_2 = P_2 \\ 4P_1 &= P_1 + P_2 = P_3 \\ 5P_1 &= P_3 + P_1 = (P_1 + P_2) + P_1) = 2P_1 + P_2 = -P_2 + P_2 = \mathcal{O}, \end{aligned}$$

d.h. $\mathrm{ord}(P_1) = 5$. Somit ist $\langle P_1 \rangle = \{\mathcal{O}, P_1, P_2, P_3, P_4\} \simeq \mathbb{Z}_5$.

Die Beispiele legen folgenden Satz nahe:

Satz 16.5 (Satz von Mordell (1922))
Sei $E(\mathbb{Q})$ eine elliptische Kurve über $\mathbb{Q}$. Dann gibt es eine endliche Liste von Punkten $P_1, \ldots, P_s \in E(\mathbb{Q})$, sodass alle (rationalen) Punkte auf $E(\mathbb{Q})$ von diesen erzeugt werden, d.h. für alle $P \in E(\mathbb{Q})$ existieren Zahlen $m_1, \ldots, m_s \in \mathbb{N}_0$ mit

$$P = m_1 P_1 + \ldots + m_s P_s$$

Mit anderen Worten: Die Gruppe $(E(\mathbb{Q}), +)$ ist endlich erzeugt. Dabei können die Erzeuger endliche Ordnung haben oder nicht. Natürlich sind die Erzeuger nicht unbedingt eindeutig bestimmt.

Bemerkung 16.6

- In Beispiel 16.3 haben wir einen endlichen Erzeuger $P_1 = (0,0), \mathrm{ord}(P_1) = 2$.
- In Beispiel 16.4 haben wir eventuell einen endlichen Erzeuger $P_1 = (0,4), \mathrm{ord}(P_1) = 5$.
- In Beispiel 16.2 haben wir eventuell einen unendlichen Erzeuger $P_1 = (-2,3)$, welcher eventuell nicht der einzige ist. Die von P_1 erzeugte Untergruppe $\mathbb{Z} \cdot P_1 := \{mP_1 : m \in \mathbb{Z}\} \subseteq E(\mathbb{Q})$ ist isomorph zu $\mathbb{Z}$ vermöge $m \cdot P_1 \mapsto m$.

Definition 16.7 (Torsionsgruppe)
Wir können in der Formuliereung von Satz 16.5 die Unterscheidung zwischen Punkten mit und ohne endliche Ordnung vornehmen. Die Teilmenge $T := \{P \in E(\mathbb{Q}) : \mathrm{ord}(P) \in \mathbb{N}\}$ aller Punkte von $E(\mathbb{Q})$ mit endlicher Ordnung ist offenbar eine Untergruppe, die **Torsionsgruppe** von $E(\mathbb{Q})$ heißt. Somit hat der Satz von Mordell auch die folgende Formulierung:

Satz 16.8 (Satz von Mordell, Gruppenstruktur)
$\mathbb{Z}^r$ mit komponentenweiser Addition
Es gibt ein $r = r(E) \in \mathbb{N}_0$ mit $E(\mathbb{Q}) \simeq \mathbb{Z}^r \times T$.

Definition 16.9 (Rang)
Die Zahl $r(E)$ heißt **Rang** von $E(\mathbb{Q})$.

Bemerkung 16.10
Die Torsionsgruppe T ist stets endlich, wie aus dem Struktursatz über endlich erzeugte abelsche Gruppen gefolgert werden kann. Allerdings bleiben Größe von T und Lage der Torsionspunkte $P \in T$ damit unbekannt. Weiter kann aber $\#E(\mathbb{Q}) = \infty \Leftrightarrow r(E) > 0$ gefolgert werden.

Beispiel 16.11

- Für $E(\mathbb{Q})$: $y^2 = x^3 - 4$ ist $E(\mathbb{Q}) \simeq \mathbb{Z}^1$, wobei z.B. $P_1 = (2,2)$ Erzeuger ist.

- Im Beispiel 16.3 und 16.4 ist $\operatorname{Rg} E(\mathbb{Q}) = 0$. (ohne Beweis)

Bemerkung 16.12
Der Rang elliptischer Kurven ist bislang schlecht verstanden. Offen, d.h. bislang unbewiesen, ist z.B. die **Rangvermutung**:

$$\limsup_{E(\mathbb{Q})} r(E) = \infty$$

D.h. man vermutet, dass es zu jedem $C \in \mathbb{R}$ eine elliptische Kurve mit $\operatorname{Rg} E(\mathbb{Q}) > C$ gibt. Der aktuelle Weltrekord (2006, von N. Elkies[14]) ist eine elliptische Kurve vom Rang ≥ 28 (da 28 "unabhängige" Punkte unendlicher Ordnung auf ihre gefunden wurden), die Kurve lautet $y^2 + xy + y = x^3 - x^2 - ax + b$ mit

$a = 20.067.762.415.575.526.585.033.208.209.338.542.750.930.230.312.178.985.502$

$b = 34.481.611.795.030.556.467.032.985.690.390.720.374.855.944.359.319.180.361.266.008.296.291.939.448.732.243.729$

Die Torsionsgruppe ist deutlich besser verstanden:

Satz 16.13 (Satz von Nagell-Lutz (Nagell 1935, Lutz 1937))
(T. Nagell[15], E. Lutz[16])
Sei $E(\mathbb{Q})$ eine elliptische Kurve mit Gleichung $y^2 = x^3 + ax^2 + bx + c$, $a, b, c \in \mathbb{Z}$, und seien $P_1, \ldots, P_s$ alle Torsionspunkte, d.h. $T = \{P_1, \ldots, P_s\}$. Schreibe die $P_i = (x_i, y_i) \in \mathbb{Q}^2$. Dann sind alle $x_i, y_i \in \mathbb{Z}$, und für $y_i \neq 0$ gilt $y_i^2 \mid \Delta(E)$.

Satz 16.14 (Satz von Mazur (1977))
(B. Mazur[17])
Sei $E(\mathbb{Q})$ eine elliptische Kurve mit Gleichung $y^2 = x^3 + ax^2 + bx + c, a, b, c \in \mathbb{Z}$, mit Torsionsuntergruppe T. Dann ist $T \simeq \mathbb{Z}_n$ mit $n \leq 12, n \neq 11$ oder $T \simeq \mathbb{Z}_2 \times \mathbb{Z}_n$ mit $n \in \{2, 4, 6, 8\}$. Andere Torsionsuntergruppen kann es nicht geben, und alle genannten kommen vor.

Das sind beachtliche, tiefe Sätze. In Beispiel 16.4 ist $T \simeq \mathbb{Z}_5$, und man kann sehen, dass der Nagell-Lutz-Satz hier korrekt ist: $4^2 \mid \Delta(E)$.
Für $a, b, c \in \mathbb{Z}$ kann es höchstens endlich viele Punkte mit ganzzahligen Koordinaten geben:

Satz 16.15 (Satz von Siegel (1926))
(C. L. Siegel[18])
Sei $E(\mathbb{Q})\colon y^2 = x^3 + ax^2 + bc + c$ mit $a, b, c \in \mathbb{Z}$ eine elliptische Kurve. Dann gibt es nur endlich viele Kurvenpunkte $(x, y) \in E(\mathbb{Q}) \cap \mathbb{Z}^2$.

Bemerkung 16.16
In Beispiel 16.2 haben genau die Punkte $\mathcal{O}, (-2, \pm -3), (-1, \pm 4), (2, \pm 5), (4, \pm 9), (8, \pm 23), (43, \pm 282), (52, \pm 375), (5324, \pm 378661)$ auf der elliptischen Kurve $E(\mathbb{Q})$ ganzzahlige Koordinaten.

Bemerkung 16.17
Elliptische Kurve über $\mathbb{Q}$ sind nicht-singuläre algebraische Kurven vom Geschlecht 1. Mordell[19] vermutete, dass jede über $\mathbb{Q}$ definierte nicht-singuläre algebraische Kurve vom Geschlecht ≥ 2 höchstens endlich viele Punkte enthält. Diese Vermutung wurde 1983 von Faltings[20] für beliebige Körper bewiesen, wofür er 1986 auf der ICM in Berkeley mit der Fields-Medaille ausgezeichnet wurde.

[14] https://de.wikipedia.org/wiki/Noam_Elkies
[15] https://de.wikipedia.org/wiki/Trygve_Nagell
[16] https://de.wikipedia.org/wiki/%C3%89lizabeth_Lutz
[17] https://de.wikipedia.org/wiki/Barry_Mazur
[18] https://de.wikipedia.org/wiki/Carl_Ludwig_Siegel
[19] https://de.wikipedia.org/wiki/Louis_Mordell
[20] https://de.wikipedia.org/wiki/Gerd_Faltings

3.2 Elliptische Kurven über $\mathbb{C}$

Bemerkung 16.18

Elliptische Kurven über $\mathbb{C}$ können einerseits über die Weierstraßgleichung dargestellt werden und zum anderen über ihre **Legendre-Normalform** mit einer Gleichung der Form $y^2 = x(x-1)(x-\lambda)$, vgl. Übungsblatt 6, Aufgabe 2(a). Wir besprechen hier kurz die dritte Darstellung mittels elliptischer Funktionen. Wir möchten hier nur erläutern, warum eine elliptische Kurve über $\mathbb{C}$ in diesem Sinne ein Torus ist.
Gegeben sei $\tau \in \mathbb{C} \setminus \mathbb{R}$. Betrachte das Gitter $\Lambda := \{a + \tau \cdot b : a, b \in \mathbb{Z}\} \subseteq \mathbb{C}$.

Definition 16.19 (Elliptische Funktion)

Eine Funktion $f\colon \mathbb{C} \setminus \mathcal{P} \to \mathbb{C}$ der Form $f(z) := \frac{f(z)}{g(z)}$, g, h holomorph, $h \neq 0$, heißt **elliptische Funktion**, falls $f(z+w) = f(z)$ für alle $z \in \mathbb{C}$ und $\omega \in \Lambda$ (sofern $f(z), f(z+\omega)$ definiert ist, wobei $\mathcal{P} = \{z : h(z) = 0\}$ die Menge der Polstellen von f ist), d.h. wenn f (doppelt-)periodische Funktion zu Λ ist. Der Körper der elliptischen Funktionen über Λ sei $\mathbb{C}(\Lambda)$.

Zeichnungen einfügen!

Eine elliptische Funktion ohne Polstellen (oder ohne Nullstellen) ist konstant (vgl. Satz von Liouville aus der Funktionentheorie).

Definition 16.20 (Weierstraß-$\wp$-Funktion, Eisensteinreihe)

Zu Λ definiere

$$\begin{aligned} \wp\colon \mathbb{C} \setminus \Lambda &\longrightarrow \mathbb{C} \\ z &\longmapsto \frac{1}{z^2} + \sum_{\omega \in \Lambda \setminus \{0\}} \left(\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2} \right) \end{aligned}$$

und $G_{2k}(\Lambda) := \sum_{\omega \in \Lambda \setminus \{0\}} \frac{1}{\omega^{2k}} \in \mathbb{C}$. Dann heißt $\wp$ die **Weierstraß-$\wp$-Funktion** und G_{2k} heißt **Eisensteinreihe** vom Gewicht $2k \in \mathbb{R}_{>2}$.

Satz 16.21

Sei Λ ein Gitter.

(a) Die Eisenstein-Reihe $G_{2k}(\Lambda)$ konvergiert absolut für $k > 1$.

(b) Die Reihe der Funktion $\wp$ konvergiert absolut und gleichmäßig auf jeder kompakten Teilmenge von $\mathbb{C} \setminus \Lambda$. Sie definiert eine elliptische Funktion mit zweifachem Pol in jedem Gitterpunkt $\omega \in \Lambda$.

Es gilt somit $\wp'(z) = -2 \sum_{\omega \in \Lambda \setminus \{0\}} \frac{1}{(z-\omega)^3}$ für $z \in \mathbb{C} \setminus \Lambda$. Die Funktionen $\wp$ und $\wp'$ liefern den "Prototyp" elliptischer Funktionen: Man kann zeigen, dass jede elliptische Funktion f schreibbar ist als

$$f(z) = \frac{P_1(\wp(z))}{Q_1(\wp(z))} + \wp'(z) \cdot \frac{P_2(\wp(z))}{Q_2(\wp(z))}, P_i, Q_i \in \mathbb{C}[Z].$$

Satz 16.22

Es gilt $(\wp'(z))^2 = 4(\wp(z))^3 - g_2 \cdot \wp(z) - g_3$, d.h. $(\wp(z), \wp'(z)) \in E(\mathbb{C})$ mit Gleichung $y^2 = 4x^3 - g_2 x - g_3$, wobei $g_2 := 60G_4, g_3 := 140G_6$. Da $\Delta(E) = g_2^3 - 27g_3^2 \neq 0$, handelt es sich bei $E(\mathbb{C})$ um eine elliptische Kurve.

Wir haben so die Abbildung

$$\begin{aligned} \varphi\colon \mathbb{C}/\Lambda &\longrightarrow \mathbb{P}^2(\mathbb{C}) \\ z + \Lambda &\longmapsto [\wp(z) : \wp'(z) : 1], \end{aligned}$$

wobei $\varphi(0+\Lambda) = [0:1:0] = \mathcal{O}$. Das Bild von φ ist genau die genannte elliptische Kurve $E(\mathbb{C})$. Die Abbildung $\varphi\colon \mathbb{C}/\Lambda \to E(\mathbb{C})$ ist bijektiv und überträgt die Addition "+" auf $\mathbb{C}/\Lambda$, gegeben durch $(x+\Lambda) + (y+\Lambda) :=$

$(x + y) + \Lambda$, auf $E(\mathbb{C})$, welche sich als unsere bisher studierte Addition $+$ auf $E(\mathbb{C})$ erweist. Der Torus $\mathbb{C}/\Lambda$ wird so mit der elliptischen Kurve $E(\mathbb{C})$ identifiziert. Umgekehrt ist auch jede elliptische Kurve $(E(\mathbb{C}), +)$ beschreibbar als Torus $(\mathbb{C}/\Lambda, +)$.

3.3 Elliptische Kurven über $\mathbb{F}_p$ und $\mathbb{F}_{p^r}$

Bemerkung 17.1

[17] Elliptische Kurven über endlichen Körpern werden vielfältig eingesetzt, zum einen in der Kryptographie, zum anderen auch in technischen Systemen mit wenig Ressourcen (eingebettete Systeme), z.B. Steuergeräte in Automobilen (elektronische Wegfahrsperren, Tuning-Schutz, Car-To-Car-Kommunikation, etc.). Manche Hardware-Implementationen arbeiten über $\mathbb{F}_{2^r}$ der Charakteristik 2, bei denen die technische Umsetzung damit günstig ist.

3.3.1 Punkte zählen, der Frobenius

Bemerkung 17.2

Wir studieren zunächst elliptische Kurven über $\mathbb{F}_p$, wo p prim, mit der "modularen Brille" modulo p. Das Verhalten dieser Kurven kann ganz anders sein als über $\mathbb{Q}$: Die elliptische Kurve $E(\mathbb{Q})\colon y^2 = x^3 + x$ aus Beispiel 16.3 etwa enthält den einzigen rationalen Punkt $(0,0)$, über $\mathbb{F}_p$ hat sie aber viele Punkte: Sei $\mathbb{N}_p := \#E(\mathbb{F}_p)$ von $y^2 = x^3+x$, d.h. N_p, die Anzahl der Punkte der elliptischen Kurve $E(\mathbb{F}_p)$, ist die Anzahl der Lösungen von $y^2 = x^3 + x$ modulo p.

Definition 17.3 (Defekt)

Numerische Daten ergeben folgende Tabelle:

p	2	3	5	7	11	13	17	19	23	29	31	37	41	43	47	…
$N_p - 1$	2	3	3	7	11	19	15	19	23	19	31	35	31	43	47	…

Offenbar gilt $p = N_p - 1$, falls $p \equiv 3 \bmod 4$. Für $p \equiv 1 \bmod 4$ sieht der so genannte **Defekt** $a_p := p + 1 - N_p$ so aus:

p	5	13	17	29	37	41	53	61	73	89	…
$a_p/2$	1	-3	1	5	1	5	-7	5	-3	5	…

Beobachtung: $p - \left(\frac{ap}{2}\right)^2$ ist stets eine Quadratzahl!

Wir halten fest:

Satz 17.4

Für $E(\mathbb{F}_p)\colon y^2 = x^3 + x$ gilt:

(a) Ist $p \equiv 3 \bmod 4$, gilt $N_p = p + 1$.

(b) Ist $p \equiv 1 \bmod 4$, ist $N_p = p + 1 \pm 2A$, wobei $p = A^2 + B^2$ mit $2 \nmid A$. Dabei gilt "+", falls $A \equiv 1 \bmod 4$ und "−", falls $A \equiv 3 \bmod 4$.

Zum Beweis benutzen wir:

Satz 17.5

Für $p > 2$ und $E(\mathbb{F}_p)\colon y^2 = x^3 + ax + b$ mit $a, b \in \mathbb{F}_p$ gilt

$$N_p := \#E(\mathbb{F}_p) = p + 1 + \sum_{x \in \mathbb{F}_p} \left(\frac{x^3 + ax + b}{\mathbb{F}_p} \right),$$

wobei

$$\left(\frac{u}{\mathbb{F}_p}\right) := \begin{cases} +1, & \text{wenn } u \not\equiv 0 \bmod p \text{ ein quadratischer Rest } \bmod p, \text{ d.h. } \exists w \in \mathbb{Z} : u \equiv w^2 \bmod p \\ -1, & \text{wenn } u \not\equiv 0 \bmod p \text{ kein quadratischer Rest } \bmod p, \text{ d.h. } \forall w \in \mathbb{Z} : u \not\equiv w^2 \bmod p \\ 0, & \text{wenn } u \equiv 0 \bmod p, \text{ d.h. } p \mid u \end{cases}$$

das verallgemeinerte **Legendre-Symbol** ist.

Beweis
Vgl. Übungsblatt 7, Aufgabe 4(a). □

Beweis 17.6 (zu Satz 17.4(a))
Für $p \equiv 3 \bmod 4$ ist $\left(\frac{-1}{\mathbb{F}_p}\right) = (-1)^{\frac{p-1}{2}} = -1$ nach dem 1. Ergänzungssatz für das Legendre-Symbol, also $\left(\frac{(-x)^3+(-x)}{\mathbb{F}_p}\right) = -\left(\frac{x^3+x}{\mathbb{F}_p}\right)$ für das $x \not\equiv 0 \bmod p$ nach den Rechenregeln für das Legendre-Symbol, sodass

Für die Rechenregeln zum Legendre-Symbol siehe EZT-Skript!

$$0 = \sum_{x \not\equiv 0 \bmod p} \left(\left(\frac{x^3+x}{\mathbb{F}_p}\right) + \left(\frac{(-x)^3+(-x)}{\mathbb{F}_p}\right)\right) = 2 \sum_{x \not\equiv 0 \bmod p} \left(\frac{x^3+x}{\mathbb{F}_p}\right)$$

folgt, denn mit x durchläuft auch $-x$ alle Restklassen $\not\equiv 0 \bmod p$. Also ist $\mathbb{N}_p = p+1$. □

Bemerkung 17.7
Für den Defekt $a_p = p + 1 - N_p$ gilt somit $a_p = -\sum_{x \in \mathbb{F}_p} \left(\frac{x^3+ax+b}{\mathbb{F}_p}\right)$. Im Absolutbetrag kann a_p nicht allzu groß werden, d.h. $N_p = \#E(\mathbb{F}_p)$ kann nicht allzu stark von p abweichen:

Satz 17.8 (Satz von Hasse (1933))
(H. Hasse[21])
Für den Defekt a_p einer elliptischen Kurve $E(\mathbb{F}_p)$ gilt $|a_p| \leq 2\sqrt{p}$. Der Satz gilt auch für $E(\mathbb{F}_{p^r}), r \geq 1$: Es ist $|p^r + 1 - N_{p^r}| \leq 2\sqrt{p^r}$.

Bemerkung 17.9
Diese Abschätzung für $|a_p|$ wurde 1920 von E. Artin[22] vermutet. Eine Verallgemeinerung zeigte A. Weil[23] in den 1940ern, und stark verallgemeinert wurde sie von P. Deligne[24] in den 1970ern, wofür er 1978 mit der Fields-Medaille ausgezeichnet wurde.

Bemerkung 17.10
Im Beispiel $y^2 = x^3 + x$ mit $p \equiv 1 \bmod 4$ folgt aus Satz 17.4(b) die Hasseschranke, da

$$|a_p| = |p + 1 - N_p| = |p + 1 - p - 1 \mp 2A| = |\mp 2A| = |\mp 2\sqrt{p - B^2}| \leq 2\sqrt{p}.$$

Bemerkung 17.11
Fragen über die Größen von N_p führen zu offenen Problemen, z.B. die schwache Vermutung von Birch[25] und Swinnerton-Dyer[26] (1963, 1965): Für $E(\mathbb{Q})$ mit Koeffizienten aus $\mathbb{Z}$ sollte

$$\prod_{p \leq x} \frac{N_p}{p} \sim C_E (\log x)^{r(E)}$$

gelten, wobei C_E eine Konstante > 0 ist, die nur von $E(\mathbb{Q})$ abhängig ist. Die Aussage $f(x) \sim g(x)$ für zwei Funktionen $f, g \colon \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ bedeutet, dass $\lim_{x \to \infty} \frac{f(x)}{g(x)} = 1$, d.h. f und g sind asymptotisch gleich. Die Zahl $r(E)$ ist der Rang von $E(\mathbb{Q})$. Numerische Untersuchungen stützen diese Vermutung bislang. Sie bedeutet: Ist die Anzahl N_p der Punkte auf $E(\mathbb{F}_p)$ bei Reduktion modulo p signifikant größer als der Erwartungswert, so sollte der Rang $r(E)$ positiv sein. Sie stellt damit ein numerisch leicht testbares Kriterium für $r(E) > 0$ dar.

[21] https://de.wikipedia.org/wiki/Helmut_Hasse
[22] https://de.wikipedia.org/wiki/Emil_Artin
[23] https://de.wikipedia.org/wiki/Andr%C3%A9_Weil
[24] https://de.wikipedia.org/wiki/Pierre_Deligne
[25] https://de.wikipedia.org/wiki/Bryan_Birch
[26] https://de.wikipedia.org/wiki/Peter_Swinnerton-Dyer

Beispiel 17.12
Weitere numerische Beobachtungen in anderen Beispielen:
Für $E\colon y^2 = x^3 + 17$ gilt $a_p = 0$ genau für $p \equiv 2 \bmod 3$, also auch $a_p = 0$ für "die Hälfte" aller Primzahlen. Das kommt für "wenige" elliptische Kurven mit Koeffizienten aus $\mathbb{Z}$ so heraus. Für die Kurve $E\colon y^2 = x^3 - 4x^2 + 16$ etwa haben wir $a_p = 0$ nur selten: Die einzigen $p < 2000$ mit $a_p = 0$ sind

$$p = 2, 19, 29, 199, 569, 809, 1289, 1439, \ldots$$

Welcher der Fälle eintritt, hängt davon ab, ob die elliptische Kurve **komplexe Multiplikation** (CM) hat.

Definition 17.13 (komplexe Multiplikation (CM))
Eine elliptische Kurve $E(k)$ hat **komplexe Multiplikation** (CM), falls sie neben den üblichen Endomorphismen $\Psi_m\colon E(k) \to E(k), P \mapsto m \cdot P$ mit $m \in \mathbb{Z}$ noch weitere hat.

Elliptische Kurven mit CM haben viele spezielle Eigenschaften, z.B.:

- Elliptische Kurven mit CM haben "ebensoviele" p mit $a_p = 0$ wie p mit $a_p \neq 0$.
- Elliptische Kurven ohne CM haben nur "wenige" p mit $a_p = 0$.

Dennoch konnte N. Elkies 1987 zeigen, dass jede elliptische Kurve $a_p = 0$ für unendlich viele p hat.

Bemerkung 17.14
Beim Übergang von $E(\mathbb{Q})$ mit Koeffizienten in $\mathbb{Z}$ zu $E(\mathbb{F}_p)$ zu einer Primzahl P reduzieren wir modulo p. Nicht immer kommt dabei wieder eine elliptische Kurve heraus, nämlich dann nicht, wenn $\Delta(E(\mathbb{F}_p)) = 0$ in $\mathbb{F}_p$ gilt, d.h. $p \mid \Delta(E(\mathbb{Q}))$ in $\mathbb{Z}$. Wir sprechen dann von **schlechter Reduktion** (**bad prime**), ansonsten von **guter Reduktion** (**good prime**). Die schlechte Reduktion kommt nur für alle Primteiler von $\Delta(E(\mathbb{Q})) \in \mathbb{Z}$ vor, also nur für endlich viele Primzahlen p. In diesen Ausnahmefällen verhält sich die kubische Kurve, die durch Reduktion der Gleichung von $E(\mathbb{Q})$ modulo p gegeben ist, oft anders; "+" gibt es dann nicht. Im Beispiel $E(\mathbb{Q})\colon y^2 = x^3 - 4x^2 + 16$ gilt etwa $N_p \equiv 4 \bmod 5$ für alle Primzahlen p außer $p = 2$ und $p = 11$. Tatsächlich sind $p = 2$ und $p = 11$ hier die Primzahlen mit schlechter Reduktion, da $\Delta(E(\mathbb{Q}))0 - 2^{12} \cdot 11$ ist.

Definition 17.15 (Spur des Frobenius)
Für eine elliptische Kurve $E(\mathbb{F}_{p^r}), r \geq 1$, wird der Defekt $a_{p^r} = p^r + 1 - N_{p^r}$ auch die **Spur des Frobenius** genannt.

Definition 17.16 (Frobeniusendomorphismus)
Der **Frobeniusendomorphismus** (kurz: Frobenius[27]) einer elliptischen Kurve $E(\mathbb{F}_{p^r})$ ist der durch die Abbildung

$$\begin{aligned} \phi\colon \mathbb{P}^2(\overline{\mathbb{F}_{p^r}}) &\longrightarrow \mathbb{P}^2(\overline{\mathbb{F}_{p^r}}) \\ [x : y : z] &\longmapsto \left[x^{p^r} : y^{p^r} : z^{p^r}\right] \end{aligned}$$

vermittelte Gruppenhomomorphiusmus $\Phi\colon E(\overline{\mathbb{F}_{p^r}}) \to E(\overline{\mathbb{F}_{p^r}})$.

Beweisskizze

- ϕ ist wirklich eine Abbildung $\mathbb{P}^2(\overline{\mathbb{F}_{p^r}})$ in sich.
- Ist $E(\mathbb{F}_{p^r})$ gegeben durch $F(X, Y, Z) = 0$, folgt auch $F(X^{p^r}, Y^{p^r}, Z^{p^r}) = 0$, weil in $\overline{\mathbb{F}_{p^r}}$ die Gleichung $(c+d)^{p^r} = c^{p^r} + d^{p^r}$ für $c, d \in \overline{\mathbb{F}_{p^r}}$ richtig ist. Damit ist durch Φ eine Abbildung von $E(\overline{\mathbb{F}_{p^r}})$ in sich definiert.

[27] https://de.wikipedia.org/wiki/Ferdinand_Georg_Frobenius

- Die Verträglichkeit der Gruppenaddition auf $E(\overline{\mathbb{F}_{p^r}})$ mit Φ, d.h. die Eigenschaft $\Phi(P_1+P_2) = \Phi(P_1)+\Phi(P_2)$, kann man nachrechnen. □

Bemerkung 17.17
Der Frobenius Φ lässt sich auf allgemeinere Strukturen (genauer: dem Tatemodul) übertragen; dieser lässt eine Matrixdarstellung zu, wobei die Spur dieser Matrix genau a_{p^r} ergibt, daher der Name. Dieser Zusammenhang liefert weitere Möglichkeiten, den Defekt a_{p^r} zu bestimmen.

Anwendung 17.18 (Text in eine elliptische Kurve einbetten)
Bei der Umsetzung des ElGamal-Verschlüsselungsverfahrens für eine elliptische Kurvengruppe $(G,+) = (E(\mathbb{F}_{p^r}),+)$, vergleiche Abschnitt 1.2.3, ist es erforderlich, dass sich die Kommunizierenden Alice und Bob darauf einigen, wie man Klartext in eine Folge von Punkten auf der elliptischen Kurve $E(\mathbb{F}_{p^r})$ übersetzt und wieder zurückerhält. Hier ein beispielhaftes Verfahren, wie dies praktisch durchgeführt werden kann:

Anwendung 17.19 (Schritt 1)
Man legt ein Alphabet mit N Buchstaben (identifiziert mit $0,1,\ldots,N-1$) fest. Der Klartext (z.B. ein Wort) habe die Blocklänge l. Die Zuordnung

$$w = (a_0\ a_1\ \ldots\ a_l) \mapsto a_0N^{l-1} + a_1N^{l-2} + \ldots + a_{l-2}N + a_{l-1} = x_w$$

liefert eine Bijektion zwischen den möglichen Klartextblöcken w und den Zahlen $0 \leq xw < N^l$. Eine Zahl x_w soll x-Koordinate eines Kurvenpunkts werden.

Anwendung 17.20 (Schritt 2)
Für eine gegebene elliptische Kurve $E(\mathbb{F}_{p^r})$ gibt es aber nicht zu jedem $x \in \mathbb{F}_{p^r}$ einen Kurvenpunkt $(x_0,y_0) \in E(\mathbb{F}_{p^r})$. Für ein $k \in \mathbb{N}$ kann man aber die nächste x-Koordinate eines Kurvenpunkts $(x_1,y_1) \in E(\mathbb{F}_{p^r})$ mit $x_0 \leq x_1 \leq x_0 + k$ schnell ermitteln; die Wahrscheinlichkeit, dass dies scheitert, d.h. dass ein solches x_1 nicht existiert, beträgt schätzungsweise nur etwa $\left(\frac{1}{2}\right)^k$ (z.B. für $k = 50$ weniger als 10^{-15}).
Wähle so ein geeignetes k fest und eine elliptische Kurve $E(\mathbb{F}_{p^r})$ mit $p^r > k \cdot N^l$, d.h. es gibt wohl Kurvenpunkte mit genügend großen x-Koordinaten.

Anwendung 17.21 (Schritt 3)
Zu $x_w \in \{0,\ldots,N^l-1\}$ bestimme $P_w \in E(\mathbb{F}_{p^r})$ mit x-Koordinate $\geq kx_w$, etwa $P_w = (kx_w + j, y)$ mit $j \geq 0$ minimal.

Bemerkung 17.22
Wir beobachten: Hat das Verfahren funktioniert, ist dabei $j < k$, sodass durch Berechnung von $x_w = \left\lfloor \frac{x}{k} \right\rfloor$ für $P_w = (x,y) \in E(\mathbb{F}_{p^r})$ der Klartext w aus P_w wieder zurückgewonnen werden kann.

Beispiel 17.23
Für das Alphabet $\{A,B,\ldots,Z\} = \{0,1,\ldots,25\}$ ist $N = 26$, wähle z.B. $l = 2, k = 10$. Dann erfüllt $p = 6833$ die Bedingung $p > kN^2 = 6760$. Ist dazu $E(\mathbb{F}_p)$ gegeben durch $E(\mathbb{F}_p)\colon y^2 = x^3 + 5984x + 1180$, kann z.B. der Text "KRYPTO" wie folgt in eine Liste von drei Punkten auf $E(\mathbb{F}_p)$ umgesetzt werden:

w	KR	YP	TO
x_w	$(10,17)_{26} = 277$	$(24,15)_{26} = 639$	$(19,14)_{26} = 508$
P_w	$(2771,353)$	$(6390,2797)$	$(5080,238)$

3.3.2 Modularitätsmuster und der große Fermatsche Satz

Bemerkung 18.1

[18] Für elliptische Kurven mit Koeffizienten in $\mathbb{Z}$ wurde ein Modularitätsmuster gefunden, welches sehr unterwartet und ungewöhnlich ist, dass es kaum vorstellbar ist, dass dieses überhaupt gefunden werden konnte. Es handelt sich (in voller Allgemeinheit) um die **Taniyama-Shimura-Vermutung**[28] von 1957, welche von A. Wiles und anderen 1995 komplett bewiesen wurde und als Baustein des Beweises des großen Fermatschen Satzes diente.

Beispiel 18.2

Im Spezialfall der Kurve $E\colon y^2 = x^3 - 4x^2 + 16$ z.B. lautet diese Vermutung, dass folgendes "Modularitätsmuster" für die Defekte a_p gilt:
Man betrachte die Potenzreihe $\Theta(T) \in \mathbb{Z}[T]$, welche durch Ausmultiplizieren des unendlichen Produkts

$$\Theta(T) := T \cdot \prod_{k=1}^{\infty}((1-T^k)(1-T^{11k}))^2 = T \cdot ((1-T)(1-T^{11}))^2 \cdot ((1-T^2)(1-T^{22}))^2 \cdot ((1-T^3)(1-T^{33}))^2 \cdots$$

entsteht. Sie beginnt mit

$$\Theta(T) = T - 2T^2 - T^3 + 2T^4 + T^5 + 2T^6 - 2T^7 - 2T^9 - 2T^{10} + T^{11} - 2T^{12} + 4T^{13} + 4T^{14} - T^{15} - 4T^{16} - 2T^{17} \ldots$$

Im Vergleich die Defekte von E:

$$a_2 = 0, a_3 = -1, a_5 = 1, a_7 = -2, a_{11} = 1, a_{13} = 4, a_{17} = -2, \ldots$$

d.h. bis auf a_2 ist a_p genau der Koeffizient vor T^p in $\Theta(T)$, $p \geq 3$ prim.

Bemerkung 18.3

Ein derartiges Muster vermuteten Taniyama und Shimura für jede elliptische Kurve mit Koeffizienten in $\mathbb{Z}$, genauer: jede elliptische Kurve ist "modular".

Bemerkung 18.4

Der große Satz von Fermat besagt, dass die Gleichung $A^n + B^n = C^n$ für $n \geq 3$ keine Lösungen in $\mathbb{Z} \setminus \{0\}$ besitzt. Fermat formulierte diese Aussage im 17. Jahrhundert und ihr Beweis galt bis 1995 als eines der größten ungelösten Probleme der Mathematik. Die Bemühungen vieler Mathematiker um diese Vermutung brachten die Mathematik, speziell die algebraische Zahlentheorie, bis heute weit voran. Die Lösung durch A. Wiles stellte 1995 eine riesigen Durchbruch dar.

Bemerkung 18.5

Bis 1980 wurden Lösungsversuche durch Faktorisierungstechniken vorgenommen. Im Jahr 1986 schlug G. Frey[29] eine Verbindung zwischen Fermats großem Satz und elliptischen Kurven vor, auf der die weiteren Erfolge beruhten: Zu einer angenommenen, nichttrivialen Lösung A, B, C des großen Fermat-Satzes zu einem primen Exponenten $n = p$ betrachtet man die zugehörige elliptische Kurve

$$E_{A,B}\colon y^2 = x(x + A^p)(x - B^p),$$

ihre Diskriminante ist $\Delta(E_{A,B}) = 16(ABC)^{2p}$, was unwahrscheinlich erscheint. Die Idee ist zu zeigen, dass eine solche Kurve nicht modular sein kann, d.h. der Taniyama-Shimura-Vermutung widerspricht. Im Jahr 1986 konnte dies gezeigt werden von K. Ribet[30]. Davon inspiriert verbrachte A. Wiles die nächsten sechs Jahre damit, die Taniyama-Shimura-Vermutung zumindest für so genannte semistabile elliptische Kurven zu zeigen, was ihm gelang. Da Frey-Kurven $E_{A,B}$ seminstabil sind, reichte dies zum Beweis des großen Fermatschen Satzes aus.

[28] https://de.wikipedia.org/wiki/Yutaka_Taniyama
https://de.wikipedia.org/wiki/Gor%C5%8D_Shimura
[29] https://de.wikipedia.org/wiki/Gerhard_Frey_(Mathematiker)
[30] https://de.wikipedia.org/wiki/Kenneth_Alan_Ribet

Bemerkung 18.6
Mittlerweile wurde die (volle) Taniyama-Shimura-Vermutung bewiesen durch C. Breuil, B. Conrad, F. Diamond und R. Taylor (2001) und heißt heute **Modularitätssatz**. Heute wird der Modularitätssatz als Spezialfall der allgemeineren **Serre-Vermutung über Galoisdarstellungen** angesehen, welche aufbauen auf den Arbeiten von A. Wiles inzwischen (im Jahr 2006) von C. Khare, J.-P. Wintenberger und M. Kisin bewiesen wurde.

3.3.3 Der Schoof-Algorithmus

Bemerkung 18.7
Für die Kryptographie-Anwendungen ist ein praktischer Weg, die Anzahl $N_{p^r} = \#E(\mathbb{F}_{p^r})$ der Punkte auf einer elliptischen Kurve über einem endlichen Körper $\mathbb{F}_{p^r}$ zu bestimmen, von Bedeutung. Dies leistet der Schoof-Algorithmus[31], den wir jetzt besprechen.

Anwendung 18.8
Sei $p > 2$ und $E(\mathbb{F}_{p^r})$ gegeben durch $y^2 = x^3 + ax + b$ mit $a, b \in \mathbb{F}_{p^r}$. Der Algorithmus bestimmt $t := a_{p^r} = p^r + 1 - \#E(\mathbb{F}_{p^r})$ nur modulo der ersten Primzahlen $p_1 = 2, p_2 = 3, p_3 = 5, \ldots$, bis damit a_{p^r} bestimmt werden kann.

Bemerkung 18.9
Für die ersten s Primzahlen $p_1, \ldots, p_s$ vermittelt der chinesische Restsatz durch die Restklassenabbildung $\mathbb{Z} \to \mathbb{Z}_{p_1} \times \ldots \times \mathbb{Z}_{p_s}$ eine Bijektion $\mathbb{Z}_{p_1 \cdots p_s} \to \mathbb{Z}_{p_1} \times \ldots \times \mathbb{Z}_{p_s}$.
Gilt $|t| < \frac{1}{2} p_1 \cdots p_s$, ist $t \bmod p_1 \cdots p_s$ durch die Restklassen $(t \bmod p_1, \ldots, t \bmod p_s)$ eindeutig bestimmt. Da nach dem Satz von Hasse $|t| < 2\sqrt{p^r}$ gilt, genügt es, r mit $p_1 \cdots p_s > 4\sqrt{p^r}$ zu wählen. Die Bestimmung von $t \bmod p_1, \ldots, t \bmod p_s$ reicht dann also zur Bestimmung von t laut chinesischem Restsatz.

Anwendung 18.10 (Schritt 1: Bestimmung von t modulo 2)
Da $t \equiv \#E(\mathbb{F}_{p^r}) \bmod 2$, muss die Parität von $\#E(\mathbb{F}_{p^r})$ bestimmt werden, d.h. ob $\#E(\mathbb{F}_{p^r})$ gerade oder ungerade ist.

- Für festes $x \in \mathbb{F}_{p^r}$ mit $x^3 + ax + b \neq 0$ in $\mathbb{F}_{p^r}$ hat $y^2 = x^3 + ax + b$ keine oder zwei Lösungen y, d.h. die Anzahl der Lösungen (x, y) mit $y \neq 0$ ist gerade und zählen modulo 2 daher nicht.
- Es bleiben $\mathcal{O}$ und die $(x_0, 0) \in E(\mathbb{F}_{p^r})$ zu zählen. Über $\overline{\mathbb{F}_{p^r}}$ faktorisiert $x^3 + ax + b$ zu $x^3 + ax + b = (x - x_0)(x - x_1)(x - x_2)$ mit $x_1, x_2 \overline{\mathbb{F}_{p^r}}$. Da E nicht-singulär, sind x_0, x_1, x_2 paarweise verschieden. Wegen $x_0 + x_1 + x_2 = 0$ (Koeffizient von x^2), folgt entweder
  a) $x_1, x_2 \in F_{p^r}$, oder
  b) $x_1, x_2 \in \overline{\mathbb{F}_{p^r}} \setminus \mathbb{F}_{p^r}$.
- Im Fall a) gibt es drei Punkte $(x_i, 0) \in E(\mathbb{F}_{p^r})$, im Fall b) nur einen Punkt, sodass $t \equiv \#E(\mathbb{F}_{p^r}) \equiv 0 \bmod 2$ folgt (wegen $\mathcal{O}$), sofern $(x_0, 0) \in E(\mathbb{F}_{p^r})$ existieren. Gibt es keine solchen Punkte, folgt (wegen $\mathcal{O}$) dann $t \equiv \#E(\mathbb{F}_{p^r}) \equiv 1 \bmod 2$.
- Ob $(x_0, 0) \in E(\mathbb{F}_{p^r})$ existieren, kann durch Überprüfen von $x - x_0 \mid x^3 + ax + b$ für alle $x_0 \in \mathbb{F}_{p^r}$ getestet werden; wegen $x^{p^r} - x = \prod_{x_0 \in \mathbb{F}_{p^r}} (x - x_0)$ also effektiv durch Überprüfen von $\mathrm{ggT}(x^3 + ax + b, x^{p^r} - x) = 1$ im Polynomring $\mathbb{F}_{p^r}[x]$ mit dem euklidischen Algorithmus.

Anwendung 18.11 (Schritt 2: Bestimmung von $t \bmod p_i > 3$)
Dies ist deutlich schwieriger, hier nur die Grundidee:

[31] https://de.wikipedia.org/wiki/Ren%C3%A9_Schoof

- Der Frobenius Φ genügt der Gleichung $\Phi^2(P) - t \cdot \Phi(P) + p^r \cdot P = \mathcal{O}$ für alle $P \in E(\overline{\mathbb{F}_{p^r}})$, da t die "Spur" des Frobenius ist (ohne Beweis).
  Zu bestimmen ist eine Zahl $\tau \in \{0, \ldots, p-1\}$, die dieser Gleichung an Stelle t für jeden Punkt $P \in E[p_i] := \{P \in E(\overline{\mathbb{F}_{p^r}}) : \mathrm{ord}(P) \mid p_i\}$ genügt.
  (Denn für so ein τ muss $(t - \tau)\Phi(P) = \mathcal{O}$ für jedes $P \in E[p_i] \setminus \mathcal{O}$ sein. Da $\Phi(P) \in E[p_i] \setminus \mathcal{O}$ ist, ist $\mathrm{ord}(\Phi(P)) = p_i$ in $E[p_i]$, es folgt $p_i = \mathrm{ord}(\Phi(P)) \mid t - \tau$, also $t \equiv \tau \bmod p_i$.

- Die Bestimmung von τ mit $\Phi^2(P) - \tau\Phi(P) + p^r \cdot P = \mathcal{O}$ für alle $P \in E[p_i]$ kann mittels der expliziten Formeln in eine Polynomgleichung übersetzt werden, für die der Reihe nach für $\tau = 0, 1, \ldots, p_i - 1$ getestet wird, ob sie gilt, bis man auf die Lösung stößt.

Bemerkung 18.12

Der Schoof-Algorithmus hat eine Laufzeit von nur $\mathcal{O}(\log^8(p^r))$, ein naiver Algorithmus zur Bestimmung von $\#E(\mathbb{F}_{p^r})$ hat eine Laufzeit von $\mathcal{O}(p^{r/4+\varepsilon})$. Damit kann $\#E(\mathbb{F}_{p^r})$ mit dem Schoof-Algorithmus effektiv und schnell bestimmt werden, wenn p^r groß ist. Mithilfe von $\#E(\mathbb{F}_{p^r}) \in \mathbb{N}$ kann dann entschieden werden, ob die (meist zufällig gewählte) elliptische Kurve kryptographisch geeignet ist oder nicht. Das behandeln wir im nächsten Abschnitt.

4 Sichere Kryptographie mit elliptischen Kurven

4.1 Bekannte Angriffe auf das DL-Problem: Überblick

Bemerkung 19.1

Die Sicherheit des ElGamal- und ECDSA-Verfahrens beruht hier auf der Schwierigkeit des DL-Problems auf elliptischen Kurven. Allerdings gibt es bestimmte Arten elliptischer Kurven, bei denen das DL-Problem algorithmisch schnell lösbar ist, sodass sich diese Kurven als kryptographisch schwach bzw. ungeeignet erweisen. Auch die Wahl eines Punktes P mit großer Ordnung ist wichtig. [19]

4.1.1 BSGS und Silver-Pohlig-Hellman

Diese beiden Methoden eignen sich zur Lösung des DL-PRoblems in einer beliebigen abelschen Gruppe G.

Definition 19.2 (DL-Problem in G)

Gegeben sei $P \in G$ mit $\mathrm{ord}(P) = n \in \mathbb{N}$, sowie $Q \in \langle P \rangle$. Gesucht ist $k \in \{0, \dots, n-1\}$ mit $kP = Q$.

Definition 19.3 (BSGS – Baby Steps Giant Steps)

Dieses Verfahren kommt in Frage, wenn P kleine Ordnung n hat. Dann kann das DL-Problem wie folgt gelöst werden; der Algorithmus hat einen Zeit- und Platzbedarf der Größenordnung $\mathcal{O}(\sqrt{n})$:

Bemerkung 19.4

Vorüberlegung: Sei $m = \lceil \sqrt{n} \rceil = \min\{l \in \mathbb{N} : l \geq \sqrt{n}\}$, schreibe $k = qm + r, r \in \{0, 1, \dots, m-1\}$ (Division mit Rest).
Ziel: Bestimme q, r.
Da $Q = kP = qmP + rP$, folgt $\underbrace{Q - rP}_{\text{"Baby step"}} = \underbrace{qmP}_{\text{"Giant step"}}$.

Bemerkung 19.5

Idee: Berechne alle möglichen Werte von "Baby step" und nach und nach die möglichen Werte von "Giant step". Trifft man auf eine Übereinstimmung, sind r und m gefunden.

Anwendung 19.6 (Schritt 1)

Berechne die Liste der "Baby steps" $B = \{(Q - rP, r) : 0 \leq r \leq m\}$.

Anwendung 19.7 (Schritt 2)

- Ist für eines der r die Gleichung $Q - rP = \mathcal{O}$ erfüllt, ist $k = r$.
- Sonst teste für den ersten "Giant step" $R = mp$, ob R in B schon vorkommt. Falls ja: $k = m + r$.
- Teste so alle "Giant steps" $2R, 3R, 4R, \dots, (m-1)R$, ob diese in B vorkommt. Wenn ja, gibt die zweite Komponente r mit $k = qm + r$.

Definition 19.8 (Silver-Pohlig-Hellman-Verfahren)

Dieses Verfahren löst das DL-Problem in einer abelschen Gruppe G, wenn die Ordnung $n = \mathrm{ord}(P)$ aus nur kleinen Primfaktoren p_i zusammengesetzt ist, d.h. **glatt** ist.

Definition 19.9 (B-glatt)

Sei $B \in \mathbb{R}_{>0}$. Dann heißt $m \in \mathbb{N}$ B-**glatt**, falls für alle $p \mid n$ gilt: $p \leq B$.

Bemerkung 19.10
Die Bestimmung von k in $\langle P\rangle$ wird auf Untergruppen von $\langle P\rangle$ der Ordnungen $p_i \mid n$ zurückgeführt. Sei $\mathrm{ord}(P) = n = \prod_{i=1}^{t} p_i^{\lambda_i}$ mit $p_1, \ldots, p_t$ paarweise verschieden und prim, $\lambda_i \in \mathbb{N}$. Der Algorithmus hat dann eine Laufzeit von $\mathcal{O}\big(\sum_{i=1}^{t}(\lambda_i(\log n + \sqrt{p_i}))\big)$.

Bemerkung 19.11
Vorüberlegung:

- Zur Bestimmung von k mit $kP = Q \in \langle P\rangle$ berechnen wir alle Restklassen $k \bmod p_1^{\lambda_1}, k \bmod p_2^{\lambda_2}, \ldots, k \bmod p_t^{\lambda_t}$. Denn laut chinesischem Restsatz ist $\mathbb{Z}/n\mathbb{Z} \simeq \mathbb{Z}/p_1^{\lambda_1}\mathbb{Z} \times \ldots \times \mathbb{Z}/p_t^{\lambda_t}\mathbb{Z}$, sodass damit dann auch die Restklasse von $k \bmod n$ bestimmt werden kann.

- Betrachte daher jedes $p = p_i, \lambda = \lambda_i$ mit $1 \le i \le t$.
  Gesucht: $z \in \{0, \ldots, p^\lambda - 1\}$ mit $z \equiv k \bmod p^\lambda$.
  Schreibe $z = z_0 + z_1 p + \ldots + z_{\lambda-1} p^{\lambda-1}$, die $z_i \in \{0, \ldots, p-1\}$, in der p-adischen Entwicklung; bestimme die $z_0, \ldots, z_{\lambda-1}$.

Anwendung 19.12 (Schritt 1)
Sei $R := \frac{n}{p} \cdot P$, dann ist $\frac{n}{p}Q = \frac{n}{p}kP = kR$ und $pR = \mathcal{O}$. Also ist $kR = zR = z_0 R$, d.h. $z_0 R = \frac{n}{p} \cdot Q$. Somit muss man in der Untergruppe $\langle R\rangle$ der (kleinen) Ordnung p ein DL-Problem lösen, um z_0 zu bestimmen – etwa mit BSGS.

Anwendung 19.13 (Schritt 2)
Seien $z_0, \ldots, z_{j-1}$ schon (rekursiv) bestimmt, wo $j \le \lambda - 1$ ist. Berechne dann $Q_j := \frac{n}{p^{j+1}}(Q - (z_0 + z_1 p + \ldots + z_{j-1}p^{j-1})P)$. Da $nP = \mathcal{O}$, ist $\frac{n}{p^{j+1}} \cdot p^\lambda P = \mathcal{O}$; da $z \equiv k \bmod p^\lambda$, ist $k = z + sp^\lambda, s \in \mathbb{Z}$, also $\frac{n}{p^{j+1}}Q = \frac{n}{p^{j+1}}kP = \frac{n}{p^{j+1}}zP + \underbrace{\frac{n}{p^{j+1}} \cdot sp^\lambda P}_{=\mathcal{O}} = \frac{n}{p^{j+1}}zP$ und somit $Q_j = \frac{n}{p^{j+1}}(z_j p^j + \ldots + z_{\lambda-1}p^{\lambda-1})P = \frac{n}{p}z_j P = z_j R$.
Zur Berechnung von z_j ist wieder ein DL-Problem in der Untergruppe $\langle R\rangle$ der Ordnung p zu lösen – etwa mit BSGS.

Bemerkung 19.14
Ist die Gruppenordnung glatt, ist der Algorithmus also sehr schnell.

4.1.2 Pollard-ρ und Pollard-λ

Bemerkung 19.15
Der **Pollard-ρ-Algorithmus** ist von der Laufzeit her vergleichbar mit BSGS, ist aber speicherplatztechnisch günstiger und lässt sich gut parallelisieren. Mit m Prozessoren wird der Algorithmus so um den Faktor m schneller.

Bemerkung 19.16
Der **Pollard-λ-Algorithmus** ist ähnlich, im Allgemeinen aber eher langsamer als Pollard-ρ. Er liefert gute Ergebnisse, wenn der diskrete Logarithmus in einem hinreichend kleinen Intervall liegt. Auch Pollard-λ ist gut parallelisierbar. (Die genauen Verfahren können in der Fachliteratur nachgeschlagen werden.)

4.1.3 MOV und SSSA

Bemerkung 19.17
Beim **MOV-Verfahren** (Autoren: Menezes, Okamoto, Vanstone) wird das DL-Problem für eine elliptische Kurve $E(\mathbb{F}_{p^r})$ auf das in der Gruppe $(\mathbb{F}_{p^{rl}}^*, \cdot)$ für ein $l \ge 1$ zurückgeführt. Es ist also speziell nur für elliptische Kurven konstruiert, nicht für allgemeine abelsche Gruppen. Zeigt sich hier, dass $l \ge 1$ so wählbar ist, dass das

DL-Problem in $(\mathbb{F}_{p^{rl}}^*, \cdot)$ leicht, d.h. schnell, zu lösen ist, ist die elliptische Kurve kryptographisch ungeeignet, etwa wenn $n = \mathrm{ord}(P)$ Teiler von $p^{rl} - 1$ ist.

Bemerkung 19.18
Generell lässt sich das DL-Problem in $(\mathbb{F}_{p^{rl}}^*, \cdot)$ in subexponentieller Zeit schnell lösen (mit so genannten Indexkalkül-Methoden), sodass Kurven, für die das DL-Problem auf ein schnelles in einem $(\mathbb{F}_{p^{rl}}^*, \cdot)$ zurückgeführt werden kann, als kryptographisch schwach bzw. ungeeignet angesehen werden. Das ist etwa bei supersingulären elliptischen Kurven der Fall, bei denen die Gruppenstruktur recht gut bekannt ist.

Definition 19.19 (supersingulär)
Eine elliptische Kurve $E(\mathbb{F}_{p^r})$ heißt **supersingulär**, falls $p = \mathrm{char}(\mathbb{F}_{p^r})$ die Spur des Frobenius teilt, d.h. $p \mid p^r + 1 - \#E(\mathbb{F}_{p^r})$.

Bemerkung 19.20

- Um zu testen, ob eine Kurve supersingulär und damit kryptographisch ungeeignet ist, muss die Gruppenordnung $\#E(\mathbb{F}_{p^r})$ der elliptischen Kurve bestimmt werden – typischerweise mit dem Schoof-Algorithmus.
- Der Begriff supersingulär hat nichts mit singulären Punkten zu tun: elliptische Kurven sind per Definition nicht-singulär.

Beispiel 19.21
Die Kurve $E(\mathbb{F}_2)\colon y^2 + y = x^3 + x + 1$ ist supersingulär, da $E(\mathbb{F}_2) = \{\mathcal{O}\}$.

Bemerkung 19.22
Supersingularität bleibt bei Übergang zu einem Erweiterungkörper erhalten: Ist $E(\mathbb{F}_{p^r})$ supersingulär, dann auch $E(\mathbb{F}_{p^{rl}})$ für alle $l \geq 1$. (ohne Beweis)

Satz 19.23 (Erstes Kriterium für Supersingularität)
Sei $p \geq 3$, $E(\mathbb{F}_p)\colon y^2 = x^3 + ax^2 + bx + c =: h(x)$ elliptische Kurve. Dann ist $E(\mathbb{F}_p)$ genau dann supersingulär, wenn der Koeffizient vor T^{p-1} ind $h(T)^{\frac{p-1}{2}} \in \mathbb{F}_p[T]$ gleich 0 ist.

Satz 19.24 (Zweites Kriterium für Supersingularität)
Sei $p = 2$, $E(\mathbb{F}_{2^r})\colon y^2 + a_1xy + y = x^3 + a_2x^2 + a_4x + a_6$. Dann ist $E(\mathbb{F}_{2^r})$ genau dann supersingulär, wenn $a_1 = 0$. (ohne Beweis)

Beispiel 19.25
Siehe Beispiel 19.21, und $E(\mathbb{F}_p)\colon y^2 = x^3 + x$ ist für $p \equiv 3 \bmod 4$ supersingulär, denn:
$(T^3 + T)^{(p-1)/2} = \sum_{j=1}^{(p-1)/2} \binom{(p-1)/2}{j} T^{3j} T^{(p-1)/2-j}$ mit $\frac{p-1}{2} + 2j = p - 1 \Leftrightarrow 2j = \frac{p-1}{2}$, d.h. wenn $2 \mid \frac{p-1}{2} \Leftrightarrow p \equiv 1 \bmod 4$
$\Rightarrow$ Koeffizient vor T^{p-1} ist $\binom{(p-1)/2}{(p-1)/4} \neq 0$ in $\mathbb{F}_p$. Für $p \equiv 3 \bmod 4$ kommt T^{p-1} nicht vor, also Koeffizient ist 0.

Bemerkung 19.26
Der MOV-Algorithmus nutzt bei einer supersingulären Kurve $E(\mathbb{F}_{p^r})$ aus, dass $t = p^r + 1 - \#E(\mathbb{F}_{p^r})$ nur einen der Werte $t \in \{0, \pm\sqrt{p^r}, \pm\sqrt{2p^r}, \pm\sqrt{3p^r}, \pm 2\sqrt{p^r}\}$ annehmen kann.

Bemerkung 19.27
Beim **SSSA-Verfahren** (Autoren: Satoh, Smart, Semaev, Araki) handelt es sich um einen schnellen Algorithmus zur Lösung des DL-Problems auf anormalen elliptischen Kurven, welche deswegen kryptographisch ungeeignet sind.

Die Grundidee ist, die elliptische Kurve über $\mathbb{F}_p$ als eine über $\mathbb{Q}_p$ zu betrachten, dem Körper der p-adischen Zahlen, und die Logarithmen auf eine Division in $\mathbb{Z}_p$ zurückzuführen (was leicht ist).

Definition 19.28 (anormal)
Eine elliptische Kurve $E(\mathbb{F}_p)$ heißt **anormal**, wenn $\#E(\mathbb{F}_p) = p$ ist. Dies lässt sich wieder durch Bestimmung von $\#E(\mathbb{F}_p)$ mit dem Schoof-Algorithmus leicht überprüfen. Der SSSA-Algorithmus kann auf Kurven über $\mathbb{F}_{p^r}$ übertragen werden. Er hat polynomielle Laufzeit.

4.1.4 Fazit: geeignete elliptische Kurven und Vergleich mit anderen Public-Key-Verfahren

Bemerkung 19.29
Eine elliptische Kurve $E(\mathbb{F}_p) : y^2 = x^3 + ax + b$ mit vorgegebener Bitzahl für p ist leicht zu finden – mit Zufallszahlengenerator und Primzahltest, was auch für große Zahlen mit mehreren hundert Dezimalstellen schnell machtbar ist; dafür kennt man ganz gute Algorithmen.

Bemerkung 19.30
Man wählt so lange die Parameter p, a, b neu, bis die Diskriminante $4a^3 + 27b^2$ nicht durch p teilbar ist und somit eine elliptische Kurve vorliegt. Ziemlich sicher liegt dann eine kryptographisch geeignete Kurve vor. Das testet man nach Berechnen der Gruppenordnung $\#E(\mathbb{F}_p)$ mit dem Schoof-Algorithmus:

Bemerkung 19.31

- Ist $\#E(\mathbb{F}_p)$ glatt, d.h. hat $\#E(\mathbb{F}_p)$ nur kleine Primteiler, ist die Kurve ungeeignet wegen Silver-Pohlig-Hellman.
- Ist $\#E(\mathbb{F}_p) = p + 1$, d.h. die Kurve ist supersingulär, dann ist die Kurve ungeeignet (MOV).
- Ist $\#E(\mathbb{F}_p) = p$, d.h. die Kurve anormal, ist die Kurve ungeeignet (SSSA).

Ob die Kurve supersingulär oder anormal ist, kann man meist leicht daran erkennen durch Wahl von Punkten $P \in E(\mathbb{F}_p)$ und dem Test, ob $(p+1)P = \mathcal{O}$ bzw. $pP = \mathcal{O}$ gilt.

Bemerkung 19.32
Die Wahl eines Punktes P mit nicht zu kleiner Ordnung n muss dann gewährleistet werden. Speziell darf n kein Teiler von $p^{rl} - 1$ sein, wenn das DL-Problem in $(\mathbb{F}^*_{p^{rl}}, \cdot)$ leicht zu lösen ist, und n darf auch kein Vielfaches von p sein (wegen SSSA). Auch sollte n nicht glatt sein; man wählt in der Praxis meist Punkte P, für die $n = \mathrm{ord}(P)$ eine hinreichend große Primzahl ist; für sie sollte etwa $n > 2^{160}$ gelten.

Bemerkung 19.33
Die für allgemeine elliptische Kurven, die in diesem sinne als kryptographisch sicher gelten, bekannten Implementationen des DL-Problems sind alle von exponentieller Komplexität. Ein Kryptographieverfahren wie ElGamal bzw. DSA gilt dann als kryptographisch sicher.

Bemerkung 19.34
Für konventionelle Kryptoverfahren (RSA und ElGamal/DSA auf $(\mathbb{F}^*_{p^r}, \cdot)$) gibt es subexponentielle Verfahren zur Lösung des DL-Problems. Dieser Vergleich schlägt sich in der Wahl der Schlüssellängen (= Bitzahl der Größe des endlichen Körpers) nieder: Die Schlüssellänge eines elliptischen Kurven-Systems wächst etwas schneller als die dritte Wurzel der Schlüssellänge eines konventionellen Krypto-Systems mit ähnlicher kryptographischer Sicherheit:

Grafik einfügen!

Bemerkung 19.35

- Man geht davon aus, dass Kurven $E(\mathbb{F}_p)$ mit $p \approx 2^{173}$. wo $\#E(\mathbb{F}_p)$ einen Primteiler $\geq 2^{160}$ hat, die gleiche Sicherheit wie ein RSA-System mit 1024 Bit bietet (für 4096 Bit beim RSA nur etwa 313 bei EC-System!).
- Durch die geringere Schlüssellänge bei Verfahren mit elliptischen Kurven kann man diese leicht auf Smart-Cards ohne Koprozessor implementieren. Solche Smart-Cards sind wesentlich billiger als Chip-Karten mit Koprozessor.

Bemerkung 19.36

Bedenken der elliptischen Kurven-Kryptographie: Die Nichteignung supersingulärer und anormaler Kurven kam schnell und überraschend. Es ist unklar, ob noch weitere ungeeignete Kurvenfamilien existieren und mit einem schnellen DL-Algorithmus angreifbar sind.

4.2 ElGamal für elliptische Kurven

Bemerkung 20.1

Erinnerung an das allgemeine ElGamal-Verschlüsselungsverfahren für eine beliebige abelsche Gruppe G aus Abschnitt 1.2.3: [20]

Alice möchte eine geheime Botschaft $m \in G$ an Bob schicken.

Anwendung 20.2 (ElGamal-Verfahren)

Das Verfahren geht wie folgt:

(1) Alice wählt eine Zufallszahl $\tilde{a} \in \{1, \dots, n-1\}$ und berechnet $\tilde{a}x$. Alice besorgt sich Bobs öffentlichen Schlüssel bx und berechnet $R = \tilde{a}(bx) + m$.

(2) Alice schickt $\tilde{a}x$ und R an Bob.

(3) Bob berechnet $b(\tilde{a}x) = \tilde{a}(bx)$ und die Nachricht durch $R - b(\tilde{a}x) = m$.

Anwendung 20.3 (ElGamal mit elliptischen Kurven)

Ist nun G die abelsche Gruppe einer kryptographisch geeigneten elliptischen Kurve, kann dieses Verfahren als sicher angesehen werden. Eine Umsetzung ist wie folgt möglich:

(i) Man wählt eine kryptographisch geeignete elliptische Kurve $E(\mathbb{F}_p)\colon y^2 = x^3 + ax + b$, d.h. eine Primzahl p und natürliche Zahlen $0 \leq a, b < p$ (und prüft die Sicherheit gemäß Abschnitt 4.1, sodass das DL-Problem schwer ist), sowie ein $P \in E(\mathbb{F}_p)$ mit großer Ordnung als Basispunkt.

(ii) A und B einigen sich, wie man Klartext als einen Punkt auf der elliptischen Kurve kodiert und wieder zurückerhält (etwa wie ab 17.18 beschrieben).

(iii) Jeder Teilnehmer wählt eine Zahl $k \in \mathbb{N}$ als privaten Schlüssel und gibt $Q = kP \in E(\mathbb{F}_p)$ als öffentlichen Schlüssel bekannt:
Alice: $a \in \mathbb{N}$ (geheim) und aP (öffentlich)
Bob: $b \in \mathbb{N}$ (geheim) und bP (öffentlich)
Danach kann das ElGamal-Verfahren wie oben beschrieben durchgeführt werden.

Beispiel 20.4

Man lege das Alphabet $\Sigma = \{A, B, \dots, Z\}$ zugrunde und nehme die elliptische Kurve $E(\mathbb{F}_p)\colon y^2 ? x^3 + Ax + B$ mit $p = 6833, A = 5984, B = 1180$ und den Basispunkt $P = (1, 2631)$.

- Teilnehmer Bob wählt den geheimen Schlüssel $b = 2465 \in \mathbb{N}$ und macht $Q = 2465 \cdot P = (4748, 2021)$ öffentlich.
- Teilnehmerin Alice schickt den geheimen Text "`INSTITUT`" etwa in der folgenden Form (Streckungsfaktor 10, $\tilde{a}$ = Zufallszahl) an Bob:

Text	`IN`	`ST`	`IT`	`UT`
w	$(8,13)_{(26)} = 221$	$(18,19)_{(26)} = 487$	$(8,19)_{(26)} = 227$	$(20,19)_{(26)} = 539$
	$\rightsquigarrow 2210$	$\rightsquigarrow 4870$	$\rightsquigarrow 2270$	$\rightsquigarrow 5390$
M_w	$(2211, 556)$	$(4872, 3315)$	$(2270, 2994)$	$(5392, 959)$
$\tilde{a}$	6794	3035	3508	2765
$\tilde{a}P$	$(687, 171)$	$(1211, 2731)$	$(2714, 2389)$	$(6818, 2527)$
$\tilde{a}Q + M_w$	$(3327, 5675)$	$(2260, 17)$	$(357, 1247)$	$(1333, 6617)$

Die Folge der Punktepaare $(\tilde{a}P, \tilde{a}Q + M_w)$ wird von Alice an Bob verschickt. Bob entschlüsselt mit $\tilde{a}Q + M_w - b\tilde{a}P = M_w$, da $Q = bP$, die x-Koordinate von $M_w \in E(\mathbb{F}_p)$ ergibt dann mit $\left\lfloor \frac{x}{10} \right\rfloor = w$ den Textblock.

4.3 ECDSA-Signaturen

Bemerkung 20.5

Das **ECDSA-Verfahren** ist das DSA-Verfahren (elektronische Unterschrift) auf elliptischen Kurven. Alice möchte dabei ein Dokument $m \in \mathcal{M}$ an Bob schicken und signieren.

Anwendung 20.6 (ECDSA-Verfahren – Schritt 1)

Zuerst müssen sich die Teilnehmer Alice und Bob darauf einigen, auf welcher elliptischen Kurve gearbeitet werden soll.

- Gewählt werden ein Grundkörper $\mathbb{F}_p$ (aber auch $\mathbb{F}_{2^r}$ möglich) mit $p > 3$ prim, $A, B \in \mathbb{F}_p$ für die elliptische Kurve $E(\mathbb{F}_p)\colon y^2 = x^3 + Ax + B$ (im Fall $\mathbb{F}_{2^r}$ nimmt man die Gleichung $y^2 + xy = x^3 + Ax^2 + B$), sodass $E(\mathbb{F}_p)$ eine kryptographisch geeignete elliptische Kurve ist.
- Gewählt wird ein Basispunkt $P = (x, y) \in E(\mathbb{F}_p)$ mit $n := \mathrm{ord}(P) \in \mathbb{N}$. Verlangt wird außerdem, dass n prim ist mit $n > 2^{160}$ und $n > 4\sqrt{p}$. Weiter soll n kein Teiler von $p - 1, p^2 - 1, \ldots, p^{30} - 1$ sein und $n \neq p$ gelten.

Bemerkung 20.7

- Die Bedingung $n > 2^{160}$ sorgt dafür, dass das DL-Problem in $\langle P \rangle$ nicht mit Pollard-ρ angreifbar ist. Ist n kein Teiler von $p^k - 1, k \leq 30$, kann man den MOV-Algorithmus nicht einsetzen. Wegen $n \neq P$ greift auch der SSSA-Algorithmus nicht.
- Es reicht, die Bedingungen an P zu erfüllen; die Kurve ist dann "von selbst" kryptographisch geeignet: Letztlich arbeitet man mit der Untergruppe $\langle P \rangle \subseteq E(\mathbb{F}_p)$.
- Dass die Kurve zufällig erzeugt wird per Zufallsgenerator für $p, A, B, P = (x, y)$, sorgt für zusätzliche Sicherheit; die zufällige Erzeugung sollte in der Praxis idealerweise überprüfbar sein, um auszuschließen, dass kryptographisch schwache Kurven durch Betrüger eingeschleust werden.

Anwendung 20.8 (weitere Schritte)

(2) Alice wählt eine Zufallszahl $a \in \{0, \ldots, n-1\}$ als privaten Schlüssel, der Punkt $aP \in E(\mathbb{F}_p)$ gibt sie als öffentlichen Schlüssel bekannt. Für ihr zu unterschreibendes Dokument $m \in \mathcal{M}$ berechnet sie $h(m) \in \{0,1\}^N$ für eine vorher festgelegte geeignete Hashfunktion; der Bitstring $(a_0, a_1, \ldots, a_{N-1}$ wird dann als $H(m) = \sum_{i=0}^{N-1} a_i 2^{N-1-i} \leq 2^N - 1$ interpretiert.

(3) Alice wählt eine Zufallszahl $\tilde{a} \in \{1, \ldots, n-1\}$ ($\tilde{a} \neq 0$) und berechnet den Punkt $\tilde{a}P = (u, v)$ sowie den Rest $\Psi(\tilde{a}P) \equiv u \bmod n$.
(Die Funktion $\Psi\colon \langle P \rangle \to \{0, 1, \ldots, n-1\}$ ist zwar nicht bijektiv, die Urbildmenge eines u ist aber klein genug, sodass unwahrscheinlich ist, dass $\Psi(R) = \Psi(kP)$ gilt, ohne dass $R = kP$ gilt. Das reicht in der Praxis.)

(4) Alice berechnet $\tilde{a}^{-1} \bmod n$ und die Restklasse $s = \tilde{a}^{-1}(H(m) - \Psi(\tilde{a}P)a) \bmod n$. Falls $s \equiv 0 \bmod n$, muss ein neues $\tilde{a}$ gewählt werden. Also zurück zu Schritt 3, da Bob bei der Prüfung der Unterschrift s invertieren wird.

(5) Alice schickt das Dokument $m \in \mathcal{M}$ zusammen mit ihrer Unterschrift $(\Psi(\tilde{a}P), s)$ an Bob.

Anwendung 20.9 (Verifikation)

Bob überprüft die Unterschrift wie folgt:

(1) Er testet, ob $\Psi(\tilde{a}P), s \in \{0, 1, \ldots, n-1\}$.

(2) Er berechnet $H(m) \in \mathbb{N}, s^{-1} \bmod n$ und den Punkt $R := s^{-1}(H(m)P - \Psi(\tilde{a}P) \cdot aP) \in E(\mathbb{F}_p)$.

(3) Für $R = \mathcal{O}$ ist die Unterschrift ungültig. Für $R = (x, y) \in E(\mathbb{F}_p) \setminus \{\mathcal{O}\}$ ist die Unterschrift gültig, wenn $x = \Psi(\tilde{a}P)$ ist, sonst nicht.

Bemerkung 20.10 (Korrektheit der Verifikation)
Denn wenn die Unterschrift von Alice stammt, ist $s = \tilde{a}^{-1}(H(m) - \Psi(\tilde{a}P)a) \bmod n$, also gilt $s^{-1}\tilde{a}^{-1}(H(m) - \Psi(\tilde{a}P)a) \equiv 1 \bmod n$, d.h. $s^{-1}(H(m) - \Psi(\tilde{a}P)a) \equiv \tilde{a} \bmod n$ und somit $R = s^{-1}(H(m)P - \Psi(\tilde{a}P)aP) = s^{-1}(H(m) - \Psi(\tilde{a}P)a)P = \tilde{a}P$, sodass die x-Koordinaten der Punkte R und $\tilde{a}P$ übereinstimmen müssen.

Index

Liste der Sätze und Definitionen