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: 22. Mai 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

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

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.

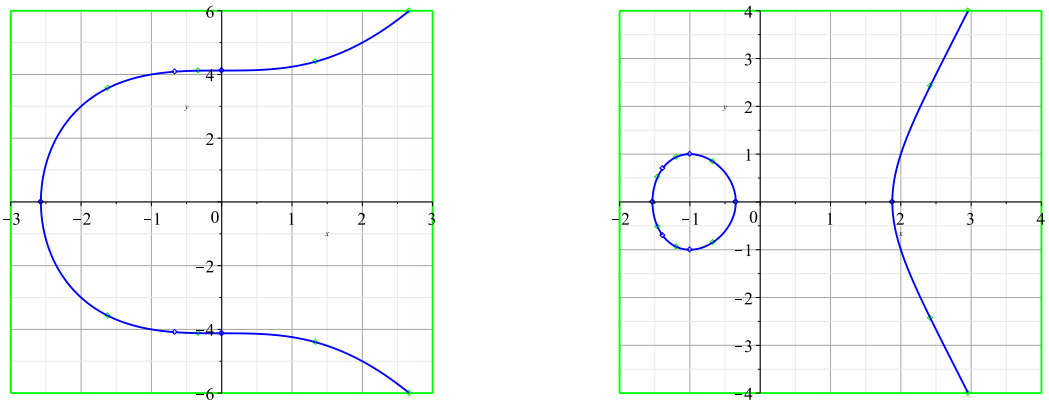


Abbildung 1: Die Kurven E_1 (links) und E_2 (rechts).

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.

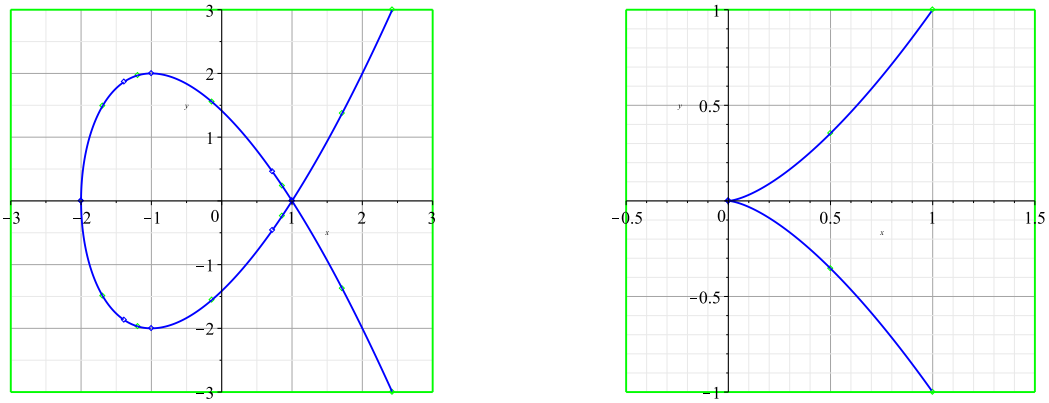


Abbildung 2: Die Kurven C_1 (links) und C_2 (rechts). C_1 ist nicht glatt im Punkt $(1, 1)$, C_2 nicht im Punkt $(0, 0)$.

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

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: [2]

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)
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 $c \cdot (a + b) = c \cdot a + c \cdot b$. Ring mit Eins

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 $0, 1, \ldots 9$ bzw. $0, 1$ bzw. $0, 1, \ldots, 9, A, \ldots, F$ 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 $c_k, c_{k-1}, \ldots, c_0 \in \{0, \ldots, g-1\}$ (genannt "Ziffern"), sodass $n = c_k g^k + c_{k-1} g^{k-1} + \ldots + c_0 = \sum_{i=0}^{k} c_i g^i$. Fordern wir $c_k \neq 0$, ist k und die Folge $c_k, \ldots, c_1, c_0$ 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 $a_{k+1} = \ldots = a_r := 0$, sonst sei $b_{r+1} = \ldots = b_k := 0$, 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 $c_k, c_{k-1}, \ldots c_0$ 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: Zum Beweis von Lemma 2.21 wurde 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

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

die **Bézout-Elemente** als $x = (-1)^{n-1}, y = (-1)^n c_{n-1}$ 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: $c_n = \frac{a}{\mathrm{ggT}(a,b)}, d_n = \frac{b}{\mathrm{ggT}(a,b)}$.
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.

Details siehe EZT-Skript Lorenz

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 $c \mid \mathrm{ggT}(ab, ac) = |a| \cdot \mathrm{ggT}(b, c) = |a|$, also $c \mid a$. Zur ersten Gleichheit: Nach Satz 2.20 existieren $x, y \in \mathbb{Z}$ mit $\mathrm{ggT}(b, c) = xb + yc$.

$|a| \cdot \mathrm{ggT}(b,c)$ 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 irgendein solcher, so teilt t auch $\mathrm{sgn}(a) \cdot (xba + yca) = xb \cdot |a| + yc \cdot |a| = |a| \cdot (xb + yc) = |a| \,\mathrm{ggT}(b,c)$. □

1.1.2 Kongruenzenrechnen und die modulare Brille

[3] 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.

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

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

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 Kongruenzensystem

$$\begin{aligned} x &\equiv a_1 \bmod m_1 \\ x &\equiv a_2 \bmod m_2 \\ &\vdots \\ x &\equiv a_r \bmod m_r \end{aligned}$$

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

$$\begin{aligned} & 8x^2 - 2x + 3 \equiv -1 \bmod 7 \\ \Leftrightarrow \quad & (x-1)^2 - 2 + 3 \equiv -1 \bmod 7 \\ \Leftrightarrow \quad & (x-1)^2 \equiv -2 \equiv 5 \bmod 7 \end{aligned}$$

man spricht auch von quadratischen Resten modulo 7

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 $x \equiv -1 \bmod 7$.

- Die Kongruenz $(x-3) \cdot 4 \equiv 1 \bmod 33$ ist schreibbar als System:

$$(x-3) \cdot 4 \equiv 1 \bmod 3$$
$$(x-3) \cdot 4 \equiv 1 \bmod 11$$

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

[4] 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.

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 $k > 0$, $k \cdot 0 := 0$

klar! 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**.

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 ↯.)
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

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. [5]
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 $(\mathbb{Z}_n^*, \cdot)$).

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 .

1.2.3 ElGamal-Verschlüsselung

[6] 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.

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 **??** 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)
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

Sei k ein beliebiger Körper. [7]

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 Polynom $1 := 1 \cdot \underline{x}^{\underline{0}} + \sum_{\underline{\nu} \neq 0} 0 \underline{x}^{\underline{\nu}}$ ist das Einselement. "Einspolynom"
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

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

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

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

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 $\underline{a}$ schreiben.

"≡" nur für $\mathbb{Z}$

doppelt unterstreichen!

Die Menge der Restklassen modulo f bezeichnen wir mit

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

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

Unterstreichungen weggelassen!

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

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

irreduzibel, da keine Nullstelle und Grad 3!

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

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

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?

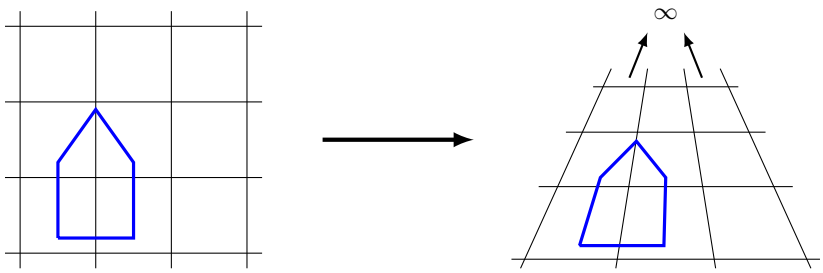


Abbildung 3: Ergänzung des zweidimensionalen affinen Raums zur projektiven Ebene.

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

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

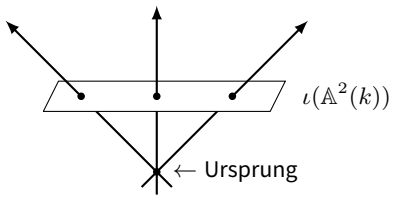
das heißt $\mathbb{P}^2(k) := (k^3 \setminus \{(0,0,0)\}/\sim$.
Wir schreiben $[y_1 : y_2 : y_3]$ für die Äquivalenzklasse, die von (y_1,y_2,y_3) repräsentiert wird und nennen sie einen **projektiven Punkt**. y_1, y_2, y_3 nennen wir **projektive Koordinaten** von $[y_1 : y_2 : y_3]$.

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

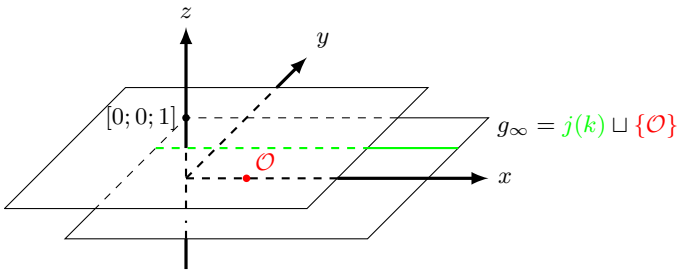
$$\iota \colon \mathbb{A}^2(k) \longrightarrow \mathbb{P}^2(k)$$
$$(x, y) \longmapsto [x : y : 1]$$

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

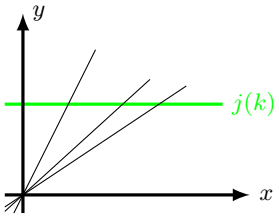


Bemerkung 8.8
Aber $\mathbb{P}^2(k)$ enthält zusätzlich noch die projektiven Punkte $[x : y : 0]$ mit $x, y \in k$ (nicht $x = y = 0$). Offenbar ist $\{[x : y : 0] : x, y \in k,$ nicht $x = y = 0\}$ eine Gerade in $\mathbb{P}^2(k)$, die wir **unendlich ferne Gerade** g_∞ nennen möchten, denn mit $j \colon k \to g_\infty, x \mapsto [x : 1 : 0]$ 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
disjunkte Vereinigung
Somit: $\mathbb{P}^2(k) = \iota(\mathbb{A}^2(k)) \sqcup \underbrace{j(k) \sqcup \{\mathcal{O}\}}_{=g_\infty}$



Ansicht auf x-y-Ebene



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 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: Für $k = \mathbb{F}_5$ können nur wenige Punkte

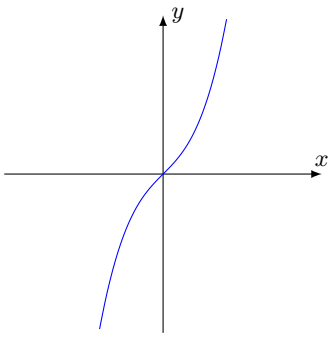


Abbildung 4: Die Menge $\{(x,y) \in \mathbb{R}^2 : y = x^3 + x\}$.

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

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

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.

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 $\frac{\partial f}{\partial x}(a,b) = -4a^3 - 4a = -4a(a^2 + 1)$, $\frac{\partial f}{\partial y}(a,b) = 2b$. 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

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

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 $P = i(\mathbb{Q}) \in \iota(\mathbb{A}^2(k))$. 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:

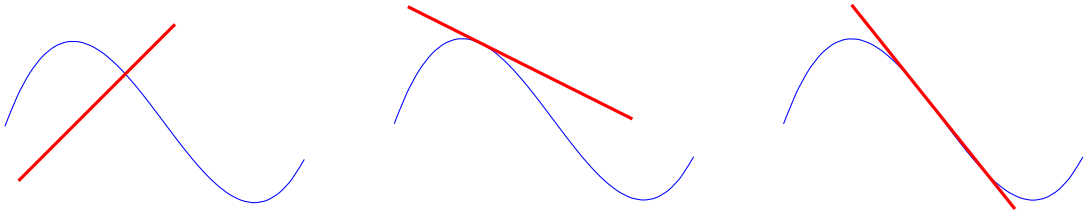


Abbildung 5: Wendepunkt-Tangenten liefern gute Approximationen an eine Kurve.

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 $a_0, \ldots, a_m, b_0, \ldots, b_n$ 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

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

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 $\deg(f_0) \leq m-1, \deg(g_0) \leq n-1$ und $f_0 g = g_0 f$

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

Beweis
(i)⇒(ii): Sei h ein gemeinsamer Teiler, $\deg(h) \geq 1$. 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 $\deg\left(\frac{f}{h}\right) \leq \deg(f_0) \leq m-1$, also $\deg(h) \geq 1$.

(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 $T^0, T^1, \ldots, T^{n+m-1}$ 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 $\deg(\mathrm{ggT}(F_1, F_2)) \geq 1$ sein müsste). Seien $P_0, P_1, \ldots, P_{d_1 d_2}$ Schnittpunkte von $\mathcal{C}_{F_1}$ und $\mathcal{C}_{F_2}$.

Beweis 10.10 (Fortsetzung)
Wir können ohne Einschränkung annehmen, dass die Punkte $P_i = (x_i, y_i), i = 0, \ldots, d_1 d_2$, 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 $\mathcal{C}_{F_2}$. □

2.4 Elliptische Kurven

2.4.1 Definition elliptischer Kurven und vereinfachte Weierstraßgleichungen

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

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 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.

lies: "Oh"

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

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

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 $b_2 = a_1^2 + 4a_2, b_4 = 2a_4 + a_1a_3, b_6 = a_3^2 + 4a_6$.

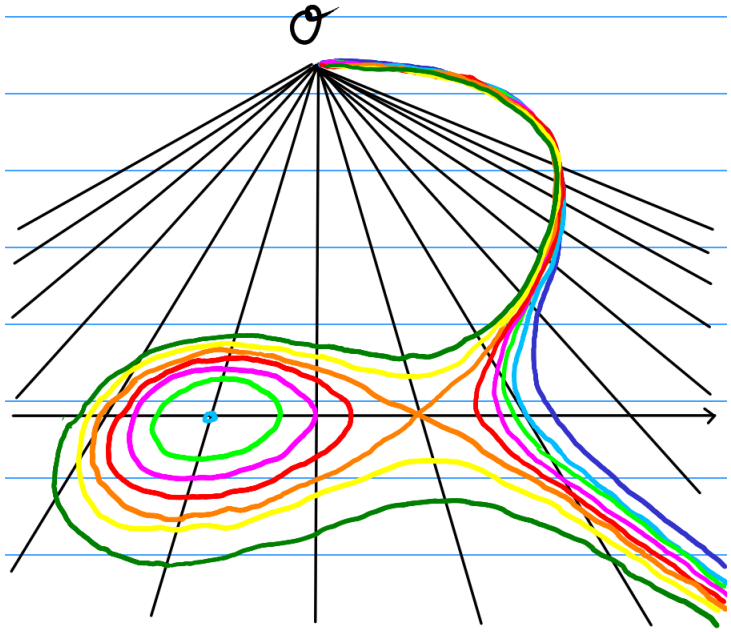


Abbildung 6: $y^2 = x^3 - 3x + s$ für $s = 5$, $s = 3$, $s = 2$, $s = 1$, $s = 0$, $s = -1$, $s = -1.999$, $s = -5$

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

$$\begin{aligned} \Psi : \mathbb{P}^2(k) &\longrightarrow \mathbb{P}^2(k) \\ [r : s : t] &\longmapsto [36r + 3b_2 t : 216s : t] \end{aligned}$$

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 $H_1(X,Y,Z) = F\big(X, Y - \frac{a_1}{2}X - \frac{a_3}{2}Z, Z\big)$:

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

- 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 $P = [r:s:t] \in \mathcal{C}_{H_1}(\overline{k})$, dann ist $\Phi^{-1}(P) = \Phi^{-1}([r:s:t])$ als Punkt der Kurve $\mathcal{C}_F(\overline{k})$ 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 $\mathcal{C}_{H_1}(\overline{k})$ nicht-singulär.

Zu (ii): Ψ hat die Inverse $[r:s:t] \mapsto \big[\frac{1}{36}r - \frac{b_2}{12}t : \frac{1}{216}s : t\big]$, 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 $H_2(X,Y,Z) = 2^6 3^6 \cdot H_1\big(\frac{1}{36}X - \frac{b_2}{12}Z, \frac{1}{216}Y, Z\big)$, 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.

Index

Liste der Sätze und Definitionen