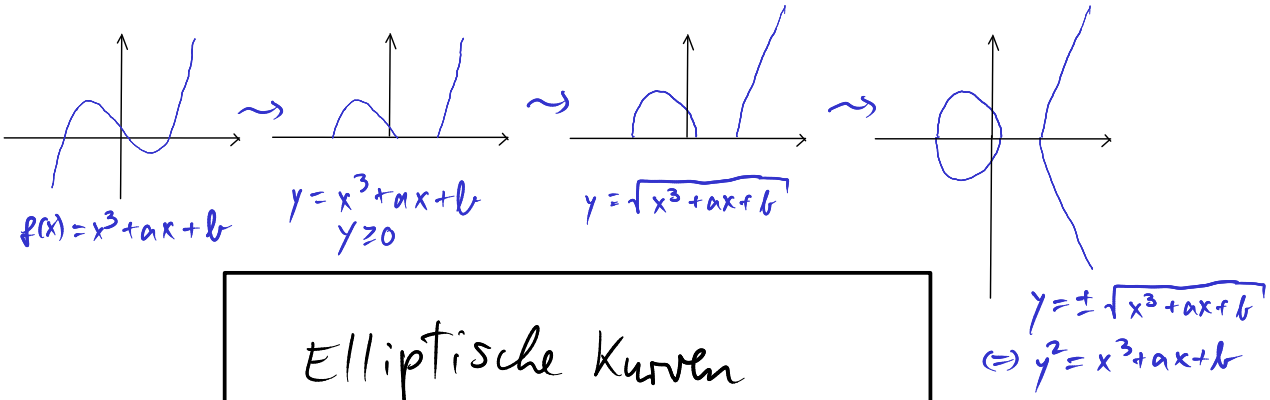


Elliptische Kurven und Kryptographie

4-std. Vorlesung im SoSe 2015 mit 2-std. Übungen

PD Dr. Karin Halupczok
Assistenz: Dipl.-Math. A. Juhas

Termin/Ort der Vorlesung: Mo/Mi 12-14 Uhr, Hörsaal M4

Vorlesungskommentar:

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.

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

Notationstabelle zur Vorlesung "Elliptische Kurven und Kryptographie"

V2:

R^* Menge der Einheiten in einem Ring mit 1 = Menge der Teiler von 1

Def.: $R^* = \{u \in R;\ \exists v \in R: uv = 1 = vu\}$

Einheit = invertierbares Ringelement = Teiler von 1

$c_k c_{k-1} \cdots c_0 (g)$ g-adische Darstellung der Zahl $n = \sum_{i=0}^{k} c_i g^i$ zur Basis g, die $c_i \in \{0, \dots, g-1\}$ heißen Ziffern

Bsp.: $2B_{(16)} = 2 \cdot 16^1 + 11 \cdot 16^0 = 43_{(10)}$

$a \mid b$ a teilt b, für $a, b \in \mathbb{Z}$

Def.: $a \mid b :\Leftrightarrow \exists c \in \mathbb{Z}: ac = b$

p prim p Primzahl, Def.: $p \in \mathbb{N}$ prim $:\Leftrightarrow \#\{t \mid p;\ t \in \mathbb{N}\} = 2$

$p^k \| n$ p^k teilt n exakt $:\Leftrightarrow p^k \mid n$ und $p^{k+1} \nmid n$

$ggT(a,b)$ größter gemeinsamer Teiler von a und b, wo $a, b \in \mathbb{Z}$

Def.: $ggT(a,b) := \max\{t \in \mathbb{N};\ t \mid a \text{ und } t \mid b\}$, falls existent

V3:

$a \equiv b \bmod m$ oder $a \equiv b\ (m)$ a kongruent zu b modulo m,

Def.: $a \equiv b\ (m) :\Leftrightarrow m \mid (b-a)$

$x + m\mathbb{Z} := \{x + ma;\ a \in \mathbb{Z}\}$ Restklasse von x mod m

kurz: $\underline{x} := x + m\mathbb{Z}$, falls $m > 1$ geg., Bsp.: $m = 10 \rightsquigarrow \underline{2} = 2 + 10\cdot\mathbb{Z} = \{\dots, -8, 2, 12, \dots\}$

$\mathbb{Z}_m := \{x + m\mathbb{Z};\ x \in \mathbb{Z}\}$ Menge der Restklassen mod m,

Bsp.: $\mathbb{Z}_{10} = \{x + 10\cdot\mathbb{Z};\ x \in \mathbb{Z}\} = \{\underline{x};\ x \in \mathbb{Z}\}$

$= \{\underline{0}, \underline{1}, \dots, \underline{9}\} = \{\underline{-4}, \underline{-3}, \dots, \underline{3}, \underline{4}, \underline{5}\} = \dots$

$\underline{x} + \underline{y} := \underline{x+y}$ Addition auf $\mathbb{Z}_m$

$\underline{x} \cdot \underline{y} := \underline{x \cdot y}$ Multiplikation auf $\mathbb{Z}_m$

$\underline{x}^*$ oder $\underline{x}^{-1}$ Inverses von $\underline{x} \in \mathbb{Z}_m$ in $\mathbb{Z}_m$, d.h. $\underline{x}^* := \underline{y} \in \mathbb{Z}_m$ mit $\underline{x} \cdot \underline{y} = \underline{1}$, definiert, falls $ggT(x,y) = 1$, explizit berechenbar mit Euklidischem Algorithmus

$\mathbb{Z}_m^*$ Menge der Einheiten im Ring $(\mathbb{Z}_m, +, \cdot)$,
Def.: $\mathbb{Z}_m^* := \{ x \in \mathbb{Z}_m ; \exists y \in \mathbb{Z}_m : x \cdot y = 1 \}$
$\mathbb{Z}_m^* = \{ x \in \mathbb{Z}_m ; \ ggT(x, m) = 1 \}$

φ Eulersche φ-Funktion, $\varphi : \mathbb{N} \to \mathbb{N}$, Def.: $\varphi(m) := \# \mathbb{Z}_m^*$

$\mathbb{F}_p$ endlicher Körper mit p Elementen, Def.: $\mathbb{F}_p := \mathbb{Z}_p$, p prim

$char(k)$ Charakteristik eines Körpers k,
Def.: $char(k) := \begin{cases} \min \{ m ; m \cdot 1 = 0 \}, & \text{falls min existiert,} \\ 0, & \text{sonst} \end{cases}$

V4:

$ord(G) := \# G$ Ordnung einer Gruppe G

$\langle a \rangle$ Erzeugnis eines Elements $a \in G$ in einer Gruppe G
= die von a in G erzeugte Untergruppe

Def.: $\langle a \rangle := \{ m \cdot a ; m \in \mathbb{Z} \}$, auch $\langle a \rangle = \mathbb{Z} \cdot a$, falls $(G, +)$ additiv geschrieben
bzw. $\langle a \rangle := \{ a^m ; m \in \mathbb{Z} \}$, auch $\langle a \rangle = a^{\mathbb{Z}}$, falls $(G, \cdot)$ multiplikativ geschrieben

$ord(a) := \# \langle a \rangle$ Ordnung eines Elements, falls $\langle a \rangle$ endlich

V7: k Körper

$f(X_1, \dots, X_n) = \sum_{\nu_1, \dots, \nu_m \geq 0} \alpha_{\nu_1, \dots, \nu_m} X_1^{\nu_1} \cdots X_m^{\nu_m}$ Polynom in n Variablen/Unbestimmten $X_1, \dots, X_n$ mit Koeffizienten $\alpha_{\nu_1, \dots, \nu_m} \in k$, wo höchstens endlich viele $\alpha_{\nu_1, \dots, \nu_m} \neq 0$

Kurz: $f(\underline{X}) = \sum_{\underline{\nu} \in \mathbb{N}_0^n} \alpha_{\underline{\nu}} \underline{X}^{\underline{\nu}} = \sum_{\underline{\nu} \in \mathbb{N}_0^n} \alpha_{\underline{\nu}} X_1^{\nu_1} \cdots X_m^{\nu_m}$ mit $\underline{X} = (X_1, \dots, X_m)$, $\underline{\nu} = (\nu_1, \dots, \nu_m)$

$\deg f$ Grad eines Polynoms f = höchste Exponentensumme eines in f vorkommenden Monoms, d.h. $\deg f := \max \{ \nu_1 + \dots + \nu_m ; \alpha_{\nu_1, \dots, \nu_m} \neq 0 \}$, falls ex. $(f \neq 0)$

$k[X_1, \dots, X_m]$ bzw. $k[\underline{X}]$ Menge aller Polynome in n Variablen über k
→ schreibe $f \in k[\underline{X}]$ für "Sei f ein Polynom in n Var. über k"

$\frac{\partial f}{\partial X_i} \in k[\underline{X}]$ formale Ableitung von $f \in k[\underline{X}]$ nach der Variablen X_i, $1 \leq i \leq n$,
Def.: $\frac{\partial f}{\partial X_i}(\underline{X}) := \sum_{\underline{\nu} \in \mathbb{N}_0^n} \alpha_{\underline{\nu}} \cdot \nu_i \cdot X_1^{\nu_1} \cdots X_{i-1}^{\nu_{i-1}} X_i^{\nu_i - 1} X_{i+1}^{\nu_{i+1}} \cdots X_m^{\nu_m}$

$f | g :\Leftrightarrow \exists h \in k[\underline{X}] : fh = g$

Betr. Polynomring $k[X]$ in einer Variablen:

a kongruent b modulo f Für $a, b, f \in k[X]$: $f \mid (b-a)$ als Polynome

$a + f \cdot k[X]$ Restklasse von a mod f, Def.: $a + f \cdot k[X] = \{a + f \cdot g;\ g \in k[X]\}$

Kurz: $\underline{a}$, falls $f \neq 0$ geg., Bsp.: $f = X^2+1 \rightsquigarrow \underline{\underline{X^3+1}} = \underline{\underline{-X+1}}$

weil X^3+1 kongr. zu $-X+1$ ist mod X^2+1

$k[X]/(f)$ Menge der Restklassen modulo f in $k[X]$,

Def. $k[X]/(f) := \{a + f \cdot k[X];\ a \in k[X]\} = \{\underline{a};\ a \in k[X]\}$

Bsp.: $k = \mathbb{R}$, $f = X^2+1 \rightsquigarrow \mathbb{R}[X]/(f) = \{\underline{g};\ g \in \mathbb{R}[X],\ g = 0 \text{ oder } \deg g \leq 1\}$

$\mathbb{F}_{p^r}$ endlicher Körper mit p^r vielen Elementen

Konstruktion: $\mathbb{F}_{p^r} := \mathbb{F}_p[X]/(f)$, wo $f \in \mathbb{F}_p[X]$ irreduzibel mit $\deg f = r \geq 1$ ist

V8: k Körper

affiner Punkt (x,y); $x, y \in k\} = k^2$ affine Ebene

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

Steigung $-\frac{a}{b}$ falls $b \neq 0$

$[x:y:z]$ projektiver Punkt mit projektiven Koordinaten $(x,y,z) \in k^3 \setminus \{(0,0,0)\}$,

Def.: $[x:y:z] := \{(u,v,w) \in k^3 \setminus \{(0,0,0)\};\ (u,v,w) \sim (x,y,z)\}$,

wobei $(u,v,w) \sim (x,y,z) :\Leftrightarrow \exists \lambda \in k \setminus \{0\}: (u,v,w) = (\lambda x, \lambda y, \lambda z) = \lambda \cdot (x,y,z)$

$\rightsquigarrow$ Kurz: $[x:y:z] := \{\lambda \cdot (x,y,z);\ \lambda \in k \setminus \{0\}\}$

$\mathbb{P}^2(k)$ $:= \{[x:y:z];\ x,y,z \in k,$ nicht $x = y = z = 0\}$

projektive Ebene: Menge aller projektiven Punkte $[x:y:z]$

Erhalten: Erweiterung von $\mathbb{A}^2(k)$ als $i(\mathbb{A}^2(k)) \subseteq \mathbb{P}^2(k)$,

$i(x,y) := [x:y:1]$.

Es gilt: $i(\mathbb{A}^2(k)) = \{[u:v:1];\ u,v \in k\}$

$= \{[x:y:z];\ x,y,z \in k;\ z \neq 0\}$

$g_\infty := \{[x:y:0];\ x,y \in k\} \subseteq \mathbb{P}^2(k)$

unendlich ferne Gerade mit $\mathbb{P}^2(k) = \underbrace{g_\infty}_{\rightsquigarrow z=0} \cup \underbrace{i(\mathbb{A}^2(k))}_{\rightsquigarrow z \neq 0}$

$G(a,b,c)$ projektive Gerade in $\mathbb{P}^2(k)$, für $a,b,c \in k$, nicht $a=b=c=0$,

Def.: $G(a,b,c) := \{[x:y:z] \in \mathbb{P}^2(k);\ ax+by+cz=0\}$.

Es ist $i(g(a,b,c)) \subseteq G(a,b,c)$

und $G(a,b,c) \setminus i(g(a,b,c)) = \{[x:y:0] \in g_\infty;\ ax+by=0\}$

$= \{[ax:ay:0] \in g_\infty;\ ax+by=0\} = \{[-by:ay:0];\ y \in k\} = [-b:a:0]$.

Falls $b \neq 0$, ist dies $= \{[x:-\frac{a}{b}x:0];\ x \in k\} = [1:-\frac{a}{b}:0]$.

$C_f(k) := \{(x,y) \in \mathbb{A}^2(k);\ f(x,y)=0\}$ zu $f \in k[x,y]$

affine Kurve in $\mathbb{A}^2(k)$, geg. als Nullstellenmenge eines Polynoms f in 2 Variablen x und y

$t_{(a,b)}(C_f)$ (affine) Tangente an eine affine Kurve $C_f(k)$ im Punkt $(a,b) \in C_f(k)$, falls existent (Tangente ex., falls C_f nicht singulär in (a,b) ist)

Def.: $t_{(a,b)}(C_f) := \{(x,y) \in \mathbb{A}^2(k);\ \frac{\partial f}{\partial x}(a,b) \cdot x + \frac{\partial f}{\partial y}(a,b) \cdot y + d = 0\}$,

mit $d \in k$ so, dass $(a,b) \in t_{(a,b)}(C_f)$

Es gilt: $t_{(a,b)}(C_f) = g\left(\frac{\partial f}{\partial x}(a,b), \frac{\partial f}{\partial y}(a,b), d\right)$, d passend

vg:

F_f Homogenisierung des Polynoms $f \in k[x,y]$, $f = \sum_{\nu,\mu \geq 0} a_{\nu,\mu} x^\nu y^\mu$, $\deg f = d$

Def.: $F_f := \sum_{\nu,\mu \geq 0} a_{\nu,\mu} X^\nu Y^\mu Z^{d-\nu-\mu} \in k[X,Y,Z]$

Bsp: $f(x,y) = y^3 - x^2 + 4xy - xy^2$

$\Rightarrow F_f(X,Y,Z) = Y^3 - X^2Z + 4XYZ - XY^2$

$C_F(k) := \{[u:v:w] \in \mathbb{P}^2(k);\ F(u,v,w)=0\}$ projektive Kurve in $\mathbb{P}^2(k)$, geg. als Nullstellenmenge eines homogenen Polynoms F in 3 Variablen X, Y und Z.

$T_P(C_F)$ (projektive) Tangente an eine projektive Kurve $C_F(k)$ im Punkt $P \in C_F(k)$, falls existent
(Tangente ex., falls C_F nicht singulär in P ist)

Def.: $T_P(C_F) := \{[x:y:z] \in \mathbb{P}^2(k);\ \frac{\partial F}{\partial x}(P)\cdot x + \frac{\partial F}{\partial y}(P)\cdot y + \frac{\partial F}{\partial z}(P)\cdot z = 0\}$

Es gilt: $T_P(C_F) = G_1\left(\frac{\partial F}{\partial x}(P), \frac{\partial F}{\partial y}(P), \frac{\partial F}{\partial z}(P)\right)$.

$m(P; G, C_F)$ $\in \mathbb{N}_0$, Schnittmultiplizität bzw. Vielfachheit, mit der sich eine Gerade G und eine Kurve C_F im Punkt P schneiden (alle Objekte in $\mathbb{P}^2(k)$ betrachtet; m existiert, wenn $G \nmid C_F$ in $k[X,Y,Z]$ ist).

Def.: $m(P; G, C_F) := 0$, falls $P \notin G \cap C_F$,
und sonst ist $m(P; G, C_F)$ die Nullstellenordnung von $t=0$ des Polynoms $\Psi(t) := F(a+ta', b+tb', c+tc')$,
wenn $P=[a:b:c]$ und $P'=[a':b':c'] \in G \setminus \{P\}$ ist;
d.h. $m(P; G, C_F)$ ist die maximale Zahl $m \in \mathbb{N}_0$, für die $t^m \in k[t]$ ein Teiler des Polynoms $\Psi(t) \in k[t]$ ist, falls ex.

V10:

$\mathrm{Res}(f,g)$ Resultante von $f, g \in k[X]$, $f = a_m X^m + \ldots + a_0$, $g = b_n X^n + \ldots + b_0$

Def:
$$\mathrm{Res}(f,g) := \det \begin{bmatrix} a_0 & & & & b_0 & & \\ \vdots & a_0 & & & b_n & \ddots & \\ a_m & & \ddots & a_0 & & \ddots & b_0 \\ & a_m & & \vdots & & & \vdots \\ & & \ddots & a_m & & & b_n \end{bmatrix}$$
(n Spalten | m Spalten)

V11:

$\mathcal{O}$ "unendlich ferner" Punkt $\mathcal{O} := [0:1:0] \in g_\infty \subseteq \mathbb{P}^2(k)$

$E(k)$ Elliptische Kurve: $C_F(k)$ zu einem kubischen, homogenen, irred. Polynom $F \in k[X,Y,Z]$, nicht-singulär mit Wendepunkt $\in \mathbb{P}^2(k)$

Δ bzw. $\Delta(C_F(k))$ Diskriminante der Kurve $C_F(k)$, wobei F ein langes Weierstraßpolynom ist

Es ist $\Delta = -16(4a^3 + 27b^2)$, falls $F(X,Y,Z) = Y^2Z - X^3 - aXZ^2 - bZ^3$

V12:

$\operatorname{disc}(\sigma) \in k$, Diskriminante eines Polynoms $\sigma \in k[X]$, $\deg \sigma = m$

Def.: $\operatorname{disc}(\sigma) := \prod_{1 \le i < j \le m} (\alpha_i - \alpha_j)^2$, falls $\alpha_1, \dots, \alpha_m \in \bar{k}$ die Nst. von σ in $\bar{k}$ sind,

d.h. wenn $\sigma(X) = c \cdot (X - \alpha_1)\cdot(X - \alpha_2)\cdots(X - \alpha_m)$ mit $\alpha_1, \dots, \alpha_m \in \bar{k}$, $c \in k$.

V13: $E(k)$ elliptische Kurve, $P, Q \in E(k)$

$G(P,Q)$ (projektive) Verbindungsgerade der Punkte $P, Q \in \mathbb{P}^2(k)$, $P \neq Q$, d.h. projektive Gerade mit $P, Q \in G(P,Q)$

$P * Q$ 3. Schnittpunkt, den die Gerade $G(P,Q)$, $P \neq Q$, mit $E(k)$ hat, gemäß Beachtung von Vielfachheiten

$P * P$ 3. Schnittpunkt, den die Tangente $T_P(E)$ mit $E(k)$ hat, gemäß Beachtung von Vielfachheiten

$P + Q$ Summe zweier Punkte $P, Q \in E(k)$,

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

$-P$ Inverses von $P \in E(k)$ bzgl. Addition "+" auf $E(k)$,

es gilt $-P = \mathcal{O} * P \rightsquigarrow P + Q = -(P * Q)$.

Ist $E(k)$ symmetrisch zur x-Achse, gilt

für $P = [a:b:c] \in E(k)$ dann $-P = [a:-b:c]$.

V15: Seien $c, d \in \mathbb{N}$

$(x:y:z)$ projektiver Punkt zu (c,d), Def.: $(x:y:z) := \{(\sigma^c x, \sigma^d y, \sigma z), \sigma \in k \setminus \{0\}\}$

$\mathbb{P}^2_{(c,d)}(k)$ projektive Ebene zu (c,d)

$m \cdot P$ m-faches von $P \in E(k)$ auf $E(k)$, Def.: $m \cdot P := \overbrace{P + \dots + P}^{m \text{ mal}}$

V16:

$r(E)$ Rang von $E(\mathbb{Q})$, d.h. $r(E) \in \mathbb{N}_0$ mit $E(\mathbb{Q}) \cong \mathbb{Z}^{r(E)} \times T$, wo $T := \{ P \in E(\mathbb{Q}) ; \operatorname{ord}(P) \in \mathbb{N} \}$ die Torsionsgruppe von E ist.

V17:

N_p Anzahl der Punkte von $E(\mathbb{F}_p)$, d.h. $N_p := \# E(\mathbb{F}_p)$

a_p Defekt, bzw. Spur des Frobenius, nämlich $a_p := p+1-N_p$ von $E(\mathbb{F}_p)$. Für $E(\mathbb{F}_{p^r})$ ist entsprechend $a_{p^r} := p^r + 1 - N_{p^r}$

$\left(\frac{u}{\mathbb{F}_p}\right)$ verallgemeinertes Legendresymbol, $= +1$ falls u ein qR mod p, $= -1$ falls u ein qNR mod p, $= 0$ falls $p \mid u$.

ϕ, Φ Frobeniusendomorphismus $\phi : \mathbb{P}^2(\overline{\mathbb{F}}_{p^r}) \to \mathbb{P}^2(\overline{\mathbb{F}}_{p^r})$, $[x:y:z] \mapsto [x^{p^r} : y^{p^r} : z^{p^r}]$
$\rightsquigarrow \Phi := \phi_{/E(\overline{\mathbb{F}}_{p^r})}$

P_w Punkt einer elliptischen Kurve, der einem Textblock w entspricht

§0: Motivation

Stichworte:

Verschlüsselungsprobleme

A/Symmetrische Verschlüsselung von Nachrichten

Anwendungen: per Internet einkaufen, Online-Banking, persönliche Daten geheimhalten... → Kommunikation über öff. Kanäle

Lösung folgender Probleme erforderlich:

- Schlüsselaustausch über öffentliche Kanäle
- Verschlüsselung ohne vorherigen Schlüsselaustausch
- Digitale Signaturen

Lösung mit elementarer Zahlentheorie, insb. RSA-Verfahren

⇝ heute: Verfahren mit elliptischen Kurven praktisch

Elliptische Kurven

▷ typische Beispiele

▷ Gruppenoperation auf elliptischen Kurven kryptographisch interessant

▷ Die Sicherheit der ECC ("elliptic curve cryptography") beruht auf dem Problem des diskreten Logarithmus auf ellipt. Kurven

---

Kryptologie

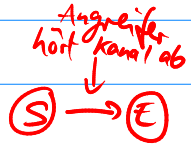
Die Kryptologie besteht aus den folgenden beiden Gebieten:

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

Kryptoloanalyse: 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, so dass 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 3. nachfolgenden im Alphabet zu ersetzen, also $A \mapsto D$, $B \mapsto E$, $C \mapsto F$, $D \mapsto G$, 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, so dass 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 / Authentifizierung

Dies können asymmetrische Verfahren leisten (auch "Public-Key-Kryptographie" genannt) und gehen zurück auf Ideen von Diffie und Hellman 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 Ansatzen 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 öffentlichen Schlüssel des Empfängers zur Verschlüsselung benutzen.
Eine derartige Funktion heißt Einwegfunktion.

- Beim RSA-Verfahren, das wir kennen lernen werden, ist diese Fkt. die Multiplikation zweier Primzahlen $(p,q) \mapsto p \cdot q$.
- Beim ECC-Verfahren ist dies die Fkt. $x \mapsto m \cdot x$ in einer abelschen Gruppe, nämlich die 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

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

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

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

I.a. 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 (↝ s. Funktionentheorie). Aufgrund dieses Zusammenhangs haben elliptische Kurven ihren Namen, sie haben ansonsten nichts mit Ellipsen zu tun.

Was sind nun elliptische Kurven? Es sind "abelsche Varietäten der Dim. 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

$$E:\ y^2 = x^3 + ax + b, \qquad a,b \in k,$$ bringen lässt.

Als Punktemenge haben wir dafür

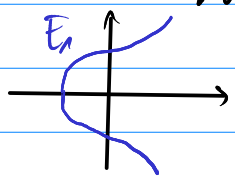
$$E(k) := \{(x,y) \in k^2 :\ y^2 = x^3 + ax + b\} \cup \{O\},$$

die Kurve hängt nur von a, b ab.

Die Rolle des zusätzlichen sog. "unendlich fernen Punkts" O werden wir dabei noch näher beleuchten.

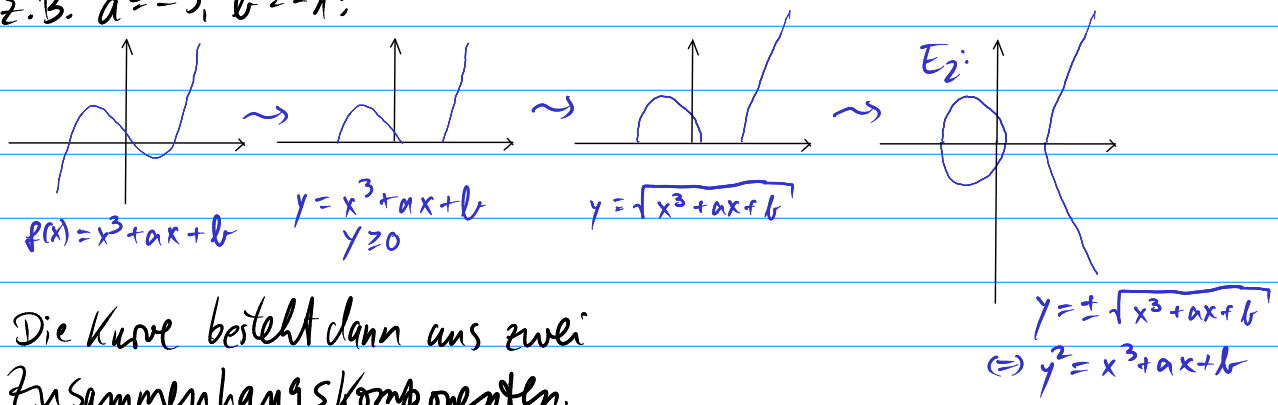
Zwei typische Beispiele für elliptische Kurven:

1.) $E_1: y^2 = x^3 + 17$, hier liegen sogar Punkte mit ganzzahligen Koordinaten auf E_1, nämlich $(-2,3)$, $(-1,4)$, $(2,5)$



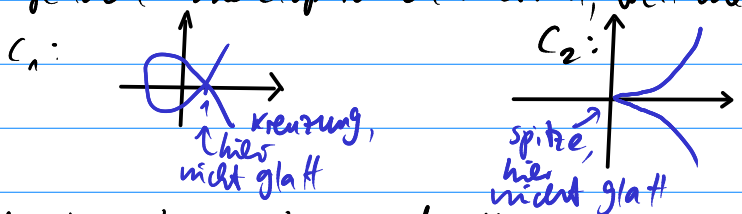
Die Kurve besteht hier aus einer Zusammenhangskomponente.

2.) $E_2: y^2 = x^3 + ax + b$, wenn $f(x) = x^3 + ax + b$ drei verschiedene Nst. hat, z.B. $a = -3$, $b = -1$:



Die Kurve besteht dann aus zwei Zusammenhangskomponenten.

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



Für die Kryptographie sind ellipt. Kurven interessant, weil sich eine Verknüpfung auf ihrer Punktemenge definieren lässt, mit der diese zu einer Gruppe wird. Dabei gerade auch endliche Körper zuzulassen, macht diese Verknüpfung auf Rechenmaschinen realisierbar. Die Sicherheit der darauf beruhenden "ECC" (elliptic curve cryptography) beruht darauf, dass das Problem des diskreten Logarithmus auf einer elliptischen Kurve E, nämlich die Umkehrung der Abb. $P \mapsto m \cdot P$ für $m \in \mathbb{N}$ fest, nach heutigem Wissensstand rechnerisch i.a. extrem schwer realisierbar ist. (Bem.: $m \cdot P := \underbrace{P + \ldots + P}_{m\text{-mal}}$, wobei "+" die Gruppenverknüpfung auf der elliptischen Kurve bezeichnet.)

§1 Allgemeines zu Kryptographie-Verfahren

§1.1 Grundlagen aus der elementaren Zahlentheorie / Gruppentheorie

Stichworte:

Def. $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$, g-adische Entw. / Binärentw.
Def. Gruppe / Ring / Körper, Teilbarkeit
Def. ggT, relativ prim / teilerfremd, PFZ
Eind. der PFZ $\rightsquigarrow$ Faktorisierungsproblem
ggT: Div. mit Rest / Eukl. Algo mit Erweiterung (↦ Bézout-El.)

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. Bzgl. den Verknüpfungen + und · sind verschiedene Axiome erfüllt, die diese Zahlbereiche zu interessanten algebraischen Strukturen machen:

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

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

Wir erinnern an die Definitionen:

1.) • **Def.:** Eine Menge $H \neq \emptyset$ mit Verknüpfung $*: H \times H \to H$, $*(a,b) = a*b$ heißt Halbgruppe, falls $*$ assoziativ ist, d.h. $\forall a,b,c \in H: a*(b*c) = (a*b)*c$

2.) • **Def.:** 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 $g*h = e = h*g$ (schreiben dann auch g^{-1} oder $\frac{1}{g}$ oder $1/g$ oder $-g$).

3.) • **Def.:** Eine Gruppe $(G,*,e)$ heißt abelsch bzw. kommutativ, falls $\forall a,b \in G: a*b = b*a$.

4.) • **Def.:** 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 neutr. El. 1, und so, dass die Distributivgesetze $(a+b)\cdot c = a\cdot c + b\cdot c$ und $c\cdot(a+b) = c\cdot a + c\cdot b$ gelten.

bedr. nur: Ring "mit 1" →

5) **Bem.:** 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.

6) • **Def.:** Die in einem Ring $(R,+,\cdot)$ bzgl. $\cdot$ invertierbaren Elemente heißen Einheiten. Die Menge der Einheiten in R wird mit R^* bezeichnet, d.h. also $R^* := \{a \in R;\ \exists\, b \in R: a\cdot b = 1 = b\cdot a\}$.
Damit ist $(R^*,\cdot,1)$ also eine Gruppe.

7.) • **Def.:** 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 / 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/kompakte Art darstellen kann. Dafür benutzen wir im Alltag das Dezimalsystem, für Rechenmaschinen ist auch das Binär- und das Hexadezimalsystem nützlich. Dafür werden die Ziffern $0,1,\dots,9$ bzw. $0,1$ bzw. $0,1,\dots,9,A,\dots,F$ verwendet. Allgemein erhalten wir die g-adische Darstellung von $n\in\mathbb{N}$ so:

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

9.) **Def.:** Die Ziffernfolge $c_k,c_{k-1},\dots,c_0$ heißt g-adische Darstellung von n.
Die Zahl a_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.

10.) **Bem.:** Wir können jede natürliche (und dann auch jede ganze) Zahl n also eind. schreiben als Linearkombination endlich vieler Potenzen von g.

11.) **Bsp.:** $163_{(10)} = 1\cdot 10^2 + 6\cdot 10^1 + 3\cdot 10^0$,

$43_{(10)} = 1\cdot 2^5 + 0\cdot 2^4 + 1\cdot 2^3 + 0\cdot 2^2 + 1\cdot 2^1 + 1\cdot 2^0 = 101011_{(2)}$

$= 2\cdot 16^1 + 11\cdot 16^0 = 2B_{(16)}$

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 $\gtrsim 2\cdot 10^{10}$) schnelle Algorithmen, die wir hier aber nicht näher behandeln möchten; etwa mit der schnellen Fouriertransformation (FFT) nach Schönhage/Strassen.

Bew. von Satz 8): ($\Leftrightarrow k\log g \le \log n < (k+1)\log g$)

Existenz: Sei $k \in \mathbb{N}_0$ so, dass $g^k \le n < g^{k+1}$ gilt, d.h. wir setzen $k := \left\lfloor \frac{\log n}{\log g} \right\rfloor$, zeigen Ex. durch vollst. Ind. nach k:

Gaußklammer: $\lfloor x \rfloor := \max\{k \in \mathbb{Z};\ k \le x\}$

$k=0$: Setze $c_0 := n$.

$k \to k+1$: Sei $g^{k+1} \le n < g^{k+2}$ und setze $n' := n - \left\lfloor \frac{n}{g^{k+1}} \right\rfloor g^{k+1}$.
Es folgt $0 \le n' < g^{k+1}$, d.h. auf n' ist die Induktionsvor. anwendbar. Nach dieser hat n' die g-ad. Darst. $n' = \sum_{i=0}^{k} c_i g^i$.
Wegen $1 \le \frac{n}{g^{k+1}} < g$ ist $1 \le \left\lfloor \frac{n}{g^{k+1}} \right\rfloor < g$, also nimm $c_{k+1} := \left\lfloor \frac{n}{g^{k+1}} \right\rfloor$, erhalten so die g-ad. Darst. $n = c_{k+1} g^{k+1} + n' = \sum_{i=0}^{k+1} c_i g^i$.

Eindeutigkeit: Sind $\sum_{i=0}^{k} a_i g^i = n = \sum_{i=0}^{r} b_i g^i$ zwei Darstellungen von $n \in \mathbb{N}$, ist $r > k$, sei $a_{k+1} = \ldots = a_r := 0$, sonst sei $b_{r+1} = \ldots = b_k := 0$ falls $r < k$.
Dann sei $\ell := \max\{i \in \mathbb{N}_0;\ i \le \max\{k,r\},\ a_i \ne b_i\}$ die größte Stelle, an der sich die Darstellungen unterscheiden. Es folgt:

$$0 = \sum_{i=0}^{\max\{r,k\}} (a_i - b_i) g^i = \sum_{i=0}^{\ell} (a_i - b_i) g^i \Rightarrow |b_\ell - a_\ell|\, g^\ell = \left| \sum_{i=0}^{\ell-1} (a_i - b_i) g^i \right|$$

($a_i - b_i = 0$ für $i > \ell$; $|b_\ell - a_\ell| \ge 1$)

$$\Rightarrow\ g^\ell \le \sum_{i=0}^{\ell-1} |a_i - b_i| g^i \le \sum_{i=0}^{\ell-1} (g-1) g^i = (g-1)\frac{g^\ell - 1}{g-1} = g^\ell - 1,\ \lightning$$

(Δ-ungl.; geom. Σ)

□

Der Beweis von Satz 8) zeigt, dass die Länge von n gleich $\left\lfloor \frac{\log n}{\log g} \right\rfloor + 1$ ist, soviele Ziffern müssen zum Hinschreiben/Eintippen von n angegeben werden. Bei verschiedenen Basen ändert sich hier nur der Faktor $\frac{1}{\log g}$ vor $\log n$. Deswegen sagt man, die Länge sei $O(\log n)$ und meint damit die Aussage $\exists C > 0 : k+1 \le C \cdot \log n$. "Landau-Symbolik" bzw. "Groß-OH-Notation"

Entscheidend für das Studium von $\mathbb{Z}$ ist der Grundbegriff der Teilbarkeit:

12) **Def.:** Für $a, b \in \mathbb{Z}$ ist a Teiler von b bzw. a teilt b, in Zeichen: $a \mid b$, falls $\exists\, c \in \mathbb{Z} : ac = b$. Ist a kein Teiler von b, schreibt man $a \nmid b$.

13) **Bsp.:** $3 \mid 12$, $4 \mid 0$, $0 \mid 0$, $7 \nmid 12$, $0 \nmid 4$. Es kann 0 nur die 0 teilen.

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

Primzahlen sind die "Bausteine" der natürlichen Zahlen:

15) **Satz von der eindeutigen Primfaktorzerlegung (PFZ) bzw. Hauptsatz der (elementaren) Arithmetik:**

Jede natürliche Zahl $n > 1$ besitzt **genau eine** Darstellung

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

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.

16) **Bem.:** 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 $p_i^{e_i+1} \nmid n$. Dafür gibt es das Symbol $p^{e_i} \| n$, und die PFZ lässt sich kompakt auch schreiben als $n = \prod_p p^{e(p)}$, wobei $e(p) := e$ mit $p^e \| n$ falls $p \mid n$, und $e(p) := 0$ falls $p \nmid n$.
Weiter ist $\omega(n) := r$ die Anzahl der versch. Primteiler von n.

Beweis des Satzes 15): Existenz: Ist n prim, ist nichts z.z., und ist n nicht prim, gibt es $k, l \in \mathbb{N} \setminus \{1\}$ mit $n = k \cdot l$.
Da $\min\{k, l\} > 1$, folgt $\max\{k, l\} < n$. Nach Induktionsvor. sind also k, l Produkte von Potenzen von PZen also auch $n = k \cdot l$.
Die Eindeutigkeit zeigen wir erst später als Anwendung von Lemma 21). □

Bsp.: die PFZ von 360 ist $360 = 2^3 \cdot 3^2 \cdot 5$, d.h. $e(2) = 3$, $e(3) = 2$, $e(5) = 1$, und sonst $e(p) = 0$ für $p \notin \{2, 3, 5\}$.

Die Eindeutigkeit der PFZ zeigt, dass auch die PFZ eine Möglichkeit zur Darstellung natürlicher Zahlen ist. Diese ist jedoch unpraktisch, weil das folgende Problem i.a. schwer zu lösen ist, worauf einige kryptographische Verfahren (insb. RSA) beruhen:

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 PFZ von n leicht bestimmt werden.
In der Praxis, wenn n nicht gerade schon von einer spezieller 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 $O(\exp(C(\log n)^{1/3}(\log\log n)^{2/3}))$, d.h. es handelt sich um ein sogenanntes subexponentiell schnelles Verfahren, weil

$$(\log n)^B \ll \exp(C(\log n)^{1/3}(\log\log n)^{2/3}) \ll \exp(D\log n) = n^D.$$

($(\log n)^B$: polynomiell in $\log n$; mittlerer Term: irgendwo dazwischen...; $\exp(D\log n)$: exponentiell in $\log n$; [Inputgröße: $O(\log n)$])

▸ P. Shor entdeckte um 1994, dass das Faktorisierungsproblem auf einem Quantencomputer mit einer Laufzeit von (meist) nur $O((\log n)^3)$ 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 Dokumenten von E. Snowden, dass die NSA an der Entwicklung eines kryptologisch nützlichen Quantencomputers arbeitet, vgl. wikipedia "Quantencomputer".

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

18.) **Def:** Seien $a, b \in \mathbb{Z}$. Der ggT von a und b (größter gemeinsamer Teiler) in $\mathbb{N}$ ist die Zahl $d := \max\{t \in \mathbb{N}: t|a \wedge t|b\}$. Notation: $ggT(a,b) := d$.
Ist $ggT(a,b)=1$, heißen a und b teilerfremd. Haben wir für a und b die PFZen $a = \prod_p p^{e(p)}$ und $b = \prod_p p^{f(p)}$ vorliegen, kann ihr ggt leicht bestimmt werden als $ggT(a,b) = \prod_p p^{\min(e(p),f(p))}$, z.B. $ggT(2^3 \cdot 3^6 \cdot 5^4, 2^4 \cdot 3^5) = 2^3 \cdot 3^5$.
Wegen dem Faktorisierungsproblem kann dies aber so nicht praktisch umgesetzt werden. Stattdessen benutzt man den (polynomiell) schnellen euklidischen Algorithmus. [Dass dieser so schnell ist, wird eine (ü)-Aufgabe]

19.) **Satz** (Teilen mit Rest): Zu $a \in \mathbb{Z}, b \in \mathbb{N}$ ex. eind. $q, r \in \mathbb{Z}$, $0 \le r < b : a = qb + r$, nämlich $q = \lfloor \frac{a}{b} \rfloor = \max\{m \in \mathbb{Z}: m \le \frac{a}{b}\}$ und $r = a - qb$. Dabei heißt r der kleinste nichtnegative Rest. Statt $0 \le r < b$ kann auch $r \in \mathbb{Z}$, $|r| < \frac{b}{2}$, erfüllt werden; r heißt dann der absolut kleinste Rest (bei Division durch b).
L.J: Gaußklammer
Bew.: ✓ Bsp.: $20 = 7 \cdot 2 + 6 = 7 \cdot 3 + (-1)$
(6: kl. nn. Rest; −1: abs. kl. Rest)

20.) **Satz** (vom euklidischen Algorithmus): Seien $a, b \in \mathbb{N}$.
Durch fortgesetztes Teilen mit Rest erhalten wir als letzten Rest $\neq 0$ den $ggT(a,b)$, sowie $x, y \in \mathbb{Z}$ mit $ggT(a,b) = xa + yb$ laut Schema.

Beschreibung des Rechenverfahrens: (a, b)
Rechnen sukzessive: $r_{-1} := a$, $r_0 := b$, $r_{-1} = q_0 r_0 + r_1$, $r_0 = q_1 r_1 + r_2$, $r_1 = q_2 r_2 + r_3$, ...
Letzte Division: $r_{n-1} = q_n r_n$
→ Das Verfahren wird fortgeführt, bis erstmals ein Rest $r_{n+1} = 0$ auftritt, was wegen $r_0 > r_1 > r_2 > \dots$ nach höchstens $b+1$ vielen Schritten der Fall sein wird. Sind die Quotienten $q_0, \dots, q_n$ bekannt, können mit den Rekursionen $c_{-2} = 0$, $c_{-1} = 1$, und $c_k = q_k c_{k-1} + c_{k-2}$, $k = 0, 1, 2, \dots, n$, sowie $d_{-2} = 1$, $d_{-1} = 0$, sowie $d_k = q_k d_{k-1} + d_{k-2}$, $k = 0, 1, 2, \dots, n$, die Bézout-Elemente als $x = (-1)^{n-1} d_{n-1}$ und $y = (-1)^n c_{n-1}$ berechnet werden:

Schematisch:

q_k			q_0	q_1	q_2	...	q_{n-1}	q_n
c_k	0	1 →	q_0	c_1	c_2	...	$c_{n-1} \rightsquigarrow \pm y$	$a/ggT(a,b)$
d_k	1	0	1	d_1	d_2	...	$d_{n-1} \rightsquigarrow \pm x$	$b/ggT(a,b)$

Bsp.: $a = 360$, $b = 84 \rightsquigarrow 360 = 4 \cdot 84 + 24$, $84 = 3 \cdot 24 + 12$, $24 = 2 \cdot 12 + 0$ (ggT = 12)

q_k			4	3	2
c_k	0	1	4	13	$30 = \frac{360}{12}$
d_k	1	0	1	3	$7 = \frac{84}{12}$

$\rightsquigarrow 13 \cdot 84 - 3 \cdot 360 = 12$ (13 = y, 3 = x)

Wir behaupten also:

(1) Es ist $ggT(a,b) = r_m$.

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

Bew.: Zu (1):

Da $r_m \mid r_{m-1}, r_m \mid r_{m-2}, \dots, r_m \mid r_0 = b, r_m \mid r_1 = a$, ist r_m Teiler von a und b ("Teilen mit Rest von unten nach oben")

Ist d irgendein Teiler ≥ 1 von a und b, folgt $d \mid r_1 = a - q_0 b \Rightarrow d \mid r_2 = r_0 - q_1 r_1 \Rightarrow d \mid r_3 = \dots$, also auch r_m, so dass $d \leq r_m$ folgt ("Teilen mit Rest von oben nach unten"). Somit ist $r_m = ggT(a,b)$.

Zu (2): Induktiv kann $c_{k-1} d_k - c_k d_{k-1} = (-1)^k$ gezeigt werden.

Daher gen. z.z.: $c_n = \frac{a}{ggT(a,b)}$, $d_n = \frac{b}{ggT(a,b)}$. [Bew.: 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. (ohne Bew.)

Einzelheiten → vgl. Lesestoff-Skript

Da beider KBE alle Brüche $\frac{c_k}{d_k}$ gekürzt sind wegen $c_{k-1} d_k - c_k d_{k-1} = (-1)^k$, folgt dies.] □

konstruktiv!

Der Satz 20.) vom euklidischen Algorithmus sichert uns ↓ also die Existenz ganzer Zahlen $x, y \in \mathbb{Z}$ mit $ggT(a,b) = xa + yb$. 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:

Bem.: Eigenschaft des ggT: $d = ggT(a,b)$, $c \mid a \wedge c \mid b \Rightarrow c \mid d$, denn $c \mid xa + yb = d$, $x, y \in \mathbb{Z}$.

21.) Lemma: $a, b, c \in \mathbb{Z}$, nicht $b = c = 0 \Rightarrow$ ($c \mid ab$ und $ggT(b,c) = 1 \Rightarrow c \mid a$)

Bew.: Vor. und $c \mid ac \overset{\text{Bem.}}{\Rightarrow} c \mid ggT(ab, ac) \overset{\otimes}{=} |a| \cdot ggT(b,c) = |a|$, also: $c \mid a$.

Noch zu ⊗: Nach Satz 20.) ex. $x, y \in \mathbb{Z}$ mit $ggT(b,c) = xb + yc$.

Haben: $|a| \cdot ggT(b,c)$ teilt $|a|b$ und $|a|c$, also auch ba und ca, d.h. die r.S. in ⊗ ist ein gem. Teiler von ba und ca. Ist t irgendein solcher, so teilt t auch $\operatorname{sign}(a)(xba + yca) = xb|a| + yc|a| = |a|(xb + yc) = |a|\, ggT(b,c)$. Es folgt ⊗. □

Bew. der Eindeutigkeit von Satz 15.):

Eindeutigkeit: Sei $n > 1$ minimal mit zwei versch. 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 versch. Zerl.], ist $ggT(p_1, q_i) = 1$, und mit ⊠ folgt $p_1 \mid q_1^{f_1 - 1} \cdot \prod_{i=2}^{s} q_i^{f_i}$ aus Lemma 21.). Die Fortsetzung des Verfahrens zeigt schließlich $p_1 \mid q_s$, was wegen $ggT(p_1, q_s) = 1$ ein ↯ ist. □

(Beachten Sie: zum Bew. wurde nie die (eind.) PFZ von Zahlen benutzt!) □

Stichworte: Kongruenz $a \equiv b \,(m)$

Zahlring $\mathbb{Z}_m := \mathbb{Z}/m\mathbb{Z} = \{x + m\mathbb{Z} : x \in \mathbb{Z}\} = \{0 + m\mathbb{Z}, 1 + m\mathbb{Z}, \dots, (m-1) + m\mathbb{Z}\}$

Einheiten: $\mathbb{Z}_m^* = \{x + m\mathbb{Z} : x \in \mathbb{Z},\ \text{ex. } y \in \mathbb{Z} : (x + m\mathbb{Z})(y + m\mathbb{Z}) = 1 + m\mathbb{Z}\}$

$= \{x + m\mathbb{Z} : x \in \mathbb{Z},\ ggT(x, m) = 1\} \rightsquigarrow \varphi(m) := \#\mathbb{Z}_m^*$

Ist m klar, schreibe $\underline{x}$ für $x + m\mathbb{Z}$

$\mathbb{Z}_m^*$ heißt auch multiplikative Gruppe von $\mathbb{Z}_m$

Invertieren von Elementen in $\mathbb{Z}_m^*$ geht mit euklidischem Algorithmus

CRS: $\mathbb{Z}_{mn} \overset{\text{Ring-iso}}{\cong} \mathbb{Z}_m \times \mathbb{Z}_n$, falls $ggT(m,n) = 1$, Version des CRS mit simultanen Kongruenzen

Rechenbeispiele zum Kongruenzenrechnen und für eine "modulare Brille"

1.1.2 Kongruenzenrechnen und die "modulare Brille"

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

1.) **Def.:** Sei $m \in \mathbb{N}$. Dann heißen $a \in \mathbb{Z}$ und $b \in \mathbb{Z}$ kongruent modulo m, wenn $m \mid (b - a)$. Kurz: $a \equiv b \bmod m$ oder $a \equiv b \,(m)$.
Die Zahl m heißt Modul der Kongruenz.

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

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

(3) $a_1 \equiv b_1 \,(m)$ und $a_2 \equiv b_2 \,(m) \Rightarrow a_1 + a_2 \equiv b_1 + b_2 \,(m)$ und
$a_1 \cdot a_2 \equiv b_1 \cdot b_2 \,(m) \rightsquigarrow$ Kor.: $a_1^m \equiv b_1^m \,(m)$

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

(5) $a \equiv b \,(m_i)$ für $i = 1, \dots, k \Rightarrow a \equiv b \,(kgV(m_1, \dots, m_k))$

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

3.) **Def.:** Die Äquivalenzklassen von $\equiv$ modulo m heißen Restklassen modulo m.
(auch: Kongruenzklassen modulo m).

4.) **Folgerungen:** 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, $i = 1, \dots, m$, angewählt, können die m Restklassen mit $x_1 + m\mathbb{Z}, x_2 + m\mathbb{Z}, \dots, x_m + m\mathbb{Z}$ angegeben werden; die Menge $\{x_1, \dots, x_m\}$ heißt dann vollständiges Restsystem modulo m. Sind $y_1, \dots, y_m \in \mathbb{Z}$ so, dass $y_i \not\equiv y_j \bmod m$ für alle $i \neq j$, $1 \leq i, j \leq m$, gilt (d.h. sind die y_i paarweise inkongruent modulo m), dann ist $\{y_1, \dots, y_m\}$ ein vollständiges RS mod 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 in der Restklasse von x mod m genau alle zu x kongruenten Zahlen liegen laut Def.

5.) **Bsp.:** $\{0, 1, 2\}$ ist vollst. RS mod 3, und vollst. Restsysteme mod 8 sind etwa $\{1, \dots, 8\}$ und $\{3, 6, 9, 12, 15, 18, 21, 24\} = \{3 \cdot a;\ 1 \leq a \leq 8\}$ da $12 \equiv 4\ (8)$, $15 \equiv 7\ (8)$, $18 \equiv 2\ (8)$, $21 \equiv 5\ (8)$, $24 \equiv 0\ (8)$.
Die Menge $\{2a;\ 1 \leq a \leq 8\}$ ist kein vollst. RS mod 8. Die Reste $0, 1, 2, \dots, m-1$ könnte man auch als "Standardrepräsentanten" mod m bezeichnen, da $\{0, 1, 2, \dots, m-1\}$ immer vollst. RS mod m ist.

6.) **Folgerungen:** Ist $\{x_1, \dots, x_m\}$ ein vollst. RS mod m und $a \in \mathbb{Z}$, $c \in \mathbb{Z}$ mit $ggT(c, m) = 1$, so sind auch $\{x_1 + a, \dots, x_m + a\}$ und $\{x_1 \cdot c, \dots, x_m \cdot c\}$ vollst. RSe mod m (vgl. (4) aus Folgerung 2.)).

Das nützliche an den Restklassen modulo m ist, dass wir nun durch folgende naheliegende Definitionen von "+" und "·" mit ihnen neue algebraische Strukturen gewinnen können:

7) **Def.:** Ist der Modul $m \in \mathbb{N}$ klar, schreiben wir auch $\underline{x} := x + m\mathbb{Z}$ für die Restklasse von x mod m.

Wir definieren für $x, y \in \mathbb{Z}$ dann $\underline{x} + \underline{y} := \underline{x+y}$ und $\underline{x} \cdot \underline{y} := \underline{x \cdot y}$, d.h. $(x+m\mathbb{Z}) \dot{+} (y+m\mathbb{Z}) := (x \dot{+} y) + m\mathbb{Z}$. Dies erklärt "+", "·".

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.

8) **Folgerungen:**

Wir addieren/multiplizieren zwei Restklassen, indem wir Repräsentanten x, y auswählen und diese addieren/multiplizieren. Das ist nur sinnvoll, wenn bei unterschiedlicher Repräsentantenwahl dieselbe Restklasse als Ergebnis herauskommt, man sagt, die Def. von $\dot{+}$ bzw. $\cdot$ ist **wohldefiniert**, da repräsentantenunabh.

Dies ist klar: $\underline{x_1} = \underline{x_2}$ und $\underline{y_1} = \underline{y_2} \Rightarrow x_1 \equiv x_2\ (m)$ und $y_1 \equiv y_2\ (m)$

$\overset{\text{Folg. 2.1 (3)}}{\Rightarrow} x_1 \dot{+} y_1 \equiv x_2 \dot{+} y_2\ (m) \Rightarrow \underline{x_1 \dot{+} y_1} = \underline{x_2 \dot{+} y_2}$,

also erhalten wir so dieselbe Restklasse für $\underline{x_1} \dot{+} \underline{y_1}$ und $\underline{x_2} \dot{+} \underline{y_2}$, wenn $\underline{x_1} = \underline{x_2}$ und $\underline{y_1} = \underline{y_2}$.

Damit kann $(\mathbb{Z}_m, +)$ oder $(\mathbb{Z}_m \setminus \{0\}, \cdot)$ auf alg. Strukturen hin untersucht werden. (Bem.: Schreiben ab jetzt die neuen, blau markierten $+, \cdot$ schwarz)

9) **Folgerung:** $(\mathbb{Z}_m, +)$ ist eine abelsche Gruppe mit neutr. El. $\underline{0} = 0 + m\mathbb{Z}$, 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 $\underline{x} + \underline{-x} = \underline{x-x} = \underline{0}$, so dass $-\underline{x} = \underline{-x} = \underline{m-x}$ für alle $x \in \mathbb{Z}$ gilt. Ebenso gilt, dass $(\mathbb{Z}_m, +, \cdot)$ ein Kommutativer Ring mit $\underline{1}$ 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 bzgl. "·" (hier $\underline{2} \neq \underline{0}$) geben kann. Der Satz 10) gibt an, welche Restklassen invertierbar sind, d.h. im Ring $\mathbb{Z}_m$ eine Einheit sind:

10) **Satz:** Zu $\underline{x} \in \mathbb{Z}_m$ ex. genau dann ein multiplikatives Inverses, d.h. ein $\underline{y} \in \mathbb{Z}_m$ mit $\underline{x} \cdot \underline{y} = \underline{1} \Leftrightarrow x \cdot y \equiv 1\ (m)$, falls $\underline{ggT(x,m) = 1}$.

Wir schreiben dann $\underline{x}^{-1}$ oder $\underline{x}^*$ für $\underline{y}$, die Bezeichnungen $\frac{\hat{1}}{\underline{x}}$ oder $\frac{1}{\underline{x}}$ oder $1/\underline{x}$ sind didaktisch ungeschickt.

Bew.: „$\Rightarrow$": Sei $\underline{x} \in \mathbb{Z}_m$ mit $\underline{x}\cdot\underline{y} = \underline{1}$, d.h. $x \cdot y \equiv 1 \,(m)$, also ex. $k \in \mathbb{Z}$ mit $1 - xy = km \Rightarrow xy + km = 1$. Wäre $d = ggT(x,m) > 1$, folgt $d \mid xy + km = 1$ ↯.

„$\Leftarrow$": Sei $ggT(x,m) = 1$. Nach V2-Satz 20), dem Satz vom eukl. Algorithmus, ex. $y, k \in \mathbb{Z}$ mit $1 = yx + km$, also folgt $\underline{x}\cdot\underline{y} = \underline{1}$. □

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

Bsp.: Gesucht: 7^{-1} mod 37, haben: $37 = 5\cdot 7 + 2$, $7 = 3\cdot 2 + 1$, $2 = 2\cdot 1 \rightsquigarrow$

q_a		1	5	3	2
c_a	0	1	5	16	37

$\rightsquigarrow +16 = 7^{-1}\,(37)$

11.) **Def.:** $\underline{x} = x + m\mathbb{Z}$ heißt prime oder reduzierte Restklasse mod m, falls $ggT(x,m) = 1$ gilt. Diese sind genau die Einheiten in $(\mathbb{Z}_m, +, \cdot)$, d.h. $\mathbb{Z}_m^* = \{\underline{x} \in \mathbb{Z}_m;\ ggT(x,m) = 1\}$.

Die Anzahl der Einheiten sei $\varphi(m) := \#\mathbb{Z}_m^* = \#\{a \in \mathbb{N};\ a \le m,\ ggT(a,m) = 1\}$, die so erklärte Fkt. $\varphi : \mathbb{N} \to \mathbb{N}$ heißt Eulersche φ-Funktion.

Jedes Repräsentantensystem $\{x_1, \dots, x_{\varphi(m)}\}$ von $\mathbb{Z}_m^*$ heißt reduziertes oder primes Restsystem modulo m.

12.) **Satz:** 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 $ggT(m,n) = 1$ (d.h. φ ist „multiplikativ").

Bew.: Unter den Zahlen $1, 2, \dots, p^k$ sind genau die Vielfachen von p zu p^k nicht teilerfremd, d.h. $p, 2p, \dots, p^{k-1}\cdot p$, was p^{k-1} viele Zahlen sind.

Den Bew. der Multiplikativität von φ verschieben wir auf später ($\rightsquigarrow$ 17.)). □

Ist $n = \prod_{p \mid n} p^{e(p)}$ die PFZ von n, folgt aus Satz 12.):

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

13.) **Folgerungen:** $(\mathbb{Z}_m^*, \cdot)$ 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 wenn $\mathbb{Z}_m^* = \mathbb{Z}_m \setminus \{0\}$ ist, ist $(\mathbb{Z}_m, +, \cdot)$ ein Körper; dies ist genau dann richtig, wenn $m = p$ eine Primzahl ist, weil genau dann alle $1, 2, \dots, m-1$ zu m teilerfremd sind. Wir bezeichnen für p prim diesen Körper mit p Elementen mit $\mathbb{F}_p$ (weitere endliche Körper folgen noch).

Der Körper $\mathbb{F}_p$ hat die Eigenschaft, dass $p \cdot a := \overbrace{a+\ldots+a}^{p\text{-mal}} = 0$ in $\mathbb{F}_p$ für alle $a \in \mathbb{F}_p$ gilt. Wir sagen, er hat die Charakteristik p.

14.) **Def.:** Sei k ein Körper. Er hat die **Charakteristik** 0, falls für alle $m \in \mathbb{N}$ gilt: $m \cdot 1 := \underbrace{1+\ldots+1}_{m\text{-mal}} \neq 0$ (z.B. $\mathbb{Q}, \mathbb{R}, \mathbb{C}$).
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, Kurz: $\mathrm{char}(k)$. Bsp.: $\mathrm{char}(\mathbb{Q}) = 0$, $\mathrm{char}(\mathbb{F}_p) = p$.
Bem.: Stets gilt: $\mathrm{char}(k) = 0$ oder $\mathrm{char}(k)$ eine PZ. [Sonst: $0 = (m \cdot n) \cdot 1 = m \cdot (n \cdot 1) = (m \cdot 1) \cdot (n \cdot 1)$ $\Rightarrow m \cdot 1 = 0 \vee n \cdot 1 = 0$, da $k^* = k \setminus \{0\}$ im Bzw. Minimalität von $m \cdot n$.]

Die Struktur der Zahlringe $(\mathbb{Z}_m, +, \cdot)$ versteht man besser, indem man sie auf "kleinere" Zahlringe zurückführt:

15.) **Chinesischer Restsatz** (Zahlring-Version):
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 $F: \mathbb{Z}/m\mathbb{Z} \to (\mathbb{Z}/m_1\mathbb{Z}) \times \ldots \times (\mathbb{Z}/m_r\mathbb{Z})$, $x + m\mathbb{Z} \mapsto (x + m_1\mathbb{Z}, \ldots, x + m_r\mathbb{Z})$ ein **Ringisomorphismus**, d.h. ein bijektiver Ringhomomorphismus.

16.) **Chinesischer Restsatz** (Simultane Kongruenzen-Version):
Seien $m_1, \ldots, m_r > 1$ paarweise teilerfremde Zahlen, und seien $a_1, \ldots, a_r \in \mathbb{Z}$. Dann ist das simultane Kongruenzensystem
$x \equiv a_1 \bmod m_1$, $x \equiv a_2 \bmod m_2, \ldots, x \equiv a_r \bmod m_r$
in x lösbar, die Lösungen sind alle kongruent modulo $m_1 \cdots m_r$.
Bem.: Aus Version 15.) folgt Version 16.) wegen der Bijektivität von F, denn $(a_1 + m_1\mathbb{Z}, \ldots, a_r + m_r\mathbb{Z})$ hat dann genau ein Urbild $x + m\mathbb{Z}$.

17.) **Zusatz zum CRS** (= chinesischer Restsatz) in Variante 16.):
Genau alle $x \equiv x_0 \bmod m_1 \cdots m_r$ lösen das System, wobei $x_0 = a_1 M_1^* M_1 + \ldots + a_r M_r^* M_r$, $M_i := \frac{m_1 \cdots m_r}{m_i}$ $(i = 1, \ldots, r)$ und $M_i^* \in \mathbb{Z}$ ein multiplikatives Inverses von M_i mod m_i repräsentiert $(i = 1, \ldots, r)$, 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.

18.) Ergänzt zum CRS in Variante 15.): Die Gruppe $\mathbb{Z}_m^*$ ist isom. zu $\mathbb{Z}_{m_1}^* \times \dots \times \mathbb{Z}_{m_r}^*$, beide Gruppen haben dann gleichviele Elemente, es folgt

$\varphi(m) = \varphi(m_1) \cdot \varphi(m_2) \cdots \varphi(m_r)$, speziell $r=2$: $\varphi(m_1 m_2) = \varphi(m_1)\varphi(m_2)$,

d.h. die Multiplikativität von φ ist ein Korollar des CRS.

Beweis des CRS: Wir beweisen Variante 16.), der von Variante 15.) ist im wesentlichen gleich.

Existenz der Lsg.: Ist $x \equiv x_0 \bmod m_1 \cdots m_r$ wie in 17.) angegeben so folgt für alle $1 \le i \le r$: $x \equiv x_0 = \underbrace{a_1 M_1^* M_1}_{\equiv 0 (m_i)} + \dots + \underbrace{a_i M_i^* M_i}_{\equiv a_i \cdot 1 \equiv a_i (m_i)} + \dots + \underbrace{a_r M_r^* M_r}_{\equiv 0 (m_i)} \equiv a_i (m_i)$ ✓

Eindeutigkeit der Lsg. mod $m_1 \cdots m_r$: Ist $y \in \mathbb{Z}$ weitere Lösung des Kongruenzensystems, so gilt $\forall j \neq i$: $y \equiv a_j (m_j)$, also $\underbrace{M_j^* \cdot M_j}_{\equiv 1 (m_j)} \cdot y \equiv a_j (m_j)$, und $M_i \cdot M_i^* \cdot a_i \equiv 0 (m_j)$, (↑ durch m_j teilbar)

und somit

$$y \equiv a_j (m_j) \equiv \sum_{j=1}^{r} M_j M_j^* a_j (m_j) \equiv x_0 (m_j) \text{ für alle } j = 1, \dots, r.$$ [→ $m_j \mid y - x_0$ gemeins. Vielf.]

Da die $m_1, \dots, m_r$ alle paarweise teilerfremd sind, folgt daraus $y \equiv x_0 (\underbrace{m_1 \cdots m_r}_{= \operatorname{kgV}(m_1, \dots, m_r)})$, vgl. Folg. 2.) Nr. (5). □

Bsp. zum CRS: Das System $x \equiv 2 (7)$, $x \equiv 3 (8)$ hat die Lösung $x \equiv 2 \cdot \underbrace{1}_{\text{Inv. von } 8 \bmod 7} \cdot 8 + 3 \cdot \underbrace{(-1)}_{\text{Inv. von } 7 \bmod 8} \cdot 7 = 16 - 21 = -5 \equiv 51 \bmod 56$.

Also: $\left\{ \begin{array}{l} x \equiv 2 (7) \\ \wedge\ x \equiv 3 (8) \end{array} \right\} \Leftrightarrow x \equiv 51 (56)$.

Bsp. dazu: Gegeben seien 2 Tüten mit gleichvielen Bonbons.

$\left\{ \begin{array}{l} \text{Verteilen 1 Tüte Bonbons gleichmäßig an 7 Kinder} \rightsquigarrow 2 \text{ übrige Bonbons} \\ \& \quad " \quad " \quad " \quad " \quad " \quad 8 \quad " \quad \rightsquigarrow 3 \quad " \quad " \end{array} \right\}$

Dann waren 51, 107, ... viele Bonbons in jeder Tüte. Können es nicht mehr als 100 sein, waren es also 51 Stück, und man kann ausrechnen, wieviele Bonbons jedes Kind erhalten hat (7 bzw. 6).

Ein paar Beispiele zum Rechnen mit Kongruenzen bzw. Restklassen:

- **Bsp.:** $5x \equiv 4\ (12) \iff 5^{-1} \cdot 5x \equiv 4 \cdot 5^{-1}\ (12)$
  $\iff x \equiv 4 \cdot 5^{-1} \equiv 4 \cdot 5 = 20 \equiv 8\ (12)$

  [$5 \cdot 5 = 25 \equiv 1\ (12) \Rightarrow 5^{-1} \equiv 5\ (12)$]

Ebenso:
Rechnen mit Restklassen mod 12:

$$\underline{5} \cdot x = \underline{4} \quad | \cdot \underline{5}^{-1} \iff x = \underline{4} \cdot \underline{5}^{-1} = \underline{4} \cdot \underline{5} = \underline{20} = \underline{8}$$

- **Bsp.:** $8x^2 - 2x + 3 \equiv -1\ (7)$
  $\iff (x-1)^2 - 1 + 3 \equiv -1\ (7) \iff (x-1)^2 \equiv -3 \equiv 4\ (7)$.

sprechen auch von „quadratischen Resten"

Da nun wegen

z	0	± 1	± 2	± 3
z^2	0	1	4	2

Die Kongruenz $(x-1)^2 \equiv 4\ (7)$ hat die 2 Lösungen $x \equiv 3\ (7)$, $x \equiv -1\ (7)$.

Die Kongruenz $(x-1)^2 \equiv 5\ (7)$ hätte keine Lösung, da 5 kein Quadrat mod 5 ist.

- **Bsp.:** Die Kongruenz $(x-3) \cdot 4 \equiv 1\ (33)$ ist als Kongruenzensystem $(x-0) \cdot 1 \equiv 1\ (3) \wedge (x-3) \cdot 4 \equiv 1\ (11)$ schreibbar.
  Man kann beide Kongruenzen einzeln lösen, also
  1.) $x \equiv 1\ (3)$ sowie 2.) $(x-3) \equiv 4^{-1} \equiv 3\ (11) \iff x \equiv 6\ (11)$,
  und wieder mit dem CRS zusammensetzen mod 33:

$$x \equiv 1 \cdot \underbrace{2}_{\text{inv. von } 11 \bmod 3} \cdot 11 + 6 \cdot \underbrace{4}_{\text{inv. von } 3 \bmod 11} \cdot 3 = 22 + 6 \cdot 12 = 94 \equiv -5 \equiv 28\ (33).$$

- **Bsp.:** 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 Bsp., wo wir die modulare Brille mod 8 aufziehen, um mehr zu sehen:
  Betr. die Glg. $8x + 7 = u^2 + v^2 + w^2$ in $u, v, w, x \in \mathbb{N}_0$.
  Sie ist unlösbar: Denn mod 8 erhalten wir $7 \equiv u^2 + v^2 + w^2\ (8)$;

z	0	± 1	± 2	± 3	4
z^2	0	1	4	1	0

  alle quadratischen Reste mod 8 sind $0, 1, 4$,
  daher ist $v^2 + w^2 \equiv 0, 1, 4, 2, 5\ (8)$,
  also $u^2 + v^2 + w^2 \equiv 0, 1, 4, 2, 5,\ 1, 2, 5, 3, 6,\ 4, 5, 0, 6, 1\ (8)$,
  d.h. $u^2 + v^2 + w^2 \equiv 0, 1, 2, 3, 4, 5, 6\ (8)$, aber nie $\equiv 7\ (8)$.
  Es kann keine Lösungen mod 8 geben, also auch keine in $\mathbb{Z}$.

Stichworte: $\mathrm{ord}(G)$, UG, Satz von Lagrange, $k \cdot a$ in $(G,+)$ und a^k in $(G,\cdot)$, $\langle a\rangle = \{k \cdot a;\ k \in \mathbb{Z}\}$ in $(G,+)$, $\langle a\rangle = \{a^k;\ k \in \mathbb{Z}\}$ in $(G,\cdot)$, $\mathrm{ord}(a) = \#\langle a\rangle$, $a^{\mathrm{ord}(G)} = 1$ in $(G,\cdot)$ ⇝ Euler-Fermat, kleiner Fermat, schnelles Potenzieren, Lösen quadratischer Kongruenzen, faires Münzwurf-Knobeln am Telefon

---

1.1.3 Gruppen

Die Gruppen $(\mathbb{Z}_m, +, 0)$ und $(\mathbb{Z}_m^*, \cdot, 1)$ sind endliche abelsche Gruppen. Wir untersuchen ein paar ihrer allgemeinen Eigenschaften und führen dabei ein paar Grundbegriffe ein.

1.) **Def.:** Die Ordnung einer endlichen Gruppe G ist die Anzahl ihrer Elemente, Kurz: $\mathrm{ord}(G) := \#G$.

2.) **Def.:** Eine Teilmenge H einer Gruppe G mit Verknüpfung $*$ heißt Untergruppe, falls auch $(H, *)$ eine Gruppe ist. Kurz: UG.

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

Bew.: Die Linksnebenklassen $a*H := \{a*h;\ h \in H\}$ für $a \in G$ sind paarweise disjunkt, d.h. stets gilt $a*H = b*H$ oder $a*H \cap b*H = \emptyset$.
(Denn: ist $c \in a*H \cap b*H$, ist $c = a*g = b*h$ für $g, h \in H$, also $a = b*(h*g^{-1})$, somit $a*H = \{a*m;\ m \in H\} = \{b*h*g^{-1}*m;\ m \in H\} = \{b*m;\ m \in H\} = b*H$.)
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 Beh. □

4.) **Def.:** Sei $(G, +)$ eine abelsche Gruppe und $a \in G$. Für $k \in \mathbb{Z}$ definieren wir $ka := a + \ldots + a$ (k-mal), falls $k > 0$, $k \cdot 0 := 0$ und $k \cdot a := -(-k)a$ falls $k < 0$.
Dann ist $\langle a\rangle := \{ka;\ k \in \mathbb{Z}\}$ eine UG von G. [Klar!]
Wir nennen $\langle a\rangle$ die von a erzeugte UG bzw. Erzeugnis von a und a einen Erzeuger. Ist $\langle a\rangle$ endliche UG, heißt ihre Ordnung die Ordnung von a, Kurz: $\mathrm{ord}(a) := \#\langle a\rangle$.
Eine Gruppe G mit Erzeuger a, d.h. $G = \langle a\rangle$, heißt zyklisch.

Schreibt man die Gruppe multiplikativ mit Verknüpfung "$\cdot$" ("mal"), so setzt man $a^k := \overbrace{a \cdots a}^{k\text{-mal}}$ falls $k>0$, $a^0 := 1$, $a^k := (a^{(-k)})^{-1}$ 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 endl. Gruppe G und $a \in G$ stets $\underline{\operatorname{ord}(a) \mid \operatorname{ord}(G)}$.

5.) Bsp.: $\mathbb{Z}/m\mathbb{Z} = \{0, 1, \dots, m-1\}$ ist "die" zyklische Gruppe G mit $\operatorname{ord}(G) = m$. Ist $m = p$ prim, können außer $\{0\}$ und $\mathbb{Z}/p\mathbb{Z}$ keine weiteren UG ex.

6.) Lemma: Sei $(G,+)$ Gruppe, $a \in G$. Es ist $\operatorname{ord}(a)$ die kleinste natürliche Zahl m mit $ma = 0$. Es gilt: $ka = 0 \Leftrightarrow \operatorname{ord}(a) \mid k$.
(Bei multiplikativer Schreibweise: $\operatorname{ord}(a) = \min\{m \in \mathbb{N} ; a^m = 1\}$ und $a^k = 1 \Leftrightarrow \operatorname{ord}(a) \mid k$.)
Bew.: Erster Teil klar, zweiter Teil: "$\Rightarrow$": Falls $k \in \mathbb{N}$ mit $ka = 0$ ist, nehme Division von k durch $\operatorname{ord}(a)$ vor: $k = q \cdot \operatorname{ord}(a) + r$ mit $0 \le r < \operatorname{ord}(a)$. Wegen $0 = ka = q \cdot \underbrace{\operatorname{ord}(a) \cdot a}_{=0} + ra$ folgt $ra = 0$, wegen der Minimalität von $\operatorname{ord}(a)$ also $r = 0$, also $\operatorname{ord}(a) \mid k$.
"$\Leftarrow$": Für $k = m \cdot \operatorname{ord}(a)$ folgt $ka = m \cdot (\underbrace{\operatorname{ord}(a) \cdot a}_{=0}) = 0$. □

7.) Folgerung: $\operatorname{ord}(G) \cdot a = 0$ bzw. multiplikativ: $a^{\operatorname{ord}(G)} = 1$ (da $\operatorname{ord}(a) \mid \operatorname{ord}(G)$ nach Lemma 6.)

8.) Folgerung: Da $\operatorname{ord}((\mathbb{Z}/m\mathbb{Z})^*) = \varphi(m)$, ist $a^{\varphi(m)} \equiv 1 \bmod m$, falls $\underline{\operatorname{ggT}(a,m) = 1}$.
Für p prim: $a^{p-1} \equiv 1 \bmod p$ für $p \nmid a$.
(Korollar aus 7.) "Kleiner Satz von Fermat"

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

10.) Bsp.: Haben $\varphi(15) = \varphi(3 \cdot 5) = \varphi(3) \cdot \varphi(5) = 2 \cdot 4 = 8$. Die möglichen Ordnungen von Zahlen a mod 15, wo $\operatorname{ggT}(a,15) = 1$ ist, sind also 1, 2, 4, 8.
Wegen $4^2 = 16 \equiv 1 \ (15)$ ist z.B. $\operatorname{ord}_{15}(4) = 2$. Bei anderen Zahlen muss man u.U. 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 mod $\varphi(m)$:
Ist $k = q \cdot \varphi(m) + r$ mit $0 \le r < \varphi(m)$, folgt $a^k = a^{\varphi(m)\cdot q + r} = (a^{\varphi(m)})^q \cdot a^r \equiv 1^q \cdot a^r = a^r \bmod m$. Ist aber auch $\varphi(m)$ bzw. r groß, hilft man sich mit folgender Methode des schnellen Potenzierens weiter:

11.) Geg. sei eine Gruppe $(G, \cdot)$, zu berechnen ist für $r \in \mathbb{N}$, $a \in G$ die Potenz $a^r := \underbrace{a \cdots a}_{r\text{-mal}}$ in der Gruppe G.

1. Schritt: Mit höchstens $d := \lfloor \frac{\log r}{\log 2} \rfloor$ vielen Verknüpfungen in G berechne durch sukzessives Quadrieren: a^2, $a^{2^2} = a^4 = (a^2)\cdot(a^2)$, $a^{2^3} = (a^{2^2})\cdot(a^{2^2})$, $a^{2^4} = (a^{2^3})\cdot(a^{2^3})$, …, a^{2^d}

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

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

Somit reichen höchstens $2d = O(\log r)$ 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.

12.) Bsp.: $5^{12} = 5^{2^2+2^3} = 5^{2^2} \cdot 5^{2^3}$, modulo 11 rechnen wir: $5^2 \equiv 3\ (11)$, $5^{2^2} \equiv 3^2 \equiv -2\ (11)$, $5^{2^3} \equiv (-2)^2 \equiv 4\ (11)$,
also $5^{12} \equiv (-2)\cdot 4 \equiv 3\ (11)$; geht schneller als $5^{12} = 244140625$ von Hand durch 11 zu teilen bzw. das $\cdot$ auszurechnen… ($\not\equiv 5^{2^3} = 5^{(2^3)} = 5^8 \not\equiv (5^2)^3 = 5^{2\cdot 3} = 5^6$)

Eine Anwendung des Kleinen Fermats

13.) Im Fall $p \equiv 3\ (4)$ prim können wir Lösungen quadratischer Kongruenzen mod p bestimmen: Sei $p = 4k+3$ prim und a mit $p \nmid a$ ein quadratischer Rest mod p, d.h. es ex. ein $b \in \mathbb{Z}$ mit $a \equiv b^2\ (p)$, und wir möchten $\pm b$ mod p ausrechnen können. Nach dem Kleinen Fermat folgt $b^{4k+2} = b^{p-1} \equiv 1\ (p)$.
Es folgt: $(a^{k+1})^2 \equiv (b^2)^{2(k+1)} = b^{(4k+2)+2} \equiv 1 \cdot b^2 \equiv a\ (p)$,
d.h. die Lösungen von $b^2 \equiv a\ (p)$ sind $b = \pm a^{k+1}\ (p)$.
Da $a^{k+1} \not\equiv -a^{k+1}\ (p) \Leftrightarrow 2a^{k+1} \not\equiv 0\ (p)$, gibt es genau 2 Lösungen mod p, die wir etwa im Restsystem $\{0, 1, \dots, p-1\}$ angeben können und mit $\pm a^{k+1}$ mod p berechnen können, z.B. mit dem schnellen Potenzieren.

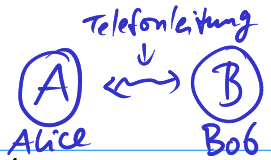
14.) Sei nun n eine zus.gesetzte Zahl, etwa $n = p \cdot q$ mit $p \equiv q \equiv 3 \,(4)$ prim, etwa $p = 4k+3$, $q = 4\ell+3$ mit $k, \ell \in \mathbb{N}_0$, und sei $p \neq q$.
Sei a mod n ein quadratischer Rest mod n, d.h. es existiere ein $b \in \mathbb{Z}$ mit $a \equiv b^2 \,(n)$.
Gesucht seien die Lösungen der Kongruenz $a \equiv x^2 \,(n)$.
Nach dem CRS gilt: $x^2 \equiv a \,(n) \iff x^2 \equiv a \,(p)$ und $x^2 \equiv a \,(q)$, und die jeweiligen Lösungen $\pm a^{k+1}$ mod p und $\pm a^{\ell+1}$ mod q kann man zusammensetzen zu (maximal) vier Lösungen mod n. Es sind genau 4 Lösungen, die explizit wie folgt bestimmt werden können:
Sind $r, s \in \mathbb{Z}$ geg. mit $rp + sq = 1$, d.h. die Bézout-Elemente von p und q, und ist $\pm b$ Lsg. von $x^2 \equiv a \,(p)$ [2 Mögl.]
$\pm c$ Lsg. von $x^2 \equiv a \,(q)$ [2 Mögl.],
so liefert die CRS-Formel (Inv. von q mod p, Inv. von p mod q)

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

genau vier Lösungen von $x^2 \equiv a \,(p \cdot q)$. Diese müssen paarweise inkongruent mod pq sein, da wir laut CRS den Ringiso $\mathbb{Z}_{pq} \cong \mathbb{Z}_p \times \mathbb{Z}_q$ haben und die 4 versch. Lösungspaare (b, c), $(-b, c)$, $(b, -c)$, $(-b, -c)$ deswegen genau 4 Restklassen in $\mathbb{Z}_{pq}$ entsprechen.

15.) Bsp.: Betr. $p = 11$, $q = 19$, d.h. $k = 2$, $\ell = 4$. Wähle $a = 47$.
Die Lösungen von $x^2 \equiv 47 \equiv 3 \,(11)$ sind $\pm 3^3 \,(11) \equiv \pm 5 \,(11)$,
die Lösungen von $x^2 \equiv 47 \equiv 9 \,(19)$ sind $\pm 3 \,(19)$.
Bézout-El. bestimmen (hier Probieren): Inv. von $19 \equiv 8$ mod 11 ist 7, Inv. von 11 mod 19 ist 7.
$\rightsquigarrow s = r = 7$ und $x \equiv \mp 5 \cdot 7 \cdot 19 \mp 3 \cdot 7 \cdot 11 \;(11 \cdot 19)$
ergibt $x \in \{\pm 16, \pm 60\}$. Probe: $16^2 \equiv 47 \,(11 \cdot 19)$ ✓, $60^2 \equiv 47 \,(11 \cdot 19)$ ✓

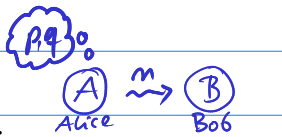
Man beachte, dass wir hier benötigen, dass a ein quadratischer Rest mod 11 und mod 19 sein muss. Würde man a zufällig wählen, wäre das nicht unbedingt der Fall; dann ist $x^2 \equiv a \,(n)$ ohnehin unlösbar, falls a kein quadratischer Rest mod $11 \cdot 19$ ist.
Wir besprechen nun eine Anwendung.



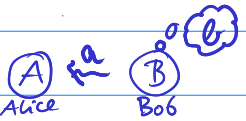
16.) Problem des fairen Münzwurfknobelns am Telefon:
Zwei Spieler, Alice (A) und Bob (B) möchten etwas ausknobeln (z.B. wer beim Fernschach beginnen soll und dann einen Vorteil hat, etc.), allerdings sprechen sie sich am Telefon oder mailen sich, und können sich daher nicht sehen. (A) wirft eine Münze, und (B) denkt vorher "Kopf" oder "Zahl", verrät das aber nicht. (*) (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) ihre 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!

(*): (Wäre (B) live dabei, würde er "Kopf" oder "Zahl" sagen und das Ergebnis sehen. Am Telefon gilt: Würde (A) seine Wahl vorher kennen, so würde (B) ihr mitgeteiltes Münzwurfergebnis u.U. anzweifeln.)

17.) Das Verfahren funktioniert wie folgt:

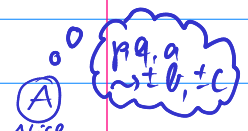


Schritt (1.) (A) wählt Primzahlen $p,q \equiv 3 (4)$, $p \neq q$, berechnet $n=p \cdot q$ und schickt n an (B)



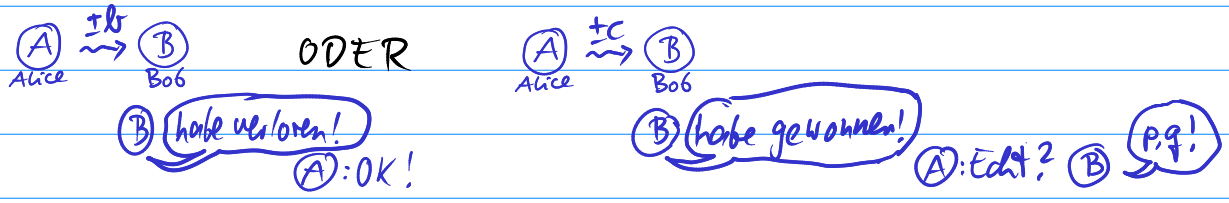
Schritt (2.) (B) wählt $1 \leq b \leq n-1$ zufällig und behält b geheim, er berechnet $a \equiv b^2 (n)$, und schickt a an (A)

Schritt (3.) (A) berechnet die 4 Lösungen von $x^2 \equiv a (n)$ mit der Berechnungsmethode aus 14.), die 4 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.) Ⓐ wählt eine der 4 Lösungen zufällig aus (etwa durch Münzwurf!), d.h. entweder $\pm b$ oder $\pm c$, und schickt Ⓑ das Ergebnis. Ⓐ kann nicht wissen, dass Ⓑ die Zahl b gewählt hat. Die Vereinbarung ist nun: Schickt Ⓐ eine der Zahlen $\pm b$, gewinnt Ⓐ,
schickt Ⓐ eine der Zahlen $\pm c$, gewinnt Ⓑ, und das verkündet Ⓑ.



Schritt (5.) Es erfolgt die Verifikation, dass Ⓐ wirklich verloren hat im 2. Fall, dann muss sich Ⓐ davon überzeugen, dass Ⓑ vorher wirklich $\pm b$ gewählt hat: Er kann Ⓐ die Lösungen $\pm b$ einfach mitteilen, da Ⓐ auch diese berechnet hat. Alternativ kann Ⓑ ihr sogar die Primfaktoren von n nennen:

Er berechnet $b+c \bmod n$ und

$d = ggT(b+c, n)$ mit dem euklidischen Algo.

Dann ist $d=p$ oder $d=q$. ⌜Denn aus $b^2 \equiv a \equiv c^2 \ (pq)$ folgt:
$pq \mid (b-c)(b+c) = b^2 - c^2$, und da $b \not\equiv c \ (p)$, $b \not\equiv c \ (q)$ folgt $q \mid b+c$ oder $p \mid b+c$,
und $d \neq n$, weil sonst $b \equiv -c \bmod n$ wäre ↯.⌟

Also kann Ⓑ, weil er c kennt, die von Ⓐ gewählten Primfaktoren bestimmen und Ⓐ mitteilen und auf diese Art Ⓐ überzeugen.

Das konnte Ⓑ nur, weil er vorher auch wirklich die nicht von Ⓐ genannte Lösung $\pm b$ hatte.

Damit ist das Spiel fair.

P.S.: 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 ev. hunderten von Stellen in Dezimaldarstellung) zusammengesetzt ist oder (sehr wahrscheinlich) prim. Daher erzeugt man solange Zufallszahlen in der gewünschten Größe, bis der Primzahltest "anschlägt".

Stichworte:
Public-Key-Kryptographie, geheime und öffentliche Schlüssel,
RSA, Kodierung von Textnachrichten,
Diskreter Logarithmus-Problem (DL-Problem),
Diffie-Hellman-Schlüsselaustausch,
Diffie-Hellman-Problem (DH-Problem),
Bsp. für Man-in-the-Middle-Attacke

---

§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 Nutzer geheim hält: Jeder kann verschlüsseln, aber nur der rechtmäßige Empfänger entschlüsseln. Möchte Nutzer (B) eine Nachricht an Nutzer (A) sendet, benutzt er zur Verschlüsselung den öffentlichen Schlüssel von (A), die Entschlüsselung gelingt aber nur (A) mit dem privaten Schlüssel.
Kerkhoffs Prinzip: Die Sicherheit eines Kryptosystems darf nicht von der Geheimhaltung des Algorithmus abhängen, sondern von der Geheimhaltung der geheimen Schlüssel.

Ein solches Szenario (auch "Protokoll" genannt) ist das RSA-Verfahren, das wir in 1.2.1 behandeln. Die Verfahren in 1.2.2 und 1.2.3 sind Kryptographie-Verfahren, die mit allgemeinen Gruppen machbar sind.

1.2.1 RSA-Verfahren

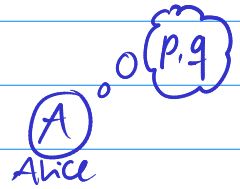
(RSA arbeitet mit $(\mathbb{Z}_m^*, \cdot)$)

Das RSA-Verfahren ist benannt nach einer Arbeit von R.L. Rivest, A. Shamir und L.M. Adleman 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 (V4_16).
$n = p \cdot q$ muss so groß sein, dass alle bekannten Faktorisierungsverfahren zu langsam wären.

1.) Wir beschreiben den Verlauf des Verfahrens:

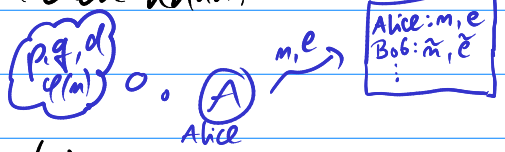
Die beiden Protagonisten heißen wieder Nutzer (A)lice und (B)ob.
Sie kommunizieren über einen unsicheren Kanal miteinander.

Schritt (1.) (Vorbereitung)

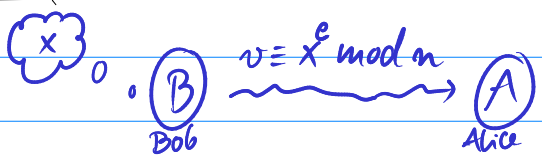


Jeder Nutzer, z.B. (A), wählt zwei große Primzahlen $p \neq q$, etwa gleichgroß mit ähnlicher Stellenanzahl, und berechnet $n = pq$ sowie $\varphi(n) = (p-1)(q-1)$.

Dann wählt (A) eine Zahl e mit $1 < e < \varphi(n)$ und berechnet $1 < d < \varphi(n)$ als Inverses von e mod $\varphi(n)$, d.h. $de \equiv 1 \bmod \varphi(n)$, unter Zuhilfenahme des euklidischen Algorithmus.
(A) hält $p, q, \varphi(n), d$ geheim und gibt n, e bekannt, z.B. durch Hinterlegung auf einen öffentlichen Schlüsselserver, wo jeder nachsehen kann.

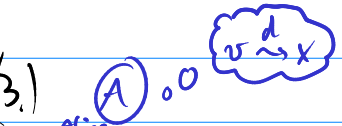


Schritt (2.)



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 $v \equiv x^e (n)$ 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

$$v^d \equiv (x^e)^d \equiv x^{1 + k \cdot \varphi(n)} \equiv x \cdot \underbrace{\left(x^{\varphi(n)}\right)^k}_{\equiv 1 \,(n) \text{ nach Euler-Fermat, falls } ggT(x,n) = 1} \equiv x \quad (n).$$

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

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

4.) Bem.: 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 Bundesnetzagentur bis Ende 2020: mind. 1976 Bit. Gegen einen Angriff mit dem Quantencomputer hätte man allerdings keine Chance.

5.) Bem.: Auch in den seltenen Fällen p|x oder q|x, d.h. ggT(x, n) > 1, arbeitet das Verfahren korrekt (ohne Beweis).

6.) Bem.: 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.

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...).

s. wikipedia "Crypto Wars"

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 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). Zimmermann stellte sein Programm 1991 kostenlos zur Verfügung. Es wurde ein Verfahren gegen ihn eröffnet, das sich über 3 Jahre lang hinzog. Vorwurf: er exportiere 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.

8.) Kodierung von Textnachrichten: Wir beschreiben hier ein Verfahren, das die Machbarkeit der Kodierung Text → 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,...,25 identifiziert, das Leerzeichen mit 26. Klartexte werden zu Blöcken aus je drei Zahlen zusammengefasst, also z.B. KLARTEXT_ → 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: KLARTEXT_ → 7587 / 12910 / 17306,
welche beim RSA-Verfahren gemäß $x^e \equiv v \ (n)$ 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 $v_1, v_2, v_3 \in \{0,\dots,28\}$,
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) → für größere n werden größere Blöcke nötig!

1.2.2 Diffie-Hellman-Verfahren

9.) Das Problem des diskreten Logarithmus (DL-Problem):
Geg. sei eine abelsche Gruppe, wir beschreiben das Problem multiplikativ und additiv:

In $(G,\cdot,1)$:	In $(G,+,0)$:
Sei $x \in G$, $n = \mathrm{ord}(x)$, $y \in \langle x\rangle = \{x^l ; l \in \mathbb{Z}\}$.	Sei $x \in G$, $n = \mathrm{ord}(x)$, $y \in \langle x\rangle = \{l \cdot x ; l \in \mathbb{Z}\}$.
Bestimme $k \bmod n$ mit $y = x^k$.	Bestimme $k \bmod n$ mit $y = k \cdot x$.
("diskreter Logarithmus")	("diskreter Logarithmus")

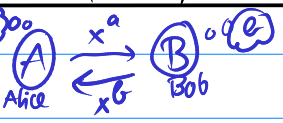
10.) Ist eine Gruppe G gegeben, in der das DL-Problem schwer ist, kann dies für ein Kryptoverfahren genutzt werden.

- Im Fall $G = (\mathbb{Z}_m^*, \cdot, 1)$ ist das DL-Problem ähnlich schwer wie das Faktorisierungsproblem. Auch dafür konnte P. Shor 1994 zeigen, dass es auf einem Quantencomputer schnell lösbar ist.
- Im Fall, dass $G = (E(\mathbb{K}), +, \mathcal{O})$ 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.

11.) **Der Diffie-Hellman-Schlüsselaustausch**

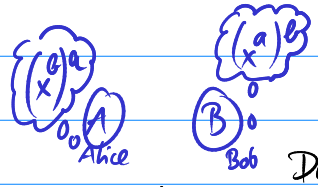
Hier vereinbaren Alice und Bob durch einen öffentlichen Kanal einen gemeinsamen geheimen Schlüssel, die sie dann für ein symmetrisches Kryptoverfahren nutzen können. Geg. sei eine Gruppe G und $x \in G$, sowie $n \in \mathbb{N}$. Diese Daten seien öffentlich bekannt.

In $(G, \cdot, 1)$:

(1.) 

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

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

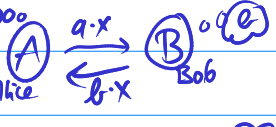
(2.) 

Alice berechnet mit a das Gruppenelement $(x^b)^a$

Bob berechnet mit b das Gruppenelement $(x^a)^b$

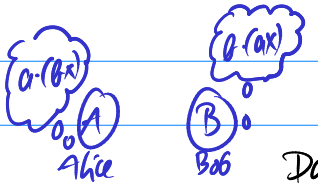
Danach besitzen beide den gemeinsamen geheimen Schlüssel $(x^b)^a = x^{ab} = x^{ba}$.

In $(G, +, 0)$:

(1.) 

Alice denkt sich eine Zahl $a \in \{1, \dots, n-1\}$ und schickt $a \cdot x \in G$ an Bob.

Bob denkt sich eine Zahl $b \in \{1, \dots, n-1\}$ und schickt $b \cdot x \in G$ an Alice.

(2.) 

Alice berechnet mit a das Gruppenelement $a \cdot (bx)$

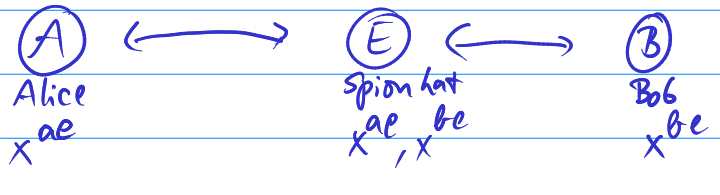
Bob berechnet mit b das Gruppenelement $b \cdot (ax)$

Danach besitzen beide den gemeinsamen geheimen Schlüssel $a \cdot (bx) = abx = b \cdot (ax)$

12.) 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, ev. leichtere Problem zu lösen:

Diffie-Hellman-Problem (DH-Problem):
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.
(Bem.: DL lösbar $\Rightarrow$ DH lösbar ist klar, „$\in^a$" ist unbekannt.)

13.) Weiter ist beim Schlüsselaustausch entscheidend, dass sich Alice und Bob sicher sein können, wirklich mit dem angegebenen Absender 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. $\vec{x^{ea}}$
Gelingt dies, braucht der Unbefugte die verschlüsselten Nachrichten zwischen Alice und Bob abzufangen:
Die Nachrichten von Alice an Bob dekodiert er mit dem Alice-Schlüssel x^{ea} schickt sie mit dem Bob-Schlüssel x^{eb} kodiert an Bob weiter, und umgekehrt.
Er kann so die gesamte geheime Kommunikation abhören.
Man nennt dies eine "Man-in-the-middle-Attacke."



Stichworte: ElGamal-Verschlüsselung und ElGamal-Signatur auf Untergruppe $\langle x \rangle$ einer beliebigen abelschen Gruppe $(G,+)$, Hashfunktion, Motivation: nimm für $(G,+)$ die Gruppe einer elliptischen Kurve

---

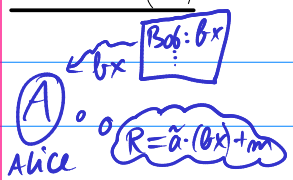
1.2.3 ElGamal-Verschlüsselung (entwickelt von T. ElGamal)

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 $d \in \{1,\dots,n-1\}$ als privaten Schlüssel und erzeugt einen öffentlichen Schlüssel $d \cdot x$:

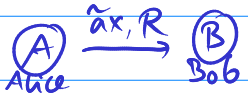
	geheim	öffentlich
Alice	a	ax
Bob	b	bx

1.) Alice möchte eine geheime Botschaft $m \in G$ an Bob schicken.
Das Verfahren geht wie folgt:

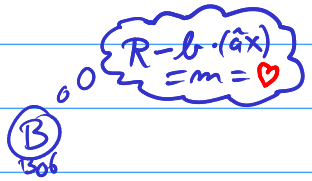
Schritt (1.) Alice wählt eine Zufallszahl $\tilde{a} \in \{1,\dots,n-1\}$ und berechnet $\tilde{a} \cdot x$.
Alice besorgt sich Bobs öffentlichen Schlüssel bx und berechnet $R = \tilde{a} \cdot (bx) + m$.



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 $R - b \cdot (\tilde{a}x) = m$.



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 $\hat{a}b \cdot x \in G$ berechnen.)

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 $\tilde{a}x$, $R_1 = \hat{a}(bx) + m_1$ und $\hat{a}x$, $R_2 = \hat{a}(bx) + m_2$ 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.

§1.3 Digitale Unterschriften

1.3.1 ElGamal- bzw. DSA-Signatur

Geg. wieder eine abelsche Gruppe $(G,+)$, $x \in G$ mit $n = \mathrm{ord}(x)$ 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 $ax \in G$.

4.) Sei $\mathcal{M}$ die Menge aller möglichen Nachrichten (etwa beliebig lange Folgen von 0 und 1), und gg. sei eine Funktion $h: \mathcal{M} \to \{0,1,\dots,n-1\}$, 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.

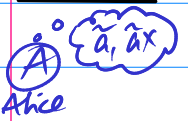
Def.: Eine solche Funktion heißt eine Hashfunktion.

5.) Bsp.: 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 $h: \mathbb{Z}_p^* \to \mathbb{Z}_p^*$, $h(z) = g^z \bmod p$ eine Hashfunktion. Das ab 7.) beschriebene Verfahren kann dann mit $G = \mathbb{Z}_p^*$, $x = g$ durchgeführt werden (in der Praxis nimmt man für p eine Sophie-Germain-PZ, d.h. p prim mit $\frac{p-1}{2}$ auch prim, denn dann ist etwa jedes zweite Element ein Erzeuger und daher leicht ein Erzeuger zu finden).

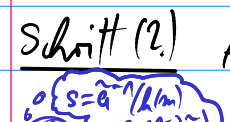
6.) Öffentlich zugänglich seien die Daten $(G,+)$, $x \in G$, $n = \mathrm{ord}(x)$, h und $ax \in G$, sowie eine Bijektion $\chi: \langle x \rangle \to \{0,1,\dots,n-1\}$, deren Werte effektiv berechenbar seien (in der Praxis reicht eine Fkt. deren Urbildmenge $\chi^{-1}(k)$ von jedem $k \in \{0,\dots,n-1\}$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

7.) Nun das Verfahren zur Signatur, wie Alice ihre Nachricht $m \in \mathcal{M}$ unterschreiben kann:

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



Schritt (2.) Alice berechnet das Inverse $\tilde{a}^{-1}$ von $\tilde{a}$ in $\mathbb{Z}_n$ (euklidischer Algo!) 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.) (Verifikation) 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 $\Psi(\tilde{a}x)\,a + s\tilde{a} \equiv h(m) \bmod n$ gewählt ist,
da ja $n = \mathrm{ord}(x)$ in G gilt.

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

9) Bem.: 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 $\Psi(\tilde{a}x)$ invertierbar in $\mathbb{Z}_n$ ist.

10.) Bem.: 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:

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

- Wäre h nicht kollisionsresistent und ein Auffinden von $m' \in M$ 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 wegen

$$\gamma(\tilde{a}x)\, a_x + s\tilde{a}x = h(m) \cdot x = h(m') \cdot x.$$

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

12.) Das beschriebene Verfahren heißt ElGamal-Signatur-Verfahren. Eine rechnerisch vorteilhafte Variante heißt DSA (= digital signature algorithm). Das mit der Gruppe einer elliptischen Kurve realisierte DSA-Verfahren heißt ECDSA (= elliptic curve digital signature algorithm), wir besprechen es später genauer.

13.) Motivation: Eine auf Koblitz/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 als für $\mathbb{Z}_m^*$ ist, gilt diese Art von Verschlüsselungstechnik heute als besonders sicher und wird vielfältig industriell angewendet; wegen der kleineren Schlüssellänge ist diese auch rechnerisch praktischer als z.B. RSA. Wir werden die Mathematik elliptischer Kurven im folgenden §2 der Vorlesung näher kennenlernen.

Stichworte:

Def. Polynom, Ableitung, Produkt-/Kettenregel
Nullstelle + Ordnung, irreduzibles Polynom,
eindeutige Zerlegung in irred. Polynome (Gauß) $\hat{=}$ PFZ in $\mathbb{Z}$
Eukl. Alg. geht im Polynomring $k[x]$ $\hat{=}$ Eukl. Alg. in $\mathbb{Z}$
Restklassenring $\underbrace{k[x]/(f)}_{\text{Körper, wenn } f \text{ irreduzibel}}$ $\hat{=}$ $\underbrace{\mathbb{Z}/m\mathbb{Z}}_{\text{Körper, wenn } m \text{ prim} \rightsquigarrow \mathbb{F}_p}$ in $\mathbb{Z}$

(neue) endliche Körper $\mathbb{F}_{p^n}$ + Rechnen darin, algebraischer Abschluss

---

§2 Elliptische Kurven (über beliebigen Grundkörper k)

§2.1 Grundlagen aus der Algebra

2.1.1 Polynome

Sei k ein (beliebiger) Körper.

1.) **Def.:** Ein Polynom über k in den n Variablen $x_1, \dots, x_n$ ist ein Ausdruck der Form

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

mit Koeffizienten $\alpha_{\nu_1, \dots, \nu_n} \in k$, 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. $\underline{x} = (x_1, \dots, x_n)$ 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
$k[x_1, \dots, x_n]$ oder noch kürzer mit $k[\underline{x}]$ bezeichnet,
schreiben dann auch kurz $f \in k[\underline{x}]$ wenn $f(\underline{x})$ ein Polynom ist.

2.) **Bem.:**
Durch komponentenweise Addition $\sum_{\nu} \alpha_\nu x^\nu + \sum_{\nu} \beta_\nu x^\nu := \sum_{\nu} (\alpha_\nu + \beta_\nu) x^\nu$
und der Multiplikation $\left(\sum_{\nu} \alpha_\nu x^\nu\right)\left(\sum_{\mu} \beta_\mu x^\mu\right) := \sum_{\nu,\mu} a_\nu b_\mu x^{\nu+\mu}$

wird $k[x] = k[x_1, \dots, x_n]$ zu einem kommutativen Ring mit 1;
das Nullpolynom $0 := \sum_{\nu} 0 \cdot x^\nu$ ist dabei das Nullelement,
das Polynom $1 := 1 \cdot x^0 + \sum_{\nu \neq 0} 0 \cdot x^\nu$ ist das Einselement. ("Einspolynom")
Der Ring
$(k[x], +, \cdot)$ heißt **Polynomring über k**.

3.) **Def.:** Für $f \in k[x]$ mit $f(x) = \sum_{\nu} \alpha_\nu x^\nu$ und $1 \leq j \leq n$

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

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

Bem.: Ist $k = \mathbb{R}$, stimmt diese Def. mit der üblichen Def. der Analysis überein.
Hier sind "Ableitungen" wieder Polynome.

Durch einfaches Nachrechnen kann man bestätigen:

4.) **Satz:** Für alle $f, g \in k[x]$ und $\gamma \in k$ gelten die Ableitungsregeln

$$\frac{\partial(\gamma f)}{\partial x_j} = \gamma \frac{\partial f}{\partial x_j} \quad \text{und} \quad \frac{\partial(f+g)}{\partial x_j} = \frac{\partial f}{\partial x_j} + \frac{\partial g}{\partial x_j},$$

$$\frac{\partial(f \cdot g)}{\partial x_j} = f \frac{\partial g}{\partial x_j} + g \frac{\partial f}{\partial x_j} \quad \text{(Produktregel)},$$

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

die **Kettenregel** $\frac{\partial f(g_1, \dots, g_m)}{\partial x_j}$

$$= \frac{\partial f}{\partial x_1}(g_1, \dots, g_m) \frac{\partial g_1}{\partial x_j} + \dots + \frac{\partial f}{\partial x_m}(g_1, \dots, g_m) \frac{\partial g_m}{\partial x_j}$$

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

5.) Ist $f \neq 0$, so heißt $\deg(f) := \max\{j \in \mathbb{N}_0;\ \alpha_j \neq 0\}$ der Grad von f.
Für $f \in k[x_1, \dots, x_n]$ in n Variablen ist $\deg(f) := \max\{\nu_1 + \dots + \nu_n;\ \alpha_{\underline{\nu}} \neq 0\}$ 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, z.B. $f(x,y) = \alpha_{(2,0)} x^2 + \alpha_{(1,1)} xy + \alpha_{(0,1)} y$, wir werden dann für die Koeffizienten einfachere Notationen wählen.

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 ("Polynomdivisionen") (daher kann man wie in $\mathbb{Z}$ z.B. den ggT von Polynomen mit dem Euklidischen Algorithmus ausrechnen). Dies ist aus den Grundvorlesungen bekannt, wir erinnern hier nur an folgendes:

7.) Def.: Geg. sei die "Einsetz"abbildung $k \to k$, $c \mapsto f(c) := \sum_{\nu \geq 0} \alpha_\nu c^\nu$.
Ein El. $c \in k$ heißt Nullstelle von f, falls $f(c) = 0$ in k ist.

8.) Bem.: $c \in k$ ist genau dann Nst., wenn $(x-c)$ ein Teiler von f im Polynomring $k[x]$ ist, d.h. falls ex. $g \in k[x]$ mit $(x-c) \cdot g = f$.

9.) Def.: Ist c eine Nullstelle von $f \neq 0$, so gibt es ein maximales $e \geq 1$, so dass $(x-c)^e$ ein Teiler von f ist. Die Zahl e heißt Ordnung der Nullstelle c. Ist $f(c) \neq 0$, def. man diese "Nullstellen"ordnung als 0.

10.) Def.: Ein Polynom $f \in k[x]$ vom Grad ≥ 1 heißt irreduzibel (oder prim), falls gilt: $\forall u, v \in k[x]: f = u \cdot v \Rightarrow \deg u = 0$ oder $\deg v = 0$, d.h. f kann nicht als Produkt zweier Polynome vom Grad ≥ 1 geschrieben werden. ($\rightsquigarrow$ vgl. 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:

11.) **Def.:** Sei $f \in k[x]$. Dann heißen $a \in k[x]$ und $b \in k[x]$ kongruent modulo f, wenn $f \mid (b-a)$, d.h. falls $g \in k[x]$ ex. mit $b = a + fg$.
(Das Kongruenzzeichen $\equiv$ möchten wir für $\mathbb{Z}$ vorbehalten.)
Die Restklassen modulo f sind Teilmengen von $k[x]$ der Gestalt
$$a + f \cdot k[x] := \{a + f \cdot g;\ g \in k[x]\} \text{ mit } a \in k[x].$$
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. (aber doppelt unterstreichen!)
Die Menge der Restklassen modulo f bezeichnen wir mit
$$k[x]/(f) := \{a + f \cdot k[x];\ a \in k[x]\} = \{\underline{\underline{a}};\ a \in k[x]\}$$
und nennen diese den Restklassenring modulo f, weil diese bzgl. der Def. $\underline{\underline{a}} \dotplus \underline{\underline{b}} := \underline{\underline{a \dotplus b}}$ für Polynome $a, b \in k[x]$ wieder zu einem kommutativen Ring mit $\underline{\underline{1}}$ als Eins wird.

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

12.) **Bem.:** Für jede Restklasse $\underline{\underline{a}} = a + f \cdot k[x] \in k[x]/(f)$ gibt es genau einen Vertreter $b \in \underline{\underline{a}} = a + f \cdot k[x]$, d.h. $\underline{\underline{b}} = \underline{\underline{a}}$ 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.

13.) **Satz:** 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.

Bew.: Körper: ✓ [Inverse findet man mit erweitertem Euklidischen Algorithmus für Polynome], p^r El.: jede Restklasse hat genau einen Vertreter $b = a_0 + \ldots + a_{r-1}X^{r-1}$. □
(p Möglichkeiten für jedes a_i)

14.) **Bem.:** Für jedes $r \in \mathbb{N}$ gibt es (mind.) ein irreduzibles Polynom $f \in \mathbb{F}_p[x]$ mit $\deg f = r$.

15.) **Bem.:** Es gibt im wesentlichen (d.h. bis auf Isomorphie) genau einen endlichen Körper mit p^r Elementen, d.h. welches irreduzible f mit $\deg f = r$ wir als Modul nehmen, ist für seine Konstruktion (bis auf Isomorphie!) egal. Wir bezeichnen diesen Körper mit $\boxed{\mathbb{F}_{p^r}}$.

16.) **Bem.:** Jeder Körper mit endlich vielen Elementen ist einer dieser Körper $\mathbb{F}_{p^r}$ mit p prim und $r \geq 1$. [Ohne Beweis, vgl. "Algebra"-Vorl.]

17.) Wegen Bem. 12.) ist nach Wahl eines irreduziblen Polynoms $f \in \mathbb{F}_p[x]$, $\deg f = r$,

also $$\mathbb{F}_{p^r} = \{(a_{r-1}x^{r-1} + \dots + a_1 x + a_0) + f \cdot \mathbb{F}_p[x] \; ; \; a_i \in \mathbb{F}_p\},$$

die Restklassenvertreter $a_{r-1}x^{r-1} + \dots + a_1 x + a_0$ lassen sich auch durch Koeffizienten-r-Tupel $(a_{r-1}, a_{r-2}, \dots, a_1, a_0) \in \mathbb{F}_p^r$ 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.

18.) **Bsp.:** Sei $p = 2$, $r = 3$, wir möchten $\mathbb{F}_8$ konstruieren.

Das Polynom $f(x) = x^3 + x + \underline{1}$ ist irreduzibel über $\mathbb{F}_2 = \{\underline{0}, \underline{1}\}$,

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

und man rechnet z.B. $(\underline{0},\underline{1},\underline{0}) \cdot (\underline{1},\underline{1},\underline{1}) = (\underline{1},\underline{0},\underline{1})$,

weil $(\underline{0}\cdot x^2 + \underline{1}\cdot x + \underline{0}) \cdot (x^2 + x + \underline{1}) = x^3 + x^2 + x = \underline{1}\cdot(x^3 + x + \underline{1}) + (\underline{x^2 + 1})$

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

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

19.) **Bsp.:** Rechnen in $\mathbb{F}_{5^3} = \mathbb{F}_{125}$: Haben wir diesen Körper mit dem irreduziblen Polynom $f = x^3 + x + \underline{1} \in \mathbb{F}_5[x]$ vom Grad 3 konstruiert (da es keine Nst. in $\mathbb{F}_5$ hat, muss es irreduzibel sein, da es Grad 3 hat!),

So rechnen wir in $\mathbb{F}_{5^3}$ z.B. $(\underline{1}, \underline{2}, \underline{4}) \cdot (\underline{-1}, \underline{3}, \underline{0})$

$= (x^2 + \underline{2}x - \underline{1})(-x^2 + \underline{3}x) = -x^4 + \underline{3}x^3 - \underline{2}x^3 + \underline{6}x + x^2 - \underline{3}x$

$= -x^4 + x^3 + x^2 + \underline{3}x = (x^3 + x + \underline{1}) \cdot (-x + \underline{1}) + \underline{2}x^2 + \underline{3}x + \underline{1}$

↑ Polydiv. durch f $\quad = (\underline{2}, \underline{3}, \underline{1}) \bmod f$,

"eigentlich" ja: $(\underline{1}, \underline{2}, \underline{-1}) \cdot (\underline{-1}, \underline{3}, \underline{0}) = (\underline{2}, \underline{3}, \underline{1})$.

20.) **Bem.:** $\mathrm{char}(\mathbb{F}_{p^r}) = p$, denn es gilt $\underline{1} + \underline{1} + \ldots + \underline{1} = \underbrace{\underline{1} + \ldots + \underline{1}}_{p\text{ mal}} = \underline{p \cdot 1} = \underline{0}$,

und p minimal so da p prim.

21.) **Def.:** Ein Körper k ist algebraisch abgeschlossen, wenn sich jedes Polynom $f \in k[x]$, $\deg f > 0$, als Produkt von Linearen Polynomen schreiben lässt, d.h. wenn $f(x) = d(x - c_1) \cdots (x - c_m)$, die $c_i, d \in k$, gilt.

22.) **Bem.:** Man kann jeden Körper k in einen algebraisch abgeschlossenen Körper einbetten. Ein bzgl. "$\subseteq$" minimaler heißt algebraischer Abschluss von k, dieser ist eindeutig und wird mit $\bar{k}$ 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].

Stichworte:

- Der affine und projektive Raum über einem Körper k
- verschiedene parallele Geraden schneiden sich nicht im Affinen
- im Projektiven schneiden sich zwei verschiedene projektive Geraden stets in genau einem Punkt
- projektive Geraden entstehen aus affinen Geraden durch Homogenisierung
- $\mathbb{P}^2(k)$ und geometrische Interpretationen
- $f \in k[x,y] \rightsquigarrow$ affine Kurve
- Tangente einer affinen Kurve
- singulärer Punkt, Beispiele

---

§2.2 Der affine Raum, affine Kurven und der zweidimensionale projektive Raum

Wir stellen den zweidimensionalen affinen und projektiven Raum vor, d.h. 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 z.B. elliptische Kurven werden dann in der projektiven Ebene interpretiert, weil es rechentechnisch einfacher und mathematisch natürlicher ist.

2.2.1 Die affine und projektive Ebene

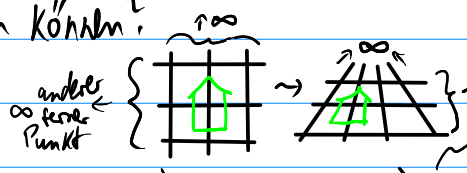
Sei k ein beliebiger Körper. Wir stellen uns meistens $\mathbb{R}$ vor, weil wir über geometrische Objekte nachdenken möchten; k ist in den Anwendungen aber meist ein endlicher Körper.

1) **Def.:** Den zweidimensionalen k-VR $k^2 = k \times k$ 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.

2) **Def.:** Eine Gerade in $\mathbb{A}^2(k)$ ist eine Teilmenge des $\mathbb{A}^2(k)$ der Form $g(a,b,c) := \{(x,y) \in \mathbb{A}^2(k);\ ax + by + c = 0\}$ für ein Tripel $(a,b,c) \in k^3 \setminus \{(0,0,c);\ c \in k\}$.

3) **Bem.:** Zwei verschiedene Geraden in $\mathbb{A}^2(k)$ schneiden sich in genau einem Punkt, es sei denn, sie sind parallel, d.h. dann haben sie keinen gemeinsamen Punkt in $\mathbb{A}^2(k)$. Soweit nichts Neues.

4) Bem.: 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, so dass bei der Schnittpunktberechnung auch die unendl. fernen Punkte erhalten werden können?



$\rightsquigarrow$ ($ax+by+c \cdot z \overset{!}{=} ax+by+d \cdot z$ lösbar $\rightarrow z=0$)

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

5) Def.: Die projektive Ebene über k ist die Menge

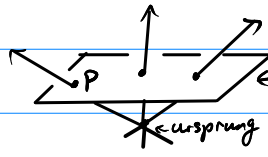
$$\mathbb{P}^2(k) = \{ [y_1 : y_2 : y_3];\ y_1, y_2, y_3 \in k, \text{ nicht } y_1 = y_2 = y_3 = 0 \}$$

mit der Vereinbarung, dass $[y_1 : y_2 : y_3] = [\tilde{y}_1 : \tilde{y}_2 : \tilde{y}_3]$ genau dann gilt, wenn es ein $\lambda \in k \setminus \{0\}$ gibt mit $y_1 = \lambda \tilde{y}_1$, $y_2 = \lambda \tilde{y}_2$, $y_3 = \lambda \tilde{y}_3$.

6) Formal: $\mathbb{P}^2(k)$ ist die Menge der Äquivalenzklassen in k^3 bzgl. der Äquivalenzrelation $(y_1, y_2, y_3) \sim (\tilde{y}_1, \tilde{y}_2, \tilde{y}_3) :\Leftrightarrow \exists \lambda \in k, \lambda \neq 0$: $y_i = \lambda \tilde{y}_i,\ i=1,2,3$
d.h. $\mathbb{P}^2(k) := (k^3 \setminus \{(0,0,0)\}) / \sim$.

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, und y_1, y_2, y_3 nennen wir projektive Koordinaten von $[y_1 : y_2 : y_3]$.

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



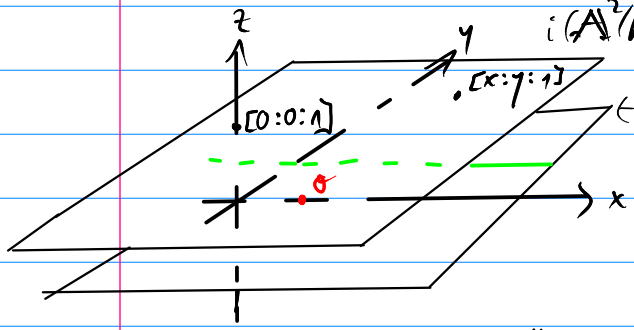
$P = [x : y : 1]$ $\quad i(\mathbb{A}^2(k)) = \{[x : y : 1];\ x, y \in k\}$

Einem projektiven Punkt $[x : y : z]$ entspricht in unserem Modell in k^3 der Ursprungsgeraden $\{(\lambda x, \lambda y, \lambda z);\ \lambda \in k\}$.

Diese Punkte sind entweder $[x:y:1]$ oder $[x:y:0]$ mit $x,y \in k$ (nicht $[0:0:0]$!)
z.B. durch die Abbildung $i: \mathbb{A}^2(k) \to \mathbb{P}^2(k)$
$(x,y) \mapsto [x:y:1]$
Kann die affine Ebene in die projektive eingebettet werden (d.h. i ist injektiv!)

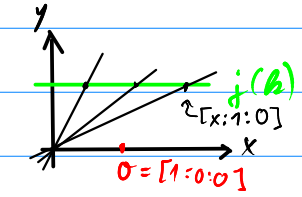
8.) Aber $\mathbb{P}^2(k)$ enthält zusätzlich noch die projektiven Punkte $[x:y:0]$, $x,y \in k$. Offenbar ist $\{[x:y:0];\ x,y \in k, \overset{\text{nicht}}{x=y=0}\}$ eine Gerade in $\mathbb{P}^2(k)$, die wir unendlich ferne Gerade g_∞ nennen möchten, denn mit
$j: k \to g_\infty,\ x \mapsto [x:1:0]$ läßt sich k darin einbetten
(d.h. j ist injektiv), wobei auffällt, daß $g_\infty \setminus \mathrm{im}(j)$ aus genau dem weiteren Punkt $O := [1:0:0]$ besteht, d.h. $g_\infty \setminus \mathrm{im}(j) = \{O\}$.

9.) Somit: $\mathbb{P}^2(k) = i(\mathbb{A}^2(k)) \dot\cup \underbrace{j(k) \dot\cup \{O\}}_{= g_\infty}$ (disjunkte Vereinigung)

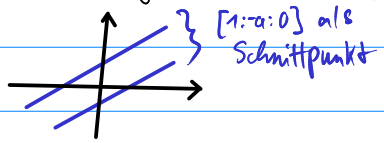


$\leftarrow$ x-y-Ebene für ∞ ferne Punkte:
$\to g_\infty = \{[x:y:0];\ x,y \in k\}$
$= j(k) \dot\cup \{O\}$
$\hat{=}$ { Ursprungsgeraden in der x-y-Ebene}

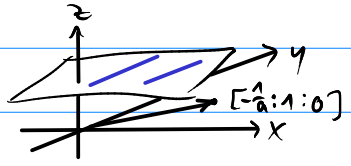
Ansicht auf die x-y-Ebene:



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



in unserem 3-dim. "Modell":



bzw. $(-\hat{a},1)$
bzw. $(-1,a)$
bzw. $(\hat{a},-1)$

Geradengl.: $\underbrace{(a,1)\cdot(x,y)}_{\text{Skalarprodukt}} + c = 0$, der Normalenvektor ist $(a,1)$ und senkrecht zum Richtungsvektor $(1,-a)$.
Die gemeinsame Richtung $(1,-a)$ wird zum gemeinsamen Schnittpunkt $[-\hat{a}:1:0]$ erklärt.

10.) **Def.:** Eine projektive Gerade ist eine Teilmenge von $\mathbb{P}^2(k)$ der Form $G(a,b,c) = \{[x:y:z];\ ax+by+cz=0\}$ 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 und 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 Form von Tangenten dann wiederbegegnen.

Bsp: Die projektiven Geraden $G(a,1,c), G(a,1,d)$ schneiden sich in $[-\hat{a}:1:0]\in g_\infty$. (a≠0)

Bem.: 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.

11.) **Def.:** 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

$$C_f(k) := \{(u,v)\in\mathbb{A}^2(k);\ f(u,v)=0\}.$$

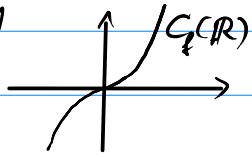
Jede solche Nullstellenmenge $C_f(k)$ nennen wir eine affine Kurve.

Ist klar, welches Polynom f vorliegt, schreiben wir auch kurz $C(k)$ für $C_f(k)$.

Geraden sind spezielle affine Kurven (zu linearen Polynomen $f(x,y,z)=ax+by+c$).

12.) **Bem.:** 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.

13.) **Bsp:** Sei $k=\mathbb{R}$ und $f(x,y)=y-x^3-x$. Die Nullstellenmenge $C_f(k)$ 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 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 $C_f(\mathbb{F}_5)=\{(0,0),(1,2),(2,0),(3,0),(4,3)\}$ ist, und

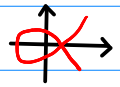
mit $f_0(x,y)=y^2-x^3-x$ haben wir $C_{f_0}(\mathbb{F}_5)=\{(0,0),(2,0),(3,0)\}$.

Ist $\tilde{k} \subseteq k$ ein Teilkörper von k (wie z.B. $\mathbb{Q} \subseteq \mathbb{R}$), so folgt auch stets $C_f(\tilde{k}) \subseteq C_f(k)$. Unsere Kurvenpunkte in $A(\mathbb{F}_5)$ finden wir deswegen z.B. in $A(\mathbb{F}_{25})$ wieder.

14.) Def.: Eine (affine) Tangente an eine affine Kurve $C_f(k)$ im Punkte $(a,b) \in C_f(k)$ ist die Gerade

$$t_{(a,b)}(C_f) = \{(x,y): \frac{\partial f}{\partial x}(a,b)x + \frac{\partial f}{\partial y}(a,b)y + d = 0\},$$

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 $(a,b) \in t_{(a,b)}(C_f)$ gilt.

15.) Es ist nicht klar, ob Tangenten stets eindeutig existieren, denn:
Affine Kurven können sich selbst schneiden oder scharfe "Spitzen" haben:
Bsp.: $f(x,y) = y^2 - x^3$: 
$f(x,y) = y^2 - x^3 + 3x^2 - 4$: 

16.) Def.: Die affine Kurve $C_f(k)$ heißt singulär im Punkt $(a,b) \in C_f(k)$, falls $\frac{\partial f}{\partial x}(a,b) = \frac{\partial f}{\partial y}(a,b) = 0$ gilt.

17.) Bem.: Affine Kurven, die in keinem Punkt singulär sind, haben überall eine wohldefinierte Tangente.

18.) Bem.: Es kann vorkommen, dass $C_f(k)$ gar keine singulären Punkte enthält, wohl aber über einem Erweiterungskörper von k, wie etwa $\bar{k}$, dem algebraischen Abschluss von k.

19.) Bsp.: $f(x,y) = y^2 - x^4 - 2x^2 - 1 \rightsquigarrow C_f(\mathbb{R})$ hat 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 $C_f(\mathbb{C})$, wo $\mathbb{C} = \bar{\mathbb{R}}$.

20.) Bsp: $f(x,y) = y^2 - x^3 - x$, $k = \mathbb{F}_p \rightsquigarrow$ Ableitungen: $\frac{\partial f}{\partial x}(x,y) = -3x^2 - 1$, $\frac{\partial f}{\partial y}(x,y) = 2y$, d.h. 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 $3a^2 = -1$
  $\rightsquigarrow 0 = a(\underbrace{3a^2}_{=-1} + 3) \rightsquigarrow 2a = 0 \overset{p \neq 2}{\rightsquigarrow} a = 0$ im $\not\downarrow$ zu $3a^2 = -1$. Also ex. keine sing. Punkte für $p \neq 2$.
- Für $p = 2$ ist $C_f(\mathbb{F}_2) = \{(0,0),(1,0)\} \rightsquigarrow \frac{\partial f}{\partial x}(1,0) = 0 = \frac{\partial f}{\partial y}(1,0)$, d.h. $(1,0)$ ist sing. Punkt.

Stichworte: Homogene Polynome definieren projektive Kurven,
Homogenisierung: affine Kurve ⇝ projektive Kurve
singuläre Punkte im Projektiven, projektive Tangenten,
nicht-singuläre projektive Kurven haben "ihre Tangenten schön",
Schnittmultiplizität im Schnittpunkt einer Gerade mit einer Kurve C_F,
deren Summe ist $\leq \deg F$, $m(P;T,C_F) \geq 2$ bei Tangente T an Kurve C_F mit $\deg F \geq 2$

---

§2.3 Projektive Kurven

2.3.1 Homogene Polynome und projektive Kurven

Durch Homogenisierung können wir affine Kurven zu projektiven Kurven machen.

1.) **Def.:** 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

$$F(X,Y,Z) = \sum_{\nu_1,\nu_2,\nu_3 \geq 0} \alpha_{\nu_1,\nu_2,\nu_3} X^{\nu_1} Y^{\nu_2} Z^{\nu_3}$$

und $\alpha_{\nu_1,\nu_2,\nu_3} \neq 0 \Rightarrow \nu_1 + \nu_2 + \nu_3 = d$,
d.h. wenn alle Monome in F den Grad d haben.

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

3.) **Bem.:** 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 $f(X,Y) = \sum_{\nu_1,\nu_2 \geq 0} \alpha_{\nu_1,\nu_2} X^{\nu_1} Y^{\nu_2}$ vom Grad d, so setze $F_f(X,Y,Z) := \sum_{\nu_1,\nu_2 \geq 0} \alpha_{\nu_1,\nu_2} X^{\nu_1} Y^{\nu_2} Z^{d-\nu_1-\nu_2}$.
Man nennt F_f dann die Homogenisierung von f. Für diese gilt $F_f(X,Y,1) = f(X,Y)$.

4.) **Lemma:** Ist $F \in k[X,Y,Z]$ homogen vom Grad d, so gilt für alle $\alpha,\beta,\gamma \in k$ und $\lambda \in k \setminus \{0\}$: $F(\alpha,\beta,\gamma) = 0 \Leftrightarrow F(\lambda\alpha, \lambda\beta, \lambda\gamma) = 0$

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

Somit können wir projektive Kurven definieren:

5.) **Def.:** Sei $F \in k[X,Y,Z]$ homogen. Dann bezeichnen wir die Nullstellenmenge mit $C_F(k) := \{[u:v:w] \in \mathbb{P}^2(k) ;\ F(u,v,w) = 0\}$.
Ist F klar, wird auch einfach $C(k)$ für $C_F(k)$ geschrieben.
Jede solche Nullstellenmenge heißt eine projektive ebene Kurve.

6.) **Bsp.:** Die affine Kurve $C_f(k)$ 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 Abb. $i: \mathbb{A}^2(k) \to \mathbb{P}^2(k)$, $(x,y) \mapsto [x:y:1]$ bildet $C_f(k)$ nach $C_{F_f}(k)$ ab.
Die projektive Kurve $C_{F_f}(k)$ hat aber noch einen weiteren (genau einen) Punkt (auf g_∞), nämlich $[0:1:0]$, d.h. $C_{F_f}(k) = i(C_f(k)) \cup \{[0:1:0]\}$.

7.) **Lemma:** $C_{F_f}(k) \cap i(\mathbb{A}^2(k)) = i(C_f(k))$ für jede affine Kurve C_f und ihre projektive Kurve C_{F_f}.

Bew.: $[x:y:1] \in C_{F_f}(k) \cap i(\mathbb{A}^2(k)) \Leftrightarrow 0 = F_f(x,y,1) = f(x,y) \Leftrightarrow [x:y:1] \in i(C_f(k))$. □

8.) **Bem.:**
- Werden hier i auch weglassen, es ist klar, was gemeint ist.
- Anstelle von i können auch die Einbettungen $i_2(x,y) = [1:x:y]$, $i_3(x,y) = [x:1:y]$ betrachtet werden, das Lemma gilt dann entsprechend.
- Geht man für eine projektive Kurve $C_F(k)$ zu einer dieser Schnitte mit $\mathbb{A}^2(k)$ über, so sagt man, man "geht zu affinen Koordinaten" über.

9.) **Def.:** Sei $F \in k[X,Y,Z]$ homogen vom Grad d.
Die projektive ebene Kurve $C_F(k)$ heißt singulär im Punkt $P = [a:b:c] \in C_F(k)$, falls alle Ableitungen von F in P verschwinden, d.h. falls $\frac{\partial F}{\partial x}(a,b,c) = \frac{\partial F}{\partial y}(a,b,c) = \frac{\partial F}{\partial z}(a,b,c) = 0$. Die Kurve $C_F(k)$ heißt nicht-singulär bzw. glatt, falls $C_F(\bar{k})$ keinen singulären Punkt enthält, wobei $\bar{k}$ den algebraischen Abschluss von k bedeutet.

10.) Diese Def. hängt nicht davon ab, welche projektiven Koordinaten a, b, c eines Punktes $P = [a:b:c]$ betrachtet werden. Sie passt auch mit der alten Def. von "singulärem Punkt" für affine Kurven zusammen, wie folgendes Lemma zeigt; nach dem Lemma genügt es, singuläre Punkte, die im Affinen liegen, auf Singularität im Affinen zu testen.

11.) **Lemma:** Sei $F(X,Y,Z) = \sum_{\nu \geq 0} \alpha_\nu X^{\nu_1} Y^{\nu_2} Z^{\nu_3}$ homogen vom Grad d und $f(x,y) = \sum_{\substack{\nu_1,\nu_2 \\ \nu_1+\nu_2 \leq d}} \alpha_{\nu_1,\nu_2,d-\nu_1-\nu_2} x^{\nu_1} y^{\nu_2} = F(x,y,1)$, d.h. $F = F_f$,

weiter sei $P \in C_F(k)$ mit $P = i(Q) \in i(\mathbb{A}^2(k))$.

Dann gilt: $C_F(k)$ singulär in $P \iff C_f(k)$ singulär in Q.

Bew.: Haben $Q \in C_f(k)$, etwa $Q = (a,b)$, dann ist $P = i(Q) = [a:b:1]$.

Es ist

$$\frac{\partial F}{\partial X}(X,Y,Z) = \sum_{\substack{\nu_1 > 0 \\ \nu_2,\nu_3 \geq 0}} \alpha_\nu \nu_1 X^{\nu_1 - 1} Y^{\nu_2} Z^{\nu_3}, \text{ also } \frac{\partial F}{\partial X}(a,b,1) = \frac{\partial f}{\partial x}(a,b),$$

entsprechend

gilt $\frac{\partial F}{\partial Y}(a,b,1) = \frac{\partial F}{\partial y}(a,b)$, sowie $\frac{\partial F}{\partial Z}(a,b,1) = \sum_{\nu_1,\nu_2,\nu_3 \geq 0} \alpha_\nu \nu_3 a^{\nu_1} b^{\nu_2}$

$$= \sum_{\nu_1,\nu_2,\nu_3 \geq 0} \alpha_{\nu_1,\nu_2,d-\nu_1-\nu_2} (d - \nu_1 - \nu_2) a^{\nu_1} b^{\nu_2} = d \cdot f(a,b) - a \frac{\partial f}{\partial x}(a,b) - b \frac{\partial f}{\partial y}(a,b).$$

Durch Vergleich der Ableitungen folgt die Beh. „$\iff$". □

12.) **Def.:** Sei $C_F(k)$ eine projektive ebene Kurve und $P = [a:b:c]$ ein nicht-singulärer Punkt auf $C_F(k)$. Die projektive Gerade

$C_T(k)$ 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 $C_F(k)$. Wir schreiben $T_P(C_F) := C_T(k)$ dafür.

13.) In nicht-singulären Punkten haben projektive ebene Kurven also eine "schöne" Tangente. Die Vor. "nichtsing." 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.

14.) **Bsp.:** $\operatorname{char} k \neq 2$, $f(x,y) := y^2 - 2x^2 - 2$, $F_f(X,Y,Z) = Y^2 - 2X^2 - 2Z^2$.

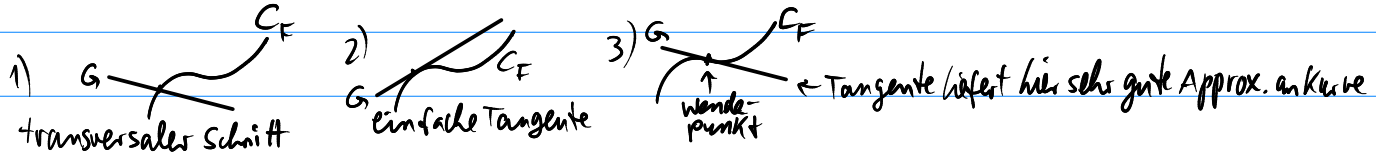
Dann: $(1,2) \in C_f(k)$, $\frac{\partial f}{\partial x}(1,2) = -4 \cdot 1 = -4$, $\frac{\partial f}{\partial y}(1,2) = 2 \cdot 2 = 4$, d.h. $(1,2)$ nichtsing.

Die (affine) Tangente von C_f in $Q = (1,2)$ ist $t_Q(C_f) = \{(x,y) \in k^2;\ -4x + 4y - 4 = 0\}$,

die (projektive) Tangente von C_F in $P = [1:2:1] = i(Q)$ ist

$$T_P(C_F) = \{[X:Y:Z] \in \mathbb{P}^2(k);\ -4X + 4Y - 4Z = 0\}.$$

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



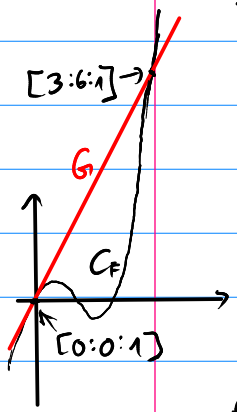
16) **Def.:** **Schnittmultiplizität** bzw. auch **Vielfachheit** genannt, mit der sich eine proj. Kurve mit einer Geraden schneidet:
Sei $C_F(k)$ eine projektive Kurve zum homogenen Polynom $F \in k[X,Y,Z]$, sei $G(\alpha,\beta,\delta)$ eine projektive Gerade und $P=[a:b:c] \in G(\alpha,\beta,\delta)$ ein Pkt. drauf.

- Ist P kein Schnittpunkt von $C_F(k)$ und G, setzen wir $m(P;G,C_F) := 0$.
- Ansonsten hat das Polynom $\Psi(t) := F(a+t\alpha', b+t\beta', c+t\gamma') \in k[t]$ eine Nullstelle in $t=0$, wobei $P'=[a':b':c'] \in G$ sind.

Dann sei: $m(P;G,C_F)$ die Ordnung der Nullstelle $t=0$ von $\Psi \in k[t]$, falls $\Psi \neq 0$.
Die Zahl $m(P;G,C_F)$ heißt **Schnittmultiplizität** bzw. **Vielfachheit**, mit der sich G und C_F im Punkt P schneiden.

17) **Bem.:** Es ist $m(P;G,C_F)$ unabhängig von der Wahl von $P' \in G(\alpha,\beta,\delta)$.

18) **Bsp.:** Sei $f(x,y) = x(x-1)(x-2) - y \in \mathbb{R}[x,y]$, d.h. $f(x,y) = x^3 - 3x^2 + 2x - y$
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 C_f in $(0,0) \in C_f$ die affine Tangente $t_{(0,0)}(C_f) = \{(x,y) \in \mathbb{R}^2 ; 2x - y = 0\}$, projektiv aufgefasst lautet die Tangente $T_{[0:0:1]}(C_F) = \{[x:y:z] \in \mathbb{P}^2(\mathbb{R}) : 2x - y + 0\cdot z = 0\} = G(2,-1,0)$.



Die Gerade $G(2,-1,0)$ schneidet C_F in $[3:6:1]$ und in $[0:0:1]$.
Dann haben wir $m([3:6:1]; G, C_F) = 1$,
weil $\Psi(t) = F(3+t\cdot 0, 6+t\cdot 0, 1+t\cdot 1)$
$= 3^3 - 3\cdot 3^2(1+t) + 2\cdot 3\cdot(1+t)^2 - 6\cdot(1+t)^2$
$= 0\cdot t^2 + (-3^3 + 6\cdot 2 - 12)t + \underbrace{(3^3 - 3^3 + 6 - 6)}_{=0} = -3^3 t^1$
eine einfache Nullstelle in $t=0$ hat, sowie $m([0:0:1]; G, C_F) = 2$, weil
$\Psi(t) = F(0+3t, 0+6t, 1+t) = (3t)^3 - 3(3t)^2(1+t) + \underbrace{2(3t)(1+t)^2 - (6t)(1+t)^2}_{=0} = -3^3 t^2$.

19.) Erläuterung: Ist $m(P; G, C_F) = 1$, liegt ein transversaler Schnitt der Geraden G mit der Kurve C_F vor. Ist $m(P; G, C_F) = 2$, so ist G eine "einfache" Tangente an C_F. Falls $m(P; G, C_F) \geq 3$, ist die Tangente eine sehr gute Approximation an C_F von "Ordnung ≥ 3" (da die Schnittmultiplizität genau die Nullstellenordnung von $\Psi(t)$ in $t=0$ ist). Ist $m(P; G, C_F)$ ungerade ≥ 3, so heißt P ein Wendepunkt von C_F.

20.) Bem.: Ist der Körper k algebraisch abgeschlossen, zerfällt Ψ vollständig in Linearfaktoren. Es folgt, dass dann die Summe der Schnittmultiplizitäten aller Schnittpunkte von G mit C_F genau $= \deg \Psi = \deg F$ ist, d.h. $\sum_{P \in G \cap C_F} m(P; G, C_F) = \deg F$. Ist k ein beliebiger Körper, folgt $\sum_{P \in G \cap C_F} m(P; G, C_F) \leq \deg F$.

21.) Bem.: 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 $G \subseteq C_F$ und für $[a:b:c], [a', b', c'] \in G$ ist dann

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

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

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

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

Beweis: Sei $T = G(\alpha, \beta, \gamma) = \{[x:y:z]; \alpha X + \beta Y + \gamma Z = 0\}$ die Tangente in $P = [a:b:c] \in G \cap C_F$, 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 bel. weiterer Punkt auf G, und $\Psi(t) = F(a+ta', b+tb', c+tc')$. Dann ist $\Psi(0) = 0$, da $P \in C_F$, und laut Kettenregel (vgl. V7_Satz 4.1)) 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 $m(P; T, C_F) \geq 2$. □

Stichworte:
- Spezialfall des Satzes von Bézout: $F_1, F_2 \in k[X]$ homogen, dann: $ggT(F_1, F_2) = 1 \Rightarrow \#(C_{F_1} \cap C_{F_2}) \leq (\deg F_1) \cdot (\deg F_2)$,
- Satz von von Bézout: $F_1, F_2 \in k[X]$ homogen, $ggT(F_1, F_2) = 1$ $\Rightarrow \sum_{P \in C_{F_1} \cap C_{F_2}} m(P; C_{F_1}, C_{F_2}) \leq (\deg F_1) \cdot (\deg F_2)$, und "=", falls k algebraisch abgeschlossen.
- Resultante zweier Polynome $\in S[X]$, S Körper oder Polynomring
- Zwei projektive Kurven C_{F_1} und C_{F_2} mit $(\deg F_1) \cdot (\deg F_2) - 1$ vielen Schnittpunkten und $ggT(F_1, F_2) = 1$, haben einen weiteren Schnittpunkt gemeinsam.

2.3.2 Der Satz von Bézout

Wir zeigen in diesem Abschnitt, dass Kurven (in der projektiven Ebene) i.a. nicht allzuviele Schnittpunkte haben:

1.) Satz: Zwei Kurven C_{F_1}, 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.
D.h.: $ggT(F_1, F_2) = 1 \Rightarrow \#(C_{F_1} \cap C_{F_2}) \leq (\deg F_1) \cdot (\deg F_2)$,
bzw. $\deg ggT(F_1, F_2) = 0 \Rightarrow \#(C_{F_1} \cap C_{F_2}) \leq (\deg F_1) \cdot (\deg F_2)$.

2.) Bem.: Der Satz 1.) ist eine sehr schwache Form des Satzes von Bézout, welcher besagt:

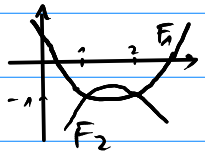
Satz von Bézout: Sei k ein algebraisch abgeschlossener Körper und seien $F_1, F_2 \in k[X, Y, Z]$ zwei homogene Polynome mit $ggT(F_1, F_2) = 1$, die zwei ebene projektive Kurven C_{F_1} und C_{F_2} definieren.
Dann ist $\sum_{P \in C_{F_1} \cap C_{F_2}} m(P; C_{F_1}, C_{F_2}) = (\deg F_1) \cdot (\deg F_2)$.

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

3.) Bem.: 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 Satz 1.), die wir hier beweisen, und insb. die spezielle Verschärfung Satz 15.) auf p. 5.

3.) **Bem.:** • Die Kurven können singuläre Punkte enthalten.
• Den Fall $\deg F_1 = 1$, d.h. wenn F_1 eine Gerade C_{F_1} erklärt, haben wir bereits in VG_Bem.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 Satz 1.): Denn für Schnittpunkte ist $m(P; C_{F_1}, C_{F_2}) \geq 1$,
also ist $\#(C_{F_1} \cap C_{F_2}) = \sum_{P \in C_{F_1} \cap C_{F_2}} 1$
$\leq \sum_{P \in C_{F_1} \cap C_{F_2}} m(P; C_{F_1}, C_{F_2}) \leq (\deg F_1) \cdot (\deg F_2)$.

4.) **Bsp:** Geg. seien die Parabeln $F_1(X,Y,Z) = X^2 - 3XZ + Z^2 - YZ$
und $F_2(X,Y,Z) = -X^2 + 3XZ - 3Z^2 - YZ$



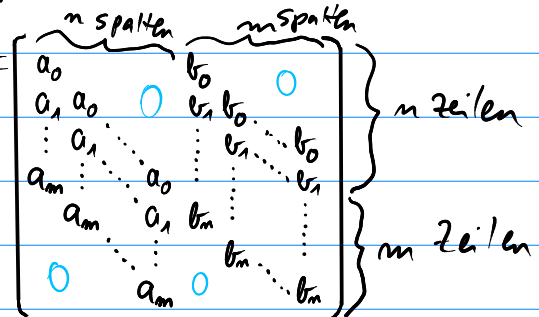
mit den beiden affinen reellen Schnittpunkten $[1:-1:1]$ und $[2:-1:1]$.
Laut Bézout-Satz haben die Parabeln noch zwei weitere Schnittpunkte über $\mathbb{C}$.
Diese sind nicht im Affinen, weil die Gleichung $F_1(X,Y,1) = F_2(X,Y,1)$ genau die Lösungen $(1,-1)$, $(2,-1)$ hat. Mit der Gleichung $F_2(X,Y,0)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

Algebraische Vorbereitung zum Beweis von Satz 1.): die Resultante

5.) **Def.:** Seien $f, g \in k[X]$ Polynome vom Grad $m = \deg f$, $n = \deg g$, etwa gegeben durch

$$f = a_m X^m + a_{m-1} X^{m-1} + \dots + a_1 X + a_0,\quad g = b_n X^n + b_{n-1} X^{n-1} + \dots + b_0 \in k[X],$$
$$a_m \neq 0 \neq b_n.$$

Sei $M(f,g) :=$



$\in k^{(m+n)\times(m+n)}$.

Dann ist $\mathrm{Res}(f,g) = \det M(f,g) \in k$ die Resultante von f und g.

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

7.) **Bsp.:** $k = \mathbb{R}$, $f(x) = x^2 + 2x - 1$, $g(x) = 4x^3 - 3x + 5$

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

"Resultantenkriterium" Wir benutzen hier nur die folgende Eigenschaft von Resultanten (genauer: (i)⇔(iii)):

8.) **Satz:** Sei S ein faktorieller Ring (z.B. Polynomring oder ein Körper), $f,g \in S[x]$ Polynome mit $\deg f = m$, $\deg g = n$. Dann sind äquivalent:
(i) $f,g \in S[x]$ haben einen gemeinsamen nichtkonstanten Teiler in $S[x]$,
(ii) es gibt $f_0, g_0 \in S[x] \setminus \{0\}$ mit $\deg f_0 \le m-1$, $\deg g_0 \le n-1$ und $f_0 g = g_0 f$,
(iii) $\mathrm{Res}(f,g) = 0$

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

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

(ii) ⇐ (iii): f_0, g_0 entsprechen der nichttriv. Lösungen des LGS

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

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

9.) **Beweis von Satz 1.):**
Wir nehmen zum Beweis OE an, dass k ein unendlicher Körper ist, andernfalls können wir z.B. zum algebraischen Abschluss $\bar{k}$ übergehen, der jedenfalls unendlich ist, vgl. dazu VL Bem. 22.1);
denn für eine Körpererweiterung könnte es mehr Schnittpunkte geben.
Sei $d_1 = \deg F_1$ und $d_2 = \deg F_2$.
Angenommen, C_{F_1} und C_{F_2} hätten (mind.) $d_1 d_2 + 1$ viele Punkte gemeinsam (wir zeigen, dass dann $\deg \mathrm{ggT}(F_1, F_2) \ge 1$ sein müsste).
Seien $P_0, P_1, \dots, P_{d_1 d_2}$ Schnittpunkte von C_{F_1} und C_{F_2}.

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

11.) 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 <u>keine</u> dieser Punkte $P_0, \dots, P_{d_1 d_2}$ geht, weil k unendlich ist. Diese Gerade sei OE g_∞, die unendlich ferne Gerade (durch eine Verschiebung/lineare Transformation lässt sich dies erreichen).

12.) 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, d.h. $f_1(x,y) := F_1(X, Y, 1)$, $f_2(x,y) := F_2(X, Y, 1)$, mit $\deg f_1 \leq d_1$, $\deg f_2 \leq d_2$. Wir können OE 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 $\left(\sum_{i+j=d_1} c_{ij} \varepsilon^i\right) Y^{d_1}$ in $\tilde{F}_1(X,Y,0)$ und $\left(\sum_{i+j=d_2} d_{ij} \varepsilon^i\right) Y^{d_2}$ in $\tilde{F}_2(X,Y,0)$).

13.) 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 Bem. 6). Sei $R(x) := R(f_1, f_2) \in k[x]$.

14.) Für jedes x_i haben die Polynome $f_1(x_i, y), f_2(x_i, y) \in k[y]$ einen Faktor $y - y_i \in k[y]$ gemeinsam. Für die $x = x_i$ muss $R(x)$ also verschwinden: $R(x_i) = 0$, $i = 0, \dots, d_1 d_2$. 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 8.), (iii) $\Rightarrow$ (i). □

15) **Satz:** 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 $ggT(F_1, F_2) = 1$, und es seien $d_1 d_2 - 1$ viele Schnittpunkte von C_{F_1} und C_{F_2} gegeben.
Dann haben sie einen weiteren Schnittpunkt $\in \mathbb{P}^2(k)$ gemeinsam.

Bew.: Wie im Beweis von Satz 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 $x_1, \dots, x_{d_1 d_2 - 1}$ laut Vor., 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 $\bar{k}$ in mind. 2 Linearfaktoren zerfallen, die dann zu zwei weiteren Schnittpunkten mit gleicher x-Koordinate a führen würden, so dass es $\geq (d_1 d_2 - 1) + 2 > d_1 d_2$ viele Schnittpunkte geben müsste im $\lightning$ zu Satz 1.).] Also gibt es nur noch genau einen weiteren Schnittpunkt (a, y) von C_{F_1} und C_{F_2}. □

16) **Bsp.:** Sei k bel. Körper, $F_1, F_2 \in k[X,Y,Z]$ homogen, $\deg ggT(F_1, F_2) = 0$, und sei $\deg F_1 = 1$, $\deg F_2 = 3$. Dann ist $\sum_{P \in C_{F_1} \cap C_{F_2}} m(P, C_{F_1}, C_{F_2}) \in \{0, 1, 3\}$.

Stichworte: • Def. elliptische Kurve, lange Weierstraßform
• $\mathcal{O} = [0:1:0]$ liegt auf allen elliptischen Kurven in langer Weierstraßform
• Kurze Weierstraßform für char $k \neq 2$ und für char $k \neq 2$ und $\neq 3$ mit Beweis
• Def. j-Invariante und Diskriminante

---

§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.

1.) **Def.:** Eine elliptische Kurve $E(k)$ ist eine nichtsinguläre, irreduzible projektive Kurve vom Grad 3, die einen (k-rationalen) Wendepunkt enthält.

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

3) **Bem.:** Durch eine sogenannte birationale Transformation kann angenommen werden, dass der Wendepunkt $\in \mathcal{O} := [0:1:0]$ ist. Eine Übungsaufgabe zeigt, dass dann die Kurven-gleichung die folgende vereinfachte Form hat:

4.) **Def.:** Eine elliptische Kurve $E_F(k)$ ist eine nicht-singuläre, projektive ebene Kurve $C_F(k) \subseteq \mathbb{P}^2(k)$, wobei F ein homogenes Polynom vom Grad 3 der Form

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

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

4) **Bem.:** • Die Monome X^2Y, Y^3, XY^2 brauchen also nicht vorzukommen.
• Die Numerierung der Koeffizienten ist historisch bedingt.
• Die affine Version lautet also:

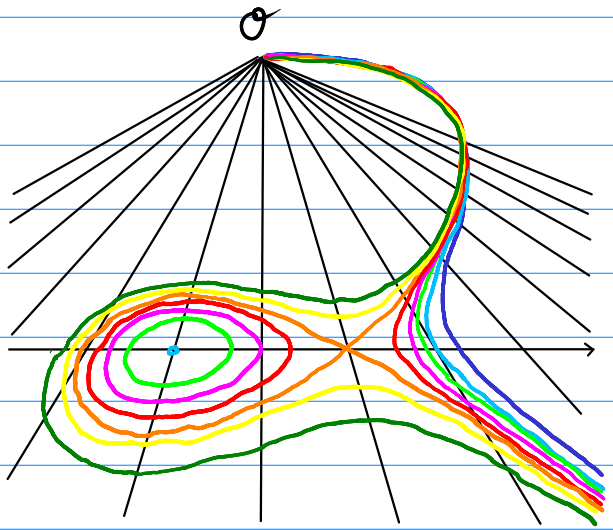
$$\circledast_{\text{affin}}: \quad Y^2 + a_1XY + a_3Y = X^3 + a_2X^2 + a_4X + a_6.$$

Die Form $\circledast$ nennen wir auch die lange Weierstraßform, das Polynom heißt langes Weierstraßpolynom.
• Wir werden sehen, dass man dies auf eine noch einfachere Form bringen kann.

5.) Bem.: Welche Punkte liegen auf $E(k)$, die nicht affin sind?
Ist $P=[x:s:0] \in \mathbb{P}^2(k) \setminus \mathbb{A}^2(k)$ ein solcher Punkt, dann ergibt Einsetzen in ⊛ dann $x^3=0$, dann muss $s\neq 0$ sein, d.h. $P=[0:s:0]=[0:1:0]$.
Diesen unendlich fernen Punkt, der allen elliptischen Kurven gemeinsam ist, nennen wir $\mathcal{O} := [0:1:0]$ ("Oh").
Dieser Punkt ist nie singulär, da $\frac{\partial F}{\partial z}(0,1,0)=1\neq 0$.
Somit genügt es, ein Polynom F der Form ⊛ die Nichtsingularität auf $C_F(k) \cap i(\mathbb{A}^2(k))$, also im Affinen zu testen.

6.) Bsp.: Sei $F(X,Y,Z)=Y^2Z-X^3-XZ$, für dieses gilt $a_1=a_2=a_3=a_6=0$
Dann ist $C_F(\mathbb{F}_p) \cap \mathbb{A}^2(\mathbb{F}_p)$ für $p\geq 3$ nicht-singulär, also eine ell. Kurve.

7.) Veranschaulichung, dass z.B. 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:



Parameterwerte:
$s=5$
$s=3$
$s=2$
$s=1$
$s=0$
$s=-1$
$s=-1.999$
$s=-5$

Das Bild ist perspektivisch so verzerrt, dass der unendlich ferne Punkt $\mathcal{O}=[0:1:0]$, der für die Richtung der y-Achse steht, am Horizont erscheint. (Das Zittern in den Kurven ist vom Abmalen per Hand.)

Vereinfachte Weierstraßgleichungen:

8.) **Satz:** Sei $E_F(k)$ eine elliptische Kurve mit
$$F(X,Y,Z) = Y^2Z + a_1XYZ + a_3YZ^2 - X^3 - a_2X^2Z - a_4XZ^2 - a_6Z^3.$$
(i) Falls $\mathrm{char}\, k \neq 2$, ist die Abb.
$$\Phi: \mathbb{P}^2(k) \to \mathbb{P}^2(k)$$
$$[r:s:t] \mapsto [r : s + \tfrac{a_1}{2}r + \tfrac{a_3}{2}t : t]$$ bijektiv und es ist
$\Phi(E_F(k)) = E_{H_1}(k)$ ebenfalls eine elliptische Kurve mit $H_1(X,Y,Z) = Y^2Z - X^3 - \frac{1}{4}b_2X^2Z - \frac{1}{2}b_4XZ^2 - \frac{1}{4}b_6Z^3$,
wobei $b_2 = a_1^2 + 4a_2$, $b_4 = 2a_4 + a_1a_3$, $b_6 = a_3^2 + 4a_6$.

(ii) Falls $\mathrm{char}\, k \neq 2$ und $\mathrm{char}\, k \neq 3$, ist die Abb.
$$\Psi: \mathbb{P}^2(k) \to \mathbb{P}^2(k)$$
$$[r:s:t] \mapsto [36r + 3b_2t : 216s : t]$$ 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 $c_4 = b_2^2 - 24b_4$, $c_6 = -b_2^3 + 36b_2b_4 - 216b_6$.

9.) **Bem.:** Wir können die lange Weierstraßgleichung im Fall $\mathrm{char}\, k \neq 2$ also stets zur affinen Glg. $y^2 = x^3 + a_2x^2 + a_4x + a_6$ 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 Glg. die kurze Weierstraßgleichung, das entsprechende Polynom dann das kurze Weierstraßpolynom.

10.) **Bem.:** Auch im Fall $\mathrm{char}\, k = 2$ lässt sich die lange Weierstraßgleichung vereinfachen, das ist nicht schwer, wenn $a_1 \neq 0$, aber auch für $a_1 = 0$ möglich. Wir behandeln das hier nicht näher.

11.) **Bew.:** Zu (i): Φ macht als Abb. nur Sinn, wenn 2 invertierbar in k ist, d.h. falls $\mathrm{char}\, k \neq 2$ ist. Φ ist dann bijektiv, da Φ die Umkehrabb. $\Phi^{-1}([r:s:t]) = [r : s - \frac{a_1}{2}r - \frac{a_3}{2}t : t]$ hat.
(klar: $\Phi^{-1}(\Phi([r:s:t])) = \Phi^{-1}([r : s + \frac{a_1}{2}r + \frac{a_3}{2}t : t]) = [r:s:t]$ ✓)

- Weiter bezeichnen wir mit Φ, Φ^{-1} auch die zugehörigen (affinen) Abbildungen $\Phi, \Phi^{-1}: k^3 \to k^3$, $\Phi(r,s,t) = (r, s + \frac{a_1}{2}r + \frac{a_3}{2}t, t)$
bzw. $\Phi^{-1}(r,s,t) = (r, s - \frac{a_1}{2}r - \frac{a_3}{2}t, t)$.

Nun können wir nachrechnen, dass $H_1(X,Y,Z) = F(X, Y - \frac{a_1}{2}X - \frac{a_3}{2}Z, Z)$:

$$\begin{aligned}
&= (Y - \tfrac{a_1}{2}X - \tfrac{a_3}{2}Z)^2 Z + a_1 X (Y - \tfrac{a_1}{2}X - \tfrac{a_3}{2}Z)Z + a_3 (Y - \tfrac{a_1}{2}X - \tfrac{a_3}{2}Z)Z^2 \\
&\quad - X^3 - a_2 X^2 Z - a_4 X Z^2 - a_6 Z^3 \\
&= Z \cdot \Big[Y^2 - 2Y\left(\tfrac{a_1}{2}X + \tfrac{a_3}{2}Z\right) + \left(\tfrac{a_1^2}{4}X^2 + 2 \cdot \tfrac{a_1 a_3}{4} XZ + \tfrac{a_3^2}{4}Z^2\right) \Big] \\
&\quad + a_1 XYZ - \tfrac{a_1^2}{2}X^2 Z - \tfrac{a_1 a_3}{2} XZ^2 + a_3 YZ^2 - \tfrac{a_1 a_3}{2} XZ^2 - \tfrac{a_3^2}{2}Z^3 \\
&\quad - X^3 - a_2 X^2 Z - a_4 XZ^2 - a_6 Z^3 \\
&= Y^2 Z - X^3 + \left(-\tfrac{a_1^2}{4} - a_2\right) X^2 Z + \left(-\tfrac{a_1 a_3}{2} - a_4\right) XZ^2 + \left(-\tfrac{a_3^2}{4} - a_6\right) Z^3 \\
&=: Y^2 Z - X^3 - \tfrac{1}{4} b_2 X^2 Z - \tfrac{1}{2} b_4 XZ^2 - \tfrac{1}{4} b_6 Z^3 = L.S.
\end{aligned}$$

mit den im Satz angegebenen Zahlen b_2, b_4, b_6. ⌋

- Es folgt $H_1(r,s,t) = F(\Phi^{-1}(r,s,t))$, also gilt: $F(r,s,t) = 0 \Leftrightarrow H_1(\Phi(r,s,t)) = 0$, so dass $\Phi(E_F(k)) = C_{H_1}(k)$ folgt. Es bleibt z.z., daß $C_{H_1}(k)$ nicht-singulär ist: Mit der Kettenregel (vgl. V7 Satz 4.1) rechnen wir nach:

$$\frac{\partial H_1}{\partial X}(r,s,t) = \frac{\partial F}{\partial X}(\Phi^{-1}(r,s,t)) - \frac{a_1}{2}\frac{\partial F}{\partial Y}(\Phi^{-1}(r,s,t)), \quad \frac{\partial H_1}{\partial Y}(r,s,t) = \frac{\partial F}{\partial Y}(\Phi^{-1}(r,s,t)),$$

$$\frac{\partial H_1}{\partial Z}(r,s,t) = -\frac{a_3}{2}\frac{\partial F}{\partial Y}\Phi^{-1}(r,s,t) + \frac{\partial F}{\partial Z}(\Phi^{-1}(r,s,t)).$$

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

Zu (ii): Ψ hat die Inverse $[r:s:t] \mapsto [\frac{1}{36}r - \frac{b_2}{12}t : \frac{1}{216}s : t]$, da wegen $\mathrm{char}\, k \neq 2, \neq 3$ die Zahlen $\frac{1}{36}, \frac{1}{12}, \frac{1}{216} = \frac{1}{2^3 3^3}$ in k existieren, und leicht zu bestätigen ist, dass $\Psi(\Psi^{-1}([r:s:t])) = [r:s:t]$ gilt.

Durch geduldiges Nachrechnen zeigt man $H_2(X,Y,Z) = 2^6 3^6 H_1(\frac{1}{36}X - \frac{b_2}{12}Z, \frac{1}{216}Y, Z)$.

Daraus folgt: $H_1(r,s,t) = 0 \Leftrightarrow H_2(\Psi(r,s,t)) = 0$, d.h. $\Psi(E_{H_1}(k)) = C_{H_2}(k)$.

Wieder mit der Kettenregel kann auch die Nicht-Singularität von $C_{H_2}(a)$ gezeigt werden. □

Wir definieren zwei wichtige Kennzahlen projektiver Kurven wie folgt.

12) **Def.:** Sei $C_F(k)$ die projektive ebene Kurve zum langen Weierstraßpolynom

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

- Dann heißt die Zahl

$$\Delta = \Delta(C_F(k)) = -b_2^2 b_8 - 8b_4^3 - 27b_6^2 + 9b_2b_4b_6$$

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

und $b_8 = a_1^2a_6 + 4a_2a_6 - a_1a_3a_4 + a_2a_3^2 - a_4^2$

die Diskriminante der Kurve $C_F(k)$.

- Die Zahl

$$j = j(C_F(k)) := \frac{(b_2^2 - 24b_4)^3}{\Delta} = \frac{c_4^3}{\Delta}$$ heißt die j-Invariante der Kurve $C_F(k)$.

13) **Bem.:** • Die j-Invariante legt die Isomorphieklasse der elliptischen Kurve über $\bar{k}$ fest: Zwei elliptische Kurven sind isomorph über $\bar{k}$ genau dann wenn sie dieselbe j-Invariante besitzen. [ohne Bew.]

- j ist unabh. von der Wahl der speziellen Kurvengleichung.

14) **Bem.:** Die Diskriminante einer Kurve $C_F(k)$ ist ein nützliches Hilfsmittel um zu testen, ob eine Kurve, die durch eine lange Weierstraßgleichung gegeben ist, nicht-singulär (und damit elliptisch) ist:

15) **Satz:** Sei die Kurve $C_F(k)$ gegeben durch das lange Weierstraßpolynom F.

Dann ist $C_F(k)$ nicht-singulär genau dann, wenn $\Delta(C_F(k)) \neq 0$ 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 in Vorlesung V12.

Stichworte: • Diskriminantenkriterium: C_F in W-Form nicht-singulär $\Leftrightarrow \Delta(C_F(\bar{k})) \neq 0$
• Beweis unterscheidet $\mathrm{char}\, k = 2$, $\mathrm{char}\, k = 3$, und sonst • Diskriminante eines Polynoms

2.4.2 Das Diskriminantenkriterium

1.) **Def.:** Sei $C_F(k)$ die projektive ebene Kurve zum langen Weierstraßpolynom

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

Dann heißt die Zahl

$$\Delta = \Delta(C_F(k)) = -b_2^2 b_8 - 8b_4^3 - 27b_6^2 + 9 b_2 b_4 b_6$$

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

und $b_8 = a_1^2 a_6 + 4a_2a_6 - a_1a_3a_4 + a_2a_3^2 - a_4^2$

die **Diskriminante** der Kurve $C_F(k)$.

2.) **Bem.:** Im Fall einer Kurve $C_F(k)$ in einer kurzen Weierstraßform

$f(x,y) = y^2 - x^3 - ax - b$ haben wir $\Delta(C_F(k)) = -8(2a)^3 - 27\cdot(4b)^2 = -16(4a^3 + 27b^2)$,

da $a_1 = 0$, $a_3 = 0$, $a_2 = 0$, $a_4 = a$, $a_6 = b$ $\rightsquigarrow$ $b_2 = 0$, $b_4 = 2a$, $b_6 = 4b$, $b_8 = -a^2$, vgl. Übungsaufgabe 3 Blatt 4.

Wir zeigen das **Diskriminantenkriterium:**

3.) **Satz:** Sei die Kurve $C_F(k)$ gegeben durch das lange Weierstraßpolynom F.
Dann ist $C_F(k)$ nicht-singulär genau dann, wenn $\Delta(C_F(k)) \neq 0$ ist.

4.) **Bem.:** • Bei diesem Kriterium, wenn $\Delta \neq 0$, erhalten wir, dass $C_F(\bar{k})$ über dem a/g. Abschluss $\bar{k}$ keine singulären Punkte enthält (insb. auch über k, aber über $\bar{k}$ ist eben noch stärker). Deswegen haben wir uns bei unserer Def. von "nicht-singulärer Kurve" auf $\bar{k}$ bezogen, was wegen Satz 3.) also mathematisch leichter wird. Für $\mathrm{char}\, k \neq 2$ kann es aber nicht sein, dass C_F über k keine singulären Punkte hat und über $\bar{k}$ hingegen schon.

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

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

5.) **Bew.:** Die Kurve $C_f(k)$ ist nicht-singulär genau dann, wenn ihre affine Kurve $C_f(k)$ mit $f(x,y) = y^2 + a_1xy + a_3y - x^3 - a_2x^2 - a_4x - a_6$ nicht-singulär ist. Wir zeigen den Satz deswegen im Affinen (der einzige nicht-affine Punkt $O = [0:1:0]$ der Kurve ist immer regulär, vgl. V11_Bem.5.)).

Nun enthält $C_f(\bar{k})$ einen singulären Punkt genau dann, wenn $\exists\, r,s \in \bar{k}$: $f(r,s) = 0$, $\frac{\partial f}{\partial x}(r,s) = 0$, $\frac{\partial f}{\partial y}(r,s) = 0$ gilt

$\frac{\partial f}{\partial x}(r,s) = a_1 s - 3r^2 - 2a_2 r - a_4$, $\frac{\partial f}{\partial y}(r,s) = 2s + a_1 r + a_3$

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

6.) **1. Fall:** $\mathrm{char}\, k = 2$ und $a_1 = 0$.

„$\Leftarrow$": Dann ist hier $b_2 = b_4 = 0$, $b_6 = a_3^2$, $\Delta = -27 a_3^4 = a_3^4$.
Weiter gilt $\frac{\partial f}{\partial y} = a_3$, so dass, falls ein sing. Pkt. ex., dann $a_3 = 0$, also $\Delta = 0$ folgt.

„$\Rightarrow$": Ist $\Delta = 0$, folgt $\frac{\partial f}{\partial y} = 0$. Nun ex. $r,s \in \bar{k}$ mit $r^2 + a_4 = 0$, $s^2 + a_3 s = r^3 + a_2 r^2 + a_4 r + a_6$, also ist $(r,s) \in \mathbb{A}^2(\bar{k})$ singulärer Punkt auf $C_f(\bar{k})$.

7.) **2. Fall:** $\mathrm{char}\, k = 2$ und $a_1 \neq 0$.

Wir haben in Charakteristik 2, dass gilt:

$$\Delta = -a_4^4(a_1^2 a_6 - a_1 a_3 a_4 + a_2 a_3^2 - a_4^2) - 27 a_3^4 + a_1^3 a_3^3$$
$$= a_1^6 a_6 + a_1^5 a_3 a_4 + a_1^4 a_2 a_3^2 + a_1^4 a_4^2 + a_1^3 a_3^3 + a_3^4.$$

„$\Leftarrow$":
Hat $C_f(\bar{k})$ einen singulären Pkt., ex. $r,s \in \bar{k}$ mit $f(r,s) = 0$,
d.h. $a_1 s + r^2 + a_4 = 0$ und $a_1 r + a_3 = 0$.
Da $a_1 \neq 0$, folgt

$$r = \frac{a_3}{a_1}, \quad s = \frac{a_3^2 + a_1^2 a_4}{a_1^3}.$$

Durch Einsetzen in $f(r,s)$ folgt $0 = f(r,s) = \Delta a_1^{-6}$, also $\Delta = 0$.

$f(r,s) = s^2 + \underbrace{a_1 rs + a_3 s}_{=0} + r^3 + a_2^2 r^2 + a_4 r + a_6 = a_1^{-6}(a_3^4 + a_1^4 a_4^2 + a_1^3 a_3^3 + a_2 a_1^4 a_3^2 + a_4 a_1^5 a_3 + a_6 a_1^6)$

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

8.) 3. Fall: $\operatorname{char} k = 3$.

Via Rechnen in Charakteristik 3 folgt $\Delta = -b_2^2 b_8 - 8 b_4^3$.

Betr. die A66. $\overline{\phi} : C_F(k) \to C_{H_1}(k)$ aus V11_Satz 8.) (i)

mit $H_1(X,Y,Z) = Y^2Z - X^3 - \frac{1}{4} b_2 X^2 Z - \frac{1}{2} b_4 X Z^2 - \frac{1}{4} b_6 Z^3$,

die die lange Weierstraßform F auf die kurze Form H_1 bringt.

Es ist $\Delta(C_{H_1}(k)) = \Delta(C_F(k))$ durch Nachrechnen, somit genügt es OE, das Diskriminantenkriterium für die kurze Form H_1 zu zeigen.

9.) Die Kurve $C_{H_1}(\bar{k})$ enthält genau dann einen singulären Punkt, wenn es $r, s \in \bar{k}$ gibt mit

$$s^2 - r^3 - \tfrac{1}{4} b_2 r^2 - \tfrac{1}{2} b_4 r - \tfrac{1}{4} b_6 = 0, \quad 3r^2 + \tfrac{1}{2} b_2 r + \tfrac{1}{2} b_4 = 0, \quad 2s = 0,$$

d.h. falls r eine doppelte Nst. des Polynoms $\sigma(X) := X^3 + \frac{1}{4} b_2 X^2 + \frac{1}{2} b_4 X + \frac{1}{4} b_6$ ist, sprich $\sigma(r) = 0 = \sigma'(r)$ ist.

über $\bar{k}$ zerfällt σ in 3 Linearfaktoren: $\sigma(X) = (X-\alpha_1)(X-\alpha_2)(X-\alpha_3)$, die $\alpha_i \in \bar{k}$.

Nun hat ein Polynom σ genau dann eine doppelte Nst., falls seine Diskriminante $\operatorname{disc}(\sigma) := (\alpha_1-\alpha_2)^2(\alpha_1-\alpha_3)^2(\alpha_2-\alpha_3)^2$ verschwindet, vgl. unten Bem. 14.)

10.) Somit ist im 3. Fall z.z.: $\Delta = 0 \iff \operatorname{disc}(\sigma) = 0$.

Wegen $\operatorname{disc}(X^3 + uX^2 + vX + w) = u^2v^2 - 4u^3w - 4v^3 - 27w^2 + 18uvw$, vgl. Kor. 16.),

erhalten wir mit $u = \frac{b_2}{4}$, $v = \frac{b_4}{2}$, $w = \frac{b_6}{4}$ dann in Charakteristik 3, dass

$$\operatorname{disc}(\sigma) = \tfrac{1}{64} b_2^2 b_4^2 - \tfrac{1}{64} b_2^3 b_6 - \tfrac{1}{2} b_4^3.$$

Wegen $4 b_8 = b_2 b_6 - b_4^2$

[r.S. $= (a_1^2 + 4a_2)(a_3^2 + 4a_6) - (2a_4 + a_1 a_3)^2 = \ldots = 4 b_8$ ✓]

$= a_1^2 a_3^2 + 4 a_1^2 a_6 + 4 a_2 a_3^2 + 4^2 a_2 a_6 - (4 a_4^2 + 4 a_4 a_1 a_3 + a_2^2 a_3^2) = 4 b_8 =$ l.S.

erhalten wir $\operatorname{disc}(\sigma) = \frac{1}{16}(-b_2^2 b_8 - 8 b_4^3) = \frac{1}{16}\Delta$.

Aus dieser Formel folgt die Beh. im 3. Fall.

11.) 4. Fall: $\operatorname{char} k > 3$, d.h. $\operatorname{char} k \geq 5$, oder $\operatorname{char} k = 0$.

Mit der Bijektion $\Psi \circ \Phi : C_F(k) \to C_{H_2}(k)$ zum kurzen Weierstraßpolynom

$$H_2(X,Y,Z) = Y^2Z - X^3 + 27 c_4 X Z^2 + 54 c_6 Z^3$$

bzw. $h_2(x,y) = y^2 - x^3 + 27 c_4 x + 54 c_6$,

wo $c_4 = b_2^2 - 24 b_4$, $c_6 = -b_2^3 + 36 b_2 b_4 - 216 b_6$,

folgt durch Untersuchung der Ableitungen wieder:

$C_F(k)$ nicht-sing. $\Leftrightarrow$ $C_{H_2}(k)$ nicht-sing.

Wir berechnen

$$\Delta(C_{H_2}(k)) = 2^6 3^9 (c_4^3 - c_6^2) = \ldots = 2^{12} 3^{12} \Delta(C_F(k)),$$

also genügt es, die Beh. für $C_{H_2}(k)$ zu zeigen.

Wie im 3. Fall:

$C_{H_2}(\bar{k})$ enthält sing. Pkt. $\Leftrightarrow$ $\sigma(x) = x^3 - \underbrace{27 c_4}_{v} x - \underbrace{54 c_6}_{w}$ hat doppelte Nst. $\Leftrightarrow \operatorname{disc}(\sigma) = 0$

Wegen der Formel in Kor. 16.) für $\operatorname{disc}(\sigma)$ folgt mit $u = 0$, $v = -27 c_4$, $w = -54 c_6$:

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

Daraus folgt die Beh. $\square$

Theoretische Ergänzungen zum Begriff "Diskriminante":

12.) Def.: Die Diskriminante eines Polynoms $\sigma \in k[X]$, $n := \deg \sigma \geq 1$, ist

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

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

13.) **Bem.:** Man vgl. dies mit der Diskriminante p^2-4q eines quadratischen Polynoms $\sigma(x) = x^2+px+q \in k[X]$, wir haben $\alpha_{1/2} = -\frac{p}{2} \pm \sqrt{\frac{p^2}{4}-q} = -\frac{p}{2} \pm \frac{1}{2}\sqrt{\underbrace{p^2-4q}_{\text{disc}(\sigma)}}$
also genau $(\alpha_1-\alpha_2)^2 = \text{disc}(\sigma)$.
Ist dies $=0$, ist $\alpha_1=\alpha_2$ eine doppelte Nullstelle von σ.

14.) **Bem.:** $\text{disc}(\sigma)$ verschwindet genau dann, wenn σ über $\bar{k}$ eine doppelte Nullstelle hat.
Dies folgt unmittelbar aus der Def. von σ.

15.) **Bem.:**
- Ist $\sigma \in k[X]$, $n=\deg\sigma \geq 1$, ein normiertes Polynom, kann die Beziehung
$$\text{disc}(\sigma) = (-1)^{n(n-1)/2}\,\text{Res}(\sigma,\sigma')$$
mit der aus V10_Def.5.) behandelten Resultante gezeigt werden.
- Aus dieser wichtigen Formel folgt wegen unser Definition für die Resultante, dass stets $\text{disc}(\sigma) \in k$ gilt.

16.) **Kor.:** Es gilt $\text{disc}(x^3+ux^2+vx+w) = u^2v^2-4vu^3-4v^3-27w^2+18uvw$.

Bew.: Wir haben $M(\sigma,\sigma') = \begin{bmatrix} \textcircled{w} & 0 & \textcircled{v} & 0 & 0 \\ v & w & 2u & v & 0 \\ u & v & 3 & 2u & v \\ 1 & u & 0 & 3 & 2u \\ 0 & 1 & 0 & 0 & \textcircled{3} \end{bmatrix}$ $\quad \sigma'(x) = 3x^2+2ux+v$

$$\Rightarrow \text{Res}(\sigma) = \det M(\sigma,\sigma') = w \det\begin{bmatrix} w & 2u & v & 0 \\ v & 3 & 2u & v \\ u & 0 & 3 & 2u \\ 1 & 0 & 0 & 3 \end{bmatrix} + v \det\begin{bmatrix} v & w & v & 0 \\ u & v & 2u & v \\ 1 & u & 3 & 2u \\ 0 & 1 & 0 & 3 \end{bmatrix}$$

$$= -w\det\begin{bmatrix} 2u & v & 0 \\ 3 & 2u & v \\ 0 & 3 & 2u \end{bmatrix} + w\cdot 3\det\begin{bmatrix} w & 2u & v \\ v & 3 & 2u \\ u & 0 & 3 \end{bmatrix}$$

$$+ v\det\begin{bmatrix} v & v & 0 \\ u & 2u & v \\ 1 & 3 & 2u \end{bmatrix} + v\cdot 3\det\begin{bmatrix} v & w & v \\ u & v & 2u \\ 1 & u & 3 \end{bmatrix}$$

$$= -w\left(2u(2u\cdot 2u-3v) - v\cdot 6u\right) + 3w\left(u(2u\cdot 2u-3v) + 3(3w-2uv)\right)$$
$$+ v\left(v(2u\cdot 2u-3v) - v(u\cdot 2u-v)\right) + 3v\left(v(3v-2u^2) - wu+vu^2-v^2\right)$$
$$= -w\cdot 8u^3 + 6uvw + 6uvw + 12wu^3 - 9uvw + 27w^2 - 18uvw$$
$$+ v^2\cdot 4u^2 - 3v^3 - 2v^2u^2 + v^3 + 9v^3 - 6v^2u^2 - 3uvw + 3u^2v^2 - 3v^3$$
$$= 4vu^3 - u^2v^2 + 27w^2 + 4v^3 - 18uvw \quad \checkmark$$

Jetzt: $(-1)^{3\cdot(3-1)/2} = -1$ beachten. $\square$

Stichworte:
- Definition Punktaddition $P+Q$ auf $E(k)$ mit 3. Schnittpunkt $P*Q$ von $G(P,Q) \cap E(k)$
- $\mathcal{O}$ ist neutrales Element • Invertieren leicht bei Kurzer Weierstraßform • explizite Formeln für "+"

2.4.3 Die Gruppenstruktur elliptischer Kurven

Sei $E(k)$ eine elliptische Kurve über einem Körper k. (Da wir auch über Tangenten sprechen möchten, muss die Kurve nicht-singulär sein.)

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

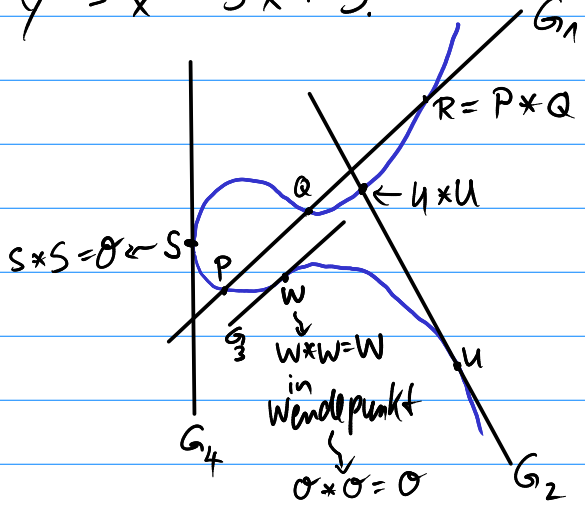
(b) Sei G die Tangente an $E(k)$ im Punkt $P \in E(k)$.
Dann hat G noch einen dritten Schnittpunkt mit $E(k)$ gemäß Vielfachheiten gezählt (d.h. ev. P selbst, falls $m(P; G, E(k)) = 3$).

2.) **Bew.:** Als Ergänzung zum Satz von Bézout haben wir Satz 15.) kennengelernt, der im Spezialfall $\deg F_1 = 1$, $\deg F_2 = 3$ dann $\sum_{P \in C_{F_1} \cap C_{F_2}} m(P, C_{F_1}, C_{F_2}) \in \{0,1,3\}$ liefert (s. V10_Bsp. 16.)).
Also gilt auch hier: $\sum_{R \in G \cap E(k)} m(R; G, E(k)) \in \{0,1,3\}$.

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

zu (b): Ist G die Tangente an $E(k)$ in $P \in E(k)$, ist $m(P; G, E(k)) \geq 2$ nach VO_Satz 20.). Es folgt wie im Beweis zu (a) wieder, dass die Vielfachensumme $= 3$ ist, also die Existenz eines $R \in G \cap E(k)$ mit $R \neq P$, falls $m(P; G, E(k)) = 2$ und $R = P$, falls $m(P; G, E(k)) = 3$ gilt. □

3) **Bsp.:** Betr. die elliptische Kurve $E(k)$ zur (kurzen) Weierstraßgleichung
$y^2 = x^3 - 3x + 3$. (OK, da $4\cdot(-3)^3 + 27\cdot 3^2 \neq 0$)



Jede Gerade, die $E(k)$ in zwei Punkten schneidet, schneidet $E(k)$ in einem dritten Punkt, gemäß Vielfachheiten gezählt. Der dritte Schnittpunkt kann auch $O \in g_\infty$ sein.

Wir möchten auf $E(k)$ eine Verknüpfung "+" erklären, also eine Punktaddition $P+Q$, bei der wiederum ein Punkt auf der elliptischen Kurve herauskommt. Den dritten Schnittpunkt, den die Gerade $G(P,Q)$ durch zwei Punkte P und Q auf $E(k)$ mit $E(k)$ hat (laut Satz 1.), bezeichnen wir mit $P*Q$.

4) **Def.:** Für $P, Q \in E(k)$, $P \neq Q$, definieren wir also

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

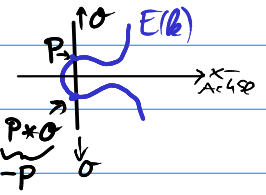
sowie $P*P := \begin{cases} R \in (T_P(E(k)) \cap E(k)) \setminus \{P\}, \text{ falls } m(P; T_P(E(k)), E(k)) = 2, \\ P, \text{ falls } m(P; T_P(E(k)), E(k)) = 3 \quad (\text{d.h. falls } P \text{ Wendepunkt}). \end{cases}$

5) **Bem.:** • Der unendlich ferne Punkt $O \in E(k)$ erfüllt $O*O = O$, da er ein Wendepunkt ist. • Weiter ist offensichtlich $P*Q = Q*P$ aufgrund der Def.
• Es gilt: $R = P*Q \Rightarrow P = Q*R \Rightarrow Q = R*P$, d.h. $P*(P*Q) = Q$ (*)
für alle $P, Q \in E(k)$.
• Es gilt: $P*Q = P*R \Leftrightarrow Q = R$, denn "$\Leftarrow$", "$\Rightarrow$": Vor. $\rightsquigarrow P*(P*Q) = P*(P*R) \overset{(*)}{\Rightarrow} Q = R$.

6.) **Bem.:** Man beachte, dass für $P=[a:b:1] \in \mathbb{P}^2(k) \cap i(\mathbb{A}^2(k))$ die Gerade

$$G(P,O) = G(c,0,-a) = \{[x:y:z] \in \mathbb{P}^2(k);\ x - az = 0\}$$

im Affinen eine Parallele zur y-Achse darstellt (Glg. $x = a$).
Für eine elliptische Kurve, die durch eine Kurze Weierstraßform gegeben und (für $\mathrm{char}\, k \neq 2$) symmetrisch zur x-Achse ist, wird typischerweise $P * O \neq P$ sein.
Wegen $(O*O)*P = O*P$ einerseits, da $O*O = O$,
und $O*(O*P) = P$ andererseits, wegen ⊠,
kann die Verknüpfung $*$ also nicht assoziativ sein.
Stattdessen setzen wir unsere Verknüpfung "+" wie folgt:



7.) **Def.:** Für $P, Q \in E(k)$ definieren wir

$$P + Q := O * (P * Q).$$

Ist $E(k)$ in Kurzer Weierstraßform und (für $\mathrm{char}\, k \neq 2$) symmetrisch zur x-Achse, erhält man $P+Q$, indem man den 3. Schnittpunkt $P*Q$ von $G(P,Q)$ mit $E(k)$ dann noch an der x-Achse spiegelt, d.h. das Negative des y-Wertes nimmt.

8.) **Bem.:** • Es gilt: $O + P = P$, d.h. O ist neutrales Element von "+",
weil $O + P = O * (O * P) = P$ nach ⊠ gilt.

• Es gilt: $-P = O * P$, da $P + (O*P) = O * (P * (O*P))$

$$= (P * (O * P)) * O = (P * (P * O)) * O \overset{⊠}{=} O * O = O.$$

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

9.) **Lemma:** Liegen drei Punkte $P, Q, R \in E(k)$ einer elliptischen Kurve auf einer projektiven Geraden G, so gilt $(P+Q)+R = O$, und umgekehrt.
Dabei müssen P, Q, R nicht notwendig verschieden sein.

Bew.: Da $R = P*Q$ nach Vor., folgt $\underbrace{O*R}_{-R} = \underbrace{O*(P*Q)}_{P+Q}$, also
$-R = P+Q$ bzw. $R + (P+Q) = O$.

Umgekehrt gilt das ebenso. □

10.) **Satz (von Poincaré, 1901)**: Sei k bel. Körper und $E(k)$ elliptische Kurve. Die Verknüpfung $+ : (P, Q) \mapsto P+Q$ macht $E(k)$ zu einer abelschen Gruppe $(E(k), +)$ mit neutralem Element $\mathcal{O}$, d.h. (i) $P + \mathcal{O} = P$ für alle $P \in E(k)$,
(ii) $\forall P \in E(k)\ \exists\, Q \in E(k)$: $P+Q = \mathcal{O} \rightsquigarrow -P := Q$
(iii) $P+Q = Q+P$ für alle $P, Q \in E(k)$
(iv) $(P+Q)+R = P+(Q+R)$ für alle $P, Q, R \in E(k)$.

Bem.: Schwierig zu beweisen ist die Assoziativität in (iv). Wir behandeln diese im Vorlesungsteil V14.

Bew. von (i)–(iii): Zu (i): s.o. Bem. 8.) ✓
Zu (ii): Sei $P \in E(k)$, setze $Q := \mathcal{O} * P$, s.o. Bem. 8.) ✓
Zu (iii): Klar aufgrund der Def von $P+Q$. ✓ □

11.) **Bem.**: • Anstelle $\mathcal{O}$ könnte man prinzipiell jeden Punkt $Q \in E(k)$ zum neutralen Element von $+$ "machen", indem man $U \boxplus V := U+V-Q$ setzt: dann ist $U \boxplus Q = U+Q-Q = U$, $U \boxplus (-U+2Q) = U-U+2Q-Q = Q$, und $(U \boxplus V) \boxplus W = U+V+W-2Q = U \boxplus (V \boxplus W)$.
Die Eigenschaft von Lemma 9.) gilt immer noch, wenn für Q ein Wendepunkt genommen wird. Nun kann eine elliptische Kurve bis zu 9 Wendepunkten haben, vgl. Ⓤ Aufgabe 4(a) – Blatt 5.

• Die Wahl des W.P. $\mathcal{O} := [0:1:0]$ als neutralem Element von $E(k)$ hat den Vorteil, dass die explizite Formel für "+" rechnerisch einfacher wird, weil dann die Gleichung von $E(k)$ in einfacher (langer oder kurzer) Weierstraßform vorliegt.
Diese explizite Formel wird in Satz 13) / Satz 14.) angegeben.

• Invertieren, d.h. Berechnen von $-P = P * \mathcal{O}$, ist dann, bei kurzer Weierstraßform in $\mathrm{char}\, k \neq 2$, besonders leicht:
Man spiegelt P an der x-Achse und erhält $-P$, d.h. ist $P = [a:b:c]$, gilt $-P = [a:-b:c]$. Schnittpunkte von $E(k)$ mit der x-Achse sind dann selbstinvers, d.h. $P = [a:0:c] \in E(k) \Rightarrow P = -P$.

12.) Bsp.: Betr. $E(\mathbb{R})$ mit der Glg. $y^2 = x^3 + 17$ bzw. $Y^2Z = X^3 + 17Z^3$.
Dann liegen $P = [-1:4:1]$ und $Q = [-2:3:1]$ auf der Kurve.
Ihre Verbindungsgerade ist $G(P,Q) = \{[x:y:z] \in \mathbb{P}^2(\mathbb{R}) : x - y + 5z = 0\}$.
Um $P+Q$ zu berechnen, bestimmen wir den dritten Schnittpunkt $P*Q$ von $G(P,Q)$ und $E(\mathbb{R})$ wie folgt: Setze $y = x+5$ ein und erhalte $(x+5)^2 = x^3+17$
$\Leftrightarrow x^3 - x^2 - 10x - 8 = 0$. Da $x=-1$, $x=-2$ Nst. der l.S. sind, führt Polynomdiv. durch $(x+1)(x+2)$ zu $x^3 - x^2 - 10x - 8 = (x+1)(x+2)(x-4)$. Der Punkt $P*Q$ hat also x-Koordinate 4, da er auf G liegt, folgt: $P*Q = [4:9:1]$, es folgt $P+Q = [4:-9:1]$.

Ist $E(k)$ gegeben durch das lange Weierstraßpolynom
$$F(X,Y,Z) = Y^2Z + a_1XYZ + a_3YZ^2 - X^3 - a_2X^2Z - a_4XZ^2 - a_6Z^3$$
bzw. $f(x,y) = y^2 + a_1xy + a_3y - x^3 - a_2x^2 - a_4x - a_6$, so möchten wir die Addition "+" für die Krypto-Anwendungen in einer expliziten Formel beschreiben:

13.) Satz: Sei $E(k)$ gegeben durch das lange Weierstraßpolynom F. Dann gilt:
(a) $P = (u,v) \in E(k) \cap A^2(k) \Rightarrow -P = (u, -v - a_1u - a_3)$
(b) $P = (u,v)$, $Q = (r,s) \in E(k) \cap A^2(k)$
- ist $u = r$ und $v + s + a_1u + a_3 = 0$, gilt $P + Q = \mathcal{O}$,
- sonst ist
$$P+Q = \Big(\underbrace{\lambda^2 + a_1\lambda - a_2 - u - r}_{=:x},\ -(\lambda + a_1)x - \mu - a_3\Big),$$
wobei $\lambda = \frac{s-v}{r-u}$, $\mu = \frac{vr - su}{r-u}$ falls $r \neq u$,
und $\lambda = \frac{3u^2 + 2a_2u + a_4 - a_1v}{2v + a_1u + a_3}$, $\mu = \frac{-u^3 + a_4u + 2a_6 - a_3v}{2v + a_1u + a_3}$ falls $r = u$.

Bew.: wäre langweiliges Nachrechnen. □

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

14.) **Satz:** Ist $E(k)$ gegeben durch $f(x,y) = y^2 - x^3 - ax - b$, so gilt:

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

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

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

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

Bem.: Der zweite Fall in b) mit $P = Q$ ist die "Punktverdopplung" $P + P = 2P$.

Bew.: (a) √, (b): • Sei $P \neq -Q$, $P \neq Q$.

Haben $g(P,Q) = \{(u + t, v + \underbrace{\frac{s-v}{r-u}}_{\lambda :=} t);\ t \in \mathbb{R}\}$,

sowie $f(u + t, v + \lambda t) = (v + \lambda t)^2 - (u + t)^3 - a(u + t) - b$

mit den Nullstellen $t = 0$, $t = r - u$. Eine weitere Nullstelle ist $t = x - u$.

Die Nullstellensumme $0 + r - u + x - u = r + x - 2u$ ergibt den Koeffizienten vor t^2 des Polynoms, nämlich $\lambda^2 - 3u$, es folgt $x = \lambda^2 - u - r$,

die y-Koordinate des Punkts auf $g(P,Q)$ ist dann $v + \lambda(x - u)$, für $P + Q$ dann das Negative.

• Ist $Q = P$, $P \neq -P$ (d.h. $v \neq 0$), nimmt man die Tangente an $E(k)$ im Punkt P,

also $t_P(E(k)) = \{(x,y);\ (-3u^2 - a)x + 2vy + v^2 - 2au - 3b = 0\}$,

$= \{(u + t, v + \lambda t);\ t \in \mathbb{R}\}$ mit Steigung $\lambda := \frac{3u^2+a}{2v}$.

Der Rest der Rechnung geht wie eben, es folgt $x = \lambda^2$ (da $u = r$) und der angegebene y-Wert für $P + P$. □

Stichworte: • Assoziativgesetz für elliptische Kurven • Neunpunktesatz als Hilfsmittel
• Bemerkungen zu anderen Beweismöglichkeiten

2.4.4 Das Assoziativgesetz

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

1.) Ziel: Wir zeigen für eine elliptische Kurve $E(k)$ das Assoziativgesetz: $\forall P,Q,R \in E(k)$: $\underline{\underline{P+(Q+R) = (P+Q)+R}}$.

2.) Lemma: Das Assoziativgesetz auf $E(k)$ gilt genau dann, wenn $\forall P,Q,R,S \in E(k)$: $(P*Q)*(R*S) = (P*R)*(Q*S)$.

Bew.: Zunächst gilt

$$P+(Q+R) = \mathcal{O} * (P * (Q+R)), \qquad (P+Q)+R = \mathcal{O} * ((P+Q) * R).$$

wegen $\mathcal{O} * (\mathcal{O} * P) = P$ folgt:

$$\mathcal{O} * U = \mathcal{O} * V \iff U = V.$$

Also ist die Assoziativität äquivalent zur Gleichung

$$P * (Q+R) \stackrel{!}{=} (P+Q) * R,$$

Jetzt zur Beh. des Lemmas.

„$\Rightarrow$": wenn das Assoziativgesetz gilt, folgt mit $P*Q = -(P+Q)$

dann l.S. $= (P*Q)*(R*S) = -((P*Q)+(R*S))$
$= -(-(P+Q)-(R+S))$
$= (P+Q)+(R+S)$
$\stackrel{\text{Assoz.}}{=} (P+R)+(Q+S)$
$\stackrel{\text{analog}}{=} (P*R)*(Q*S) =$ r.S.

„$\Leftarrow$": Gilt die Relation, folgt mit $\tilde{P} = \mathcal{O}$, $\tilde{Q} = P*Q$, $\tilde{R} = Q*R$, $\tilde{S} = Q$

daraus $(\tilde{P}*\tilde{Q})*(\tilde{R}*\tilde{S}) = (\tilde{P}*\tilde{R})*(\tilde{Q}*\tilde{S})$,

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

was nach obigem die Assoziativität impliziert. □

Als Beweishilfsmittel benötigen wir den

3) **Neunpunktesatz**: Sei k ein algebraisch abgeschlossener Körper. Seien C_F, C_{F_1} und C_{F_2} drei kubische Kurven in $\mathbb{P}^2(k)$ zu paarweise teilerfremden homogenen Polynomen F, F_1, F_2 vom Grad 3, und C_F enthalte 8 der 9 Schnittpunkte von $C_{F_1} \cap C_{F_2}$. Dann liegt auch der 9. Schnittpunkt auf C_F.

4) Beweis (nicht vollständig, nur Beweisidee): Dass $C_{F_1} \cap C_{F_2}$ (mit Vielfachheiten gezählt) aus genau $9 = 3 \cdot 3$ Punkten besteht, besagt der Satz von Bézout, vgl. Satz 2.)_V10. Damit bestehen auch $C_F \cap C_{F_1}$ und $C_F \cap C_{F_2}$ aus 9 Punkten, von denen laut Vor. 8 Punkte identisch sind.

Eine allgemeine projektive kubische Kurve $\mathcal{C}$ (Singularitäten egal) ist definiert über 10 Koeffizienten:

$$\mathcal{C}: a_1X^3 + a_2X^2Y + a_3X^2Z + a_4XY^2 + a_5XYZ + a_6XZ^2 + a_7Y^3 + a_8Y^2Z + a_9YZ^2 + a_{10}Z^3 = 0$$

- Die Kurve verändert sich nicht, wenn das kubische Polynom mit einem Skalar $s \neq 0$ multipliziert wird. Deswegen sehen wir eine kubische Kurve als einen projektiven Punkt $\underline{a} = [a_1 : a_2 : a_3 \ldots : a_{10}] \in \mathbb{P}^9(k)$ an.

Betrachte nun alle kubischen Kurven, die durch die 8 vorgegebenen Punkte verlaufen. Sei $M \in k^{8\times 10}$ die Matrix, für die die nichttrivialen Lösungen des LGS $M \cdot \underline{a} = \underline{0} \in k^8$ genau die kubischen Kurven $\underline{a} \in \mathbb{P}^9(k)$ ergeben, die durch die 8 Punkte verlaufen. Dann ist $\operatorname{rg} M \leq 8$, also folgt wegen der Dimensionsformel $\dim \ker M = 10 - \operatorname{rg} M \geq 2$.

- I.a. ist der Lösungsraum zweidimensional, d.h. $\dim \ker M = 2$, sofern die 8 Punkte in allgemeiner Lage liegen, was wir hier zur Vereinfachung annehmen möchten; liegen die Punkte in spezieller Lage, kann der Lösungsraum mind. dreidimensional werden. Zur Beweisführung sind dann umständliche Fallunterscheidungen nötig, die wir hier nicht weiter ausführen möchten.
- Ist $\dim \ker M = 2$, wird der Lösungsraum von den Koeffizienten von F_1 und F_2 aufgespannt, d.h. es ex. $r, s \in k$ mit $F = rF_1 + sF_2$. Für den 9. Schnittpunkt $[x:y:z]$ gilt $F_1(x,y,z) = 0 = F_2(x,y,z)$, und somit auch $F(x,y,y) = 0$, d.h. $[x:y:z] \in C_F$. □

5) Bew. der Assoziativität von "+" bei einer elliptischen Kurve $E(k)$:

OE sei k ein algebraisch abgeschlossener Körper (die Assoz. folgt dann für Teilkörper). Weiter genügt es, die Relation des Lemmas 2.) nachzuweisen. Seien $P,Q,R,S \in E(k)$ und $E(k)$ durch das kubische homogene Polynom $F \in k[X,Y,Z]$ definiert.

- Wir betrachten dann die folgenden 8 Punkte:

$$P,\ Q,\ P*Q,\quad R,\ S,\ R*S,\quad P*R,\ Q*S,\qquad \oplus$$

diese definieren 6 Geraden G_1, G_2, G_3 und H_1, H_2, H_3 so, dass die Schnittpunkte der Geraden durch folgende Tabelle gegeben sind:

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

Nun ist z.z., dass der Schnittpunkt $T \in H_3 \cap G_3$ ebenfalls auf $E(k)$ liegt, denn daraus folgt dann $(P*R)*(Q*S) = T = (P*Q)*(R*S)$.

- Seien mit $G_1, G_2, G_3, H_1, H_2, H_3$ auch die linearen homogenen Polynome $\in k[X,Y,Z]$ bezeichnet, die diese Geraden definieren. Wir betrachten dann die beiden kubischen Polynome $G_1G_2G_3$ und $H_1H_2H_3$. Diese enthalten jeweils alle 9 angegebenen Punkte der Tabelle. Die elliptische Kurve trifft 8 dieser Punkte des Schnittes

$$C_{G_1G_2G_3} \cap C_{H_1H_2H_3},$$ nämlich die Punkte der Liste $\oplus$.

- Nach dem Neunpunktesatz liegt dann auch der 9. Punkt des Schnittes, nämlich $T \in H_3 \cap G_3$, auf der elliptischen Kurve, wie z.z. war. $\square$

6) Bem.:
- Man kann die Assoziativität auch mithilfe der expliziten Formeln aus Satz 13.) _ V13 direkt nachrechnen, was mühsam ist.
- Algebraiker betrachten zu beliebigen algebraischen Varietäten die sogenannte Divisorklassengruppe, welche eine abelsche Gruppe laut Definition ist, sowie eine bestimmte Abb. von $E(k)$ auf ihre Divisorklassengruppe. Diese ist ein Isomorphismus nach dem tiefen Satz von Riemann-Roch, d.h. die Gruppenstruktur überträgt sich, insbesondere also die Assoziativität der Verknüpfung "+".

Stichworte: • Varianten projektiver Koordinaten • Jakobinische Koordinaten
• Rechnerischer Vorteil bei Punktaddition/-verdopplung auf $E(k)$

2.4.5 Schnelle Arithmetik auf elliptischen Kurven

Die Gruppenoperation "+" auf einer Elliptischen Kurve $E(k)$ soll rechnerisch praktisch mit den expliziten Formeln V-13_Satz13.)/14.) auf den Computer umgesetzt werden.

1.) Bem.: Ein Blick auf die expliziten Formeln zeigt:
- Bei kurzer Weierstraßform $y^2 = x^3 + ax + b$ spielt der Koeffizient b keine Rolle. Also ist es rechnerisch günstig, Kurven mit kleinem a und großem b zu benutzen.
- Für Kurven mit $\operatorname{char} k = 2$ ergeben sich mit V13_Satz13) Formeln für "+", die sich besonders gut für Hardwareimplementierungen eignen.
- sowohl bei der Addition "+" als auch bei der Punkteverdopplung $P \rightsquigarrow 2 \cdot P$ (affin) wird eine Division in k benötigt. Das ist z.B. für $k = \mathbb{R}$ unpässlich bzw. zu ungenau. Das Problem lässt sich mit projektiven Koordinaten beheben: Ist $P = (\frac{x}{r}, \frac{y}{s}) \in \mathbb{A}^2(k)$ mit $r, s \in k \setminus \{0\}$, ist $P = [xs : yr : rs]$ ohne Division berechenbar, aber es müssen mehr Multiplikationen durchgeführt werden. Durch die Einführung einer Variante von projektiven Koordinaten – den sogenannten Jakobinischen Koordinaten – kann man in Vergleich dazu Multiplikationen einsparen, was den Rechenaufwand vermindert. Wir besprechen dies in diesem Abschnitt zur schnellen Punktaddition.

2.) Def.: Sei k ein Körper und $c, d \in \mathbb{N}$, dann definieren wir die Relation $\sim$ auf $k^3 \setminus \{(0,0,0)\}$ durch

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

3.) Bem.: • $\sim$ ist eine Äquivalenzrelation auf $k^3 \setminus \{(0,0,0)\}$, ihre Äquivalenzklassen bezeichnen wir mit

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

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

4) Bem.: • Wenn $z \neq 0$, gilt durch Normierung $(x/z^c, y/z^d, 1) \sim (x,y,z)$, vermöge $\sigma = \hat{z}$, damit kann man in der Menge
$$\mathbb{P}^2_{(c,d)}(k) := \{(x:y:z);\ (x,y,z) \in k^3 \setminus \{(0,0,0)\}\}$$
die Punkte mit $z \neq 0$ wieder mit $\mathbb{A}^2(k)$ identifizieren.
- Die Punkte mit $z=0$ bilden wieder die unendlich ferne Gerade.
- Die projektive Form der Weierstraßgleichung erhält man durch Einsetzen von $\frac{x}{z^c}$ und $\frac{y}{z^d}$ in die Gleichung und Entfernung der Nenner durch Multiplikation:
$$y^2 - x^3 - ax - b = 0 \rightsquigarrow \left(\frac{y}{z^d}\right)^2 - \left(\frac{x}{z^c}\right)^3 - a\frac{x}{z^c} - b = 0$$
$$\rightsquigarrow y^2 - x^{3-3c+2d} - a x z^{2d-c} - b z^{2d} = 0, \text{ falls etwa } 2d \geq 3c,$$
was offenbar i.a. nicht mehr homogen sein muss.

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

Schnelle Punktaddition mit Jakobinischen Koordinaten

5) Sei $E(k): y^2 = x^3 + ax + b$ gegeben mit $4a^3 + 27b^2 \neq 0$.
Anhand der affinen Version für die explizite Punkteverdopplung zeigen wir nun, dass man Rechenaufwand sparen kann, wenn man mit Jakobinischen Koordinaten $c=2$, $d=3$ arbeitet.
Ist $P=(u,v)$, $P \neq -P$, ist $2P = P+P = (\underbrace{\lambda^2 - 2u}_{=:x}, \underbrace{\lambda(u-x) - v}_{=:y})$, wo
$\lambda := \frac{3u^2+a}{2v}$, die affine Version der expliziten Formel in Standard-Darstellung.

Für die Koordinaten x, y von $2P = (x:y:1) = [x:y:1]$ bei $P = [u:v:z]$ erhält man durch Einsetzen von $\frac{u}{z^2}$ für u und $\frac{v}{z^3}$ für v dann

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

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

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

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

und $y_0 = \frac{3u^2 + az^4}{2vz} \cdot 8v^3z^3 \cdot \left(\frac{u}{z^2} - x\right) - \frac{v}{z^3} \cdot 8v^3z^3$

$= (3u^2 + az^4) \cdot 4v^2 \cdot (u - z^2x) - 8v^4$

$= (3u^2 + az^4) \cdot (4uv^2 - x_0) - 8v^4$.

explizite Formeln zur Punkteverdoppelung in jakobinischen Koordinaten

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

$$A := v^2, \quad B := 4u \cdot A, \quad C := 8A^2, \quad D := 3u^2 + a \cdot z^4$$

(2 Quadrierungen)

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

Das sind insg. 6 Quadrierungen und 4 Multiplikationen im Basiskörper k, es sind keine Divisionen nötig! (Die Skalaren Vielfachen mit 2, 3, 4, 8 zählen wie Additionen: $2 \cdot s = s + s$ usw.)

Analog gewinnt man die folgenden effizienten, expliziten Formeln zur Punktaddition $P + Q$ mit $P = (u, v)$, $Q = (r, s)$ in jakobinischen Koordinaten:

$$x_0 = (sz^3 - v)^2 - (rz^2 - u)^2(u + rz^2)$$

$$y_0 = (sz^3 - v)(u(rz^2 - u)^2 - x_0) - v(rz^2 - u)^3$$

$$z_0 = (rz^2 - u)z$$

Als Rechenverfahren dient dann:

$$A := z^2,\quad B := z \cdot A,\quad C := r \cdot A,\quad D := s \cdot B,\quad E := C - M,$$
$$F := D - N,\quad G := E^2,\quad H := G \cdot E,\quad I := u \cdot G,$$
$$x_0 := F^2 - (H + 2I),\quad y_0 := F \cdot (I - x_0) - v \cdot H,\quad z_0 := z \cdot E$$

7.) Das sind insg. 3 Quadrierungen und 8 Produkte in k, keine Divisionen!
Aufstellung des Rechenaufwands für eine elliptische Kurve $y^2 = x^3 - 3x + b$:
[Idee: Ansatz $3u^2 + a \cdot z^4$ in expl. Formel bei Punkteverdopplung braucht 3Q, 1M und wird mit $a = -3$ zu $3u^2 - 3z^4 = 3 \cdot (u - z^2) \cdot (u + z^2)$ mit 1Q, 1M]

	Punkteverdopplung $2P = P + P$	Punkteaddition $P + Q$
affin	1D, 2M, 2Q	1D, 2M, 1Q
standard-projektiv	7M, 3Q	12M, 2Q
Jakobinische Koordinaten	4M, 4Q	12M, 4Q
Chudnovski-Koordinaten	5M, 4Q	11M, 3Q

D = Divisionen, M = Multiplikationen, Q = Quadrierungen

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

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

1.) Sei $d = \lfloor \frac{\log m}{\log 2} \rfloor$, berechne durch sukzessives Verdoppeln: $P, 2P, 4P, 8P, \ldots, 2^d P$

2.) Schreibe m als Binärzahl: $m = \sum_{i=0}^{d} c_i 2^i$, $c_i \in \{0, 1\}$.

3.) Berechne $m \cdot P = (c_0 P) + (c_1 \cdot 2P) + (c_2 \cdot 4P) + \ldots + (c_d \cdot 2^d P)$ mit maximal d weiteren Additionen von Punkten auf $E(k)$.

Stichworte: $E(\mathbb{Q})$: Beispiele, Satz von Mordell, Rangvermutung, Sätze von Mazur, Nagell-Lutz, Siegel, Faltings, $E(\mathbb{C})$: elliptische Kurve über $\mathbb{C}$ = Torus

§ 3 Elliptische Kurven über verschiedenen Körpern

Bem.: die Fälle $k = \mathbb{R}, \mathbb{Q}, \mathbb{C}$ sind für die Theorie von Bedeutung, nicht so sehr für die Kryptographie-Anwendungen, wo endl. Körper $\mathbb{F}_p$ praktisch sind.

§3.1 Elliptische Kurven über $\mathbb{Q}$

1.) Motivation: Betr. die elliptische Kurve $E(k)$ jetzt über $k = \mathbb{Q}$.
Wie findet man möglichst viele rationale Punkte (d.h. $P = (x,y) \in \mathbb{A}^2(\mathbb{Q})$) auf der Kurve $E(\mathbb{Q})$? $\mathcal{O} = [0:1:0]$ ist immer einer, wie findet man affine P?
Die expliziten Formeln zeigen:

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

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

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

$$P + Q = (4, -9) \quad \text{[Schnitt } G(P,Q) \text{ mit } E(k), \text{ an x-Achse spiegeln]}$$
$$2P + Q = (2, 5) \quad \text{[Schnitt } G(P, P+Q) \text{ mit } E(k), \text{ an x-Achse spiegeln]}$$
$$3P + Q = \left(\tfrac{1}{4}, -\tfrac{33}{8}\right) \quad \text{[Schnitt } G(P, 2P+Q) \text{ mit } E(k), \text{ an x-Achse spiegeln]}$$
$$4P + Q = \left(\tfrac{106}{9}, \tfrac{1097}{27}\right)$$
$$5P + Q = \left(-\tfrac{2228}{961}, -\tfrac{63465}{29791}\right)$$
$$6P + Q = \left(\tfrac{76271}{289}, -\tfrac{2106328}{4913}\right)$$
$$\vdots$$

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

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

4.) Bsp.: Sei $E(\mathbb{Q}): y^2 = x^3 - 4x^2 + 16$. Dann ist $\Delta(E) = -16\cdot(4(-4)^3 + 27\cdot 16^2) \neq 0$
Eine kurze Suche liefert die 4 rationalen Punkte
$$P_1=(0,4),\ P_2=(4,4),\ P_3=(0,-4)=-P_1,\ P_4=(4,-4)=-P_2.$$
Können hier wie in Bsp. 2.) beliebig viele rationale Punkte konstruiert werden?
Hier ist die Gerade durch P_1 und P_2 die Tangente an $E(k)$ in P_1, weil $4^2 = x^3-4x^2+16 \Leftrightarrow 0 = x^2(x-4)$ ist und $x=0$ doppelte Nst.
Damit ist $-P_1 = P_1+P_2 = P_3$, also kann so kein weiterer rationaler Punkt konstruiert werden. Auch mit anderen Paaren P_i und P_j der 4 Punkte passiert dies. Vermutlich gibt es außer den 4 angegebenen rationalen Punkten keine weiteren auf $E(\mathbb{Q})$.
Wir haben $P_1=(0,4)$, $2P_1 = -P_2 = P_4$, $3P_1 = P_1 + \widetilde{-P_4} = P_2$, $4P_1 = P_1+P_2 = P_3$,
$5P_1 = P_3 + P_1 = (P_1+P_2)+P_1 = 2P_1+P_2 = -P_2+P_2 = \mathcal{O}$, d.h. $\operatorname{ord}(P_1)=5$.
$$\rightsquigarrow \langle P_1\rangle = \mathbb{Z}_5,\ \langle P_1\rangle = \{\mathcal{O}, P_1, P_2, P_3, P_4\}$$

5.) Die Beispiele legen folgenden Satz nahe:
Satz von Mordell (1922):
Sei $E(\mathbb{Q})$ eine elliptische Kurve über $\mathbb{Q}$.
Dann gibt es eine endliche Liste von Punkten $P_1, \dots, P_s \in E(\mathbb{Q})$, so dass alle (rationalen) Punkte auf $E(\mathbb{Q})$ von diesen erzeugt werden, d.h. $\forall P \in E(\mathbb{Q})\ \exists m_1, \dots, m_s \in \mathbb{N}_0: P = m_1P_1 + \dots + m_sP_s$.
M.a.W.: die Gruppe $(E(\mathbb{Q}), +)$ ist endlich erzeugt.
- Dabei können die Erzeuger endliche Ordnung haben oder nicht.
- Natürlich sind die Erzeuger nicht unbedingt eindeutig bestimmt.

6.) Bem.: In Bsp. 3.) haben wir einen endlichen Erzeuger $P_1 = (0,0)$, $\operatorname{ord}(P_1) = 2$,
in Bsp. 4.) haben wir ev. einen endlichen Erzeuger $P_1 = (0,4)$, $\operatorname{ord}(P_1) = 5$.
In Bsp. 2.) haben wir ev. einen unendlichen Erzeuger $P_1 = (-2,3)$, welcher
ev. nicht der einzige ist. Die von P_1 erzeugte Untergruppe
$\mathbb{Z} \cdot P_1 := \{ m P_1 ; m \in \mathbb{Z} \} \subseteq E(\mathbb{Q})$ ist isomorph zu $\mathbb{Z}$
vermöge $\mathbb{Z} \cdot P_1 \cong \mathbb{Z}$, $m \cdot P_1 \mapsto m$.

7.) Wir können in der Formulierung von 5.) die Unterscheidung zwischen Punkten
mit und ohne endlicher Ordnung vornehmen.
Die Teilmenge $T := \{ P \in E(\mathbb{Q}) ; \operatorname{ord}(P) \in \mathbb{N} \}$ aller Punkte von $E(\mathbb{Q})$
mit endlicher Ordnung ist offenbar eine Untergruppe, die
Torsionsgruppe von $E(\mathbb{Q})$ heißt. Somit hat der Satz von Mordell
auch die folgende Formulierung:

8.) Satz von Mordell, Formulierung als Aussage über die Gruppenstruktur:
Es gibt ein $r \in \mathbb{N}_0$ mit $\underbrace{E(\mathbb{Q})}_{\text{Gruppe bzgl. } +} \cong \underbrace{\mathbb{Z}^r}_{\text{komponentenweise Addition}} \times T$.

9.) Def.: Die Zahl $r(E)$ heißt Rang von $E(\mathbb{Q})$.

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

11.) Bsp.: • $E(\mathbb{Q})$: $y^2 = x^3 - 4 \rightsquigarrow E(\mathbb{Q}) \cong \mathbb{Z}^1$, wobei z.B. $P_1 = (2,2)$ Erzeuger ist.
• Im Bsp. 3.) und 4.): $\operatorname{Rg} E(\mathbb{Q}) = 0$. [Ohne Beweis]

12.) Der Rang elliptischer Kurven ist bislang schlecht verstanden.
Offen, d.h. bislang unbewiesen ist z.B. die
Rangvermutung: $\limsup_{E(\mathbb{Q})} r(E) = \infty$.

D.h. man vermutet, dass es zu jedem $C \in \mathbb{R}$ eine elliptische Kurve mit $\operatorname{rg} E(\mathbb{Q}) > C$ gibt. Der aktuelle Weltrekord (2006, von N. Elkies) ist eine elliptische Kurve vom Rang ≥ 28 (da 28 "unabhängige" Punkte unendlicher Ordnung auf ihr gefunden wurde, die Kurve lautet

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

mit $a =$ 20 067 762 415 575 526 585 033 208 209 338 542 750 930 230 312 178 956 502

und $b =$ 34 481 611 795 030 556 467 032 985 690 390 720 374 855 944 359 319 180 361
266 008 296 291 939 448 732 243 429

Die Torsionsgruppe ist deutlich besser verstanden:

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

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

Das sind beachtliche, tiefe Sätze. In Bsp. 4.) ist $T \cong \mathbb{Z}_5$, und man kann sehen, dass der Nagell-Lutz-Satz hier korrekt ist: $4^2 \mid \Delta(E)$.

Für $a,b,c \in \mathbb{Z}$ kann es höchstens endlich viele Punkte mit ganzzahligen Koordinaten geben:

15.) **Satz von Siegel (1926)**: Sei $E(\mathbb{Q})$: $y^2 = x^3 + ax^2 + bx + c$ mit $a,b,c \in \mathbb{Z}$ eine elliptische Kurve. Dann gibt es nur endlich viele Kurvenpunkte $(x,y) \in E(\mathbb{Q}) \cap \mathbb{Z}^2$.
[Historische Bem.: Siegel veröffentlichte den ersten Beweis 1926 unter dem Pseudonym "X"]

16.) Im Bsp. 2.) haben genau die Punkte $\mathcal{O}$, $(-2,\pm 3)$, $(-1,\pm 4)$, $(2,\pm 5)$, $(4,\pm 9)$, $(8,\pm 23)$, $(43,\pm 282)$, $(52,\pm 375)$, $(5234,\pm 378661)$ auf der elliptischen Kurve $E(\mathbb{Q})$ ganzzahlige Koordinaten.

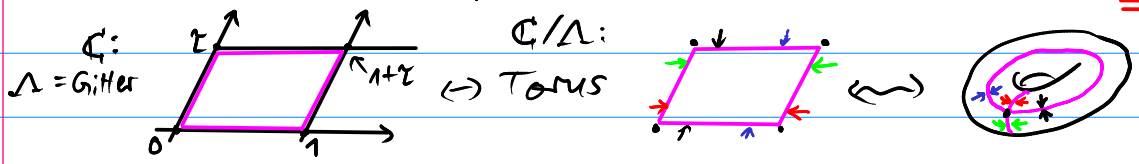
17.) Elliptische Kurven über $\mathbb{Q}$ sind nicht-sing. glg. Kurven vom Geschlecht 1. Mordell vermutete, dass jede über $\mathbb{Q}$ def. nicht-sing. glg. Kurve vom Geschlecht ≥ 2 höchstens endl. viele Punkte enthält. Diese Vermutung wurde 1983 von G. Faltings für bel. Körper bewiesen, wofür er 1986 auf der ICM in Berkeley mit der Fieldsmedaille ausgezeichnet wurde.

§3.2 Elliptische Kurven über $\mathbb{C}$

18.) Elliptische Kurven über $\mathbb{C}$ können einerseits über die Weierstraßglg. dargestellt werden und zum anderen über ihre Legendre-Normalform mit einer Glg. der Form $y^2 = x(x-1)(x-\lambda)$, vgl. (11) Übungsaufgabe 2a), Blatt 6.
Wir besprechen hier kurz die dritte Darstellung mittels elliptischer Funktionen; ein ganzer Teil der Funktionentheorie behandelt die Theorie elliptischer Funktionen. Wir möchten hier nur erläutern, warum eine elliptische Kurve über $\mathbb{C}$ in diesem Sinne ein Torus ("Donghnut") ist.

Geg. sei $\tau \in \mathbb{C}\setminus\mathbb{R}$, betr. das Gitter $\Lambda := \{a+\tau\cdot b;\ a,b\in\mathbb{Z}\} \subseteq \mathbb{C}$.

19.) Def.: Eine Funktion $f : \mathbb{C}\setminus\mathcal{P} \to \mathbb{C}$, der Form $f(z) := \frac{g(z)}{h(z)}$, g,h holomorph, $h\not\equiv 0$, heißt elliptische Funktion, falls $f(z+\omega) = f(z)$ für alle $z\in\mathbb{C}$ und $\omega\in\Lambda$, (sofern $f(z)$, $f(z+\omega)$ definiert ist, wobei $\mathcal{P} = \{z;\ h(z)=0\}$ die Menge der Polstellen von f ist), d.h. wenn f (doppelt-)periodische Funktion zu Λ ist.
Der Körper der elliptischen Funktionen über Λ sei $\mathbb{C}(\Lambda)$.



Konstruktion eines Torus zum Gitter $\Lambda \subseteq \mathbb{C}$: $\mathbb{C}/\Lambda := \{z+\Lambda;\ z\in\mathbb{C}\}$.
Jedes $z+\Lambda$ kann repräsentiert werden als $z+\Lambda = z'+\Lambda$ mit $z' = u+v\tau$, $u,v\in[0,1[$ (hier rosa Bereich "Fundamentalparallelogramm"; die Randverklebung ergibt Torus).

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

20.) **Def.:** Zu Λ def. $\wp(z) := \frac{1}{z^2} + \sum_{\substack{\omega\in\Lambda\\ \omega\neq 0}} \left(\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2}\right)$, $\wp: \mathbb{C}\setminus\Lambda \to \mathbb{C}$,

und $G_{2k}(\Lambda) := \sum_{\substack{\omega\in\Lambda\\ \omega\neq 0}} \frac{1}{\omega^{2k}} \in \mathbb{C}$. Dann heißt $\wp$ die Weierstraß-$\wp$-Funktion und G_{2k} heißt Eisensteinreihe vom Gewicht $2k \in \mathbb{R}_{>2}$.
(Englische Anleitung zur Aussprache: "pay-function")

21.) **Satz:** Sei Λ ein Gitter.
a) Die Eisenstein-Reihe $G_{2k}(\Lambda)$ konvergiert absolut für $k>1$.
b) Die Reihe der Fkt. $\wp$ konvergiert absolut und gleichmäßig auf jeder kompakten Teilmenge von $\mathbb{C}\setminus\Lambda$. Sie definiert eine elliptische Funktion mit zweifachem Pol in jedem Gitterpunkt $\omega\in\Lambda$.

Es gilt somit $\wp'(z) = -2\sum_{\substack{\omega\in\Lambda\\ \omega\neq 0}} \frac{1}{(z-\omega)^3}$ für $z\in\mathbb{C}\setminus\Lambda$. Die Funktionen $\wp$ und $\wp'$ liefern den "Prototyp" elliptischer Funktionen: Man kann zeigen, dass jede elliptische Fkt. f schreibbar ist als $f(z) = \frac{P_1(\wp(z))}{Q_1(\wp(z))} + \wp'(z)\cdot\frac{P_2(\wp(z))}{Q_2(\wp(z))}$, $P_i, Q_i \in \mathbb{C}[Z]$.

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

Haben so die Abb.: $\varphi: \mathbb{C}/\Lambda \longrightarrow \mathbb{P}^2(\mathbb{C})$
$$z+\Lambda \mapsto [\wp(z) : \wp'(z) : 1],$$
wobei $\varphi(0+\Lambda) = [0:1:0] = \mathcal{O}$.

Das Bild von φ ist genau die genannte elliptische Kurve $E(\mathbb{C})$. Die Abb. $\varphi: \mathbb{C}/\Lambda \to E(\mathbb{C})$ ist bijektiv und überträgt die Addition "+" auf $\mathbb{C}/\Lambda$, gegeben durch $(x+\Lambda)+(y+\Lambda) := (x+y)+\Lambda$, auf $E(\mathbb{C})$, welche sich als unsere bisher studierte Addition + auf $E(\mathbb{C})$ erweist. Der Torus $\mathbb{C}/\Lambda$ wird so mit der elliptischen Kurve $E(\mathbb{C})$ identifiziert. Umgekehrt ist auch jede elliptische Kurve $(E(\mathbb{C}), +)$ beschreibbar als Torus $(\mathbb{C}/\Lambda, +)$.

Stichworte: Elliptische Kurven über endlichen Körpern, $N_p = \#E(\mathbb{F}_{p^r})$, Defekt, Satz von Hasse, Birch & Swinnerton-Dyer-Vermutung, komplexe Multiplikation (CM), gute/schlechte Reduktion, Spur des Frobenius, Text in elliptische Kurve einbetten

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

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

3.3.1 Punkte zählen, der Frobenius

2.) Wir studieren zunächst elliptische Kurven über $\mathbb{F}_p$, wo p prim, mit der "modularen Brille" mod p. Das Verhalten dieser Kurven kann ganz anders sein als über $\mathbb{Q}$: Die elliptische Kurve $E(\mathbb{Q}): y^2 = x^3 + x$ aus Bsp 3.)_V16 etwa enthält den einzigen rationalen Punkt $(0,0)$, über $\mathbb{F}_p$ hat sie aber viele Punkte: Sei $N_p := \#E(\mathbb{F}_p)$ von $y^2 = x^3 + x$, d.h. N_p, die Anzahl der Punkte der elliptischen Kurve $E(\mathbb{F}_p)$, ist die Anzahl der Lösungen von $y^2 = x^3 + x$ modulo p.

3.) Numerische Daten ergeben folgende Tabelle:

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

Offenbar gilt $p = N_p - 1$, falls $p \equiv 3\ (4)$. Für $p \equiv 1\ (4)$ sieht der sogenannte Defekt $a_p := p + 1 - N_p$ aus:

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

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

Wir halten fest:

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

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

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

Zum Bew. benötigen wir:

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

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

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

das verallgemeinerte Legendresymbol ist.

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

6.) **Bew. von 4.) (a):** Für $p\equiv 3(4)$ ist $\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}} = -1$ nach 1. EG für das Legendresymbol, also $\left(\frac{(-x)^3+(-x)}{p}\right) = -\left(\frac{x^3+x}{p}\right)$ für $x\not\equiv 0(p)$ nach den Rechenregeln für das Legendresymbol, so dass $0 = \sum_{x\not\equiv 0(p)}\left(\left(\frac{x^3+x}{p}\right)+\left(\frac{(-x)^3+(-x)}{p}\right)\right) = 2\sum_{x\not\equiv 0(p)}\left(\frac{x^3+x}{p}\right)$ folgt, denn mit x durchläuft auch $-x$ alle Restklassen $\not\equiv 0(p)$ mod p. Also ist $N_p = p+1$. □

Bem.: den Bew. von 4.)(b) können wir hier nicht bringen.

7.) **Bem.:** Für den Defekt $a_p = p+1-N_p$ gilt somit $a_p = -\sum_{x\in\mathbb{F}_p}\left(\frac{x^3+ax+b}{\mathbb{F}_p}\right)$.

Im Absolutbetrag kann a_p nicht allzu groß werden, d.h. $N_p = \#E(\mathbb{F}_p)$ kann nicht allzu stark von p abweichen:

8.) **Satz von Hasse (1933):** Für den Defekt a_p einer elliptischen Kurve $E(\mathbb{F}_p)$ gilt

$$|a_p| \leq 2\sqrt{p}.$$

Der Satz gilt auch für $E(\mathbb{F}_{p^r})$, $r\geq 1$: Es ist $|p^r+1-N_{p^r}| \leq 2\sqrt{p^r}$.

9.) Diese Abschätzung für $|a_p|$ wurde 1920 von E. Artin vermutet. Eine Verallgemeinerung zeigte A. Weil in den 1940ern, und stark verallgemeinert wurde sie von P. Deligne in den 1970ern, wofür P. Deligne 1978 mit der Fieldsmedaille ausgezeichnet wurde.

10.) **Bem.:** Im Bsp. $y^2 = x^3+x$ mit $p\equiv 1(4)$ folgt aus 4.)(b) die Hasseschranke, da

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

11.) Bem.: Fragen über die Größe von N_p führen zu offenen Problemen, z.B. die schwache Vermutung von Birch und Swinnerton-Dyer (1963/65): Für $E(\mathbb{Q})$ mit Koeff. aus $\mathbb{Z}$ sollte

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

gelten, wobei C_E eine Konstante >0 ist, die nur von $E(\mathbb{Q})$ abhängig ist. Die Zahl $r(E)$ ist der Rang von $E(\mathbb{Q})$, vgl. V16. Numerische Untersuchungen stützen diese Vermutung bislang. Sie bedeutet: ist die Anzahl N_p der Punkte auf $E(\mathbb{F}_p)$ bei Reduktion mod p signifikant größer als der Erwartungswert, so sollte der Rang $r(E)$ positiv sein. Sie stellt damit ein numerisch leicht testbares Kriterium für $r(E)>0$ dar.

Zur Schreibweise $\sim$: Die Aussage $f(x) \sim g(x)$ für zwei Funktionen $f,g: \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ bedeutet, dass $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$, d.h. f,g sind asymptotisch gleich.

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

$$p = 2, 19, 29, 199, 569, 809, 1289, 1439, \quad \text{usw.}$$

Welcher der Fälle eintritt, hängt davon ab, ob die elliptische Kurve "komplexe Multiplikation" (Kurz: CM für complex multiplication) hat.

13.) Def.: Eine elliptische Kurve $E(k)$ hat komplexe Multiplikation (CM), falls sie neben den üblichen Endomorphismen $\psi_m: E(k) \to E(k)$, $P \mapsto m \cdot P$, wo $m \in \mathbb{Z}$, noch weitere hat.

Elliptische Kurven mit CM haben viele spezielle Eigenschaften, z.B.:
- Elliptische Kurven mit CM haben "ebensoviele" p mit $a_p = 0$ wie p mit $a_p \neq 0$.
- Elliptische Kurven ohne CM haben nur "wenige" p mit $a_p = 0$.

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

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

15.) Für eine elliptische Kurve $E(\mathbb{F}_{p^r})$, $r \geq 1$, wird der Defekt $a_{p^r} = p^r + 1 - N_{p^r}$ auch die "Spur des Frobenius" genannt.
Wir erklären kurz diesen Begriff:

16.) Satz & Def.: Der Frobeniusendomorphismus (kurz: Frobenius) einer elliptischen Kurve $E(\mathbb{F}_{p^r})$ ist der durch die Abb.
$$\phi : \mathbb{P}^2(\overline{\mathbb{F}_{p^r}}) \to \mathbb{P}^2(\overline{\mathbb{F}_{p^r}}), \quad [x:y:z] \mapsto [x^{p^r} : y^{p^r} : z^{p^r}]$$
vermittelte Gruppenhomomorphismus $\Phi : E(\overline{\mathbb{F}_{p^r}}) \to E(\overline{\mathbb{F}_{p^r}})$.

Beweisskizze:
- ϕ ist wirklich eine Abb. von $\mathbb{P}^2(\overline{\mathbb{F}_{p^r}})$ in sich. ✓
- ist $E(\mathbb{F}_{p^r})$ gegeben durch $F(X,Y,Z) = 0$, folgt auch $F(X^{p^r}, Y^{p^r}, Z^{p^r}) = 0$, weil in $\mathbb{F}_{p^r}$ die Gleichung $(c+d)^{p^r} = c^{p^r} + d^{p^r}$ für $c, d \in \overline{\mathbb{F}_{p^r}}$ richtig ist.
  Damit ist durch Φ eine Abb. von $E(\overline{\mathbb{F}_{p^r}})$ in sich definiert.
- die Verträglichkeit der Gruppenadd. auf $E(\overline{\mathbb{F}_{p^r}})$ mit Φ, d.h. die Eigenschaft $\Phi(P_1 + P_2) = \Phi(P_1) + \Phi(P_2)$, kann man nachrechnen. □

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

18) **Text in eine elliptische Kurve einbetten**

Bei der Umsetzung des ElGamal-Verschlüsselungsverfahrens für eine elliptische Kurvengruppe $(G,+)=(E(\mathbb{F}_{p^n}),+)$, vgl. V6, ist erforderlich, dass sich die Kommunizierenden Ⓐlice und Ⓑob darauf einigen, wie man Klartext in eine Folge von Punkten auf der elliptischen Kurve $E(\mathbb{F}_{p^n})$ übersetzt und wieder zurückerhält. Hier ein beispielhaftes Verfahren, wie dies praktisch durchgeführt werden kann:

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

$$w=(a_0\,a_1\,\dots\,a_\ell)\mapsto a_0N^{\ell-1}+a_1N^{\ell-2}+\dots+a_{\ell-2}N+a_{\ell-1}=x_w$$

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

20) **2. Schritt:** Für eine gegebene elliptische Kurve $E(\mathbb{F}_{p^n})$ gibt es aber nicht zu jedem $x_0\in\mathbb{F}_{p^n}$ einen Kurvenpunkt $(x_0,y_0)\in E(\mathbb{F}_{p^n})$. Für ein $k\in\mathbb{N}$ kann man aber die nächste x-Koordinate eines Kurvenpunkts $(x_1,y_1)\in E(\mathbb{F}_{p^n})$ mit $x_0\le x_1<x_0+k$ schnell ermitteln; die Wahrscheinlichkeit, dass dies scheitert, d.h. dass ein solches x_1 nicht ex., beträgt schätzungsweise nur $\approx(\frac{1}{2})^k$. (Bsp.: $k=50\rightsquigarrow(\frac{1}{2})^{50}<10^{-15}$)
Wähle so ein geeignetes k fest und eine elliptische Kurve $E(\mathbb{F}_{p^n})$ mit $p^n>k\cdot N^\ell$, d.h. es gibt wohl Kurvenpunkte mit genügend großen x-Koordinaten.

21) **3. Schritt:** Zu $x_w\in\{0,\dots,N^\ell-1\}$ bestimme $P_w\in E(\mathbb{F}_{p^n})$ mit x-Koordinate $\ge kx_w$, etwa $P_w=(kx_w+j,y)$ mit $j\ge0$ minimal.

22) **Beobachtung:** Hat das Verfahren funktioniert, ist dabei $j<k$, so dass durch Berechnung von $x_w=\lfloor\frac{x}{k}\rfloor$ für $P_w=(x,y)\in E(\mathbb{F}_{p^n})$ der Plaintext w aus P_w wieder zurückgewonnen werden kann.

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

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

Stichworte: Modularitätsmuster ⇝ Taniyama-Shimura-Vermutung = Modularitätssatz, Frey-Kurven, großer Fermatscher Satz, Schoof-Algorithmus zur Bestimmung von $\#E(\mathbb{F}_{p^n})$ in der Praxis für eine gegebene elliptische Kurve mit Koeff. $\in \mathbb{Z}$

3.3.2 Modularitätsmuster und der große Fermatsche Satz

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

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

$$\Theta(T) := T\cdot\left((1-T)(1-T^{11})\right)^2\cdot\left((1-T^2)(1-T^{22})\right)^2\cdot\left((1-T^3)(1-T^{33})\right)^2\cdot\left((1-T^4)(1-T^{44})\right)^2\cdot\ldots$$

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

Im Vergleich die Defekte von E: $a_2 = 0$, $a_3 = -1$, $a_5 = 1$, $a_7 = -2$, $a_{11} = 1$, $a_{13} = 4$, $a_{17} = -2, \ldots$,
d.h. bis auf a_2 ist a_p genau der Koeffizient vor T^p in $\Theta(T)$, $p \geq 3$ prim.

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

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

5.) Bis 1980 wurden Lösungsversuche durch Faktorisierungstechniken vorgenommen. Im Jahr 1986 schlug G. Frey eine Verbindung zwischen Fermats großem Satz und elliptischen Kurven vor, auf der die weiteren Erfolge beruhten:
Zu einer angenommenen, nichttrivialen Lösung A, B, C des großen Fermat-Satzes zu einem primen Exponenten $n=p$ betr. man die zugehörige elliptische Kurve
$$E_{A,B}: y^2 = x(x+A^p)(x-B^p),$$
ihre Diskriminante ist $\Delta(E_{A,B}) = 16(ABC)^{2p}$, was unwahrscheinlich erscheint. Die Idee ist, z.z., dass eine solche Kurve nicht modular sein kann, d.h. der Taniyama-Shimura-Vermutung widerspricht. Im Jahr 1986 konnte dies gezeigt werden von K. Ribet. Davon inspiriert verbrachte A. Wiles die nächsten 6 Jahre damit, die Taniyama-Shimura-Vermutung zumindest für sogenannte semistabile elliptische Kurven zu zeigen, was ihm gelang.
Da Frey-Kurven $E_{A,B}$ semistabil sind, reichte dies zum Beweis des großen Fermatschen Satzes aus.

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

3.3.3 Der Schoof-Algorithmus

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

8.) Vor.: Sei $p>2$ und $E(\mathbb{F}_{p^r})$ gegeben durch $y^2 = x^3 + ax + b$, wo $a, b \in \mathbb{F}_{p^r}$ ist.
Der Algorithmus bestimmt $t := a_{p^r} = p^r + 1 - \#E(\mathbb{F}_{p^r})$ nur modulo der ersten Primzahlen $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, $p_4 = 7$, ... bis damit a_{p^r} bestimmt werden kann.

9.) Für die ersten s Primzahlen $p_1, \dots, p_s$ vermittelt der CRS durch die Restklassenabbildung $\mathbb{Z} \longrightarrow \mathbb{Z}_{p_1} \times \dots \times \mathbb{Z}_{p_s}$ eine Bijektion $\mathbb{Z}_{p_1 \cdots p_s} \xrightarrow{\cong} \mathbb{Z}_{p_1} \times \dots \times \mathbb{Z}_{p_s}$.

Gilt $|t| < \frac{1}{2} p_1 \cdots p_s$, ist $t \bmod p_1 \cdots p_s$ durch die Restklassen $(t \bmod p_1, \dots, t \bmod p_s)$ eindeutig bestimmt. Da nach dem Satz von Hasse $|t| < 2\sqrt{p^r}$ gilt, genügt es, r mit $p_1 \cdots p_s > 4\sqrt{p^r}$ zu wählen. Die Bestimmung von $t \bmod p_1, \dots, t \bmod p_s$ reicht dann also zur Bestimmung von t laut CRS.

10.) 1. Schritt: Bestimmung von $t \bmod 2$.

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

- Für festes $x \in \mathbb{F}_{p^r}$ mit $x^3 + ax + b \neq 0$ in $\mathbb{F}_{p^r}$ hat $y^2 = x^3 + ax + b$ keine oder 2 Lsgn. y, d.h. die $\#$ Lsgn. (x,y) mit $y \neq 0$ ist gerade und zählen mod 2 daher nicht.
- Es bleiben $\mathcal{O}$ und die $(x_0, 0) \in E(\mathbb{F}_{p^r})$ zu zählen. Über $\overline{\mathbb{F}}_{p^r}$ faktorisiert $x^3 + ax + b$ zu $x^3 + ax + b = (x - x_0)(x - x_1)(x - x_2)$ mit $x_1, x_2 \in \overline{\mathbb{F}}_{p^r}$. Da E nicht-singulär, sind x_0, x_1, x_2 p.w.v. Wegen $x_0 + x_1 + x_2 = 0$ ($=$ Koeff. vor x^2) folgt
  a) $x_1, x_2 \in \mathbb{F}_{p^r}$ oder b) $x_1, x_2 \in \overline{\mathbb{F}}_{p^r} \setminus \mathbb{F}_{p^r}$.
- Im Fall a) gibt es 3 Punkte $(x_i, 0) \in E(\mathbb{F}_{p^r})$, im Fall b) nur einen Punkt, so dass $t \equiv \#E(\mathbb{F}_{p^r}) \equiv 0\ (2)$ folgt (wegen $\mathcal{O}$), sofern $(x_0, 0) \in E(\mathbb{F}_{p^r})$ existieren. Gibt es keine solchen Punkte, folgt (wegen $\mathcal{O}$) dann $t \equiv \#E(\mathbb{F}_{p^r}) \equiv 1\ (2)$.
- Ob $(x_0, 0) \in E(\mathbb{F}_{p^r})$ existieren, kann durch Überprüfen von $x - x_0 \mid x^3 + ax + b$ für alle $x_0 \in \mathbb{F}_{p^r}$ getestet werden; wegen $x^{p^r} - x = \prod_{x_0 \in \mathbb{F}_{p^r}} (x - x_0)$ also effektiv durch Überprüfen von

  $\operatorname{ggT}(x^3 + ax + b,\ x^{p^r} - x) = 1$ im Polynomring $\mathbb{F}_{p^r}[x]$ mit dem euklidischen Algorithmus.

[Erläuterung zu $x^{p^r} - x \overset{!}{=} \prod_{x_0 \in \mathbb{F}_{p^r}} (x - x_0)$: $\forall\, b \in \mathbb{F}_{p^r}$ ist $b^{p^r - 1} = 1$, d.h. $b^{p^r} = b$ bzw. $b^{p^r} - b = 0$, denn $p^r - 1$ ist die Gruppenordnung von $(\mathbb{F}_{p^r}^*, \cdot)$.]

11.) 2. Schritt: Bestimmung von $t \bmod p_i \geq 3$.

Dies ist deutlich schwieriger, hier nur die Grundidee:

- Der Frobenius $\Phi: E(\overline{\mathbb{F}}_{p^r}) \to E(\overline{\mathbb{F}}_{p^r}), [x:y:z] \mapsto [x^{p^r}:y^{p^r}:z^{p^r}]$ genügt der Gleichung
$\Phi^2(P) - t \cdot \Phi(P) + p^r \cdot P = \mathcal{O}$ für alle $P \in E(\overline{\mathbb{F}}_{p^r})$,
da t die "Spur" des Frobenius ist. [ohne Beweis]
Zu bestimmen ist eine Zahl $\tau \in \{0, \dots, p_i - 1\}$, die dieser Glg. (an Stelle t) für jeden Punkt $P \in E[p_i] := \{P \in E(\overline{\mathbb{F}}_{p^r}); \underbrace{\operatorname{ord}(P) \mid p_i}_{\Leftrightarrow p_i \cdot P = \mathcal{O}}\}$ genügt.

⌜Denn für so ein τ muss $(t - \tau)\Phi(P) = \mathcal{O}$ für jedes $P \in E[p_i] \setminus \mathcal{O}$ sein.
Da $\Phi(P) \in E[p_i] \setminus \mathcal{O}$ ist, ist $\operatorname{ord}(\Phi(P)) = p_i$ in $E[p_i]$,
es folgt $p_i = \operatorname{ord}(\Phi(P)) \mid t - \tau$, also $t \equiv \tau \bmod p_i$.⌟

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

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

Stichworte: Bekannte Angriffe auf das DL-Problem bei speziellen elliptischen Kurven, supersinguläre/anomale Kurven, Vergleich mit konventionellen Kryptoverfahren

§4 Sichere Kryptographie mit elliptischen Kurven

§4.1 Bekannte Angriffe auf das DL-Problem: Überblick

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

4.1.1 BSGS und Silver-Pohlig-Hellman

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

2.) DL-Problem in G: Geg. sei $P \in G$ mit $\operatorname{ord}(P) = n \in \mathbb{N}$, sowie $Q \in \langle P \rangle$.
Gesucht ist $k \in \{0, \dots, n-1\}$ mit $kP = Q$.

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

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

5.) Idee: Berechne alle möglichen Werte der l.S. = "Babystep" und nach und nach die möglichen Werte der r.S. = "Giant step". Trifft man auf eine Übereinstimmung, sind r und m gefunden.

6.) 1. Schritt: Berechne die Liste der "Babysteps" $B = \{(Q - rP, r) : 0 \le r \le m\}$.

7.) 2. Schritt:
- Ist für eines der r die Glg. $Q - rP = \mathcal{O}$ erfüllt, ist $k = r$. ✓
- Sonst teste für den ersten "Giantstep" $R = mP$, ob R in der Babystepliste B schon vorkommt. Falls ja: $k = m + r$. ✓
- Teste so alle "Giantsteps" $2R, 3R, 4R, \dots, (m-1)R$, ob diese in B vorkommt, wennja, gibt die 2. Komponente r mit $k = qm + r$. ✓

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

9.) Def.: Sei $B \in \mathbb{R}_{>0}$. Dann heißt $n \in \mathbb{N}$ B-glatt, falls $\forall p \mid n : p \le B$.

10.) Die Bestimmung von k in $\langle P \rangle$ wird auf Untergruppen von $\langle P \rangle$ der Ordnungen $p_i \mid n$ zurückgeführt.

Sei $\text{ord}(P) = n = \prod_{i=1}^{t} p_i^{\lambda_i}$ mit $p_1, \dots, p_t$ p.w.v. prim, $\lambda_i \in \mathbb{N}$.

Der Algorithmus hat dann eine Laufzeit von $\mathcal{O}\left(\sum_{i=1}^{t} (\lambda_i (\log n + \sqrt{p_i}))\right)$.

11.) Vorüberlegung:
- Zur Bestimmung von k mit $kP = Q \in \langle P \rangle$ berechnen wir alle Restklassen $k \bmod p_1^{\lambda_1}$, $k \bmod p_2^{\lambda_2}$, …, $k \bmod p_t^{\lambda_t}$. Denn laut CRS ist $\mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}/p_1^{\lambda_1}\mathbb{Z} \times \dots \times \mathbb{Z}/p_t^{\lambda_t}\mathbb{Z}$, so dass damit dann auch die Restklasse von $k \bmod n$ bestimmt werden kann.
- Betr. daher jedes $p = p_i$, $\lambda = \lambda_i$ mit $1 \le i \le t$.
  Gesucht: $z \in \{0, \dots, p^\lambda - 1\}$ mit $z \equiv k \bmod p^\lambda$. $(\rightsquigarrow z \equiv k \equiv z_0 \ (p))$

Schreibe $z = z_0 + z_1 p + \dots + z_{\lambda-1} p^{\lambda-1}$, die $z_i \in \{0, \dots, p-1\}$, in der p-adischen Entwicklung; bestimme die $z_0, \dots, z_{\lambda-1}$.

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

13) 2. Schritt: Seien $z_0, \dots, z_{j-1}$ schon (rekursiv) bestimmt, wo $j \leq \lambda - 1$ ist.
Berechne dann $Q_j := \frac{n}{p^{\lambda+1}} \left(Q - (z_0 + z_1 p + \dots + z_{j-1} p^{j-1}) P \right)$.
Da $nP = \mathcal{O}$, ist $\frac{n}{p^{\lambda+1}} \cdot p^\lambda P = \mathcal{O}$, da $z \equiv k \bmod p^\lambda$ ist $k = z + s p^\lambda$, $s \in \mathbb{Z}$,

$$\text{also } \frac{n}{p^{\lambda+1}} Q = \frac{n}{p^{\lambda+1}} kP = \frac{n}{p^{\lambda+1}} zP + \underbrace{\frac{n}{p^{\lambda+1}} \cdot s p^\lambda P}_{=\mathcal{O}} = \frac{n}{p^{\lambda+1}} zP$$

$$\text{und somit } Q_j = \frac{n}{p^{\lambda+1}} \left(z_j p^j + \dots + z_{\lambda-1} p^{\lambda-1} \right) P = \frac{n}{p} z_j P = z_j R.$$

→ Zur Bestimmung von z_j ist wieder ein DL-Problem in der Untergruppe $\langle R \rangle$ der Ordnung p zu lösen – etwa mit BSGS.

14) Bem.: Ist die Gruppenordnung glatt, ist der Algorithmus also sehr schnell.

4.1.2 Pollard-S und Pollard-λ

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

16) Der Pollard-λ-Algorithmus ist ähnlich, i.a. eher langsamer als Pollard-S. Er liefert gute Ergebnisse, wenn der diskrete Logarithmus k in einem hinreichend kleinen Intervall liegt. Auch Pollard-λ ist gut parallelisierbar.
(Die genauen Verfahren können in der Fachliteratur nachgeschlagen werden.)

4.1.3 MOV und SSSA

17) Beim MOV-Verfahren [Autoren: Menezes, Okamoto, Vanstone] wird das DL-Problem für eine elliptische Kurve $E(\mathbb{F}_{p^r})$ auf das in der Gruppe $(\mathbb{F}_{p^{rl}}^*, \cdot)$ für ein $l \geq 1$ zurückgeführt. Es ist also speziell nur für elliptische Kurvengruppen konstruiert, nicht für allgemeine abelsche Gruppen. Zeigt sich hier, dass $l \geq 1$ so wählbar ist, dass das DL-Problem in $(\mathbb{F}_{p^{rl}}^*, \cdot)$ leicht, d.h. schnell, zu lösen ist, ist die elliptische Kurve kryptographisch ungeeignet, etwa wenn $n = \mathrm{ord}(P)$ Teiler von $p^{rl} - 1$ ist.

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

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

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

21.) **Bsp.:** Die Kurve $E(\mathbb{F}_2): y^2 + y = x^3 + x + 1$ ist supersingulär, da $E(\mathbb{F}_2) = \{\mathcal{O}\}$.

22.) **Bem.:** Supersingularität bleibt bei Übergang zu einem Erweiterungskörper erhalten: Ist $E(\mathbb{F}_{p^n})$ supersingulär, dann auch $E(\mathbb{F}_{p^{n\ell}})$ für alle $\ell \geq 1$. [Ohne Bew.]

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

24.) **2. Kriterium für Supersingularität:** Sei $p = 2$, $E(\mathbb{F}_{2^n}): y^2 + a_1 xy + y = x^3 + a_2 x^2 + a_4 x + a_6$.
Dann ist $E(\mathbb{F}_{2^n})$ genau dann supersingulär, wenn $a_1 = 0$. [Ohne Bew.]

25.) **Bsp:** s.o. 21.), und $E(\mathbb{F}_p): y^2 = x^3 + x$ ist für $p \equiv 3 \,(4)$ supersingulär.
[Denn: $(T^3 + T)^{(p-1)/2} = \sum_{j=1}^{(p-1)/2} \binom{(p-1)/2}{j} T^{3j} T^{(p-1)/2 - j}$
mit $\frac{p-1}{2} + 2j = p-1 \iff 2j = \frac{p-1}{2}$, d.h. wenn $2 \mid \frac{p-1}{2} \iff p \equiv 1 \,(4)$
$\rightsquigarrow$ Koeff. vor T^{p-1} ist $\binom{(p-1)/2}{(p-1)/4} \neq 0$ in $\mathbb{F}_p$.
Für $p \equiv 3 \,(4)$ kommt T^{p-1} nicht vor $\rightsquigarrow$ Koeff. $= 0$.]

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

27.) Beim SSSA-Verfahren [Autoren: Satoh, Smart, Semaev, Araki] handelt es sich um einen schnellen Algorithmus zur Lösung des DL-Problems auf anomalen elliptischen Kurven, welche deswegen kryptographisch ungeeignet sind. Die Grundidee ist, die elliptische Kurve über $\mathbb{F}_p$ als eine über $\mathbb{Q}_p$ zu betrachten, dem Körper der p-adischen Zahlen, und die Logarithmenberechnung auf eine Division in $\mathbb{Z}_p$ zurückführen (was leicht ist).

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

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

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

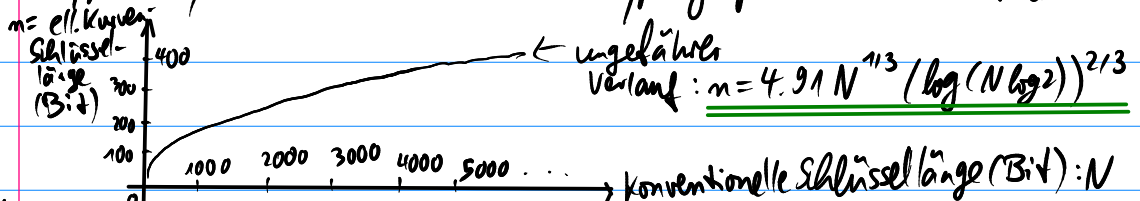
30.) Man wählt solange die Parameter p, a, b neu, bis die Diskriminante $4a^3 + 27b^2$ nicht durch p teilbar ist und somit eine elliptische Kurve vorliegt. Ziemlich sicher liegt dann eine kryptographisch geeignete Kurve vor. Das testet man nach Berechnender Gruppenordnung $\#E(\mathbb{F}_p)$ mit dem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

31.) • Ist $\#E(\mathbb{F}_p)$ glatt, d.h. hat $\#E(\mathbb{F}_p)$ nur kleine Primteiler, ist die Kurve ungeeignet wegen Silver-Pohlig-Hellman.
• Ist $\#E(\mathbb{F}_p) = p+1$, d.h. die Kurve supersingulär, ist die Kurve ungeeignet (MOV).
• Ist $\#E(\mathbb{F}_p) = p$, d.h. die Kurve anomal, ist die Kurve ungeeignet (SSSA).
Ob die Kurve supersingulär/anomal ist, kann man meist leicht erkennen durch Wahl von Punkten $P \in E(\mathbb{F}_p)$ und dem Test, ob $(p+1)P = O$ bzw $pP = O$ gilt.

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

33) Die für allgemeine elliptische Kurven, die in diesem Sinne als kryptographisch sicher gelten, bekannte Implementationen des DL-Problems sind alle von exponentieller Komplexität. Ein Kryptographieverfahren wie Elgamal bzw. DSA gilt dann als kryptographisch sicher.

34) Für konventionelle Kryptoverfahren (RSA und Elgamal / DSA auf $(\mathbb{F}_{p^n}^*, \cdot)$) gibt es subexponentielle Verfahren zur Lösung des DL-Problems. Dieser Vergleich schlägt sich in der Wahl der Schlüssellängen (= Bitzahl der Größe des endl. Körpers) nieder: Die Schlüssellänge eines elliptischen Kurven-Systems wächst etwas schneller als die 3. Wurzel der Schlüssellänge eines konventionellen Krypto-Systems mit ähnlicher kryptographischer Sicherheit:



35)
- Man geht davon aus, dass Kurven $E(\mathbb{F}_p)$ mit $p \approx 2^{173}$, wo $\#E(\mathbb{F}_p)$ einen Primteiler $\geq 2^{160}$ hat, die gleiche Sicherheit wie ein RSA-System mit 1024 Bit bietet (für 4096 Bit beim RSA nur etwa 313 bei EC-System!).
- Durch die geringere Schlüssellänge bei Verfahren mit elliptischen Kurven kann man diese leicht auf Smart-Cards ohne Koprozessor implementieren. Solche Smart-Cards sind wesentlich billiger als Chip-Karten mit Koprozessor.

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

Stichworte: ElGamal für elliptische Kurven mit Beispiel,
ECDSA: elektronische Unterschriften mit elliptischen Kurven

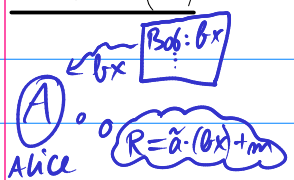
---

§4.2 ElGamal für elliptische Kurven

1.) Erinnerung an das allgemeine ElGamal-Verschlüsselungsverfahren für eine beliebige abelsche Gruppe G aus V6:
Alice möchte eine geheime Botschaft $m \in G$ an Bob schicken.

2.) Das Verfahren geht wie folgt:

Schritt (1.) Alice wählt eine Zufallszahl $\tilde{a} \in \{1, \dots, n-1\}$ 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 $R - b \cdot (\tilde{a}x) = m$.

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

1. Man wählt eine kryptographisch geeignete elliptische Kurve $E(\mathbb{F}_p): y^2 = x^3 + ax + b$, d.h. eine Primzahl p und natürliche Zahlen $0 \le a, b < p$ (und prüft die Sicherheit gemäß V19, so dass das DL-Problem schwer ist), sowie ein $P \in E(\mathbb{F}_p)$ mit großer Ordnung als Basispunkt.
2. (A) und (B) einigen sich, wie man Klartext als einen Punkt auf der elliptischen Kurve kodiert und wieder zurückerhält (etwa wie in V17 beschrieben).
3. Jeder Teilnehmer wählt eine Zahl $k \in \mathbb{N}$ als privaten Schlüssel und gibt $Q = kP \in E(\mathbb{F}_p)$ als öffentlichen Schlüssel bekannt:
   (A)lice: $a \in \mathbb{N}$ (geheim) und $a \cdot P$ (öffentlich),
   (B)ob: $b \in \mathbb{N}$ (geheim) und $b \cdot P$ (öffentlich).

Danach kann das ElGamal-Verfahren wie oben beschrieben durchgeführt werden.

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

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

Text	IN	ST	IT	UT
w	$(8,13)_{(26)} = 221 \rightsquigarrow 2210$	$(18,19)_{(26)} = 487 \rightsquigarrow 4870$	$(8,19)_{(26)} = 227 \rightsquigarrow 2270$	$(20,19)_{(26)} = 539 \rightsquigarrow 5390$
M_w	$(2211, 556)$	$(4872, 3315)$	$(2270, 2994)$	$(5392, 959)$
$\tilde{a}$	6794	3035	3508	2765
$\tilde{a}P$	$(687, 1771)$	$(1211, 2731)$	$(2714, 2389)$	$(6818, 2527)$
$\tilde{a}Q + M_w$	$(3327, 5675)$	$(2260, 17)$	$(357, 1247)$	$(1333, 6617)$

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

§4.3 ECDSA-Signaturen

5.) ECDSA ist das DSA-Verfahren (elektronische Unterschrift) auf elliptischen Kurven. (A)lice möchte dabei ein Dokument $m \in M$ an (B)ob schicken und signieren.

6.) Schritt 1.): Zuerst müssen sich die Teilnehmer (A)lice und (B)ob darauf einigen, auf welcher elliptischen Kurve gearbeitet werden soll.

- Gewählt werden ein Grundkörper $\mathbb{F}_p$ (aber auch $\mathbb{F}_{2^r}$ möglich) mit $p > 3$ prim, $A, B \in \mathbb{F}_p$ für die elliptische Kurve $E(\mathbb{F}_p): y^2 = x^3 + Ax + B$
  (im Fall $\mathbb{F}_{2^r}$ nimmt man die Glg. $y^2 + xy = x^3 + Ax^2 + B$),
  so dass $E(\mathbb{F}_p)$ eine kryptographisch geeignete elliptische Kurve ist.
- Gewählt wird ein Basispunkt $P = (x, y) \in E(\mathbb{F}_p)$ mit $n := \mathrm{ord}(P) \in \mathbb{N}$. Verlangt wird außerdem, dass n prim ist mit $n > 2^{160}$ und $n > 4\sqrt{p}$. Weiter soll n kein Teiler von $p-1, p^2-1, \dots, p^{30}-1$ sein und $n \neq p$ gelten.

7.) Bem.: • Die Bedingung $n > 2^{160}$ sorgt dafür, dass das DL-Problem in $\langle P \rangle$ nicht mit Pollard-S angreifbar ist. Ist n kein Teiler von $p^k - 1$, $k \leq 30$, kann man den MOV-Algorithmus nicht einsetzen. Wegen $n \neq p$ greift auch der SSSA-Algorithmus nicht.

• Es reicht, die Bedingungen an P zu erfüllen; die Kurve ist dann "von selbst" kryptographisch geeignet: Letztlich arbeitet man mit der Untergruppe $\langle P \rangle \subseteq E(\mathbb{F}_p)$.

• Dass die Kurve zufällig erzeugt wird per Zufallsgenerator für $p, A, B, P=(x,y)$, sorgt für zusätzliche Sicherheit; die zufällige Erzeugung sollte in der Praxis idealerweise überprüfbar sein, um auszuschließen, dass kryptographisch schwache Kurven durch Betrüger eingeschleust werden.

8.) Schritt 2.): Alice wählt eine Zufallszahl $a \in \{0, \dots, n-1\}$ als privaten Schlüssel, der Punkt $aP \in E(\mathbb{F}_p)$ gibt sie als öffentlichen Schlüssel bekannt.

Für ihr zu unterschreibendes Dokument $m \in \mathcal{M}$ berechnet sie $h(m) \in \{0,1\}^N$ für eine vorher festgelegte geeignete Hashfunktion; der Bitstring $(a_0, a_1, \dots, a_{N-1})$ wird dann als $H(m) = \sum_{i=0}^{N-1} a_i 2^{N-1-i} \leq 2^N - 1$ interpretiert.

Schritt 3.): Alice wählt eine Zufallszahl $\hat{a} \in \{1, \dots, n-1\}$ $(\hat{a} \neq 0)$ und berechnet den Punkt $\hat{a}P = (u, v)$ sowie den Rest $\Psi(\hat{a}P) \equiv u \bmod n$.

(Die Funktion $\Psi : \langle P \rangle \to \{0, 1, \dots, n-1\}$ ist zwar nicht bijektiv, die Urbildmenge eines u ist aber klein genug, so dass unwahrscheinlich ist, dass $\Psi(R) = \Psi(kP)$ gilt, ohne dass $R = kP$ gilt. Das reicht in der Praxis.)

Schritt 4.): Alice berechnet $\hat{a}^{-1} \bmod n$ und die Restklasse

$$s = \hat{a}^{-1}\left(H(m) - \Psi(\hat{a}P)\, a\right) \bmod n.$$

Falls $s \equiv 0 \bmod n$, muss ein neues $\hat{a}$ gewählt werden $\rightsquigarrow$ zurück zu Schritt 3.), da Bob bei der Prüfung der Unterschrift s invertieren wird.

Schritt 5.): Alice schickt das Dokument $m \in \mathcal{M}$ zusammen mit ihrer Unterschrift $(\Psi(\hat{a}P), s)$ an Bob.

9.) (B)ob überprüft die Unterschrift wie folgt:

1. Schritt: er testet, ob $\Psi(\hat{a}P)$, $s \in \{0, 1, \dots, n-1\}$.

2. Schritt: er berechnet $H(m) \in \mathbb{N}$,
er berechnet $s^{-1} \bmod n$
und den Punkt $R := s^{-1}(H(m)P - \Psi(\hat{a}P) \cdot \underbrace{aP}_{\text{(A)lices öffentl. Schlüssel}}) \in E(\mathbb{F}_p)$.

3. Schritt: Für $R = \mathcal{O}$ ist die Unterschrift ungültig.
Für $R = (x, y) \in E(\mathbb{F}_p) \setminus \{\mathcal{O}\}$ ist die Unterschrift gültig,
wenn $x = \Psi(\hat{a}P)$ ist, sonst nicht.

10.) Begründung der Korrektheit dieser Verifikation:
Denn wenn die Unterschrift von (A)lice stammt, ist
$$s = \hat{a}^{-1}(H(m) - \Psi(\hat{a}P)\, a) \bmod n,$$
also gilt $s^{-1}\hat{a}^{-1}(H(m) - \Psi(\hat{a}P)\, a) \equiv 1 \bmod n$,
d.h. $s^{-1}(H(m) - \Psi(\hat{a}P)\, a) \equiv \hat{a} \bmod n$
und somit $R = s^{-1}(H(m)P - \Psi(\hat{a}P)\, aP) = s^{-1}(H(m) - \Psi(\hat{a}P)\, a)P = \hat{a}P$,
so dass die x-Koordinaten der Punkte R und $\hat{a}P$ übereinstimmen müssen.

Stichworte: Verwendung elliptischer Kurven, Patente, Standards, Sicherheitsbedenken, NSA-Hintertür, Angriffe durch Quantencomputer u.a.

---

§4.4 Patente, Standards, Schlussbemerkungen

Verwendung elliptischer Kurven

1.) Die sehr praktische Nutzbarkeit elliptischer Kurven ist mittlerweile sehr vielseitig. Elliptische Kurven können außer für kryptographische Anwendungen auch für pseudo-Zufallsgeneratoren oder das Faktorisierungsproblem eingesetzt werden. Kryptographie-Anwendungen gehören aber zu ihrem Haupteinsatzgebiet. Man bezeichnet die Kryptographie-Anwendungen elliptischer Kurven zusammenfassend mit ECC (= "elliptic curve cryptography").

2.) Die in Österreich gängigen Bürgerkarten (e-card oder Bankomat-karte) verwenden schon seit 2004/2005 ECC. Die meisten Reisepässe europäischer Staaten verwenden ECC zumindest als Zugriffsschutz für den Chip, manche Länder (u.a. Deutschland und Schweiz) auch, um die gespeicherten Daten mit "Passive Authentification" zu schützen. Auch die deutsche Gesundheits-Karte hat auf ihrem Speicherchip ECC implementiert.

3.) Die Firma Sony benutzt ECDSA zur digitalen Signierung von Software für die Playstation 3.

Patente

4.) Die allgemeine Idee zu ECC wurde nicht patentiert, d.h. ECC selbst ist prinzipiell patentfrei. (Im Gegensatz zu RSA oder DH.) Es gibt aber eine Reihe von Patenten zu effizienten Implementationen. Daher sind Implementierungen mit Patentproblemen konfrontiert.

5.) Die kanadische Firma Certicom (vgl. www.certicom.com) besitzt über 350 Patente, die für ECC oder Public-key-Kryptographie benötigt werden. Davon wurden 26 von der NSA (die US-amerikanische National Security Agency) im Wert von 25 Millionen US-Dollar lizenziert, um ECC-Verfahren zu Zwecken der nationalen Sicherheit zu implementieren; davon Kurven über $\mathbb{F}_p$ für Primzahlen p mit 256, 384 und 521 Bit. ($2^{512} \approx 10^{154}$) Angeblich sind einiger dieser Lizenzen abgelaufen, weil die NSA die Lizenzbeträge dafür nicht mehr bezahlt hat.

6.) Die patent-bedingte Unsicherheit bzgl. der ECC ist mit ein Grund dafür, dass die ECC nicht in jederlei Hinsicht als empfehlenswert akzeptiert wird.

7.) Auf der Konferenz "ECC 2015" September 2015 in Bordeaux wird (wieder) über Standards der ECC diskutiert werden.

8.) Patentstreit certicom gegen Sony 2007: Certicom klagte Sony wegen der Verwendung zweier ihrer US-Patente zur ECC an. Die Anklage wurde 2009 abgewiesen.

9.) Die Firma certicom wurde 2009 von RIM (Research in Motion, heute: blackberry) zum Preis von 130 Millionen US-Dollar als Tochterfirma übernommen.

Standards und Sicherheit

10.) Eine elliptische Kurve wird zur Benutzung für gewöhnlich nicht jedesmal neu erzeugt: Die Berechnung von $\#E(\mathbb{F}_p)$ ist zeitaufwendig und kompliziert zu implementieren. Daher werden Kurvenparameter geeigneter elliptischer Kurven zur praktischen Verwendung von verschiedenen Organisationen veröffentlicht ("Standardkurven"/"benannte Kurven"), z.B. von der NIST (US National Institute of Standards and Technology) oder SECG (Standards for Efficient Cryptography Group).

11.) Man kann Kurven selbst zur Nutzung erzeugen, wenn u.a. die in V19 genannten Kriterien für die kryptographische Eignung nachprüfbar erfüllt sind. Diese Liste ist nicht vollständig; die Kriterien sollten dem aktuellen Stand der Forschung angepasst sein. Auch Patentfragen werden dann wichtig.

12.) Verschiedene Organisationen geben Sicherheitsstandards für den Gebrauch des ECC heraus. In Deutschland gibt etwa das BSI (Bundesamt für Sicherheit in der Informationstechnik) technische Vorgaben und Empfehlungen zur ECC-Implementierung auf der Basis des ISO/IEC 15946 - Standards heraus. So halten sich z.B. deutsche Banken an diese Vorgaben.

13.) Im Jahr 2013 meldete die New York Times auf Grundlage eines Snowden-Dokuments, dass der vom NIST als Standard gesetzte ECC-Algorithmus "Dual_EC_DRBG" zur Erzeugung von Pseudo-Zufallszahlen eine von der NSA eingeschleuste Schwäche im Algorithmus und der empfohlenen elliptischen Kurve besitzen würde. Die Firma RSA-Security gab daraufhin die Empfehlung, darauf basierende Algorithmen nicht weiter zu verwenden. Das NIST hat die Empfehlung des Algorithmus inzwischen zurückgezogen. Kryptographie-Experten äußerten aufgrund dieser "NSA Hintertür" auch Bedenken gegenüber mancher der vom NIST empfohlenen elliptischen Kurven und rieten wieder zur Nutzung EC-freier Kryptographie.
→ vgl. Artikel "Nach Snowden wenig Schlaf für Kryptoforscher" 17.09.2014 auf www.heise.de/security

Pikant ist, dass certicom bereits 2006 eine Patentanmeldung eingereicht hatte, in der die Hintertür beschrieben wurde.
→ vgl. Artikel "NSA-Skandal: So funktionieren Kryptographie-Hintertüren" auf Spiegel online
→ vgl. Artikel "Konkurrent für die NIST: Bernsteins elliptische Kurven auf dem Weg zum Standard", auch: wikipedia "krypto-Handy": letzter Absatz

Mögliche und bekannte Angriffe auf ECC

14.) Das DL-Problem auf elliptischen Kurven ist mittlerweile auch gegenüber einem Angriff mit einem Quantencomputer nicht mehr resistent: der Shor-Algorithmus konnte auf elliptische Kurven-Gruppen übertragen werden, vgl. Proos und Zalka, z.B. arXiv:quant-ph/0301141 (2004). Ein solcher Angriff braucht nur etwa halb so lange wie der klassische Shor-Algorithmus zur Lösung des Faktorisierungsproblems bei einem RSA-Verfahren mit vergleichbarer Sicherheit. Auch der Speicheraufwand eines Quantencomputers (Anzahl benötigter Qubits) ist dabei um (grob) den Faktor 2/3 geringer.
Bei Aufkommen von Quantencomputern werden die ECC-Verfahren daher Jahre vor den entsprechenden RSA-Verfahren geknackt sein.
An Kryptographie-Alternativen wird derzeit geforscht ("post-quantum-cryptology").

15.) Mai 2011 veröffentlichten Brumley/Tuveri eine Arbeit zu einem erfolgreichen Timing-Angriff auf ECDSA in Form eines "Seitenkanalangriffs": Weil Ver- und Entschlüsseln mit verschiedenen Schlüsseln unterschiedlich viel Zeit in Anspruch nimmt, konnte durch Abhören der verschlüsselten Kommunikation (über einen "Seitenkanal") auf die privaten Schlüssel geschlossen werden.

16.) Im Jahr 2010 gelang es einer Hackergruppe, den private key bei der von Sony für die Playstation 3 benutzten ECDSA zu erbeuten und damit das Sicherheitssystem fast vollständig zu unterwandern. Dies war aber vor allem auf Implementierungsfehler von Sony zurückzuführen und beruhte nicht auf etwaige Sicherheitslücken des verwendeten ECC-Systems.

Ende der Vorlesung