

Symbolisches Rechnen

Zusammenfassung der Vorlesung des Wintersemesters 2009/10

Einführung

Definition. Im engeren Sinn ist Symbolisches Rechnen das Umformen (Vereinfachen ?) von mathematischen Ausdrücken nach gewissen Regeln.

Beispiele sind das Kürzen von Brüchen, das Rechnen modulo n , das Umformen von rationalen Funktionen (Kürzen, Faktorisieren,...)

Definition. Im weiteren Sinn gehört zum Symbolischen Rechnen auch das Bestimmen von Stammfunktionen und Ableitungen und anderer von den Eingabedaten abhängiger mathematischer Objekte, – sofern auf dem Computer realisierbare Algorithmen dazu existieren – sowie deren Analyse.

So bekommt man beispielsweise mit der Partialbruchzerlegung und der Kenntnis der Stammfunktion von $\frac{1}{x-a}$

$$\int \frac{1}{x^3 - x} dx = \frac{1}{2} \int \frac{1}{x-1} dx + \frac{1}{2} \int \frac{1}{x+1} dx - \int \frac{1}{x} dx = \log \left(\sqrt{\left| 1 - \frac{1}{x^2} \right|} \right)$$

(falls man nicht über eine der Singularitäten $-1, 0, 1$ integriert).

Auch endliche Summen können (mit dem Algorithmus von Gosper) berechnet werden,

$$\sum_{k=0}^m (-1)^k \binom{n}{k} = (-1)^{m-1} \binom{n-1}{m-1}.$$

Gleichungssysteme $f_i(x_1, \dots, x_n) = 0$, $i = 1, \dots, s$, (mit Polynomen $f_i \in K[x_1, \dots, x_n]$ und einem auf dem Rechner realisierbaren Körper K) können symbolisch gelöst werden. Das wurde in der Vorlesung über das symbolisch/numerische Lösen von Gleichungssystemen im SS09 behandelt.

In dieser Vorlesung geht es hauptsächlich um Zahlen, d.h., wir betrachten die Darstellung der Zahlen und das Rechnen mit Zahlen aus den Mengen

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}.$$

Auch Elemente aus endlichen Körpern und aus Körpern zwischen \mathbb{Q} und \mathbb{C} werden betrachtet und die zugehörigen arithmetischen Operationen.

Die Grenzen der Zahldarstellung auf dem Computer wird exemplarisch an $\sqrt{2}$ gezeigt. Man führt eine neue "Variable" w_2 ein und ersetzt immer in der Rechnung w_2^2 durch 2. (Man vergleiche dazu das Rechnen mit der imaginären Einheit i .) Sämtliche Rechnungen bleiben korrekt, wenn ob man nun überall $-\sqrt{2}$ oder überall $\sqrt{2}$ für w_2 einsetzt!

Dasselbe gilt auch bei Polynomen höheren Grads (w_2 ist Nullstelle des Polynoms $x^2 - 2$). Man betrachte etwa $x^3 - x + \frac{1}{4}$ und nenne eine (beliebige) seiner Nullstellen w_3 ,

$$w_3^3 - w_3 + \frac{1}{4} = 0.$$

Wenn man bei Rechnungen immer w_3^3 durch $w_3 - \frac{1}{4}$ ersetzt, bekommt man am Ende immer nur Ausdrücke, die w_3 höchstens quadratisch enthalten. Und egal, welche der drei Nullstellen von $x^3 - x + \frac{1}{4}$ man einsetzt, die Rechnung ist immer korrekt. (Bei Rechnung mit zehn Dezimalstellen bekommt man im übrigen als Nullstellen $-1, 107159872, 0, 2695944364, 0, 8375654353$.)

Numerisches Rechnen: Gleitkommazahlen erlauben schnelle Rechnungen, liefern aber nur gerundete Werte. Insbesondere ist $a = 0$ nicht mehr erkennbar, wenn a nur gerundet ist. (Dagegen ist $a > b$ entscheidbar, wenn $a \neq b$ bekannt ist und man mit beliebig hoher Genauigkeit rechnen kann.) Die Ergebnisse von numerischen Rechnungen benötigen Fehlerabschätzungen. Algorithmen erfordern Stabilitätsuntersuchungen.

Symbolisches Rechnen: Zahldarstellung ist problemangepasst. Man kann ganzzahlig und rationalzahlig rechnen, mit Parametern und mit algebraischen Zahlen, wie etwa $\sqrt{2}$ oder $i = \sqrt{-1}$. Es treten keine Rundefehler auf. Dafür sind die Zahldarstellung im Rechner speicherintensiv und die Rechnung deutlich langsamer als die numerische Rechnung. Hier ist $a = 0$ entscheidbar, dagegen $a > b$ nicht.

§1 Ganze und rationale Zahlen

Anstelle der Menge der natürlichen Zahlen \mathbb{N} betrachte wir meistens

$$\mathbb{N}_0 := \mathbb{N} \cup \{ 0 \} = \{ 0, 1, 2, 3, \dots \}$$

\mathbb{N}_0 ist mit der natürlichen Ordnung $<$ wohlgeordnet, d.h.,

$$\mathcal{M} \subseteq \mathbb{N}_0, \mathcal{M} \neq \emptyset \Rightarrow \exists a \in \mathcal{M} : a \leq b \text{ für alle } b \in \mathcal{M}.$$

Die Menge der ganzen Zahlen

$$\mathbb{Z} := \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$$

ist nicht wohlgeordnet bzgl. $<$, dagegen wohlgeordnet mit \prec ,

$$a \prec b :\Leftrightarrow |a| < |b| \vee (|a| = |b| \wedge a < 0 < b).$$

Hier gilt also

$$0 \prec -1 \prec 1 \prec -2 \prec 2 \prec -3 \prec \dots$$

Satz 1.1 (Transfinite Induktion)

Sei $M \neq \emptyset$ mit $<$ wohlgeordnet. a sei kleinstes Element von M . Die Abbildung (Aussage)

$$p : M \rightarrow \{w, f\}$$

besitze die beiden Eigenschaften

i) $p(a) = w$

ii) Für alle $m \in M$ gilt: Wenn $p(n) = w$ für alle $n < m$, dann gilt auch $p(m) = w$.

Dann ist

$$p(m) = w \quad \forall m \in M.$$

Im *Beweis* betrachtet man die Menge $\mathcal{M} := \{m \in M \mid p(m) = f\}$ und führt die Annahme $\mathcal{M} \neq \emptyset$ zu einem Widerspruch mit Hilfe des kleinsten Elements von \mathcal{M} .

Sei $p \in \mathbb{N} \setminus \{1\}$. Jedes $z \in \mathbb{Z}$ kann geschrieben werden als

$$z = \varepsilon \sum_{i=1}^N z_i p^i, \quad \varepsilon \in \{-1, 1\} \quad z_i \in \{0, 1, \dots, p-1\}.$$

Dieses ist die sogenannte *p-adische Darstellung von z*. p heißt in diesem Zusammenhang *Basis* oder *Basiszahl* und die z_i sind *Ziffern*. In Computern wird oft $p = 2$ oder $p = 16$ verwendet, mitunter auch $p = 8$ oder sogar $p = 10$.

Bei der Addition von *p*-adischen Zahlen ist wenig zeitaufwändig. Einzig der *Übertrag* (auch *Überlauf* genannt) führt zu kleinen Verzögerungen. Bei der Multiplikation von zwei Zahlen in *p*-adischer Darstellung

$$a = \sum_{i=0}^{k-1} a_i p^i, \quad b = \sum_{i=0}^{k-1} b_i p^i,$$

müssen alle Produkte von Ziffern $a_i b_j$ berechnet werden und passend (mit Übertrag) zur p -adischen Darstellung von ab aufsummiert werden. Hier sind also k^2 Multiplikationen von Ziffern nötig (plus etwa genausoviele Additionen).

Sind a und b grosse Zahlen, so ist diese (Schul-)multiplikation zu aufwändig. Man nimmt bei sehr großen ℓ gern die FFT-Methode, die später erläutert wird. Aber auch die Methode von Karatsuba reduziert den Aufwand. Hierbei werden die p -adischen Darstellungen zerlegt (unter der Annahme, dass $k = 2\ell$ gilt) $a = A_1 p^\ell + A_2$, $b = B_1 p^\ell + B_2$ mit

$$A_1 = \sum_{i=0}^{\ell-1} a_{\ell+i} p^i, \quad A_2 = \sum_{i=0}^{\ell-1} a_i p^i, \quad B_1 = \sum_{i=0}^{\ell-1} b_{\ell+i} p^i, \quad B_2 = \sum_{i=0}^{\ell-1} b_i p^i.$$

Dann gilt

$$ab = A_1 B_1 p^{2\ell} + (A_1 B_2 + A_2 B_1) p^\ell + A_2 B_2$$

Wenn man nur drei Mal Zahlen der Länge ℓ multipliziert,

$$A_1 B_1, \quad A_2 B_2, \quad (A_1 + A_2)(B_1 + B_2)$$

bekommt man das fehlende $A_1 B_2 + A_2 B_1$ nur durch (preiswerte) Additionen,

$$A_1 B_2 + A_2 B_1 = (A_1 + A_2)(B_1 + B_2) - A_1 B_1 - A_2 B_2.$$

Mit diesem Trick von Karatsuba hat man nur $3\ell^2$ Ziffern-Multiplikationen (bei Zahlen der Länge ℓ) und nicht $k^2 = 4\ell^2$ Ziffern-Multiplikationen (bei Zahlen der Länge $k = 2\ell$).

Nach Addition und Multiplikation jetzt die Division. Die Division von ganzen Zahlen führt zu Brüchen, die immer in der Form

$$\frac{a}{b}, \quad a \in \mathbb{Z}, \quad b \in \mathbb{N},$$

dargestellt werden können. Weil man möglichst schnell erkennen möchte, ob zwei Brüche den gleichen Wert haben, speichert man am liebsten statt eines Bruches $\frac{a}{b}$ die ausgekürzte Form $\frac{a'}{b'}$ mit $g.g.T.(a', b') = 1$ und $\frac{a}{b} = \frac{a'}{b'}$.

Satz 1.2 (Eindeutigkeitssatz für Brüchen)

Sei $\frac{a}{b} = \frac{c}{d}$ mit $a, c \in \mathbb{Z}$ und $b, d \in \mathbb{N}$. Sind beide Brüchen ausgekürzt, dann gilt $a = c$ und $b = d$.

Die Berechnung des größten gemeinsamen Teilers zweier Zahlen $a, b \in \mathbb{N}$, $g.g.T.(a, b)$, erfolgt üblicherweise mit dem Euklidischen Algorithmus. Eine Alternative ist der

Algorithmus (Binärer g.g.T.-Algorithmus)

Eingabe: $a, b \in \mathbb{N}$.

Ausgabe: $g.g.T.(a, b)$

Rechnung: Bestimme das größte $m \in \mathbb{N}$, sodass 2^m Teiler von a und b ist.

Ersetze a durch $2^{-m}a$ und b durch $2^{-m}b$.

(Ab jetzt sind a und b nicht gleichzeitig gerade).

Falls a ungerade, dann teile aus b alle Zweierpotenzen.

Falls a gerade, dann teile aus a alle Zweierpotenzen.

(*) Wenn $a = b$, dann fertig; $2^m a$ ist der g.g.T.

Sonst ersetze a durch $\frac{|a-b|}{2}$ und b durch $\frac{a+b}{2}$.

Teile aus a und b alle Zweierpotenzen und geh zu (*).

Der Algorithmus ist korrekt, weil 2^m die richtige Zweierpotenz des g.g.T. ist und ansonsten immer der g.g.T. des jeweiligen (a, b) unverändert bleibt. Der Algorithmus terminiert, weil durch das Herauskürzen von Zweierpotenzen die Größe $\max\{a, b\}$ nicht vergrößert wird, aber bei jedem Ersetzen von a, b durch $\frac{|a-b|}{2}, \frac{a+b}{2}$ echt verkleinert wird und nach unten durch 1 beschränkt ist.

Problem: Wie bekommt man effizient die ausgekürzte Form eines Produkts von Brüchen?

Lösung: Seien $\frac{a}{b}$ und $\frac{c}{d}$ ausgekürzt. Man berechnet dann

$$g_1 := g.g.T.(c, b), \quad \text{und} \quad g_2 := g.g.T.(a, d).$$

Dann ist

$$\frac{(a/g_2)(c/g_1)}{(b/g_1)(d/g_2)}$$

der ausgekürzte Bruch zum Produkt $\frac{ac}{bd}$. Die naive Methode braucht nur eine g.g.T.-Berechnung von etwa doppelt so großen Zahlen, $g.g.T.(ac, bd)$. Hier benötigt man zwei g.g.T.-Berechnungen mit normallangen Zahlen.

Problem: Wie bekommt man effizient die ausgekürzte Form einer Summe von Brüchen?

Lösung: Seien $\frac{a}{b}$ und $\frac{c}{d}$ ausgekürzt. Man berechnet dann

$$g_1 := g.g.T.(b, d).$$

Ist $g_1 = 1$, dann ist $\frac{ad+bc}{bd}$ schon ausgekürzt. Im Fall $g_1 \neq 1$ berechnet man

$$t := a \frac{d}{g_1} + c \frac{b}{g_1} (\in \mathbb{N}) \quad \text{und dann} \quad g_2 := g.g.T.(t, g_1).$$

Dann ist

$$\frac{t/g_2}{(bd)/(g_1 g_2)}$$

der ausgekürzte Bruch zur Summe $\frac{ad+bc}{bd}$. Auch hier nur zwei g.g.T.-Berechnungen mit normallangen Zahlen gegenüber einer g.g.T.-Berechnung von etwa doppelt so großen Zahlen, $g.g.T.(ad + bc, bd)$.

Zum Abschluss noch eine (hier unbewiesene) Aussage über die Häufigkeit, wie oft zwei Zahlen teilerfremd sind.

Satz 1.3 (Cesàro)

Sei $q_n := \#\{(a, b) \in \mathbb{N}^2 \mid a \leq n, b \leq n, g.g.T.(a, b) = 1\}$. Dann gilt

$$\lim_{n \rightarrow \infty} \frac{q_n}{n^2} = \frac{6}{\pi^2} \approx 0,6079.$$

§2 Der Euklidische Algorithmus und Kettenbrüche

Definition 2.1 (Ring)

Ein Ring (mit Eins) ist ein algebraisches System mit

- i) $[R; +, 0]$ ist abelsche Gruppe,
- ii) $[R, \cdot, 1]$ mit $1 \neq 0$ ist Monoid,
- iii) Für beliebige $a, b, c \in R$ gelten die Distributivgesetze

$$a \cdot (b + c) = a \cdot b + a \cdot c, \quad (b + c) \cdot a = b \cdot a + c \cdot a.$$

Wenn das Monoid $[R; \cdot, 1]$ kommutativ ist, spricht man von einem kommutativen Ring (mit Eins).

In dieser Vorlesung soll “Ring” stets für “kommutativer Ring mit Eins” stehen.

Definition 2.2 (Integritätsring)

Ein Ring R heißt Integritätsring, wenn für alle $a, b \in R$ gilt

$$a \cdot b = 0, \quad a \neq 0 \Rightarrow b = 0.$$

Definition 2.3 (Körper)

Ein Ring R heißt Körper, wenn für alle $a \in R \setminus \{0\}$ gilt

$$\exists b \in R : a \cdot b = 1.$$

Körper sind insbesondere Integritätsringe.

Beispiel. In $\mathbb{Z}/m\mathbb{Z}$ für $m \in \mathbb{N} \setminus \{1\}$ bezeichnen wir Elemente mit \underline{a}_m , $a \in \mathbb{Z}$. Dabei gilt also

$$\underline{a}_m = \underline{b}_m$$

genau dann, wenn $a - b$ durch m teilbar ist. (Meistens nehmen wir aber die Bezeichnungen \underline{a}_m mit $a \in \{0, 1, 2, \dots, m - 1\}$.) Mit dieser Schreibweise gilt dann

$$\underline{a}_m + \underline{b}_m = \underline{a + b}_m, \quad \underline{a}_m \cdot \underline{b}_m = \underline{a \cdot b}_m.$$

Ist $m \in \mathbb{N}$ mit $m > 1$ keine Primzahl, dann ist $\mathbb{Z}/m\mathbb{Z}$ ein Ring, aber kein Integritätsring, insbesondere kein Körper. Ist dagegen m eine Primzahl, dann ist $\mathbb{Z}/m\mathbb{Z}$ sogar ein Körper.

Definition 2.4 (Euklidischer Ring)

Sei D ein Integritätsring und $d : D \setminus \{0\} \rightarrow \mathbb{N}_0$ mit

i) $d(ab) \geq d(a)$ für alle $a, b \in D \setminus \{0\}$,

ii) Für alle $a, b \in D$ mit $b \neq 0$ gibt es $q, r \in D$ mit

$$a = qb + r \quad \text{mit } r = 0 \text{ oder } d(r) < d(b). \quad (1)$$

Dann heißt D Euklidischer Ring (und d wird Gradfunktion genannt). Man nennt die Darstellung (1) Euklidische Division. Findet man ein $q \in D$, so dass (1) gilt mit $r = 0$, dann sagt man b teilt a , in Zeichen $b|a$.

Das klassische Beispiel für Euklidische Ringe ist \mathbb{Z} mit $d(a) = |a|$. Man beachte, dass bei der Euklidischen Division hier q und r nicht eindeutig sind,

$$7 = 2 \cdot 3 + 1, \quad 7 = 3 \cdot 3 - 2.$$

Beispiel. Sei K ein Körper und $D := K[x]$ (der Ring der Polynome in x mit Koeffizienten aus K). Man definiert dann

$$d\left(\sum_{k=0}^n a_k x^k\right) := \max\{k \mid a_k \neq 0\} \quad (\text{Polynomgrad}).$$

Die Euklidische Division ist die aus der Schule bekannte Polynomdivision mit Rest. Hier sind q und r eindeutig bestimmt bei vorgegebenen a und b .

Definition 2.5 (Einheiten)

Ist D Euklidischer Ring und $u \in D$, dann heißt u Einheit, wenn es ein $v \in D$ gibt mit $vu = 1$.

Bemerkung 1: Ist u eine Einheit, dann gilt für alle $a \in D \setminus \{0\}$

$$d(ua) = d(a).$$

Bemerkung 2: In D ist eine Äquivalenzrelation gegeben durch

$$a \sim b :\Leftrightarrow \exists \text{ Einheit } u : ua = b.$$

Definition 2.6 (größter gemeinsamer Teiler)

Sei D Euklidischer Ring und $a, b \in D \setminus \{0\}$. Größter gemeinsamer Teiler von a und b , kurz: $\text{g.g.T.}(a, b)$, ist ein Element $g \in D$, wenn gilt

i) $g|a$ und $g|b$,

ii) $\forall d \in D : d|a, d|b \Rightarrow d|g$.

Satz 2.1 (Eindeutigkeitsatz für den g.g.T.)

Sind g_1 und g_2 beide $\text{g.g.T.}(a, b)$, dann gilt $g_1 \sim g_2$.

Beispiel In $D = \mathbb{Z}$ sind 1 und -1 die Einheiten. Dementsprechend ist der g.g.T. nur bis auf das Vorzeichen eindeutig. Wir definieren, **der** g.g.T. ist der positive g.g.T.

Beispiel In $D = K[x]$ sind alle $c \in K \setminus \{0\}$ Einheiten. Die Eindeutigkeit des g.g.T. bekommt man daher nur durch Normieren des Höchstkoeffizienten. Wir definieren, **der** g.g.T. ist der g.g.T. mit Höchstkoeffizient 1.

Euklidischer Algorithmus (Euklid, ca. 300 v. Chr.)

Eingabe: $a, b \in D, b \neq 0$. D Euklidischer Ring mit Gradfunktion d .

Ausgabe: Ein $\text{g.g.T.}(a, b)$.

Rechnung: $(a_0, a_1) := (a, b)$

while $a_1 \neq 0$ **do**

$(a_0, a_1) := (a_1, r)$ **where** $a_0 = qa_1 + r$

return a_0

Die *Korrektheit* folgt daraus, dass stets $\text{g.g.T.}(a, b) = \text{g.g.T.}(a_0, a_1)$ gilt und $\text{g.g.T.}(a_0, 0) = a_0$. Das *Terminieren* ist klar, weil bei jedem Übergang von einem (a_0, a_1) zum nächsten das $d(a_1)$ echt kleiner wird, aber in \mathbb{N}_0 bleibt.

In $D = K[x]$ werden höchstens $\min\{d(a), d(b)\}$ Euklidische Divisionen benötigt.

Satz 2.2. (Aufwand des Euklidischen Algorithmus' in \mathbb{Z})

Wenn $a, b \in \mathbb{Z}$ mit $|a| \geq |b| > 0$, dann erfordert die Berechnung eines $\text{g.g.T.}(a, b)$ mit dem Euklidischen Algorithmus höchstens $2 \log_2(|a|)$ Euklidische Divisionen.

Beim Euklidischen Algorithmus zur Berechnung des g.g.T.'s von $a, b \in D$ berechnet man der Reihe nach mit $a_0 := a, a_1 := b$ lauter Euklidische Divi-

sionen,

$$\begin{aligned}
 a_0 &= q_1 a_1 + a_2 && \text{mit } d(a_2) < d(a_1) \\
 a_1 &= q_2 a_2 + a_3 && \text{mit } d(a_3) < d(a_2) \\
 &\vdots \\
 a_{n-2} &= q_{n-1} a_{n-1} + a_n && \text{mit } d(a_n) < d(a_{n-1}) \\
 a_{n-1} &= q_n a_n + 0 && \text{mit } a_n \sim \text{g.g.T.}(a, b) \text{ und } a_{n+1} = 0
 \end{aligned}$$

Mit diesen Werten kann man Matrizen M_i bilden

$$M_i := \begin{pmatrix} 0 & 1 \\ 1 & -q_i \end{pmatrix}, \quad i = 1, \dots, n.$$

Damit kann man die Euklidischen Divisionen auch schreiben in der Form

$$\begin{pmatrix} a_i \\ a_{i+1} \end{pmatrix} = M_i \begin{pmatrix} a_{i-1} \\ a_i \end{pmatrix}, \quad i = 1, \dots, n.$$

Wir definieren jetzt

$$\begin{pmatrix} s_0 \\ s_1 \end{pmatrix} := \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} t_0 \\ t_1 \end{pmatrix} := \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

und dann für $i = 1, \dots, n$

$$\begin{pmatrix} s_i \\ s_{i+1} \end{pmatrix} := M_i \begin{pmatrix} s_{i-1} \\ s_i \end{pmatrix}, \quad \begin{pmatrix} t_i \\ t_{i+1} \end{pmatrix} := M_i \begin{pmatrix} t_{i-1} \\ t_i \end{pmatrix}.$$

Satz 2.3 (Die Kofaktoren beim Euklidischen Algorithmus)

Für $i = 1, \dots, n$ gilt $s_i a + t_i b = a_i$ und

$$\begin{vmatrix} s_i & t_i \\ s_{i+1} & t_{i+1} \end{vmatrix} = (-1)^i.$$

Algorithmus EEA (Erweiterter Euklidischer Algorithmus)

Eingabe $a, b \in D \setminus \{0\}$ mit $d(a) \geq d(b)$.

Ausgabe $g, s, t \in D$ mit $sa + tb = g$ und $g \sim \text{g.g.T.}(a, b)$.

Rechnung $a_0 := a, a_1 := b, \begin{pmatrix} s_0 & t_0 \\ s_1 & t_1 \end{pmatrix} := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

while $a_1 \neq 0$ **do**

let $a_0 = qa_1 + r$ (Euklid. Div.)

$$\begin{pmatrix} a_0 \\ a_1 \end{pmatrix} := \begin{pmatrix} 0 & 1 \\ 1 & -q \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}$$

$$\begin{pmatrix} s_0 & t_0 \\ s_1 & t_1 \end{pmatrix} := \begin{pmatrix} 0 & 1 \\ 1 & -q \end{pmatrix} \begin{pmatrix} s_0 & t_0 \\ s_1 & t_1 \end{pmatrix}$$

return $g := a_0, s := s_0, t := t_0$.

Satz 2.4 (Resultat des EEA)

Der erweiterte Euklidische Algorithmus terminiert mit $s_n a + t_n b \sim \text{g.g.T.}(a, b)$ und mit $s_{n+1} a + t_{n+1} b = 0$, wobei $-\frac{t_{n+1}}{s_{n+1}}$ die ausgekürzte Form von $\frac{a}{b}$ ist.

Zum Abschluss zeigen wir, was man aus den beiden definierenden Eigenschaften i) und ii) des Euklidischen Rings noch folgern kann.

Bemerkung 1. Ist $b \in D \setminus \{0\}$ eine Einheit, dann gilt für alle $a \in D \setminus \{0\}$

$$d(ab) = d(a)$$

(denn $d(a) \leq d(ab) \leq d(abb^{-1}) = d(a)$), insbesondere (für $a = 1$) gilt $d(b) = d(1)$. Andererseits gibt es kein $r \in D \setminus \{0\}$ mit $d(r) < d(1)$, weil $d(r) = d(r \cdot 1) \geq d(1)$. Daher ist die Euklidische Division von 1 durch ein b mit $d(b) = d(1)$ nur möglich in der Form $1 = qb + 0$. Damit sind also genau die $b \in D \setminus \{0\}$ mit $d(b) = d(1)$ Einheiten und genau diese erfüllen $d(ab) = d(a)$ für alle $a \in D \setminus \{0\}$.

Bemerkung 2. Gilt $a|b$ und $d(a) = d(b)$ für zwei Elemente $a, b \in D \setminus \{0\}$, dann unterscheiden sich a und b nur um eine Einheit, $a \sim b$. Wegen $a|b$ gibt es nämlich ein $w \in D$ mit $b = wa$. Und bei der Euklidischen Division $a = qb + r$ folgt dann $r = (1 - qw)a$. Angenommen, r wäre nicht 0, dann folgt $d(a) = d(b) > d(r) = d((1 - qw)a) \geq d(a)$, ein Widerspruch. Also $r = 0$ und $1 - qw = 0$, d.h. w ist Einheit.

Satz 2.5 (Alternative Charakterisierungen des g.g.T.)

Sei D Euklidischer Ring D mit Gradfunktion d . Zu $a, b \in D \setminus \{0\}$ sei $\mathcal{I} := \{u \in D \mid \exists s, t \in D : u = sa + tb\}$ und $\mathcal{T} := \{u \in D \mid u|a, u|b\}$. Für $g \in D$ sind dann äquivalent

- i) g ist g.g.T. von a und b ,
- ii) $g \in \mathcal{T}$, $d(g) = \max\{d(u) \mid 0 \neq u \in \mathcal{T}\}$,
- iii) $g \in \mathcal{I}$, $d(g) = \min\{d(u) \mid 0 \neq u \in \mathcal{I}\}$.

Beweis¹ Aus Eigenschaft i) der Gradfunktion d folgt

$$a|b \Rightarrow \exists u \in D : b = ua \Rightarrow d(a) \leq d(ua) = d(b).$$

Für $0 \neq u \in \mathcal{T}$ und $0 \neq v \in \mathcal{I}$ gilt daher

$$\exists s, t \in D : v = sa + tb \Rightarrow u|v \text{ (denn } u|a \text{ und } u|b) \Rightarrow d(u) \leq d(v).$$

Ist g^* ein g.g.T. (a, b) , dann folgt wegen $g^* \in \mathcal{T} \cap \mathcal{I}$

$$d(u) \leq d(g^*) \leq d(v) \quad \forall u \in \mathcal{T}, v \in \mathcal{I}, u \neq 0, v \neq 0.$$

¹Ausnahmsweise, weil hier schönerer Beweis als in der Vorlesung

Also gelten die Implikationen $i) \Rightarrow ii)$ und $i) \Rightarrow iii)$.

Sei g^* ein größter gemeinsamer Teiler von a und b . Dann gilt wie gezeigt $d(g^*) = \max\{d(u) \mid 0 \neq u \in \mathcal{T}\}$. Ist $0 \neq g \in \mathcal{T}$ mit $d(g) = d(g^*)$, dann gilt $g \mid g^*$, weil g Teiler von a und b , und mit Bemerkung 2 folgt $g \sim g^*$.

Ist $g \in \mathcal{I}$ mit $d(g) = d(g^*)$, dann betrachte die Euklidische Division $g = qg^* + r$. Wegen $g = sa + tb$ und $g^* = s^*a + t^*b$ folgt

$$r = g - qg^* = (s - qs^*)a + (t - qt^*)b, \quad (s - qs^*), (t - qt^*) \in D.$$

r liegt also auch in \mathcal{I} . $r \neq 0$ ist nicht möglich, weil dann $d(r) < d(g^*)$, aber

$$d(g^*) = \min\{d(v) \mid 0 \neq v \in \mathcal{I}\}.$$

Also ist $r = 0$ und $g = qg^*$ mit $d(g) = d(g^*)$. Wie eben folgt $g \sim g^*$. \square

Vergleich

	\mathbb{Z}	$K[x]$
Einheiten	± 1	Alle konstanten Polynome $\neq 0$
Gradfunktion	Betrag	Polynomgrad
g.g.T.	eindeutig, wenn positiv	Eindeutig, wenn Höchstkoeff. 1.
Komplexität des EA	$\leq 2 \log_2(a)$ Euklid.Div.	$\leq \deg(a)$ Euklid.Div.

Satz 2.6 (Quotientenkörper)

Sei D ein Integritätsring. Zu $(a, b) \in D \times (D \setminus \{0\})$ sei

$$[a, b] := \{(c, d) \in D \times (D \setminus \{0\}) \mid ad = bc\}.$$

Dann gilt

- i) $[a, b]$ ist eine Äquivalenzklasse in $D \times (D \setminus \{0\})$.
- ii) Mit $[a, b] + [c, d] := [ad + bc, bd]$ und $[a, b] \cdot [c, d] := [ac, bd]$ sind Addition und Multiplikation von Äquivalenzklassen wohldefiniert.
- iii) Die Menge der Äquivalenzklassen

$$K := \{[a, b] \mid (a, b) \in D \times (D \setminus \{0\})\}$$

ist mit der Addition und Multiplikation aus ii) ein Körper.

iv) D ist eingebettet in K mittels $a \mapsto [a, 1]$.

v) Ist F ein Körper, in den D eingebettet ist, dann ist K eingebettet in F .

Definition 2.7 (Quotientenkörper)

Der so konstruierte Körper K heißt Quotientenkörper von D , kurz $Q(D)$.

Wir schreiben ab sofort $\frac{a}{b}$ statt $[a, b] \in Q(D)$ (in Übereinstimmung mit der Schreibweise der rationalen Zahlen in $\mathbb{Q} = Q(\mathbb{Z})$ und der rationalen Funktionen im Körper der rationalen Funktionen $Q(K[x])$).

Hat man im Euklidischen Ring D mit dem Euklidischen Algorithmus die Sequenz der Euklidischen Divisionen

$$\begin{aligned} a &= q_0 b + r_1 \\ b &= q_1 r_1 + r_2 \\ r_1 &= q_2 r_2 + r_3 \\ &\vdots \\ r_{n-2} &= q_{n-1} r_{n-1} + r_n \\ r_{n-1} &= q_n r_n \end{aligned}$$

berechnet, dann gilt in $Q(D)$

$$\frac{a}{b} = q_0 + \frac{1}{q_1 + \frac{1}{q_2 + \dots + \frac{1}{q_n}}}$$

Wir schreiben für diese ineinandergeschachtelten Brüche kurz

$$\frac{a}{b} =: q_0 + \underline{1} \overline{q_1} + \underline{1} \overline{q_2} + \dots + \underline{1} \overline{q_n}.$$

Definition 2.8 (endliche Kettenbrüche)

Sei $q_0 \in Q(D)$ und $q_1, \dots, q_n \in Q(D) \setminus \{0\}$. Dann heißt

$$q_0 + \underline{1} \overline{q_1} + \underline{1} \overline{q_2} + \dots + \underline{1} \overline{q_n}$$

endlicher Kettenbruch (der Länge n).

Definition 2.9 (Näherungsbrüche)

Sei $\frac{a}{b} = q_0 + \underline{1} \overline{q_1} + \underline{1} \overline{q_2} + \dots + \underline{1} \overline{q_n}$ und

$$\begin{pmatrix} Q_{-1} & P_{-1} \\ Q_0 & P_0 \end{pmatrix} := \begin{pmatrix} 0 & 1 \\ 1 & q_0 \end{pmatrix},$$

$$\begin{pmatrix} Q_{k-1} & P_{k-1} \\ Q_k & P_k \end{pmatrix} := \begin{pmatrix} 0 & 1 \\ 1 & q_k \end{pmatrix} \begin{pmatrix} Q_{k-2} & P_{k-2} \\ Q_{k-1} & P_{k-1} \end{pmatrix}, \quad k = 1, \dots, n.$$

Dann wird $\frac{P_k}{Q_k}$ als k -ter Näherungsbruch an $\frac{a}{b}$ bezeichnet.

Der Name Näherungsbruch rührt daher, dass zumindest für $D = \mathbb{Z}$ die Brüche $\frac{P_k}{Q_k}$ für wachsendes k immer bessere Näherungen an die Zahl darstellen, die durch den (endlichen) Kettenbruch gegeben ist.

Satz 2.7 (Näherungsbrüche als Kettenbrüche)

Für beliebige $q_0 \in Q(D)$, $q_1, \dots, q_n \in Q(D) \setminus \{0\}$ ist der k -te Näherungsbruch an $q_0 + \underbrace{1}_{\sqrt{\quad}} q_1 + \underbrace{1}_{\sqrt{\quad}} q_2 + \dots + \underbrace{1}_{\sqrt{\quad}} q_n$

$$\frac{P_k}{Q_k} = q_0 + \underbrace{1}_{\sqrt{\quad}} q_1 + \underbrace{1}_{\sqrt{\quad}} q_2 + \dots + \underbrace{1}_{\sqrt{\quad}} q_k$$

für $k = 0, 1, \dots, n$.

Bemerkung. Beim erweiterten Euklidischen Algorithmus zur Bestimmung eines g.g.T. (a, b) berechnet man die Kofaktoren s_k und t_k . Diese hängen eng mit den Zählern P_k und Nennern Q_k der Näherungsbrüche an $\frac{a}{b}$ zusammen. Es gilt $s_k = (-1)^k Q_k$ und $t_k = (-1)^{k+1} P_k$ für $k = 1, \dots$.

Satz 2.8 (Eigenschaften der Näherungsbrüche)

Es gilt jeweils für $k = 0, 1, \dots$,

- i) $P_k Q_{k-1} - P_{k-1} Q_k = (-1)^{k-1}$
- ii) $P_k Q_{k-2} - P_{k-2} Q_k = (-1)^k q_k$.

Ab jetzt werden wir bis zum Ende dieses Paragraphen nur noch den speziellen Euklidischen Ring $D = \mathbb{Z}$ betrachten. In diesem Ring sind Kettenbrüche sehr intensiv in Büchern der Zahlentheorie² studiert worden. Wir beschränken uns hier auf wenige Aussagen.

Ein wichtiges Hilfsmittel zum Abschätzen des Nennerwachstums bei Näherungsbrüchen sind die Fibonaccizahlen $\{F_n\}_{n=0}^\infty$ mit

$$F_0 := F_1 := 1, \quad F_n := F_{n-1} + F_{n-2}, \quad \text{für } n \geq 2.$$

Diese Folge beginnt mit $1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, \dots$. Man kann F_n erhalten durch Rundung von

$$\frac{1}{\sqrt{5}} \left(\frac{1 + \sqrt{5}}{2} \right)^{n+1}$$

zur nächsten natürlichen Zahl (durch Auf- oder Abrunden, je nachdem welche natürliche Zahl dichter an dieser irrationalen Zahl liegt).

²Etwa Godfrey H. Hardy, Einführung in die Zahlentheorie, Oldenbourg, 1958, b140/Hard

Satz 2.9 (Einschachtelungseigenschaften der Naherungsbruche)

Seien $\frac{P_k}{Q_k}$, $k = 0, \dots, n$, die Naherungsbruche an eine Zahl $\frac{a}{b} \in \mathbb{Z}$. Dann gilt

$$\frac{a}{b} = \frac{P_n}{Q_n} \text{ und}$$

i) $Q_k \geq F_k$, $k = 1, \dots, n$.

ii) $\frac{P_0}{Q_0} < \frac{P_2}{Q_2} < \frac{P_4}{Q_4} < \dots$ und $\frac{P_1}{Q_1} > \frac{P_3}{Q_3} > \frac{P_5}{Q_5} > \dots$

iii) $\frac{P_{n-2}}{Q_{n-2}} < \frac{a}{b} = \frac{P_n}{Q_n} < \frac{P_{n-1}}{Q_{n-1}}$ fur n gerade,
 $\frac{P_{n-1}}{Q_{n-1}} < \frac{a}{b} = \frac{P_n}{Q_n} < \frac{P_{n-2}}{Q_{n-2}}$ fur n ungerade.

Satz 2.10 (Kettenbruchapproximation irrationaler Zahlen)

Fur jedes irrationale x , also $x \in \mathbb{R} \setminus \{ \mathbb{Q} \}$, gibt es eine Folge $\{q_n\}_{n=0}^\infty$ mit $q_0 \in \mathbb{Z}$, $q_n \in \mathbb{N}$ fur alle $n \in \mathbb{N}$, mit

$$x = \lim_{n \rightarrow \infty} (q_0 + \underbrace{1}_{\sqrt{\quad}} q_1 + \underbrace{1}_{\sqrt{\quad}} q_2 + \dots + \underbrace{1}_{\sqrt{\quad}} q_n)$$

und fur beliebige $n \in \mathbb{N}$

$$|x - (q_0 + \underbrace{1}_{\sqrt{\quad}} q_1 + \underbrace{1}_{\sqrt{\quad}} q_2 + \dots + \underbrace{1}_{\sqrt{\quad}} q_n)| \leq \frac{1}{Q_n Q_{n+1}} < \frac{1}{Q_n^2}.$$

Man schreibt dann

$$x = q_0 + \underbrace{1}_{\sqrt{\quad}} q_1 + \underbrace{1}_{\sqrt{\quad}} q_2 + \underbrace{1}_{\sqrt{\quad}} q_3 + \dots$$

und nennt es (einfachen) unendlichen Kettenbruch. Jeder endliche (einfache) Kettenbruch stellt eine rationale Zahl dar. Bei den unendlichen kann man die irrationalen Zahlen noch etwas feiner unterscheiden. Ist der Kettenbruch namlich periodisch, gilt also fur ein $k \in \mathbb{N}$ ab einem $n_0 \in \mathbb{N}$ $q_n = q_{n+k}$, $n \geq n_0$, dann kann man zeigen, dass der Wert des Kettenbruchs Nullstelle eines quadratischen Polynoms ist.

Beispiel Mit der etwas platzsparenderen Schreibweise

$$[q_0; q_1, q_2, q_3, q_4, \dots] \text{ fur } q_0 + \underbrace{1}_{\sqrt{\quad}} q_1 + \underbrace{1}_{\sqrt{\quad}} q_2 + \underbrace{1}_{\sqrt{\quad}} q_3 + \underbrace{1}_{\sqrt{\quad}} q_4 + \dots$$

hat man

$$\begin{aligned} \sqrt{2} &= [1; 2, 2, 2, 2, 2, 2, 2, \dots] \quad (\text{Periode mit } k = 1), \\ e &= [2; 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, 1, 1, 10, \dots], \\ \pi &= [3; 7, 15, 1, 292, 1, 1, 1, 2, 1, 3, 1, 14, 2, \dots]. \end{aligned}$$

Die ersten Naherungsbruche an π bekommt man dann³

i	Q_i	P_i	q_i	$\frac{P_i}{Q_i}$
-1	0	1	-	-
0	1	3	3	3
1	7	22	7	3,142857...
2	106	333	15	3,141509...
3	113	355	1	3,141529...
4	33102	103993	292	3,1415265...

Die ersten Naherungsbruche an e ergeben sich analog,

i	Q_i	P_i	q_i	$\frac{P_i}{Q_i}$
-1	0	1	-	-
0	1	2	2	2
1	1	3	1	3
2	3	8	2	2,666666...
3	4	11	1	2,75
4	7	19	1	2,7142857...
5	32	87	4	2,71875
6	39	106	1	2,7179487...

Die Zahlen q_n des unendlichen Kettenbruchs fur $x \in \mathbb{R} \setminus \{ \mathbb{Q} \}$ kann man naturlich nicht alle mit einem (endlichen!) Algorithmus berechnen. Zumindest kann man aber zu jedem vorgegebenen N die Zahlen q_0, \dots, q_N und damit den N -ten Naherungsbruch bekommen.

Algorithmus (Kettenbruchberechnung fur reelle x)

Eingabe: Eine reelle Zahl x und eine Schranke $N \in \mathbb{N}$.

Ausgabe: Entweder der endliche Kettenbruch der Lange $< N$ fur $x \in \mathbb{Q}$

oder Zahlen $q_0 \in \mathbb{Z}$ und $q_1, \dots, q_N \in \mathbb{N}$, sodass

$q_0 + \frac{1}{q_1 + \frac{1}{q_2 + \dots + \frac{1}{q_N}}}$ N -ter Naherungsbruch an x .

Rechnung: $q_0 := \lfloor x \rfloor$, $k := 0$, $x_0 := x - q_0$,

while $k < N$ **and** $x_k \neq 0$ **do**

$r_k := \frac{1}{x_k}$,

$k := k + 1$,

$q_k := \lfloor r_{k-1} \rfloor$,

$x_k := r_{k-1} - q_k$,

return q_0, q_1, \dots, q_k

³mit dem MAPLE-Befehlen `convert(Pi, confrac,5,'cvgts')`; und `cvgts`;

Die Korrektheit des Algorithmus' ist unmittelbar einzusehen, denn man berechnet

$$x = q_0 + x_0 = q_0 + \frac{1}{r_0} = q_0 + \frac{1}{q_1 + x_1} = q_0 + \frac{1}{q_1 + \frac{1}{r_1}} = q_0 + \frac{1}{q_1 + \frac{1}{q_2 + x_2}} = \dots$$

Hierbei ist x_i im Intervall $[0, 1)$ und für $x_i \neq 0$ ist dann $r_i > 1$, sodass $q_{i+1} \in \mathbb{N}$ und wieder $x_{i+1} \in [0, 1)$.

Wir haben jetzt gesehen, dass zu jeder reellen Zahl ein einfacher Kettenbruch existiert. Umgekehrt gehört auch zu jedem einfachen Kettenbruch eine reelle Zahl, denn für beliebige $q_0 \in \mathbb{Z}$, $q_n \in \mathbb{N}$, $n \in \mathbb{N}$, konvergiert die Folge der Näherungsbrüche

$$\frac{P_n}{Q_n} = q_0 + \underbrace{1}_{\text{Zähler}} \sqrt{q_1} + \underbrace{1}_{\text{Zähler}} \sqrt{q_2} + \dots + \underbrace{1}_{\text{Zähler}} \sqrt{q_n}$$

gegen ein $x \in \mathbb{R}$, denn die Teilfolge $\{\frac{P_{2n}}{Q_{2n}}\}_{n=0}^\infty$ wächst streng monoton, die Teilfolge $\{\frac{P_{2n+1}}{Q_{2n+1}}\}_{n=0}^\infty$ fällt streng monoton und es gilt

$$\left| \frac{P_k}{Q_k} - \frac{P_{k-1}}{Q_{k-1}} \right| \leq \frac{1}{F_k F_{k-1}} \rightarrow 0.$$

Satz 2.11 (Näherungsbrüche als beste Näherungen)

Sei $\frac{P_n}{Q_n}$ der n -te Näherungsbruch an $x \in \mathbb{R}$. Für alle Brüche $\frac{P}{Q}$ mit $0 < Q \leq Q_n$ und $\frac{P}{Q} \neq \frac{P_n}{Q_n}$ gilt dann

$$\left| x - \frac{P_n}{Q_n} \right| < \left| x - \frac{P}{Q} \right|.$$

§3 Rechnen mit homomorphen Bildern.

Definition 3.1 (Ringhomomorphismus)

Sind $[R; +, \cdot, 1]$ und $[\tilde{R}, \tilde{+}, \tilde{\cdot}, \tilde{1}]$ Ringe und gilt für $\Phi : R \rightarrow \tilde{R}$

$$\begin{aligned} \Phi(a + b) &= \Phi(a) \tilde{+} \Phi(b) \quad \text{für alle } a, b \in R, \\ \Phi(a \cdot b) &= \Phi(a) \tilde{\cdot} \Phi(b) \quad \text{für alle } a, b \in R, \\ \Phi(1) &= \tilde{1}, \end{aligned}$$

dann ist Φ ein Ringhomomorphismus.

Beispiel 1. Sei K ein Körper und $a \in K$ vorgegeben. Dann ist mit $R = K[x]$ und $\tilde{R} = K$ durch

$$\Phi_a(f) := f(a) \quad \text{für alle } f \in K[x]$$

ein Ringhomomorphismus $\Phi_a : R \rightarrow \tilde{R}$ (Einsetzhomomorphismus) gegeben.

Beispiel 2. Sei D ein Euklidischer Ring und $m \in D$. Dann ist

$$D/mD := \{\underline{a}_m \mid a \in D\} \quad \text{mit } \underline{a}_m := \{x \in D \mid m \mid (x - a)\}$$

ebenfalls ein Ring mit $\underline{a}_m + \underline{b}_m = \underline{a + b_m}$ und $\underline{a}_m \cdot \underline{b}_m = \underline{a \cdot b_m}$ und $\underline{1}_m$ als Eins. Die kanonische Abbildung $\Phi_m : D \rightarrow D/mD$ mit

$$\Phi_m(a) := \underline{a}_m \quad \text{für alle } a \in D$$

ist ein Ringhomomorphismus.

Insbesondere für $D = \mathbb{Z}$ und $m = 9$ hat man

$$\Phi_9(a) := \underline{a}_9 = \{a + 9n \mid n \in \mathbb{Z}\}.$$

Wenn $a = \sum_{k=0}^s a_k 10^k$, dann gilt wegen $\Phi_9(10^k) = \underline{1}_9$

$$\Phi_9(a) = \sum_{k=0}^s \Phi_9(a_k) \Phi_9(10^k) = \sum_{k=0}^s \Phi_9(a_k) = \Phi_9\left(\sum_{k=0}^s a_k\right) \quad (\text{Quersumme})$$

Wegen $\Phi_9(9) = \underline{0}_9$ heisst diese schnelle Berechnung von $\Phi_9(a)$ auf Englisch *casting out nines*.

Ringhomomorphismen kann man zur Kontrolle von Rechnungen verwenden. So hat man in der Vor-Taschenrechner-Zeit gern mit den Quersummen, also dem Rechnen modulo 9, die Richtigkeit längerer ganzzahliger Rechnungen kontrolliert. Der mathematische Hintergrund ist dabei folgender.

Hat man zwei Ringe, $[R; +, \cdot, 1]$ und $[\tilde{R}; \tilde{+}, \tilde{\cdot}, \tilde{1}]$, und einen Ringhomomorphismus $\Phi : R \rightarrow \tilde{R}$, so kann man eine beliebige Rechnung in R , etwa

$$a \cdot (x_1 + x_2 + \dots + x_n) \quad \text{mit } a, x_1, \dots, x_n \in R,$$

auf Richtigkeit prüfen. Das Resultat der Rechnung in R sei z , man hat aber z_1 herausbekommen und möchte gern wissen, ob $z = z_1$ gilt. Unter der Annahme, dass das Rechnen in \tilde{R} einfacher ist, berechnet man dann

$$\Phi(z) = \Phi(a) \tilde{\cdot} (\Phi(x_1) \tilde{+} \Phi(x_2) \tilde{+} \dots \tilde{+} \Phi(x_n)).$$

Das erhaltene Resultat $\Phi(z)$ vergleicht man mit $\Phi(z_1)$. Gilt $\Phi(z) \neq \Phi(z_1)$, dann war das erhaltene Resultat z_1 falsch. Dagegen weiss man im Fall $\Phi(z) = \Phi(z_1)$ nicht, ob $z = z_1$ gilt. Hier kann man mit einem weiteren Ringhomomorphismus eine neue Probe machen, oder gar mehrere Ringhomomorphismen

zur Kontrolle verwenden, um die Sicherheit zu erhöhen, dass $z = z_1$ gilt.

Zumindest in den Fällen $R = \mathbb{Z}$ und $R = K[x]$, K ein Körper, kann man auf die geschilderte Weise einmal entscheiden, ob $z = z_1$ gilt. Man braucht dazu nur eine obere Schranke für $d(z)$, wie im folgenden gezeigt wird.

Wir betrachten im folgenden der Einfachheit zuliebe nur die Euklidischen Ringe $D = \mathbb{Z}$ und $D = K[x]$, K ein Körper. Weil in \mathbb{Z} der Rest bei der Euklidischen Division nicht eindeutig ist, führen wir folgende Definition ein.

Definition 3.2 (Restmenge, Restabbildung)

Sei $D = \mathbb{Z}$ oder $D = K[x]$, K ein Körper. Für jedes $m \in D \setminus \{0\}$ definieren wir als Restmenge

$$R_m := \begin{cases} \{0, 1, 2, \dots, |m| - 1\} & \text{für } D = \mathbb{Z}, \\ \{0 \neq p \in K[x] \mid \deg(p) < \deg(m)\} \cup \{0\} & \text{für } D = K[x] \end{cases}$$

und die Restabbildung $\Psi_m : D \rightarrow R_m$ durch $\Psi_m(a) := r$, wenn $a = qm + r$ (Euklidische Division mit eindeutigem Rest).

Bemerkung. Die Abbildungen Ψ_m sind nicht die Ringhomomorphismen, weil ihre Bilder nicht Elemente des Rings D/mD sind. Die Ringhomomorphismen sind die Abbildungen $\Phi_m : D \rightarrow D/mD$ mit $\Phi_m(x) = \underline{\Psi_m(x)}_m$. Dementsprechend gelten die Eigenschaften der Ringhomomorphismen,

$$\Phi_m(a + b) = \Phi_m(a) + \Phi_m(b), \quad \Phi_m(a \cdot b) = \Phi_m(a) \cdot \Phi_m(b) \quad \text{für alle } a, b \in D$$

nicht für die Ψ_m . Allerdings kann man sie in abgeschwächter Form benutzen, weil für alle $a, b \in D$ $\Psi_m(a + b)$ und $\Psi_m(a) + \Psi_m(b)$ in derselben Äquivalenzklasse von D/mD liegen. Entsprechendes gilt für $\Phi_m(a \cdot b)$ und $\Phi_m(a) \cdot \Phi_m(b)$.

Satz 3.1 (Chinesischer Restsatz, einfache Version)

Sei $D = \mathbb{Z}$ oder $D = K[x]$, K ein Körper. Die Elemente $m, n \in D \setminus \{0\}$ seien teilerfremd. Ψ_m und Ψ_n seien die dazu gehörenden Restabbildungen. Dann gibt es zu beliebigen $a \in R_m$, $b \in R_n$ genau ein $x \in R_{mn}$ mit

$$\Psi_m(x) = a, \quad \Psi_n(x) = b. \quad (*)$$

Die Menge aller Lösungen $x \in D$ von (*) ist dann

$$\mathcal{M} := \{x + c \cdot mn \mid c \in D\}.$$

Bemerkung. In den Beweis geht entscheidend die Relation $sm + tn = 1$ ein. Wegen $\Psi_m(tn) = 1$, $\Psi_n(tn) = 0$, und $\Psi_m(sm) = 0$, $\Psi_n(sm) = 1$, kann man tn und sm als Lagrange-Grundpolynome der "Interpolationsaufgabe"

$$\text{Suche } x \text{ mit } \Psi_m(x) = a, \quad \Psi_n(x) = b,$$

auffassen. Eine Lösung ist $x = a \cdot tn + b \cdot sm$, weil $\Psi_m(x)$ in der selben Äquivalenzklasse wie $a \cdot \Psi_m(tn) + b \cdot \Psi_m(sm) = a \in R_m$ liegt (und damit gleich a ist) und $\Psi_n(x)$ in der Äquivalenzklasse von $a \cdot \Psi_n(tn) + b \cdot \Psi_n(sm) = b \in R_n$, also gleich b ist.

Satz 3.2 (Chinesischer Restsatz, allgemeine Version)

Sei $D = \mathbb{Z}$ oder $D = K[x]$, K ein Körper. Die Elemente $m_1, \dots, m_s \in D$ seien paarweise teilerfremd. Mit den Restabbildungen Ψ_{m_i} und Restmengen R_{m_i} gilt dann für beliebige $a_i \in R_{m_i}$, $i = 1, \dots, s$:

Es gibt genau ein $x \in R_{m_1 \cdots m_s}$ mit

$$\Psi_{m_i}(x) = a_i \quad i = 1, \dots, s. \quad (**)$$

Die Menge der Lösungen $x \in D$ von $(**)$ ist dann

$$\mathcal{M} := \{x + c \cdot m_1 \cdots m_s \mid c \in D\}.$$

Die Lösung $x \in R_{m_1 \cdots m_s}$ aus $(**)$ kann wie ein Interpolationspolynom auf verschiedene Weisen konstruiert werden. Man kann nach *Lagrange* erst Elemente $\ell_i \in R_N$ mit $N = m_1 m_2 \cdots m_s$ bestimmen mit

$$\Psi_{m_j}(\ell_i) = \delta_{ij} \quad (\text{Kroneckersymbol}).$$

Dann ist $\sum_{i=1}^N a_i \ell_i$ eine Lösung von $(**)$ und $\Psi_N(\sum_{i=1}^N a_i \ell_i)$ ist die Lösung in R_N . Nach *Newton* berechnet man sukzessive für $k = 1, 2, \dots, s$ die Lösung $x_k \in R_{m_1 \cdots m_k}$ von

$$\Psi_{m_1}(x) = a_1, \dots, \Psi_{m_k}(x) = a_k.$$

Ist x_{k-1} schon bekannt, dann bekommt man $x_k = \Psi_{m_1 \cdots m_k}(x)$ aus

$$x := x_{k-1} + \left(a_k - \Psi_{m_k}(x_{k-1}) \right) \cdot u \cdot m_1 \cdots m_{k-1}$$

wobei u der erste Kofaktor aus der Darstellung $u \cdot (m_1 \cdots m_{k-1}) + v \cdot m_k = 1$ ($= \text{g.g.T.}(m_1 \cdots m_{k-1}, m_k)$) ist.

Ein Polynom p des Grads $\leq n$ ist durch Angabe seiner Koeffizienten a_0, \dots, a_n eindeutig bestimmt, aber auch durch seine Werte an $n+1$ verschiedenen Stellen (wg. Eindeutigkeit der Interpolation). Speichert man statt der Koeffizienten nur die Werte an $n+1$ verschiedenen Stellen x_0, \dots, x_n von Polynomen des Grads $\leq n$, dann kann man die Summe zweier Polynome durch Addition der Werte an den Stellen x_k bekommen. Die Multiplikation zweier Polynome

p, q vom Grad $\leq n$ gibt ein Polynom pq vom Grad $\leq 2n$. Kennt man aber p und q an $2n+1$ paarweise Stellen x_0, \dots, x_{2n} , dann ist pq eindeutig bestimmt durch die $2n+1$ Produkte $p(x_k) \cdot q(x_k)$. Um die Koeffizienten von pq aus den Koeffizienten von p und q zu bekommen, braucht man dagegen insgesamt $(n+1)^2$ Multiplikationen,

$$p(x) = \sum_{k=0}^n a_k x^k, \quad q(x) = \sum_{k=0}^n b_k x^k \Rightarrow pq(x) = \sum_{k=0}^{2n} \left(\sum_{i=0}^k a_i b_{k-i} \right) x^k$$

Das Problem bleibt, wie man aus dem Koeffizientenvektor (a_0, a_1, \dots, a_n) schnell die Werte $p(x_k)$ eines Polynoms $p = \sum_{k=0}^n a_k x^k$ bekommt und umgekehrt, wie man aus dem Werten $p(x_k)$ den Koeffizientenvektor bekommt. Hier hilft der *FFT-Algorithmus*, der (bei genügend großem n) so schnell ist, dass der Umweg von p und q über die Funktionswerte $p(x_k)$ und $q(x_k)$ zu den Funktionswerten $p(x_k) \cdot q(x_k)$ von pq und dann zum Koeffizientenvektor von pq schneller ist, als die direkte Berechnung mittels der Cauchyprodukte $\sum_{i=0}^k a_i b_{k-i}$.

$$\begin{array}{ccc} p, & q & \implies h = p \cdot q \\ \downarrow & \downarrow & \uparrow \\ p(x_k), & q(x_k) & h(x_k) = p(x_k)q(x_k) \\ k = 0, \dots, 2n & & k = 0, \dots, 2n \end{array}$$

Definition 3.3 (Einheitswurzeln)

Sei $n \in \mathbb{N}$. Die n Lösungen von $x^n - 1 = 0$ in \mathbb{C} heißen n -te Einheitswurzeln. Ist ω eine n -te Einheitswurzel und gilt

$$1 \notin \{\omega^1, \omega^2, \dots, \omega^{n-1}\},$$

dann heißt ω primitive n -te Einheitswurzel.

Die n -ten Einheitswurzeln sind $e^{\frac{2\ell\pi i}{n}}$, $\ell = 0, 1, 2, \dots, n-1$. $e^{\frac{2\ell\pi i}{n}}$ ist genau dann primitiv, wenn ℓ und n teilerfremd sind.

Satz 3.3 (Eigenschaften der Einheitswurzeln)

Sei $n \in \mathbb{N}$ gerade, also $n = 2m$, $m \in \mathbb{N}$, und sei ω eine n -te primitive Einheitswurzel. Dann gelten

- i) ω^2 ist eine primitive m -te Einheitswurzel,
- ii) $\omega^{m+k} = -\omega^k$ für $k = 0, \dots, m-1$,
- iii) $\bar{\omega} = \omega^{2m-1}$ ist eine primitive n -te Einheitswurzel.

Wir sagen, eine Menge $\{x_0, \dots, x_{2m-1}\} \subset \mathbb{C} \setminus \{0\}$ habe die Eigenschaft S , wenn (evtl. nach Ummumerierung) gilt

$$x_{m+k} = -x_k, \quad k = 0, 1, \dots, m-1.$$

Wenn $n \in \mathbb{N}$ eine Zweierpotenz ist, $n = 2^r$, und wenn ω eine primitive n -te Einheitswurzel ist, dann hat die Menge $\{\omega^\ell \mid \ell = 0, \dots, n-1\}$ die Eigenschaft S nach Satz 3.3 ii). ω^2 ist primitive m -te Einheitswurzel mit $m = \frac{n}{2}$ nach Satz 3.3 i). Also hat $\{\omega^{2^\ell} \mid \ell = 0, \dots, m-1\}$ die Eigenschaft S , ω^4 ist ebenfalls primitive Einheitswurzel usw. Wir können also aus der Ausgangsmenge $\{\omega^0, \dots, \omega^{n-1}\}$ (mit $n = 2^r$) fortwährend Mengen mit einer Eigenschaft S gewinnen, indem wir jeweils in der Vorgängermenge die Elemente quadrieren (und doppelte streichen).

Wir betrachten jetzt das Problem $P(n)$, wie man ein Polynom des Grads $\leq n-1$ mit Koeffizienten aus einem Körper $K \subseteq \mathbb{C}$ in den Punkten einer n -elementigen Menge $\{x_0, \dots, x_{n-1}\} \subset \mathbb{C} \setminus \{0\}$ mit Eigenschaft S effizient auswertet, wobei n eine gerade Zahl ist, $n = 2m$.

$$f(x) = \sum_{k=0}^{2m-1} a_k x^k = \sum_{k=0}^{m-1} a_{2k} x^{2k} + x \cdot \sum_{k=0}^{m-1} a_{2k+1} x^{2k} = g(x^2) + x \cdot h(x^2).$$

Setzt man jetzt die Punkte x_k ein und benutzt o.B.d.A. $x_{m+k} = -x_k$, $k = 0, 1, \dots, m-1$, dann hat man

$$\begin{aligned} f(x_k) &= g(x_k^2) + x_k \cdot h(x_k^2) & k = 0, \dots, m-1, \\ f(x_{m+k}) &= g(x_k^2) - x_k \cdot h(x_k^2) & k = 0, \dots, m-1. \end{aligned}$$

Hat jetzt die Menge $\{x_1^2, \dots, x_m^2\}$ auch die Eigenschaft S , dann haben wir das Problem aus $P(n)$ auf zwei Probleme aus $P(\frac{n}{2})$ zurückgeführt.

Satz 3.4 (Schnelle Auswertung von Polynomen in Einheitswurzeln)

Sei ω eine primitive n -te Einheitswurzel mit $n = 2^r$ und p ein Polynom des Grads $\leq n-1$ mit Koeffizienten aus \mathbb{C} . Dann bekommt man die Werte $p(\omega^k)$, $k = 0, \dots, n-1$, mit insgesamt höchstens $\frac{1}{2} \log_2(n)n$ Multiplikationen.

Entscheidend für den Beweis ist die Identität $A(n) = 2A(\frac{n}{2}) + \frac{n}{2}$, wobei $A(n)$ die Anzahl der Multiplikationen bedeutet, die man braucht, um das Problem aus $P(n)$ zu lösen.

Algorithmus FFT(R, ω, f)

Eingabe: Eine primitive n -te Einheitswurzel ω , $n = 2^r$, und der Koeffizientenvektor $(a_0, a_1, \dots, a_{n-1})$ eines Polynoms f .

Ausgabe: $(F_0, F_1, \dots, F_{n-1})$ mit $F_k = f(\omega^k)$, $k = 0, 1, \dots, n-1$.

Rechnung: **if** $n = 1$ **then return** $F = a_0$

else $m := n/2$

$(G_0, G_1, \dots, G_{m-1}) := \text{FFT}(m, \omega^2, g)$

$(H_0, H_1, \dots, H_{m-1}) := \text{FFT}(m, \omega^2, h)$

(wobei $g = \sum_{k=0}^{m-1} a_{2k} x^k$ und $h = \sum_{k=0}^{m-1} a_{2k+1} x^k$)

for $k := 0, \dots, m - 1$ **do**
 $F_k := G_k + \omega^k H_k$
 $F_{m+k} := G_k - \omega^k H_k$
return $(F_0, F_1, \dots, F_{2m-1})$

Zusammenhang zwischen den Koeffizienten a_k von $f = \sum_{k=0}^{n-1} a_k x^k$ und den Werten $f(\omega^k)$:

$$\begin{pmatrix} f(1) \\ f(\omega) \\ \vdots \\ f(\omega^{n-1}) \end{pmatrix} = \begin{pmatrix} 1 & \omega^0 & \omega^{0 \cdot 2} & \dots & \omega^{0 \cdot (n-1)} \\ 1 & \omega^1 & \omega^{1 \cdot 2} & \dots & \omega^{1 \cdot (n-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega^{n-1} & \omega^{(n-1) \cdot 2} & \dots & \omega^{(n-1) \cdot (n-1)} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{pmatrix}.$$

Satz 3.5 (Die Einheitswurzelmatrix)

Sei ω eine primitive n -te Einheitswurzel. Die Einheitswurzelmatrix zu ω ,

$$V(\omega) := \left(\omega^{j \cdot k} \right)_{j,k=0}^{n-1},$$

ist symmetrisch und erfüllt

$$V(\omega)V(\bar{\omega}) = nE_n \quad (E_n \text{ } n\text{-te Einheitsmatrix}).$$

Folge: Die Koeffizienten a_k eines Polynoms f vom Grad $< n$ können aus $f(\omega^0), f(\omega), \dots, f(\omega^{n-1})$ berechnet werden mittels

$$n \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{pmatrix} = V(\bar{\omega}) \begin{pmatrix} f(1) \\ f(\omega) \\ \vdots \\ f(\omega^{n-1}) \end{pmatrix}.$$

Diesen Zusammenhang nutzt man beim Inversen FFT-Algorithmus aus.

Algorithmus IFFT($n, \omega, (F_0, F_1, \dots, F_{n-1})$)

Eingabe: Eine primitive n -te Einheitswurzel ω , $n = 2^r$, und ein Vektor aus komplexen Zahlen $(F_0, F_1, \dots, F_{n-1}) \in \mathbb{C}^n$.

Ausgabe: Ein Polynom f des Grads $< n$ mit $F_k = f(\omega^k)$, $k = 0, 1, \dots, n - 1$.

Rechnung: $(a_0, a_1, \dots, a_{n-1}) := \text{FFT}(n, \bar{\omega}, \sum_{k=0}^{n-1} F_k x^k)$

return $\frac{1}{n} \sum_{k=0}^{n-1} a_k x^k$

Satz 3.5 (Aufwand der Polynommultiplikation)

Sind f und g zwei Polynome vom Grad $< m$ durch ihre Koeffizientenvektoren gegeben, dann kostet die Berechnung des Koeffizientenvektors des Polynoms fg mit FFT

$$3m \log_2(m) + \mathcal{O}(m)$$

Multiplikationen. Dagegen kostet die Polynommultiplikation nach der Schulmethode m^2 Multiplikationen.

Die schnelle Polynommultiplikation wird bei der Multiplikation großer ganzer Zahlen eingesetzt, denn man kann z.B. eine Dezimalzahl als Wert eines Polynoms an der Stelle 10 auffassen (und eine p -adische Zahl entsprechend als Wert an der Stelle p).

§4 Einfache Körpererweiterungen

Erinnerung. Wenn L ein Körper ist mit Unterkörper $K \subset L$, dann wird L Körpererweiterung oder Erweiterungskörper über K genannt und auch mit $L : K$ bezeichnet. Ist $a \in L : K$ für einen Erweiterungskörper von K , so wird mit $K(a)$ der kleinste Erweiterungskörper von K bezeichnet, der a enthält. Dieser Erweiterungskörper heißt algebraische Erweiterung (und a heißt dann algebraisch über K), wenn a Nullstelle eines Polynoms aus $K[x] \setminus \{0\}$ ist, andernfalls heißt die Körpererweiterung $K(a) : K$ transzendent.

Bei einer Körpererweiterung $L : K$ ist L in natürlicher Weise ein Vektorraum über K . Dessen Dimension wird mit $[L : K]$ bezeichnet. Ist $[K(a) : K]$ unendlich, dann ist die Körpererweiterung transzendent. $K(a)$ ist dann der Quotientenkörper zum Integritätsring $K[a]$, d.h., $K(a)$ besteht aus den rationalen Funktionen $\frac{f}{g}$, wobei f und $g \neq 0$ Polynome in a mit Koeffizienten aus K sind.

Ist $[K(a) : K]$ endlich, dann ist a algebraisch. Ist a Nullstelle eines Polynoms $p^* \in K[x] \setminus \{0\}$ mit $\deg(p^*) = m$ und nicht Nullstelle eines Polynoms kleineren Grads, dann gilt $m = [K(a) : K]$ und jedes $c \in K(a)$ kann in eindeutiger Weise geschrieben werden als

$$c = c_0 + c_1 a + c_2 a^2 + \dots + c_{m-1} a^{m-1}, \quad c_0, c_1, \dots, c_{m-1} \in K.$$

Das Polynom p^* heißt in diesem Zusammenhang Minimalpolynom von a .

Satz 4.1 Ist $K(a) : K$ eine algebraische Körpererweiterung und $p^* \in K[x]$ ein Minimalpolynom von a , dann ist $K(a)$ isomorph zu

$$K[x]/\langle p^* \rangle$$

mit $\langle p^* \rangle := \{q \cdot p^* \mid q \in K[x]\}$.

Die arithmetischen Operationen in einem Körper $K(a)$, a transzendent über K , lassen sich effizient mit schon behandelten Verfahren realisieren. Auskürzen

von Brüchen mit Satz 2.4, die Addition mit der Methode aus §1, die Multiplikation von Brüchen erfolgt z.B. mit der schnellen Multiplikation der Zähler- und Nennerpolynome (FFT). (Die Inversion ist problemlos, weil nur Vertauschen von Zähler und Nenner.)

Das Rechnen in $K(a)$, a algebraisch über K , wird mit Satz 4.1 zurückgeführt auf das Rechnen mit Restklassen

$$[p] := \{g \in K[x] \mid g - p \in \langle p^* \rangle\}.$$

Die Nullteilerfreiheit des Körpers $K(a)$ ist gleichbedeutend zur Irreduzibilität des Minimalpolynoms p^* . Damit gilt g.g.T. $(g, p^*) = 1$ für jedes $g \in K[x] \setminus \langle p^* \rangle$ und aus

$$sg + tp^* = 1 \quad \text{mit } s, t \in K[x]$$

folgt $[s]$ ist das (multiplikativ) Inverse zu $[g]$. Die übrigen arithmetischen Operationen kann man realisieren durch gewöhnliche Polynomoperationen evtl. gefolgt von einer Euklidischen Division durch p^* , etwa

$$[p] \cdot [q] = [p \cdot q]$$

und Euklidische Division von $p \cdot q$ durch p^* (zum effizienten Speichern der Restklasse $[p \cdot q]$).

Eine alternative Methode zum Speichern der Elemente von $K(a)$ und der Realisierung der arithmetischen Operationen ist das Rechnen mit Multiplikationsmatrizen. Zur Vereinfachung beschränken wir uns auf den Fall, dass der Körper K die Charakteristik 0 hat, also \mathbb{Q} (oder ein isomorphes Bild davon) als Unterkörper enthält.

Definition 4.1 (Frobenius-Begleitmatrix)

Sei $p^* \in K[x]$ (nicht notwendig irreduzibel) mit $p^*(x) = \sum_{k=0}^m p_k x^k$ und $\deg(p^*) = m$. Dann heißt

$$F(p^*) := \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & \vdots & \ddots & \ddots & \\ 0 & 0 & & 0 & 1 \\ \frac{-p_0}{p_m} & \frac{-p_1}{p_m} & \dots & & \frac{-p_{m-1}}{p_m} \end{pmatrix}$$

Frobenius-Begleitmatrix zu p^* .

Definition 4.2 (Multiplikationsmatrizen)

Zu vorgegebenem $p^* \in K[x]$ mit $\deg(p^*) = m$ sind die Endomorphismen Φ_g , $g \in K[x]$, definiert durch

$$\Phi_g : K[x]/\langle p^* \rangle \rightarrow K[x]/\langle p^* \rangle, \quad \Phi_g : [f] \mapsto [g][f] \quad (= [gf]).$$

Die zum Endomorphismus Φ_g und zur Basis $\{[1], [x], \dots, [x^{m-1}]\}$ gehörende Abbildungsmatrix B_g wird Multiplikationsmatrix genannt.

Es gilt für beliebige $f, g \in K[x]$

$$\begin{aligned} f \in [g] &\Rightarrow \Phi_f = \Phi_g, \\ \Phi_f + \Phi_g &= \Phi_{f+g}, \\ \Phi_f \circ \Phi_g &= \Phi_{f \cdot g} = \Phi_{g \cdot f} = \Phi_g \circ \Phi_f. \end{aligned}$$

Hieraus folgt für beliebige $g = \sum_{k=0}^{m-1} c_k x^k \in K[x]$

$$\Phi_g = \sum_{k=0}^{m-1} c_k (\Phi_x)^k.$$

Die zu Φ_x gehörende Abbildungsmatrix ist $B_x = F(p^*)^T$.

Satz 4.2 (Eigenschaften der Multiplikationsmatrizen)

Die Multiplikationsmatrizen B_g , $g \in K[x]$, erfüllen

- i) $h \in [g] \Rightarrow B_h = B_g$.
- ii) $B_g B_h = B_h B_g$ für alle $g, h \in K[x]$.
- iii) Gilt $g = \sum_{k=0}^n b_k x^k$, dann gilt mit $B_x := F(p^*)^T$

$$B_g = \sum_{k=0}^n b_k (B_x)^k.$$

Bemerkung 1. Wegen ii) und iii) nennt man $\{B_g \mid g \in K[x]\}$ eine Familie kommutierender Matrizen, die von B_x erzeugt wird.

Bemerkung 2. Wegen i) kann man sich in iii) auf Polynome vom Grad $< m$, m der Grad des Minimalpolynoms, beschränken.

Die arithmetischen Operationen in $K(a)$, a algebraisch mit Minimalpolynom p^* , können daher auch wie folgt realisiert werden. Es sei dabei $m = \deg(p^*)$.

Speicherung: Man speichert von der Zahl $\sum_{k=0}^{m-1} c_k a^k$ bzw. von der Äquivalenzklasse $\sum_{k=0}^{m-1} c_k [x^k]$ nur den Koeffizientenvektor (c_0, \dots, c_{m-1}) .

Addition: Man addiert die Koeffizientenvektoren.

Multiplikation: Sind (c_0, \dots, c_{m-1}) und (d_0, \dots, d_{m-1}) die Koeffizientenvektoren der beiden Faktoren, dann bildet man die Multiplikationsmatrix $M_g := \sum_{k=0}^n c_k (B_x)^k$ und erhält mit $M_g (d_0, \dots, d_{m-1})^T$ den Koeffizientenvektor des Produkts.

Division: Ist (c_0, \dots, c_{m-1}) der Koeffizientenvektoren des Dividenden und (d_0, \dots, d_{m-1}) der des Divisors, dann bildet man die Multiplikationsmatrix zum Divisor, $M_h = \sum_{k=0}^{m-1} d_k(B_x)^k$, und löst das Gleichungssystem

$$M_h(w_0, \dots, w_{m-1})^T = (c_0, \dots, c_{m-1})^T.$$

Der Vektor (w_0, \dots, w_{m-1}) ist der Koeffizientenvektor des Quotienten.

§5 Polynomideale

Sei K ein Körper. Mit $\mathcal{P} := K[x_1, \dots, x_n]$ bezeichnen wir den Polynomring in x_1, \dots, x_n über K . Man kann \mathcal{P} als K -Vektorraum auffassen mit Basis

$$T := \{x_1^{i_1} \cdot x_2^{i_2} \cdots x_n^{i_n} \mid i_1, i_2, \dots, i_n \in \mathbb{N}_0\}.$$

Zu jedem $f \in \mathcal{P}$ gibt es daher eine endliche Teilmenge $A = A(f) \subset T$ mit

$$f = \sum_{(a_1, \dots, a_n) \in A} c_{(a_1, \dots, a_n)} x_1^{a_1} \cdots x_n^{a_n}, \quad c_{(a_1, \dots, a_n)} \in K.$$

Definition 5.1 (Polynomgrad)

Für $x_1^{a_1} \cdots x_n^{a_n} \in T$ ist

$$\deg(x_1^{a_1} \cdots x_n^{a_n}) := a_1 + \dots + a_n.$$

Für $f = \sum_{(a_1, \dots, a_n) \in A} c_{(a_1, \dots, a_n)} x_1^{a_1} \cdots x_n^{a_n}$ ist

$$\deg(f) := \max\{\deg(x_1^{a_1} \cdots x_n^{a_n}) \mid c_{(a_1, \dots, a_n)} \neq 0\}.$$

Definition 5.2 (Ideale)

Sei R ein Ring⁴. Eine Menge $\mathfrak{a} \subset R$ heißt Ideal, wenn gilt

- i) $0 \in \mathfrak{a}$,
- ii) $f, g \in \mathfrak{a} \Rightarrow f + g \in \mathfrak{a}$,
- iii) $f \in \mathfrak{a}, h \in R \Rightarrow h \cdot f \in \mathfrak{a}$.

Ist $R = \mathcal{P}$, dann wird \mathfrak{a} auch Polynomideal genannt.

⁴Erinnerung: Hier sind alle Ringe kommutativ mit $1 \in R$

§5.1 Idealbasen

Definition 5.3 (Idealbasis)

Ist \mathfrak{a} ein Ideal (im Ring R) mit $\mathcal{F} := \{f_1, \dots, f_s\} \subset \mathfrak{a}$ und

$$\forall f \in \mathfrak{a} : \exists h_1, \dots, h_s \in R : f = \sum_{k=1}^s h_k f_k,$$

dann heißt \mathcal{F} Idealbasis, kurz: Basis, von \mathfrak{a} . Man schreibt

$$\langle \mathcal{F} \rangle = \mathfrak{a} \text{ oder auch } \langle f_1, \dots, f_s \rangle = \mathfrak{a}.$$

Ist $\mathcal{F} := \{f_1, \dots, f_s\} \subset R$, dann ist $\langle f_1, \dots, f_s \rangle$ das kleinste Ideal, das f_1, \dots, f_s enthält,

$$\langle f_1, \dots, f_s \rangle = \left\{ \sum_{k=1}^s h_k f_k \mid h_1, \dots, h_s \in R \right\}.$$

Satz 5.1 (Hilberts Basissatz)

Sei R ein Ring, in dem jedes Ideal eine Basis besitzt. Dann hat auch jedes Ideal in $R[x]$ eine Basis.

Bemerkung: Man sagt, *im Ring R gilt der Basissatz*, wenn in R jedes Ideal eine Basis besitzt.

Korollar (Basen bei Polynomidealen)

In $\mathcal{P} = K[x_1, \dots, x_n]$, K ein Körper, hat jedes Ideal eine Basis.

Satz 5.2 (Aufsteigende Kettenbedingung)

Im Ring R hat jedes Ideal genau dann eine Basis, wenn jede aufsteigende Kette von Idealen $\mathfrak{a}_k \in R$, $k \in \mathbb{N}$, also $\mathfrak{a}_k \subseteq \mathfrak{a}_{k+1}$ für alle $k \in \mathbb{N}$, stationär wird, m.a.W.

$$\exists s \in \mathbb{N} : i > s \Rightarrow \mathfrak{a}_i = \mathfrak{a}_s.$$

Bemerkung: Man sagt, *im Ring R gilt die aufsteigende Kettenbedingung*, wenn in R jedes aufsteigende Kette von Idealen stationär wird. Die Gültigkeit der aufsteigenden Kettenbedingung und die Gültigkeit des Basissatzes sind also äquivalent. Ebenfalls äquivalent dazu ist (in dieser Vorlesung ohne Beweis!) die Maximalbedingung, die besagt, dass jede nichtleere Menge von Idealen ein maximales Ideal besitzt, also ein Ideal, dass nicht in einem anderen der Menge enthalten ist.

Definition 5.4 (noetherscher Ring)

Ein Ring heißt noethersch, wenn in ihm jedes Ideal eine Basis besitzt.

Definition 5.5 (Terme)

Die von x_1, \dots, x_n erzeugte multiplikative Halbgruppe

$$[x_1, \dots, x_n] := \{x_1^{a_1} \cdots x_n^{a_n} \mid a_1, \dots, a_n \in \mathbb{N}_0\}$$

heißt Termmenge, ihre Elemente heißen Terme oder Potenzprodukte. Das neutrale Element $x_1^0 \cdots x_n^0$ wird mit 1 bezeichnet.

Definition 5.6 (Termordnung)

Eine vollständige Ordnung \leq_T auf T ist zulässig, wenn gilt

- (i) $1 \leq_T t$ für alle $t \in T$,
- (ii) $t_1 \leq_T t_2 \Rightarrow tt_1 \leq_T tt_2$ für alle $t \in T$.

(Wir sagen kurz “Termordnung” statt “vollständige zulässige Ordnung auf T ”.) Eine Termordnung \leq_T ist gradverträglich (oder graduiert), wenn

$$\deg(t_1) <_T \deg(t_2) \Rightarrow t_1 <_T t_2.$$

Definition 5.7 (Die Leitbegriffe)

Für jedes $f = \sum_{k=1}^s a_k t_k$ mit $a_k \in K \setminus \{0\}$ und $t_k \in T$, $k = 1, \dots, m$, mit $t_1 <_T t_2 < \dots <_T t_s$ definiert man

$$\begin{aligned} \mathbf{lt}(f) &:= t_s && \text{(Leitterm)} \\ \mathbf{lc}(f) &:= a_s && \text{(Leitkoeffizient)} \\ \mathbf{lm}(f) &:= a_s t_s && \text{(Leitmonom)} \end{aligned}$$

Aus Bequemlichkeit $\mathbf{lt}(0) := 0$ und $0 <_T t$ für alle $t \in T$.

Es folgt $\mathbf{lt}(fg) = \mathbf{lt}(f)\mathbf{lt}(g)$ und $\mathbf{lt}(f+g) \leq_T \max\{\mathbf{lt}(f), \mathbf{lt}(g)\}$ für alle $f, g \in \mathcal{P}$.

Lemma von Dickson

Sei U eine nicht-leere Teilmenge von T . Dann gibt es $m_1, \dots, m_r \in U$ mit

$$u \in U \Rightarrow \exists i \in \{1, \dots, r\} : m_i \mid u.$$

Korollar.

Jede Termordnung \leq_T ist eine Wohlordnung.

Im folgenden ist R immer ein noetherscher Ring (und kommutativ mit 1). Wir wollen Hilberts Basissatz in der folgenden verschärften Form beweisen:

Ist R ein noetherscher Ring und \leq_T eine Termordnung, dann hat jedes Ideal $\mathfrak{a} \subseteq R[x_1, \dots, x_n]$ mit $\mathfrak{a} \neq \langle 0 \rangle$ eine Gröbnerbasis bzgl. \leq_T .

Definition 5.8 (Monomiales Ideal)

Ein Ideal $\mathfrak{a} \subseteq R[x_1, \dots, x_n]$ heißt monomial, wenn es eine Familie \mathcal{F} von Monomen gibt, $\mathcal{F} \subseteq \{c \cdot t \mid c \in R, t \in T\}$, mit $\mathcal{F} \subset \mathfrak{a}$ und

$$f \in \mathfrak{a} \Rightarrow \exists m \in \mathbb{N}, \exists c_1 t_1, \dots, c_m t_m \in \mathcal{F} : f = \sum_{k=1}^m c_k t_k.$$

Man nennt dann \mathfrak{a} von \mathcal{F} erzeugt.

Satz 5.3 (Eigenschaften monomialer Ideale)

Sei $\mathfrak{a} \subseteq R[x_1, \dots, x_n]$ ein Ideal. Dann sind äquivalent

- i) \mathfrak{a} ist monomial,
- ii) $\exists \{t_1^*, \dots, t_s^*\} \subset T, \{c_1, \dots, c_s\} \subset R : \mathfrak{a} = \langle c_1 t_1^*, \dots, c_s t_s^* \rangle$,
- iii) $\sum_{j=1}^r b_j t_j \in \mathfrak{a} \Rightarrow b_j t_j \in \mathfrak{a} \forall j = 1, \dots, r$.

Definition 5.9 (Gröbnerbasis)

Sei $\mathfrak{a} \subseteq R[x_1, \dots, x_n]$ ein Ideal mit $\mathfrak{a} \neq \langle 0 \rangle$ und \leq_T eine Termordnung. $\{f_1, \dots, f_r\} \subset \mathfrak{a}$ heißt Gröbnerbasis von \mathfrak{a} bezüglich \leq_T , wenn

$$\langle \mathbf{lm}(f_1), \dots, \mathbf{lm}(f_r) \rangle$$

das von $\{\mathbf{lm}(f) \mid f \in \mathfrak{a}\}$ erzeugte monomiale Ideal ist.

Korollar (von Satz 5.3)

Jedes Ideal $\mathfrak{a} \subseteq R[x_1, \dots, x_n]$ mit $\mathfrak{a} \neq \langle 0 \rangle$ hat eine Gröbnerbasis bzgl. \leq_T .

Satz 5.4 (G-Darstellung)

$\{f_1, \dots, f_r\} \subset \mathfrak{a}$ ist genau dann Gröbnerbasis von \mathfrak{a} bzgl. \leq_T , wenn für alle $f \in \mathfrak{a}$ eine sogenannte G-Darstellung existiert,

$$\exists g_1, \dots, g_r \in R[x_1, \dots, x_n] : f = \sum_{k=1}^r g_k f_k \quad \text{mit } \mathbf{lt}(f) = \max_k \mathbf{lt}(g_k) \mathbf{lt}(f_k).$$

Bemerkung 1. Hiermit ist eine Gröbnerbasis tatsächlich eine Idealbasis und das Korollar von Satz 5.3 eine Verschärfung von Hilberts Basissatz.

Bemerkung 2. Die G-Darstellung von Polynomen des Ideals nach Satz 5.4 erlaubt eine effektive Entscheidung, ob ein Polynom f zu einem Ideal \mathfrak{a} gehört, sobald eine Gröbnerbasis von \mathfrak{a} bekannt ist. Für allgemeine Termordnungen wird das in §5.2 im Detail gezeigt. Hier können wir es uns aber schon im Fall einer gradverträglichen Termordnung und R ein Körper klar machen, denn in diesem Fall liegt ein Polynom f im (endlichdimensionalen!)

R -linearen Raum aller Polynome vom Grad $\leq \deg(\mathbf{1t}(f)) =: d$ und der Unterraum aller Polynome vom Grad $\leq d$ in \mathfrak{a} wird erzeugt von den Polynomen $\{t \cdot f_i \mid t \in T, \deg(t \cdot f_i) \leq d, 1 \leq i \leq r\}$, wobei $\{f_1, \dots, f_r\}$ eine Gröbnerbasis von \mathfrak{a} ist. Das Problem, ob ein Element zu einem endlichdimensionalen Vektorraum gehört ist aber ein (einfach lösbares) Problem der linearen Algebra, wenn (wie hier) vom Vektorraum ein endliches Erzeugendensystem bekannt ist.

Satz 5.5 (Gröbnerbasen beim Übergang von R zu $Q(R)$)

Sei R ein noetherscher Integritätsring, \leq_T eine Termordnung und $\{f_1, \dots, f_r\}$ eine Gröbnerbasis von $\mathfrak{a} := \langle f_1, \dots, f_r \rangle$ in $R[x_1, \dots, x_n]$ bzgl. \leq_T . $K := Q(R)$ sei der Quotientenkörper. Dann ist

$$\mathfrak{a}^* := \{f \in K[x_1, \dots, x_n] \mid \exists s \in R \setminus \{0\} : sf \in \mathfrak{a}\}$$

ein Ideal in $K[x_1, \dots, x_n]$ mit Gröbnerbasis $\{f_1, \dots, f_r\}$ bzgl. \leq_T .

Definition 5.10 (starke und schwache Gröbnerbasen)

Sei R Integritätsring und \leq_T Termordnung. $\{f_1, \dots, f_r\} \subset R[x_1, \dots, x_n]$ wird schwache Gröbnerbasis bzgl. \leq_T genannt, wenn $\{f_1, \dots, f_r\}$ in $Q(R)[x_1, \dots, x_n]$ Gröbnerbasis bzgl. \leq_T ist. Sie heißt starke Gröbnerbasis, wenn sie Gröbnerbasis ist mit

$$\mathbf{1m}(f) \in \langle \mathbf{1m}(f_1), \dots, \mathbf{1m}(f_r) \rangle \Rightarrow \exists i \in \{1, \dots, r\} : \mathbf{1t}(f_i) \mid \mathbf{1t}(f), \mathbf{1c}(f_i) \mid \mathbf{1c}(f).$$

Bemerkung. Nach Satz 5.5 sind Gröbnerbasen insbesondere schwache Gröbnerbasen. Wenn R ein Körper ist, sind Gröbnerbasen auch starke Gröbnerbasen. Starke Gröbnerbasen erhält man aus schwachen mittels einer Vervollständigungsverfahren. Mehr über schwache und starke Gröbnerbasen findet man z.B. im Buch von William W. Adams, An introduction to Groebner bases, AMS 1994.

§5.2 Gröbnerbasen

Ab sofort ist der Koeffizientenring ein Körper, also $\mathcal{P} = K[x_1, \dots, x_n]$ mit K Körper. Daher ist jetzt z.B. die Normierung $\mathbf{1c}(f) = 1$ möglich und eine Menge $\{f_1, \dots, f_r\}$ ist genau dann Gröbnerbasis, wenn gilt

$$f \in \langle f_1, \dots, f_r \rangle \Rightarrow \exists i \in \{1, \dots, r\} : \mathbf{1t}(f_i) \mid \mathbf{1t}(f).$$

Definition 5.11 (Reduktion modulo \mathcal{F})

Sei $\mathcal{F} := \{f_1, \dots, f_m\} \subset \mathcal{P} \setminus \{0\}$ und $f, g \in \mathcal{P}$. Man sagt, f reduziert auf g

modulo \mathcal{F} , kurz: $f \rightarrow_{\mathcal{F}} g$, wenn $f \neq 0$ und $\exists f_i \in \mathcal{F} : \mathbf{lt}(f_i) | \mathbf{lt}(f)$ und

$$g = f - \frac{\mathbf{lc}(f)}{\mathbf{lc}(f_i)} f_i.$$

Gilt $f \rightarrow_{\mathcal{F}} g$, dann ist $\mathbf{lt}(g) <_T \mathbf{lt}(f)$ sowie $f - g \in \langle f_1, \dots, f_m \rangle$.

Wir schreiben $f \rightarrow_{\mathcal{F}}^* g$ für den transitiven und reflexiven Abschluß der Relation $\rightarrow_{\mathcal{F}}$. Weil aus

$$f \rightarrow_{\mathcal{F}} g_1 \rightarrow_{\mathcal{F}} g_2 \rightarrow_{\mathcal{F}} g_3 \rightarrow_{\mathcal{F}} \dots$$

folgt, dass $\exists s \in \mathbb{N} : g_s = g_{s+1} = \dots$ (denn \leq_T ist eine Wohlordnung), ist jede Reduktionskette $f \rightarrow_{\mathcal{F}}^* g$ endlich.

Definition 5.12 (Irreduzibel modulo \mathcal{F})
 $f \in \mathcal{P}$ heißt irreduzibel modulo \mathcal{F} , wenn

$$f \rightarrow_{\mathcal{F}}^* g \Rightarrow f = g.$$

Definition 5.13 (Vollständiges Reduzieren modulo \mathcal{F})

Ist $f = \sum_{k=1}^m c_k t_k$ mit $c_k \in K \setminus \{0\}$ und $t_k \in T$, $k = 1, \dots, m$, und $\mathcal{F} = \{f_1, \dots, f_r\} \subset \mathcal{P} \setminus \{0\}$, dann sagt man, f reduziert mittels vollständiger Reduktion modulo \mathcal{F} auf $g \in \mathcal{P}$, wenn

$$\exists f_i \in \mathcal{F} : \exists k \in \{1, \dots, m\} : \mathbf{lt}(f_i) | t_k, \quad g = f - \frac{c_k}{\mathbf{lc}(f_i)} \frac{t_k}{\mathbf{lt}(f_i)} f_i,$$

in Zeichen $f \xrightarrow{\mathcal{V}}_{\mathcal{F}} g$. Mit $\xrightarrow{\mathcal{V}}_{\mathcal{F}}^*$ wird wieder der reflexive und transitive Abschluß von $\xrightarrow{\mathcal{V}}_{\mathcal{F}}$ bezeichnet. Und f heißt vollständig irreduzibel modulo \mathcal{F} , wenn $f \xrightarrow{\mathcal{V}}_{\mathcal{F}}^* g \Rightarrow f = g$.

Definition 5.14 (Minimale Gröbnerbasis)

Sei $G := \{g_1, \dots, g_r\}$ eine Gröbnerbasis mit

i) $\mathbf{lc}(g_i) = 1$, $i = 1, \dots, r$,

ii) jedes g_i sei vollständig irreduzibel modulo $G \setminus \{g_i\}$, $i = 1, \dots, r$.

Dann heißt G minimale Gröbnerbasis.

Satz 5.6 (Existenz minimaler Gröbnerbasen)

Jedes Ideal $\mathfrak{a} \subseteq \mathcal{P} \setminus \{0\}$ besitzt bzgl. jeder Termordnung \leq_T eine minimale Gröbnerbasis.

Satz 5.7 (Eindeutigkeit minimaler Gröbnerbasen)

Sind G_1 und G_2 minimale Gröbnerbasen desselben Ideals \mathfrak{a} und bzgl. der selben Termordnung \leq_T , dann gilt $G_1 = G_2$.

Satz 5.8 (Zusammenhang zwischen Reduktion und G-Basisdarstellung)

Sei $\mathcal{F} \subset \mathcal{P} \setminus \{0\}$ und $f \in \mathcal{P} \setminus \{0\}$ (und \leq_T eine Termordnung). Dann sind äquivalent

i) $f \xrightarrow{\mathcal{F}}^* 0$,

ii) Es gibt $c_i \in K \setminus \{0\}$, $t_i \in T$, $h_i \in \mathcal{F}$, $i = 1, \dots, s$, mit $f = \sum_{i=1}^s c_i t_i h_i$ und

$$t_s \cdot \text{lt}(h_s) <_T t_{s-1} \cdot \text{lt}(h_{s-1}) <_T \dots <_T t_2 \cdot \text{lt}(h_2) <_T t_1 \cdot \text{lt}(h_1).$$

Die Darstellung aus Satz 5.8 ii) lässt sich für $\mathcal{F} = \{f_1, \dots, f_m\}$ auch schreiben als

$$f = \sum_{k=1}^m g_k f_k \quad \text{mit} \quad \text{lt}(f) = \max_{\leq_T} \{\text{lt}(g_1)\text{lt}(f_1), \dots, \text{lt}(g_m)\text{lt}(f_m)\},$$

wobei jedes g_k Summe der $c_i t_i$ ist, bei denen $h_i = f_k$ gilt. Genau für ein $\text{lt}(g_k)\text{lt}(f_k)$ wird das Maximum $\max_{\leq_T} \{\text{lt}(g_1)\text{lt}(f_1), \dots, \text{lt}(g_m)\text{lt}(f_m)\}$ angenommen.

Zu jeder endlichen Polynommenge $\mathcal{F} := \{f_1, \dots, f_m\} \subset \mathcal{P} \setminus \{0\}$ und jedem $f \in \mathcal{P}$ betrachte man die Menge aller Darstellungen

$$D_{\mathcal{F}}(f) := \{(g_1, \dots, g_m) \in \mathcal{P}^m \mid f = \sum_{k=1}^m g_k f_k\}.$$

Die Elemente von $D_{\mathcal{F}}(0)$ heißen *Syzygien*,

$$\text{Syz}(\mathcal{F}) := D_{\mathcal{F}}(0) = \{(g_1, \dots, g_m) \in \mathcal{P}^m \mid \sum_{k=1}^m g_k f_k = 0\}.$$

Genau dann, wenn $f \in \langle \mathcal{F} \rangle$ gilt, ist $D_{\mathcal{F}}(f)$ nicht leer. In diesem Fall ist

$$D_{\mathcal{F}}(f) = (g_1^*, \dots, g_m^*) + \text{Syz}(\mathcal{F})$$

mit einem beliebigen $(g_1^*, \dots, g_m^*) \in D_{\mathcal{F}}(f)$.

Im freien Modul \mathcal{P}^m lassen sich auch (Modul-)Terme definieren. Dann kann man mit Hilfe von $\mathcal{F} := \{f_1, \dots, f_m\}$ auch eine (Modul-)Termordnung definieren. Sind e_1, \dots, e_m die Einheitsvektoren in K^m , dann sind die (Modul-)

Terme alle $t \cdot e_i$, $t \in T$, $i \in \{1, \dots, m\}$. Die Menge \tilde{T} dieser Terme ordnet man linear mittels

$$t \cdot e_i < t' \cdot e_j \Leftrightarrow \left(t \cdot \mathbf{lt}(f_i) <_T t' \cdot \mathbf{lt}(f_j) \vee (t \cdot \mathbf{lt}(f_i) = t' \cdot \mathbf{lt}(f_j), i < j) \right).$$

In \mathcal{P}^m lassen sich mit Hilfe von $\mathcal{F} := \{f_1, \dots, f_m\}$ auch Leiterteile definieren. Sind e_1, \dots, e_m die Einheitsvektoren in K^m , dann sind hier die Terme alle $t \cdot e_i$, $t \in T$, $i \in \{1, \dots, m\}$. Die Menge \tilde{T} dieser Terme ordnet man linear mittels

$$t \cdot e_i < t' \cdot e_j \Leftrightarrow \left(t \cdot \mathbf{lt}(f_i) <_T t' \cdot \mathbf{lt}(f_j) \vee (t \cdot \mathbf{lt}(f_i) = t' \cdot \mathbf{lt}(f_j), i < j) \right).$$

Man definiert für beliebige $G := (g_1, \dots, g_m) \in \mathcal{P}^m$ mit $G \neq 0$

$$\begin{aligned} M(G) &:= \max_{i=1}^m \mathbf{lt}(g_i) \mathbf{lt}(f_i) \in T && \text{(Maximalterm)} \\ J(G) &:= \{j \in \mathbb{N} \mid j \leq m, \mathbf{lt}(g_j) \mathbf{lt}(f_j) = M(G)\} && \text{(Menge der Maximalindizes)} \\ \mathbf{lt}(G) &:= M(G) \cdot e_k \text{ mit } k = \max J(G) && \text{(Modul-Leiterteil)} \end{aligned}$$

Mit diesen Begriffen ist $f = \sum_{k=1}^m g_k f_k$ genau dann eine G-Darstellung, wenn $M(g_1, \dots, g_m) = \mathbf{lt}(f)$. Bei der G-Darstellung aus Satz 5.8 hat $J(G)$ nur ein Element. Ist G eine Syzygie, dann hat $J(G)$ mindestens zwei Elemente.

Satz 5.9 (Charakterisierung von Gröbnerbasen)

Für $\mathcal{F} := \{f_1, \dots, f_m\} \subset \mathcal{P} \setminus \{0\}$ und $\mathfrak{a} := \langle f_1, \dots, f_m \rangle$ sind äquivalent

- i) \mathcal{F} ist Gröbnerbasis von \mathfrak{a} .
- ii) $f \in \mathfrak{a} \Rightarrow f \xrightarrow{*}_{\mathcal{F}} 0$.
- iii) $\forall 1 \leq i < j \leq m : S(f_i, f_j) \xrightarrow{*}_{\mathcal{F}} 0$.

Hierbei ist

$$S(f, g) := \frac{kgV(\mathbf{lt}(f), \mathbf{lt}(g))}{\mathbf{lc}(f)\mathbf{lt}(f)} f - \frac{kgV(\mathbf{lt}(f), \mathbf{lt}(g))}{\mathbf{lc}(g)\mathbf{lt}(g)} g$$

das sogenannte S-Polynom von f und g .

Bemerkung 1. Im Beweisteil iii) \Rightarrow i) wird konstruktiv gezeigt, wie man aus einer beliebigen Darstellung $f = \sum_{i=1}^m g_i f_i$ mit Hilfe von S-Polynomen eine G-Darstellung für f bekommen kann. Dabei kann man sich klarmachen, dass nicht alle S-Polynome gebraucht werden, man also in iii) einige S-Polynome weglassen darf. Sind nämlich zwei S-Polynomen $S(f_i, f_k)$ und $S(f_j, f_k)$ mit $i < j < k$ und $\varphi_{ik} := k.g.V.(\mathbf{lt}(f_i), \mathbf{lt}(f_k))$, $\varphi_{jk} := k.g.V.(\mathbf{lt}(f_j), \mathbf{lt}(f_k))$, dann braucht man $S(f_j, f_k)$ nicht, wenn $\varphi_{ik} \mid \varphi_{jk}$ bzw. $S(f_i, f_k)$ nicht, wenn $\varphi_{jk} \mid \varphi_{ik}$.

Bemerkung 2. Ist \mathcal{F} eine Gröbnerbasis und hat man für ein $f \in \langle \mathcal{F} \rangle$ eine Darstellung $f = \sum_{i=1}^m g_i f_i$, die noch keine G-Darstellung ist, so kann man die im Beweisschritt iii) \Rightarrow i) vorgestellte Methode benutzen, um eine G-Darstellung von f zu erhalten.

Satz 5.10 (Eindeutigkeit der irreduziblen Elemente)

Seien G_1 und G_2 beide Gröbnerbasen von \mathfrak{a} bezüglich der Termordnung \leq_T und $f \in \mathcal{P}$. Wenn

$$f \xrightarrow{V}_{G_1}^* f_1, f \xrightarrow{V}_{G_2}^* f_2,$$

wobei f_1 und f_2 jeweils vollständig irreduzibel modulo G_1 bzw. G_2 sind, dann gilt $f_1 = f_2$.

Definition 5.15 (Normalform, Normalmenge)

Sei G eine Gröbnerbasis von \mathfrak{a} und $f \xrightarrow{V}_G^* f_1$ mit vollständig irreduziblem f_1 modulo G . Dann heißt f_1 Normalform von f bzgl. \mathfrak{a} ,

$$\text{NF}(f, \mathfrak{a}) := f_1.$$

Die Menge der modulo G irreduziblen Terme,

$$\mathcal{N}(\mathfrak{a}) := \{t \in T \mid t \notin \text{lt}(\mathfrak{a})\},$$

wird Normalmenge von \mathfrak{a} genannt.

Die Normalform $\text{NF}(f, \mathfrak{a})$ hängt nur von f und vom Ideal \mathfrak{a} ab (und von der Termordnung), nicht aber von der Gröbnerbasis, die Normalmenge nur vom Ideal (und der Termordnung). Bei festem Ideal \mathfrak{a} ist $\text{NF}(\cdot, \mathfrak{a})$ eine Projektion von \mathcal{P} auf $\text{span}_K \mathcal{N}(\mathfrak{a})$.

Ist G die minimale Gröbnerbasis eines Ideals \mathfrak{a} bzgl. \leq_T , dann hat jedes $g \in G$ die Gestalt

$$g = t + \sum_{k=1}^N c_k t_k \quad \text{mit } t \in \text{lt}(\mathfrak{a}), t_1, \dots, t_N \in \mathcal{N}(\mathfrak{a}), c_1, \dots, c_N \in K.$$

Satz 5.11 (Normalform als Repräsentant einer Äquivalenzklasse)

Für die Äquivalenzklasse $[f]$ im Faktoring \mathcal{P}/\mathfrak{a} gilt

$$[f] \cap \text{span}_K \mathcal{N}(\mathfrak{a}) = \{\text{NF}(f, \mathfrak{a})\}.$$

§5.3 Die Berechnung von Gröbnerbasen

Der Algorithmus von Buchberger berechnet zu gegebener Termordnung \leq_T und gegebener Basis $\mathcal{F} = \{f_1, \dots, f_m\}$ eines Ideals $\mathfrak{a} \subset R[x_1, \dots, x_n]$ eine Gröbnerbasis von \mathfrak{a} bzgl. \leq_T . Wir beschränken uns hier auf den Fall, dass der Koeffizientenring R ein Körper ist, obwohl Buchberger auch den Fall $R = \mathbb{Z}$ behandelt hat, den man leicht auf beliebige berechenbare Euklidische Ringe verallgemeinern kann.

Die Grundidee des Algorithmus' von Buchberger besteht darin, die Äquivalenz i) \Leftrightarrow iii) von Satz 5.9 auszunutzen. Wenn $S(f_i, f_j) \xrightarrow{*}_{\mathcal{F}} h \neq 0$ gilt und dabei h irreduzibel modulo \mathcal{F} ist, dann erweitert man einfach \mathcal{F} um h , setzt also $\mathcal{F}' := \mathcal{F} \cup \{h\}$. Dann gilt bei Reduktion mit dem erweiterten \mathcal{F}' ,

$$S(f_i, f_j) \xrightarrow{*}_{\mathcal{F}'} 0.$$

Die Grundversion des Algorithmus' von Buchberger ist wie folgt.

Algorithmus von Buchberger

Eingabe: $f_1, \dots, f_m \in K[x_1, \dots, x_n] \setminus \{0\}$, K Körper, Termordnung \leq_T .

Ausgabe: Eine Gröbnerbasis \mathcal{F} von $\langle f_1, \dots, f_m \rangle$ bzgl. \leq_T .

Rechnung: $B := \{(i, j) \mid 1 \leq i < j \leq m\}$

$$\mathcal{F} := \{f_1, \dots, f_m\}$$

MARKE: **if** $B = \emptyset$ **then return** \mathcal{F}

$$(I, J) := \text{selectfrom}(B)$$

$$B := B \setminus \{(I, J)\}$$

Berechne h , irreduzibel modulo \mathcal{F} mit

$$S(f_I, f_J) \xrightarrow{*}_{\mathcal{F}} h$$

Wenn $h = 0$, dann spring zurück zu MARKE.

$$f_{m+1} := h$$

$$\mathcal{F} := \mathcal{F} \cup \{f_{m+1}\}$$

$$B := B \cup \{(i, m+1) \mid 1 \leq i \leq m\}$$

$$m := m + 1$$

Spring zurück zu MARKE

Hierbei ist **selectfrom** eine Prozedur, die aus einer (nicht-leeren) Menge ein Element nach bestimmten Kriterien auswählt.

Damit dieses Verfahren zu Recht als Algorithmus bezeichnet werden kann, muss präzisiert werden, wie der Algorithmus **selectfrom** definiert ist und in welche Teilschritte bei der Reduktion $S(f_I, f_J) \xrightarrow{*}_{\mathcal{F}} h$ verwendet werden. Bei **selectfrom** wird gern die von Buchberger vorgeschlagene Auswahl

$$(I, J) \in B, k.g.V.(\text{lt}(f_I), \text{lt}(f_J)) = \min_{\leq_T} \{k.g.V.(\text{lt}(f_i), \text{lt}(f_j)) \mid (i, j) \in B\}$$

getroffen. Falls hierdurch (I, J) noch nicht eindeutig bestimmt ist, wählt man unter den infrage kommenden Paaren das Paar (i, j) mit kleinster zweiter Komponente. Falls das immer noch nicht eindeutig ist, dann unter diesen wiederum das mit kleinster erster Komponente. Bei der Reduktionsreihenfolge $S(f_I, f_J) \rightarrow_{\mathcal{F}}^* h$ variieren die Methoden. Wenn schon $S(f_I, f_J) \rightarrow_{\mathcal{F}}^* h_i$ berechnet wurde, aber für h_i mehrere $f \in \mathcal{F}$ existieren mit $\mathbf{1t}(f) | \mathbf{1t}(h_i)$, dann nimmt man bei manchen Methoden das f mit minimalem $\mathbf{1t}(f)$. Alternativ kann man das f wählen mit möglichst einfachen Koeffizienten oder bevorzugt die f , die schon bei der Eingabe in \mathcal{F} waren o.ä.

Definition 5.16 (redundant)

Sei \mathcal{F} eine Gröbnerbasis. Zu $f \in \mathcal{F}$ existiere ein $\tilde{f} \in \mathcal{F}$, $f \neq \tilde{f}$, mit $\mathbf{1t}(\tilde{f}) | \mathbf{1t}(f)$. Dann heißt f redundant in \mathcal{F} .

Bemerkung. $\mathcal{F} \setminus \{f\}$ ist dann eine Gröbnerbasis desselben Ideals.

Satz 5.12 (Kriterium von Buchberger)

Sind $\mathbf{1t}(f_i)$ und $\mathbf{1t}(f_j)$ teilerfremd, dann gilt $S(f_i, f_j) \rightarrow_{\{f_i, f_j\}}^* 0$.

Für Handrechnung empfiehlt sich, statt die Paare in B zu sammeln, besser zu jedem Paar (i, j) , dessen S-Polynom $S(f_i, f_j)$ noch in Buchbergers Algorithmus reduziert werden muss, den Term $\varphi_{ij} := k.g.V.(\mathbf{1t}(f_i), \mathbf{1t}(f_j))$ zu notieren. Man streicht danach solche Terme φ_{ij} , für die entweder $S(f_i, f_j) \rightarrow_{\mathcal{F}}^* 0$ wegen Bemerkung 1 nach Satz 5.9 nicht gebraucht wird oder $S(f_i, f_j) \rightarrow_{\mathcal{F}}^* 0$ gilt wegen Buchbergers Teilerfremdheitskriterium (Satz 5.12). Wegen $i < j$ für Paare (i, j) aus B hat man also ein Dreiecksschema

	f_2	f_3	\dots	f_m
f_1	φ_{12}	φ_{13}	\dots	φ_{1m}
f_2	-	φ_{23}	\dots	φ_{2m}
\vdots			\ddots	\vdots
f_{m-1}	-	-	-	$\varphi_{m-1,m}$

in dem spaltenweise einige Terme φ_{ij} gestrichen werden. Nimmt man an, das f_m gerade in \mathcal{F} eingefügt worden, so hat man die letzte Spalte zusammenzustreichen wie oben angegeben. Unter den noch nicht gestrichenen φ_{ij} im Schema sucht man dann das kleinste bzgl. \leq_T . Gibt es mehrere, dann nimmt man nach Buchbergers Empfehlung das, was in der am weitesten links stehenden Spalte des Schemas steht. (Man streicht dann dieses φ_{ij} und reduziert dann $S(f_i, f_j)$ modulo \mathcal{F} , bis man bei einem irreduziblen Polynom h aufhören muss.)

§6 Rechnen in nulldimensionalen Idealen

Hier ist immer noch K ein Körper mit $\mathbb{Q} \subseteq K \subseteq \mathbb{C}$ und $\mathcal{P} = K[x_1, \dots, x_n]$.

Definition 6.1 (Varietät)

Sei $\mathfrak{a} \subseteq \mathcal{P}$ ein Ideal. Dann wird

$$V(\mathfrak{a}) := \{y \in \mathbb{C}^n \mid f(y) = 0 \quad \forall f \in \mathfrak{a}\}$$

die Varietät von \mathfrak{a} genannt.

Bemerkung. Gilt $\mathfrak{a} = \langle f_1, \dots, f_m \rangle$, dann auch

$$V(\mathfrak{a}) = \{y \in \mathbb{C}^n \mid f_1(y) = f_2(y) = \dots = f_m(y) = 0\}.$$

Satz 6.1 (Hilberts Nullstellensatz)

Sei $\mathfrak{a} \subseteq \mathcal{P}$ ein Ideal. Dann gilt für jedes $f \in \mathcal{P}$

$$f(y) = 0 \quad \forall y \in V(\mathfrak{a}) \Rightarrow \exists s \in \mathbb{N} : f^s \in \mathfrak{a}.$$

(Ohne Beweis!)

Definition und Satz 6.2 (Radikal)

Sei M eine Teilmenge von \mathbb{C}^n und

$$\mathfrak{a} := \{f \in \mathcal{P} \mid f(y) = 0 \quad \forall y \in M\}.$$

Dann ist \mathfrak{a} ein Ideal und wird Radikal genannt. Man schreibt $\mathfrak{a} = I(M)$.

Bemerkung 1. Es gilt $M \subseteq V(I(M))$.

Bemerkung 2. Aus Hilberts Nullstellensatz folgt, dass \mathfrak{a} genau dann ein Radikal ist, wenn für beliebige $f \in \mathcal{P}$, $s \in \mathbb{N}$ gilt $f^s \in \mathfrak{a} \Rightarrow f \in \mathfrak{a}$.

Bemerkung 3. Zum Ideal \mathfrak{a} gehört das Radikal $\sqrt{\mathfrak{a}} := \{f \in \mathcal{P} \mid \exists s \in \mathbb{N} : f^s \in \mathfrak{a}\}$ mit $V(\mathfrak{a}) = V(\sqrt{\mathfrak{a}})$. Wegen $\sqrt{I(M)} = I(M)$ kann jedes Radikal als $\sqrt{\mathfrak{a}}$ geschrieben werden.

Definition 6.3 (irreduzible und reduzible Punktmenge)

Eine Menge $M \subseteq \mathbb{C}^n$ heißt reduzibel, wenn es zwei Teilmengen $M_1, M_2 \subset M$ gibt mit $I(M_1) \neq I(M) \neq I(M_2)$ und $M_1 \cap M_2 = \emptyset, M_1 \cup M_2 = M$. Sonst heißt sie irreduzibel.

Jede reduzible Varietät ist der endliche Durchschnitt von irreduziblen.

Die Dimension einer irreduziblen Varietät $V(\mathfrak{a})$ wird nach van der Waerden wie folgt bestimmt. Ist $M \subset \mathbb{C}^n$ so, dass $I(M) = \mathfrak{a}$ gilt und benötigt man zur Beschreibung der Punkte von M d Parameter, d.h. gibt es einen Punkt $y^* \in \mathbb{C}(t_1, \dots, t_d)^n$ (d.i. n -fache Kopie einer d -fachen transzendenten Körpererweiterung von \mathbb{C}), so dass jedes $y \in M$ durch Spezifikation der t_1, \dots, t_d aus y^* gewonnen wird, dann heißt das kleinste derartige d die Dimension von $V(\mathfrak{a})$. Die Dimension einer reduziblen Varietät ist das Maximum der Dimensionen seiner irreduziblen Bestandteile. Die Dimension eines Ideals \mathfrak{a} ist die Dimension von $V(\mathfrak{a})$.

Folge: Ein Ideal \mathfrak{a} wird nulldimensional genannt, wenn seine Varietät aus endlich vielen Punkten in \mathbb{C}^n besteht.

§6.1 Nulldimensionale Ideale

Satz 6.3 (Charakterisierung nulldimensionaler Ideale)

Sei \mathfrak{a} ein Ideal in $\mathcal{P} = K[x_1, \dots, x_n]$. Dann sind äquivalent

- i) \mathfrak{a} ist nulldimensional,
- ii) Zu jedem $i \in \{1, \dots, n\}$ gibt es ein $0 \neq f_i \in \mathfrak{a} \cap K[x_i]$,
- iii) Ist G eine Gröbnerbasis von \mathfrak{a} , dann gibt es zu jedem $i \in \{1, \dots, n\}$ ein $g \in G$ und ein $m_i \in \mathbb{N}$ mit $\text{lt}(g) = x_i^{m_i}$,
- iv) $\dim \mathcal{P}/\mathfrak{a} < \infty$.

Bemerkung 1. Im Beweisschritt iv) \Rightarrow i) des Satzes wird beschrieben, wie man ein univariates Polynom in einem nulldimensionalen Ideal \mathfrak{a} findet: Weil \mathcal{P}/\mathfrak{a} endlichdimensional ist, gibt es ein kleinstes $s \in \mathbb{N}$, so dass die Äquivalenzklassen

$$[1], [x_k], [x_k^2], \dots, [x_k^s]$$

lin. abh. sind. Aus $[x_k^s] = \sum_{i=0}^{s-1} c_i [x_k^i]$ folgt dann, dass $x_k^s - \sum_{i=0}^{s-1} c_i x_k^i \in \mathfrak{a}$ gilt.

Bemerkung 2. Ist \mathfrak{a} ein Ideal mit $\mathfrak{a} \cap K[x_k] \neq \{0\}$, und ist G eine Gröbnerbasis von \mathfrak{a} bezüglich einer lexikographischen Termordnung \leq_T mit $x_k <_T x_i$ für $i \neq k$, dann enthält G ein Polynom, das nur von x_k abhängt. (Diese Methode ist bei nulldimensionalen Idealen viel zu aufwändig, erfordert sie doch die Berechnung von n Gröbnerbasen für \mathfrak{a} .)

Definition 6.4 (einfache Nullstellen eines Ideals)

Sei \mathfrak{a} nulldimensionales Ideal. $y \in V(\mathfrak{a})$ heißt einfache Nullstelle von \mathfrak{a} , wenn

es keine Richtungsableitung $\sum_{i=1}^n c_i \frac{\partial}{\partial x_i} \neq 0$ gibt mit

$$\sum_{i=1}^n c_i \frac{\partial f}{\partial x_i}(y) = 0 \quad \text{für alle } f \in \mathfrak{a}.$$

Ist $\mathfrak{a} = \langle f_1, \dots, f_m \rangle$ nulldimensional und $y \in V(\mathfrak{a})$, dann ist y genau dann einfache Nullstelle von \mathfrak{a} , wenn gilt

$$\text{Rang} \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(y) & \cdots & \frac{\partial f_1}{\partial x_n}(y) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(y) & \cdots & \frac{\partial f_m}{\partial x_n}(y) \end{pmatrix} = n.$$

Im Fall $n = 1$ ist \mathfrak{a} ein Hauptideal, $\mathfrak{a} = \langle p \rangle$, $V(\mathfrak{a})$ sind die Nullstellen von p und eine Nullstelle y ist genau dann einfach, wenn mit dem eben hergeleiteten Kriterium $\text{Rang}(p'(y)) = 1$ gilt, wenn also $p'(y) \neq 0$.

Bemerkung. Im Fall $n = 1$ ist die Varietät eines Ideals $\langle p \rangle$ die Menge der Nullstellen von p . Ist $s(y)$ die Vielfachheit von y als Nullstelle von p , dann gilt

$$f \in \langle p \rangle \Leftrightarrow f(y) = f'(y) = \dots = f^{(s(y)-1)}(y) = 0 \quad \text{für alle } y \in V(\langle p \rangle).$$

Im Fall $n > 1$ gibt es eine analoge Beschreibung. Setzt man

$$D(\alpha_1, \dots, \alpha_n) := \begin{cases} \frac{1}{\alpha_1! \cdots \alpha_n!} \frac{\partial^{\alpha_1 + \dots + \alpha_n}}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}} & \text{falls } \alpha_1, \dots, \alpha_n \in \mathbb{N}_0, \\ 0 & \text{falls } \alpha_1, \dots, \alpha_n \in \mathbb{Z} \text{ aber } \exists i : \alpha_i < 0, \end{cases}$$

dann nennt man eine (endliche) Linearkombination L (Koeffizienten aus K) von Operatoren $D(\alpha_1, \dots, \alpha_n)$ einen Ableitungsoperator. Ein endlich-dimensionaler K -Vektorraum W von solchen Ableitungsoperatoren heißt abgeschlossen, wenn mit $L = \sum_{\alpha_1, \dots, \alpha_n} c_{\alpha_1, \dots, \alpha_n} D(\alpha_1, \dots, \alpha_n)$ auch

$$\sum_{\alpha_1, \dots, \alpha_n} c_{\alpha_1, \dots, \alpha_n} D(\alpha_1 - \beta_1, \dots, \alpha_n - \beta_n) \quad \text{für jedes } (\beta_1, \dots, \beta_n) \in \mathbb{N}_0^n$$

zu W gehört. Ist \mathfrak{a} nun ein nulldimensionales Ideal in $\mathcal{P} = K[x_1, \dots, x_n]$, dann gehört zu jedem Punkt $y \in V(\mathfrak{a})$ ein abgeschlossener K -Vektorraum $W(y)$ (sog. Max-Noether-Raum), sodass

$$f \in \mathfrak{a} \Leftrightarrow Df(y) = 0 \quad \text{für alle } D \in W(y) \text{ und alle } y \in V(\mathfrak{a}).$$

In dieser Äquivalenz kann man auch “ $Df(y) = 0$ für alle $D \in W(y)$ ” durch “ $Df(y) = 0$ für alle D aus einer (Vektorraum-)Basis von $W(y)$ ” ersetzen. Es gilt

$$\dim \mathcal{P}/\mathfrak{a} = \sum_{y \in V(\mathfrak{a})} \dim W(y).$$

Ist $y \in V(\mathfrak{a})$ eine einfache Nullstelle, dann gilt $\dim W(y) = 1$ und wenn alle $y \in V(\mathfrak{a})$ einfache Nullstellen von \mathfrak{a} sind, folgt $\dim \mathcal{P}/\mathfrak{a} = |V(\mathfrak{a})|$ und sonst $\dim \mathcal{P}/\mathfrak{a} > |V(\mathfrak{a})|$.

Ist p ein univariates Polynom mit mehrfachen Nullstellen, dann hat

$$\frac{p}{g.g.T.(p, p')}$$

dieselben Nullstellen wie p , aber alle Nullstellen sind einfach.

Satz 6.4 (Charakterisierung nulldimensionaler Radikale)

Sei \mathfrak{a} ein nulldimensionales Ideal. Dann sind äquivalent

- i) \mathfrak{a} ist ein Radikal
- ii) alle $y \in V(\mathfrak{a})$ sind einfache Nullstellen von \mathfrak{a} .

Korollar (Konstruktion von $\sqrt{\mathfrak{a}}$ aus \mathfrak{a})

Ist $\mathfrak{a} = \langle f_1, \dots, f_m \rangle$ nulldimensional und p_i Basis des Hauptideals $\mathfrak{a} \cap K[x_i]$, dann bekommt man mit

$$\tilde{p}_i := \frac{p_i}{g.g.T.(p_i, \frac{\partial p_i}{\partial x_i})}, \quad i = 1, \dots, n,$$

das Radikal von \mathfrak{a} aus $\sqrt{\mathfrak{a}} = \langle f_1, \dots, f_m, \tilde{p}_1, \dots, \tilde{p}_n \rangle$.

Definition 6.5 (maximale Ideale)

Ein Ideal \mathfrak{a} heißt maximal, wenn $\mathfrak{a} \neq \mathcal{P}$ und wenn für jedes Ideal \mathfrak{b} gilt

$$\mathfrak{a} \subseteq \mathfrak{b} \subseteq \mathcal{P} \Rightarrow (\mathfrak{a} = \mathfrak{b} \text{ oder } \mathfrak{b} = \mathcal{P}).$$

Satz 6.5 (algebraische Körpererweiterung)

Sei \mathfrak{a} ein Ideal in $\mathcal{P} := K[x_1, \dots, x_n]$. \mathcal{P}/\mathfrak{a} ist genau dann ein Körper, wenn \mathfrak{a} maximal ist.

Satz 6.6 (Charakterisierung maximaler Ideale)

Ein Ideal \mathfrak{a} ist genau dann maximal, wenn es nulldimensionales Radikal ist mit irreduzibler Varietät $V(\mathfrak{a})$.

Bemerkung. Ist \mathfrak{a} Ideal im Hauptidealring $K[x]$ (also $n = 1$), dann gilt: $\mathfrak{a} = \langle f \rangle$ ist genau dann maximal, wenn f irreduzibel ist, d.h., keinen echten Teiler in $K[x]$ hat.

Konstruktion maximaler Ideale \mathfrak{m} mit $\mathfrak{a} \subseteq \mathfrak{m}$ (und $\dim(\mathfrak{a}) = 0$)

Sei $\{g_1, \dots, g_r\}$ minimale lexikographische Gröbnerbasis von \mathfrak{a} mit $x_1 <_{LEX} x_2 <_{LEX} \dots <_{LEX} x_n$ und $\mathbf{lt}(g_1) <_{LEX} \mathbf{lt}(g_2) <_{LEX} \dots <_{LEX} \mathbf{lt}(g_r)$.

Dann hat das Ideal $\mathfrak{a} \cap K[x_1]$ in $K[x_1]$ die Basis g_1 . Die Zerlegung von g_1 in Primfaktoren sei

$$g_1 = p_1^{\sigma_1} \cdot p_1^{\sigma_1} \cdots p_m^{\sigma_m},$$

p_i irreduzibel, also ohne echte Teiler in $K[x_1]$, und $\text{g.g.T.}(p_i, p_j) = 1$ für $i \neq j$, sowie $\sigma_1, \dots, \sigma_m \in \mathbb{N}$. Dann gilt

$$V(\mathfrak{a}) = \bigcup_{i=1}^m V(\langle p_i, g_1, g_2, \dots, g_m \rangle).$$

Für jedes $i \in \{1, \dots, m\}$ ist

$$\mathfrak{a} \subseteq \mathfrak{a}_i := \langle p_i, g_1, g_2, \dots, g_m \rangle = \langle p_i, g_2, \dots, g_m \rangle.$$

Wir suchen jetzt maximale Ideale \mathfrak{m} mit $\mathfrak{a}_i \subseteq \mathfrak{m}$.

Ist $y = (\xi_1, \dots, \xi_n) \in V(\mathfrak{a}_i)$, dann gilt $p_i(\xi_1) = 0$. $K(\xi_1)$ ist ein Körper. Die weiteren Koordinaten der Punkte $y \in V(\mathfrak{a})$ mit erster Koordinate ξ_1 bekommt man als gemeinsame Nullstellen der Polynome

$$\begin{aligned} \tilde{g}_2(x_2, \dots, x_n) &:= g_2(\xi_1, x_2, \dots, x_n) \in K(\xi_1)[x_2, \dots, x_n], \\ \tilde{g}_3(x_2, \dots, x_n) &:= g_3(\xi_1, x_2, \dots, x_n) \in K(\xi_1)[x_2, \dots, x_n], \\ &\vdots \\ \tilde{g}_r(x_2, \dots, x_n) &:= g_r(\xi_1, x_2, \dots, x_n) \in K(\xi_1)[x_2, \dots, x_n]. \end{aligned}$$

Diese Polynome erzeugen das Ideal $\mathfrak{a}_{x_1=\xi_1} := \{p(\xi_1, x_2, \dots, x_n) \mid p \in \mathfrak{a}\} \subset K(\xi_1)[x_2, \dots, x_n]$. Sei $\{h_1, \dots, h_s\}$ minimale lexikographische Gröbnerbasis von $\mathfrak{a}_{x_1=\xi_1}$ mit $x_2 <_{LEX} \dots <_{LEX} x_n$ und $\text{lt}(h_1) <_{LEX} \text{lt}(h_2) <_{LEX} \dots <_{LEX} \text{lt}(h_s)$. Dann hat das Ideal $\mathfrak{a}_{x_1=\xi_1} \cap K(\xi_1)[x_2]$ in $K(\xi_1)[x_2]$ die Basis h_1 . Man nimmt dann die Primfaktorzerlegung von h_1 (im Polynomring $K(\xi_1)[x_2]$) und bestimmt von einem irreduziblen Faktor eine Nullstelle ξ_2 . Jetzt kann man die maximalen Oberideale von \mathfrak{a}_i bekommen, deren Varietät nur Punkte mit erster Komponente ξ_1 und zweiter Komponente ξ_2 enthält usw.

Das Verfahren endet mit dem maximalen Ideal $\langle p_i, p_{2,i_2}, \dots, p_{n,i_n} \rangle$, wobei p_{2,i_2} ein irreduzibler Faktor des Basiselements von $K(\xi_1)[x_2] \cap \mathfrak{a}_{x_1=\xi_1}$ ist. Eine Nullstelle von p_{2,i_2} ist ξ_2 . Dann ist p_{3,i_3} ein irreduzibler Faktor des Basiselements von $\mathfrak{a}_{x_1=\xi_1, x_2=\xi_2} \cap K(\xi_1, \xi_2)[x_3]$ usw.

Man macht sich leicht klar, dass man jedes maximale Oberideal \mathfrak{m} von \mathfrak{a} bekommt, wenn man bei den auftretenden Primfaktorzerlegungen jeden möglichen Faktor zur Bestimmung der nächsten Koordinate nimmt.

Lemma (Das Substitutionslemma)

Sei \mathfrak{a} ein nulldimensionales Ideal in $\mathcal{P} = K[x_1, \dots, x_n]$ und $G := \{g_1, \dots, g_r\}$ eine Gröbnerbasis bzgl. \leq_{LEX} mit $x_1 <_{LEX} x_2 <_{LEX} \dots <_{LEX} x_n$. Sei g_1 das Polynom kleinsten Grads in $G \cap K[x_1]$ und $\xi \in \mathbb{C}$ eine Nullstelle von g_1 . Dann ist die Menge $\tilde{G} := \{g(\xi, x_2, \dots, x_n) \mid g \in G\}$ lexikographische Gröbnerbasis des Ideals $\mathfrak{a}_{x_1=\xi} = \{p(\xi, x_2, \dots, x_n) \mid p \in \mathfrak{a}\} \subset K(\xi)[x_2, \dots, x_n]$, $K(\xi)$ ein Körper. Insbesondere gibt es in $\tilde{G} \cap K(\xi)[x_2]$ ein Polynom, das alle anderen Polynome aus $\tilde{G} \cap K(\xi)[x_2]$ teilt.

Satz 6.8 (Charakterisierung mehrfacher algebraischer Körpererweiterungen)

L ist genau dann eine n -fache algebraische Körpererweiterung von K , wenn $L = K(\xi_1, \dots, \xi_n) \cong K[x_1, \dots, x_n]/\langle q_1, \dots, q_n \rangle$, wobei

$$\begin{aligned} q_i &\in K[x_1, \dots, x_i], \quad \exists m_i \in \mathbb{N} : \text{lt}(q_i) = x_i^{m_i}, \\ q_i(\xi_1, \dots, \xi_{i-1}, x_i) &\text{ irreduzibel in } K(\xi_1, \dots, \xi_{i-1})[x_i], \\ q_i(\xi_1, \dots, \xi_{i-1}, \xi_i) &= 0, \quad i = 1, \dots, n. \end{aligned}$$

(Für $i = 1$ heißt dies nur: $q_1 \in K[x_1]$ irreduzibel in $K[x_1]$ mit Nullstelle ξ_1 .)

Definition 6.6 (Spezielle Gröbnerbasen)

Man nennt eine Polynommenge $G := \{g_1, \dots, g_n\}$ eine Dreiecksbasis, wenn für $i = 1, \dots, n$ gilt

$$\exists m_i \in \mathbb{N} : \text{lt}(g_i) = x_i^{m_i},$$

wobei die lexikographische Termordnung mit $x_1 <_{LEX} x_2 <_{LEX} \dots <_{LEX} x_n$ zugrunde liegt. Eine Dreiecksbasis hat die shape-Lemma-Gestalt, wenn $g_1 \in K[x_1]$ und

$$g_i = x_i - h_i(x_1) \text{ mit } h_i \in K[x_1], \deg(h_i) < \deg(g_1), \quad i = 2, \dots, n.$$

In diesem Fall ist $\{g_1, \dots, g_n\}$ eine minimale Gröbnerbasis bzgl. \leq_{LEX} , wenn noch $\text{lc}(g_1) = 1$ gefordert wird. Dreiecksbasen sind lexikographische Gröbnerbasen (mit $x_1 <_{LEX} x_2 <_{LEX} \dots <_{LEX} x_n$).

Ist $G = \{g_1, x_2 - h_2, \dots, x_n - h_n\}$ eine Dreiecksbasis in shape-Lemma-Gestalt und $y = (\xi_1, \dots, \xi_n)$ aus der Varietät $V(\langle G \rangle)$, dann gilt für $i = 2, \dots, n$

$$\xi_i = h_i(\xi_1).$$

Hat g_1 nur einfache und paarweise verschiedene Nullstellen $\xi_1^{(1)}, \dots, \xi_1^{(d)}$, dann besteht die Varietät $V(\langle G \rangle)$ aus d verschiedenen Punkten

$$(\xi_1^{(j)}, \dots, \xi_n^{(j)}) \in \mathbb{C}, \quad j = 1, \dots, d,$$

und h_i , $i = 2, \dots, n$, erfüllt die Interpolationsbedingungen

$$h_i(\xi_1^{(j)}) = \xi_i^{(j)}, \quad j = 1, \dots, d.$$

Satz 6.9 (“shape lemma”)

Sei \mathfrak{a} ein nulldimensionales Radikal in $K[x_1, \dots, x_n]$. Für Punkte der Varietät $V(\mathfrak{a})$ gelte

$$(\xi_1, \dots, \xi_n), (\eta_1, \dots, \eta_n) \in V(\mathfrak{a}), \xi_1 = \eta_1 \Rightarrow \forall i \in \{1, \dots, n\} : \xi_i = \eta_i.$$

Dann ist die minimale Gröbnerbasis von \mathfrak{a} bzgl. \leq_{LEX} mit $x_1 <_{LEX} x_2 <_{LEX} \dots <_{LEX} x_n$ Dreiecksbasis in shape-Lemma-Gestalt.

Korollar

Ist L eine n -fache algebraische Körpererweiterung von K , dann gibt es ein $a \in L$, so dass L eine einfache Körpererweiterung, $L = K(a)$, ist.

§6.2 Probleme lösbar mit linearen Methoden

Im folgenden ist \mathfrak{a} immer ein Ideal aus $\mathcal{P} = K[x_1, \dots, x_n]$.

Problem 1 (Hauptproblem der Idealtheorie lt. van der Waerden)

Entscheide $f \in \mathfrak{a}$.

Lösung: $f \in \mathfrak{a} \Leftrightarrow \text{NF}(f, \mathfrak{a}) = 0$.

Problem 2 (Unterideale)

Sei $\mathfrak{B} = \langle b_1, \dots, b_r \rangle$.

Entscheide $\mathfrak{B} \in \mathfrak{a}$.

Lösung: $\mathfrak{B} \in \mathfrak{a} \Leftrightarrow \forall i \in \{1, \dots, r\} : \text{NF}(b_i, \mathfrak{a}) = 0$.

Problem 3 (univariate Polynome in \mathfrak{a})

Sei \mathfrak{a} nulldimensional und $i \in \{1, \dots, r\}$.

Finde Polynom kleinsten Grads in $\mathfrak{a} \cap K[x_i]$.

Lösung: Sei $s \in \mathbb{N}_0$ kleinste Zahl, so dass

$$\text{NF}(1, \mathfrak{a}), \text{NF}(x_i, \mathfrak{a}), \text{NF}(x_i^2, \mathfrak{a}), \dots, \text{NF}(x_i^s, \mathfrak{a})$$

linear abhängig sind, $\text{NF}(x_i^s, \mathfrak{a}) = \sum_{j=0}^{s-1} c_j \text{NF}(x_i^j, \mathfrak{a})$. Dann ist $x_i^s - \sum_{j=0}^{s-1} c_j x_i^j$ das gesuchte Polynom.

Problem 4 (FGLM-Algorithmus)

Gegeben: nulldimensionales Ideal \mathfrak{a} , Termordnungen \leq_1, \leq_2 , Gröbnerbasis

G_1 von \mathfrak{a} bzgl. \leq_1 .

Gesucht: Minimale Gröbnerbasis G_2 von \mathfrak{a} bzgl. \leq_2 und zugehörige Normalmenge \mathcal{N}_2 .

(Bemerkung: $\text{NF}_1(f, \mathfrak{a})$ ist die Normalform von f bzgl. \leq_1)

Rechnung: Initialisierung mit

$$T_0 := \{x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n} \mid i_1, \dots, i_n \in \mathbb{N}_0\}, \mathcal{N}_2 := \emptyset, G_2 := \emptyset.$$

SCHLEIFE $t := \min_{\leq_2} T_0$,

$$T_0 := T_0 \setminus \{t\}.$$

Wenn $\text{NF}_1(t, \mathfrak{a}) = \sum_{t_i \in \mathcal{N}_2} c_i \text{NF}(t_i, \mathfrak{a})$,

dann $G_2 := G_2 \cup \{t - \sum_{t_i \in \mathcal{N}_2} c_i t_i\}$ und
streich in T_0 alle Vielfachen von t .

Sonst $\mathcal{N}_2 := \mathcal{N}_2 \cup \{t\}$.

Wenn $T_0 = \emptyset$, dann fertig. Sonst weiter bei SCHLEIFE.

Problem 5 (Eliminationsideale)

Gegeben: Ideal \mathfrak{a} mit Gröbnerbasis G und

$$\exists m_1, \dots, m_i \in \mathbb{N}_0 : \text{lt}(x_1^{m_1}), \dots, \text{lt}(x_i^{m_i}) \in \text{lt}(\mathfrak{a}).$$

Finde lexikographische Gröbnerbasis für $\mathfrak{a} \cap K[x_1, \dots, x_i]$

Lösung: Nimm FGLM-Algorithmus, wobei \leq_1 die Termordnung der Gröbnerbasis G ist, \leq_2 die lexikographische mit $x_1 <_{\text{LEX}} x_2 <_{\text{LEX}} \dots <_{\text{LEX}} x_n$ und brich den FGLM-Algorithmus ab, sobald die Terme aus $[x_1, \dots, x_i]$ abgearbeitet sind, d.h., sobald $\min_{\leq_2} T_0 \notin [x_1, \dots, x_i]$.

Definition 6.7 (Quotientenideal, Saturation)

Ist \mathfrak{a} ein Ideal im Ring R und $f \in R$, dann wird die Menge

$$\mathfrak{a} : f := \{g \in R \mid g \cdot f \in \mathfrak{a}\}$$

Quotientenideal (von \mathfrak{a} durch f) genannt, kurz: \mathfrak{a} durch f . Die Menge

$$\mathfrak{a} : f^* := \{g \in R \mid \exists s \in \mathbb{N}_0 : g f^s \in \mathfrak{a}\}$$

wird als Saturation von $\mathfrak{a} : f$ bezeichnet.

Bemerkung 1. $\mathfrak{a} : f$ und $\mathfrak{a} : f^*$ sind Ideale. Es gilt $(\mathfrak{a} : f^k) : f = \mathfrak{a} : f^{k+1}$ für alle $k \in \mathbb{N}$ und

$$\mathfrak{a} \subseteq \mathfrak{a} : f \subseteq \mathfrak{a} : f^2 \subseteq \mathfrak{a} : f^3 \subseteq \dots \subseteq \mathfrak{a} : f^*.$$

Bemerkung 2. Ist $R = K[x_1, \dots, x_n]$ und \mathfrak{a} ein Radikal, dann gilt

$$\mathfrak{a} : f = \mathfrak{a} : f^2 = \mathfrak{a} : f^3 = \dots = \mathfrak{a} : f^* = I(\{y \in V(\mathfrak{a}) \mid f(y) \neq 0\}).$$

Insbesondere ist $\mathfrak{a} : f$ dann auch ein Radikal. Ist \mathfrak{a} nulldimensionales Polynomideal, dann gilt $V(\mathfrak{a} : f^*) = \{y \in V(\mathfrak{a}) \mid f(y) \neq 0\}$.

Problem 6 (Basis des Quotientenideals)

Gegeben: nulldimensionales Ideal \mathfrak{a} , Termordnungen \leq_1, \leq_2 , Gröbnerbasis G_1 von \mathfrak{a} bzgl. \leq_1 und ein $f \in \mathcal{P}$.

Gesucht: Minimale Gröbnerbasis G_2 von $\mathfrak{a} : f$ bzgl. \leq_2 und zugehörige Normalmenge \mathcal{N}_2 .

Lösung: Nimm FGLM-Algorithmus, aber statt $\text{NF}_1(t, \mathfrak{a})$ bzw. $\text{NF}(t_i, \mathfrak{a})$ nimm jeweils $\text{NF}_1(t \cdot f, \mathfrak{a})$ bzw. $\text{NF}(t_i \cdot f, \mathfrak{a})$. Die Relation

$$t \cdot f - \sum_{t_i \in \mathcal{N}_2} c_i t_i \cdot f \in \mathfrak{a}$$

gilt genau dann, wenn $t - \sum_{t_i \in \mathcal{N}_2} c_i t_i \in \mathfrak{a} : f$.

Satz 6.10 (Saturation als Eliminationsideal)

Sei $\mathfrak{a} = \langle f_1, \dots, f_r \rangle \subseteq K[x_1, \dots, x_n](= \mathcal{P})$ und $f \in \mathcal{P}$. Für das Ideal

$$J := \langle f_1, \dots, f_r, 1 - t \cdot f \rangle \subseteq K[x_1, \dots, x_n, t]$$

gilt dann $J \cap K[x_1, \dots, x_n] = \mathfrak{a} : f^*$.

Bemerkung. Mit Satz 6.10 kann man aufbauend auf der Lösung von Problem 5 eine Gröbnerbasis des Saturationsideals bekommen. Man muss zunächst (mit Buchbergers Algorithmus) eine Gröbnerbasis des Ideals J aus Satz 6.10 berechnen. Wenn \mathfrak{a} nulldimensional ist, sind die im Problem 5 genannten Voraussetzungen erfüllt. Dann bekommt man mit dem dort genannten Verfahren eine minimale Gröbnerbasis von $J \cap K[x_1, \dots, x_n] = \mathfrak{a} : f^*$.

Problem 7 (Basis der Saturation)

Gegeben: \mathfrak{a}, f und \leq_1, \leq_2 wie bei Problem 6.

Gesucht: Minimale Gröbnerbasis G_2 von $\mathfrak{a} : f^*$ bzgl. \leq_2 und zugehörige Normalmenge \mathcal{N}_2 .

Lösung: Berechne der Reihe nach mit dem Verfahren von Problem 6 minimale Gröbnerbasen von $\mathfrak{a} : f, \mathfrak{a} : f^2, \mathfrak{a} : f^3$, usw. bis für ein $k \in \mathbb{N}$ gilt $\mathfrak{a} : f^k = \mathfrak{a} : f^{k-1}$. Dann ist $\mathfrak{a} : f^k = \mathfrak{a} : f^*$.

Bemerkung. Ein Ideal in $\mathcal{P} = K[x_1, \dots, x_n]$ betrachtet man dann als gegeben, wenn man eine Basis davon kennt. Hat man zwei Ideale $\mathfrak{a} := \langle f_1, \dots, f_r \rangle$, $\mathfrak{b} := \langle g_1, \dots, g_s \rangle$, dann kann man neue Ideale bilden,

$$\begin{aligned} \mathfrak{a} + \mathfrak{b} &:= \{a + b \mid a \in \mathfrak{a}, b \in \mathfrak{b}\}, \\ \mathfrak{a} \cdot \mathfrak{b} &:= \{\sum_{k=1}^m a_k b_k \mid m \in \mathbb{N}, a_k \in \mathfrak{a}, b_k \in \mathfrak{b}\}, \\ \mathfrak{a} \cap \mathfrak{b} &:= \{c \mid c \in \mathfrak{a}, c \in \mathfrak{b}\}, \\ \mathfrak{a} : \mathfrak{b} &:= \{c \mid b \in \mathfrak{b} \Rightarrow c \cdot b \in \mathfrak{a}\} \end{aligned}$$

Eine Basis von $\mathfrak{a} + \mathfrak{b}$ ist $\{f_1, \dots, f_r, g_1, \dots, g_s\}$, eine von $\mathfrak{a} \cdot \mathfrak{b}$ ist $\{f_i \cdot g_j \mid 1 \leq i \leq r, 1 \leq j \leq s\}$. Wegen $\mathfrak{a} : \mathfrak{b} = \bigcap_{j=1}^s \mathfrak{a} : g_j$ kann man mit Problem

6, zumindest für nulldimensionales \mathfrak{a} , die Berechnung einer Basis von $\mathfrak{a} : \mathfrak{b}$ zurückführen auf die Berechnung von Basen für Idealdurchschnitte.

Problem 8 (Basis vom Idealdurchschnitt)

Gegeben: nulldimensionale Ideale $\mathfrak{a} := \langle f_1, \dots, f_r \rangle$, $\mathfrak{b} := \langle g_1, \dots, g_s \rangle$.

Gesucht: Eine Gröbnerbasis von $\mathfrak{a} \cap \mathfrak{b}$ bzgl. \leq_{LEX} .

Lösung: Berechne eine Gröbnerbasis G vom Ideal $J := \langle (1-t)f_1, \dots, (1-t)f_r, tg_1, \dots, tg_s \rangle \subset K[x_1, \dots, x_n, t]$. Mit dem Verfahren von Problem 5 berechne man dann eine Gröbnerbasis von $J \cap K[x_1, \dots, x_n] = \mathfrak{a} \cap \mathfrak{b}$ bzgl. \leq_{LEX} .

Die Verfahren zur Lösung der Probleme 3 - 8 basieren im Grund auf der Idee des FGLM-Algorithmus. Der historisch erste Algorithmus dieser Art ("Die Mutter aller Algorithmen"), also die Idee, mit einem Orakel zu entscheiden, ob ein $t - \sum_{t_i \in \mathcal{N}} c_i t_i$ im gegebenen Ideal \mathfrak{a} liegt und dann mit der gewünschten Termordnung der Reihe nach alle Terme t durchzumustern, ob sie zu \mathcal{N} gehören oder ob ein Polynom aus \mathfrak{a} mit dem Leitern t existiert, stammt schon aus dem Jahr 1982 und berechnete zu vorgegebenen Punkten das zugehörige Verschwindungsideal.

Problem 9 (Der BM-Algorithmus)

Gegeben: Punktmenge $M := \{y_1, \dots, y_N\} \subset K^n$ und Termordnung \leq_T .

Gesucht: Eine Gröbnerbasis G bzgl. \leq_T für das Verschwindungsideal $I(M)$ und zugehörige Normalmenge \mathcal{N}_2 .

Lösung: Initialisierung mit

$$T_0 := \{x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n} \mid i_1, \dots, i_n \in \mathbb{N}_0\}, \quad \mathcal{N}_2 := \emptyset, \quad G := \emptyset.$$

SCHLEIFE $t := \min_{\leq_T} T_0$,

$$T_0 := T_0 \setminus \{t\}.$$

Berechne $v(t) := (t(y_1), \dots, t(y_N))$

Wenn $v(t) = \sum_{t_i \in \mathcal{N}_2} c_i \cdot v(t_i)$,

dann $G := G \cup \{t - \sum_{t_i \in \mathcal{N}_2} c_i t_i\}$ und

streich in T_0 alle Vielfache von t .

Sonst $\mathcal{N}_2 := \mathcal{N}_2 \cup \{t\}$.

Wenn $T_0 = \emptyset$, dann fertig. Sonst weiter bei SCHLEIFE.

Beim FGLM- und BM-Algorithmus tritt jeweils das Problem auf, das Minimum der abzählbaren Menge T_0 bzgl. einer Termordnung \leq_T zu finden. Das Komplement des aktuellen T_0 in der Menge aller Potenzprodukte, $T = [x_1, \dots, x_n]$, ist aber das aktuelle \mathcal{N}_2 (abgesehen von Termen, die Vielfache von Leitern der schon gefundenen Gröbnerbasiselemente sind). Damit ist das gesuchte Minimum vom Typ $x_i \cdot t$ mit $t \in \mathcal{N}_2$ und $x_i \in \{x_1, x_2, \dots, x_n\}$, so dass man nur das Minimum in der endlichen Menge $\{x_i \cdot t \mid 1 \leq i \leq n, t \in \mathcal{N}_2\}$, suchen muss.

Den Test, ob beim BM-Algorithmus $v(t)$ Linearkombination der $v(t_i)$ mit $t_i \in \mathcal{N}_2$ ist, - entsprechend beim FGLM-Algorithmus, ob $\mathbf{NF}(t, \mathfrak{a})$ Linearkombination der $\mathbf{NF}(t_i, \mathfrak{a})$ ist, - kann man effizient mit Gaußelimination lösen. Als Nebenprodukt bekommt man Polynome, die in allen Punkten der Varietät bis auf einen Null sind. Das gibt dann die Lagrange-Grundpolynome für die Polynominterpolation. (Beim FGLM-Algorithmus sind die entsprechenden Polynome diejenigen, deren Normalform aus genau einem Monom besteht, denn man speichert statt $\mathbf{NF}(t_i, \mathfrak{a})$ den Koeffizientenvektor zur Basis \mathcal{N}_2 .)

Zum Abschluß greifen wir das Rechnen in algebraischen Körpererweiterungen wieder auf. Nach Korollar von Satz 6.9 ist jede (mehrfache) algebraische Körpererweiterung auch einfache algebraische Körpererweiterung. In diesen Körpererweiterungen kann man die Addition leicht durchführen. Die Multiplikation und die Division lassen sich mit Multiplikationsmatrizen realisieren, wie am Ende von §4 gezeigt wurde. Der Umweg über einfache Körpererweiterungen ist aber nicht nötig:

Sei \mathfrak{a} ein maximales Ideal in $K[x_1, \dots, x_n]$, K ein Körper mit $\mathbb{Q} \subseteq K$. Nach Satz 6.5 ist $L := K[x_1, \dots, x_n]/\mathfrak{a}$ ein Körper. Zu einer vorgegebenen Termordnung \leq_T sei $\mathcal{N} = \{t_1, \dots, t_N\}$ die Normalmenge von \mathfrak{a} . Dann ist $\{[t_1], \dots, [t_N]\}$ eine Basis des K -linearen Raums L , vgl. Beweis von Satz 6.3. Betrachtet man zu vorgegebenem $f \in K[x_1, \dots, x_n]$ die Abbildung

$$\Phi_f : L \rightarrow L, \quad \Phi_f[g] := [f] \cdot [g] \quad (= [f \cdot g]),$$

dann gehört zu Φ_f und der Basis $\{[t_1], \dots, [t_N]\}$ die Abbildungsmatrix $B_f := (b_{ij})_{i,j=1}^N$ mit

$$\Phi_f([t_j]) = \sum_{i=1}^N b_{ij}[t_i], \quad j = 1, \dots, N.$$

Die Transponierte dazu, die Matrix $M_f := B_f^T$ mit $M_f = (m_{ij})_{i,j=1}^N$ erfüllt ($m_{ij} = b_{ji}$ und)

$$\Phi_f([t_j]) = \sum_{k=1}^N m_{jk}[t_k], \quad j = 1, \dots, N.$$

In der Literatur wird M_f als Multiplikationsmatrix (zur Multiplikation mit f) bezeichnet, im Widerspruch zu Definition 4.2.

Für die Matrizen $B_f, f \in K[x_1, \dots, x_n]$, gilt

$$\begin{aligned} f \in [g] &\quad \Rightarrow B_f = B_g, \\ B_f + B_g &\quad = B_{f+g}, \\ B_f B_g = B_{fg} &= B_{gf} = B_g B_f. \end{aligned}$$

Diese Matrizen bilden also eine Familie kommutierender Matrizen. Sie wird von $B_{x_1}, B_{x_2}, \dots, B_{x_n}$ erzeugt, denn es gilt

$$\begin{aligned} t = x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n} &\Rightarrow B_t = B_{x_1}^{i_1} B_{x_2}^{i_2} \cdots B_{x_n}^{i_n}, \\ f = \sum_{i=1}^N c_i t_i &\Rightarrow B_f = \sum_{i=1}^N c_i B_{t_i}. \end{aligned}$$

Die Elemente des Körpers L besitzen eine eindeutige Darstellung

$$\sum_{j=1}^N c_j [t_j], \quad c_1, \dots, c_N \in K,$$

denn L hat als K -linearer Raum $[t_1], \dots, [t_N]$ als Basis. Wegen der Eindeutigkeit speichert man statt $\sum_{j=1}^N c_j [t_j]$ nur den Koeffizientenvektor $(c_1, \dots, c_N)^T \in K^N$. Die Addition in L ist dann die Addition der Koeffizientenvektoren.

Für $[g] = \sum_{j=1}^N d_j [t_j] \in L$ ist

$$[f][g] = \Phi_f([g]) = \sum_{j=1}^N d_j \Phi_f([t_j])$$

Also ist der Koeffizientenvektor von $[fg] = [f][g]$

$$B_f \begin{pmatrix} d_1 \\ \vdots \\ d_N \end{pmatrix},$$

was aus den Eigenschaften der Abbildungsmatrix B_f direkt folgt (lin. Algebra 1 !!!).

Ist $f = \sum_{i=1}^N c_i t_i \neq 0$, dann ist B_f invertierbar, denn

$$0 = B_f \begin{pmatrix} d_1 \\ \vdots \\ d_N \end{pmatrix}$$

mit $(d_1, \dots, d_N) \neq 0$ bedeutet, dass $[f][\sum_{j=1}^N d_j t_j] = 0$ gilt, $[f]$ also Nullteiler ist. Daher ist für $f \neq 0$ das Gleichungssystem

$$B_f \begin{pmatrix} u_1 \\ \vdots \\ u_N \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_N \end{pmatrix}$$

stets eindeutig lösbar und bedeutet, dass es zu jedem $[f] \neq 0$ und jedem $[g] = \sum_{k=1}^N b_k [t_k] \in L$ ein Element $[u] := \sum_{i=1}^N u_i [t_i] \in L$ gibt mit $[f] \cdot [u] = [g]$. Auf diese Weise wird zu $[f], [g] \in L$ mit $[f] \neq 0$ der Koeffizientenvektor von $[u] = [f]^{-1}[g]$ konstruiert.