Wolfgang K. Seiler

Algebraische Statistik

Vorlesung im Herbstsemester 2018
an der Universität Mannheim

Dieses Skriptum entsteht parallel zur Vorlesung und soll mit möglichst geringer Verzögerung erscheinen. Es ist daher in seiner Qualität auf keinen Fall mit einem Lehrbuch zu vergleichen; insbesondere sind Fehler bei dieser Entstehensweise nicht nur möglich, sondern **sicher**. Dabei handelt es sich wohl leider nicht immer nur um harmlose Tippfehler, sondern auch um Fehler bei den mathematischen Aussagen. Da mehrere Teile aus anderen Skripten für Hörerkreise der verschiedensten Niveaus übernommen sind, ist die Präsentation auch teilweise ziemlich inhomogen.

Das Skriptum sollte daher mit Sorgfalt und einem gewissen Mißtrauen gegen seinen Inhalt gelesen werden. Falls Sie Fehler finden, teilen Sie mir dies bitte persönlich oder per e-mail (seiler@math.uni-mannheim.de) mit. Auch wenn Sie Teile des Skriptums unverständlich finden, bin ich für entsprechende Hinweise dankbar.

Falls genügend viele Hinweise eingehen, werde ich von Zeit zu Zeit Listen mit Berichtigungen und Verbesserungen zusammenstellen. In der online Version werden natürlich alle bekannten Fehler korrigiert.

Biographische Angaben von Mathematikern beruhen größtenteils auf den entsprechenden Artikeln im *MacTutor History of Mathematics archive* (`www-history.mcs.st-andrews.ac.uk/history/`), von wo auch die meisten abgedruckten Bilder stammen. Bei noch lebenden Mathematikern bezog ich mich, soweit möglich, auf deren eigenen Internetauftritt.

Kapitel 0
Einführung

Kapitel 1
Gröbner-Basen

Die klassische Aufgabe der Algebra besteht in der Lösung von Gleichungen und Gleichungssystemen. Im Falle eines Systems von Polynomgleichungen in mehreren Veränderlichen kann die Lösungsmenge sehr kompliziert sein und, sofern sie unendlich ist, möglicherweise nicht einmal explizit angebbar: Im Gegensatz zum Fall linearer Gleichungen können wir hier im allgemeinen keine endliche Menge von Lösungen finden, durch die sich alle anderen Lösungen ausdrücken lassen. Trotzdem gibt es Algorithmen, mit denen sich nichtlineare Gleichungssysteme deutlich vereinfachen lassen, und zumindest bei endlichen Lösungsmengen lassen sich diese auch konkret angeben – sofern wir die Nullstellen von Polynomen einer Veränderlichen explizit angeben können.

§1: Algebraische Vorbereitungen

Wenn wir lineare Gleichungssysteme mit dem GAUSS-Algorithmus lösen, verändern wir das Gleichungssystem sukzessive, indem wir Gleichungen so durch Linearkombinationen mit anderen Gleichungen ersetzen, daß sich an der Lösungsmenge nichts ändert. Indem wir eine lineare Gleichung

$$a_1 X_1 + \cdots + a_n X_n = b$$

über einem Körper k mit dem $(n+1)$-Tupel $(a_1, \ldots, a_n, b) \in k^{n+1}$ identifizieren, sehen wir leicht, daß die sämtlichen linearen Gleichungen in n Unbekannten über einem Körper k einen $(n+1)$-dimensionalen Vektorraum bilden; die Gleichungen eines konkreten linearen Gleichungssystems erzeugen darin einen Untervektorraum. Dieser besteht aus allen Linearkombinationen der gegebenen Gleichungen, und das sind gleichzeitig alle linearen Gleichungen, die auf der Lösungsmenge des linea-

ren Gleichungssystems verschwinden. Zwei lineare Gleichungssysteme haben somit genau dann die gleiche Lösungsmenge, wenn sie den gleichen Untervektorraum erzeugen.

Wenn wir Systeme nichtlinearer Gleichungen betrachten, ist es sinnvoll, die Menge aller möglicher Gleichungen nicht mehr nur als Vektorraum zu betrachten, sondern auch die Multiplikation mit Polynomen zuzulassen: Zur Lösung des Gleichungssystems

$$X^2Y^2 + 2X^3 - 3X^2 - X = 0 \quad \text{und} \quad Y^2 + X - 3 = 0$$

bietet sich etwa an, die zweite Gleichung mit X^2 zu multiplizieren und das Produkt $X^2Y^2 + X^3 - 3X^2 = 0$ von der ersten Gleichung zu subtrahieren; die Differenz $X^3 - X$ hängt nur noch von X ab und verschwindet bei 0 und ± 1. Setzen wir dies in die zweite Gleichung ein, erhalten wir die Lösungsmenge

$$\left\{ \left(0, \sqrt{3}\right), \left(0, -\sqrt{3}\right), \left(1, \sqrt{2}\right), \left(1, -\sqrt{2}\right), (-1, 2), (-1, -2) \right\}.$$

Wir sollten die Menge aller möglicher Gleichungen daher nicht mehr nur als einen Vektorraum betrachten, sondern als einen *Ring* im Sinne der folgenden Definition:

Definition: *a)* Ein Ring ist eine Menge R zusammen mit zwei Rechenoperationen „+“ und „$\cdot$“ von $R \times R$ nach R, so daß gilt:

1.) R bildet bezüglich „+“ eine abelsche Gruppe, d.h. für die Addition gilt das Kommutativgesetz $f + g = g + f$ sowie das Assoziativgesetz $(f + g) + h = f + (g + h)$ für alle $f, g, h \in \mathbb{R}$, es gibt ein Element $0 \in R$, so daß $0 + f = f + 0 = f$ für alle $f \in R$, und zu jedem $f \in R$ gibt es ein Element $-f \in R$, so daß $f + (-f) = 0$ ist.

2.) Die Verknüpfung „$\cdot$“: $R \times R \to R$ erfüllt das Assoziativgesetz $f(gh) = (fg)h$, und es gibt ein Element $1 \in R$, so daß $1f = f1 = f$.

3.) „+“ und „$\cdot$“ erfüllen die Distributivgesetze $f(g + h) = fg + fh$ und $(f + g)h = fh + gh$.

b) Ein Ring heißt *kommutativ,* falls zusätzlich noch das Kommutativgesetz $fg = gf$ der Multiplikation gilt.

c) Ein Ring heißt *nullteilerfrei* wenn gilt: Falls ein Produkt $fg = 0$ verschwindet, muß mindestens einer der beiden Faktoren f, g gleich Null sein. Ein nullteilerfreier kommutativer Ring heißt *Integritätsbereich.*

Natürlich ist jeder Körper ein Ring; für einen Körper werden schließlich genau dieselben Eigenschaften gefordert und zusätzlich auch noch die Kommutativität der Multiplikation sowie die Existenz multiplikativer Inverser. Ein Körper ist somit insbesondere auch ein Integritätsbereich.

Das bekannteste Beispiel eines Rings, der kein Körper ist, sind die ganzen Zahlen; auch sie bilden einen Integritätsbereich.

Für die Betrachtung nichtlinearer Gleichungssysteme interessieren uns allerdings vor allem Polynomringe. Da auch diese kommutativ sind, vereinbaren wir:

Wenn nicht explizit etwas anderes gesagt wird, soll *Ring* im folgenden stets für einen *kommutativen Ring* stehe.

Definition: R sei ein Ring, und $X_1, \ldots, X_n$ seien n Symbole, die nicht in R liegen.
a) Ein *Monom* ist ein Produkt $X_1^{\alpha_1} \cdots X_n^{\alpha_n}$ mit nichtnegativen ganzen Zahlen $\alpha_1, \ldots, \alpha_n$. Die Summe der α_i bezeichnen wir als den *Grad* des Monoms.
b) Ein *Polynom* über R in den Variablen $X_1, \ldots, X_n$ ist eine endliche Linearkombination f von Monomen mit Koeffizienten aus R. Falls diese nicht Null ist, bezeichnen wir den größten Grad eines in f vorkommenden Monoms als den *Grad* $\deg f$ von f. Für das Polynom $f = 0$ definieren wir keinen Grad. *c)* Die Menge aller Polynome über R in den Variablen $X_1, \ldots, X_n$ bezeichnen wir als den *Polynomring* $R[X_1, \ldots, X_n]$ über R in den Variablen $X_1, \ldots, X_n$.

Es ist klar, daß $R[X_1, \ldots, X_n]$ mit der offensichtlichen Addition und Multiplikation ein Ring ist. Wir nehmen dabei natürlich an, daß die X_i untereinander kommutieren.

Wir interessieren uns vor allem für Polynomringe über Körpern; für Induktionsbeweise ist es aber oft nützlich, beispielsweise den Polynomring $k[X, Y]$ aufzufassen als den Polynomring in Y über dem Ring $R = k[X]$; daher die allgemeinere Definition.

Wie wir beim obigen Beispiel eines nichtlinearen Gleichungssystems gesehen haben, kann es bei der Lösung nützlich sein, nicht nur skalare Linearkombinationen der Gleichungen zu betrachten, sondern auch solche mit beliebigen Polynomen als Koeffizienten. Anstelle von Untervektorräumen des Polynomrings $k[X_1, \ldots, X_n]$ sollten wir daher Strukturen betrachten, in denen man Linearkombinationen mit beliebigen Ringelementen als Koeffizienten bilden kann, die sogenannten Ideale:

Definition: Eine nichtleere Teilmenge I eines Rings R heißt *Ideal,* in Zeichen $I \lhd R$, wenn gilt:
1.) Für je zwei Elemente $f, g \in I$ ist auch $f + g \in I$
2.) Für jedes $f \in I$ und jedes $r \in R$ liegt auch rf in I.

Bei den Produkten verlangen wir also, daß sie bereits dann in I liegen, wenn nur *ein* Faktor in I liegt.

Die Bedingung, daß ein Ideal mindestens ein Element enthalten muß, können wir auch ersetzen durch die Bedingung, daß es die Null von R enthalten muß, denn wenn es irgendein Element $f \in R$ enthält, muß es gemäß der zweiten Bedingung auch $0 \cdot f = 0$ enthalten.

Um mit dem Idealbegriff vertraut zu werden, betrachten wir zunächst Ideale im Ring der ganzen Zahlen:

Lemma: Zu jedem Ideal $I \lhd \mathbb{Z}$ gibt es eine ganze Zahl $n \in \mathbb{Z}$, so daß $I = \{nq | q \in \mathbb{Z}\}$.

Beweis: I ist nach Definition nicht leer, enthält also mindestens ein Element. Falls I nur aus der Null besteht, können wir $n = 0$ setzen und sind fertig. Wenn es ein Element $m \neq 0$ gibt, enthält das Ideal auch dessen sämtliche ganzzahlige Vielfachen, insbesondere also gibt es in I dann positive Zahlen. Die kleinste dieser Zahlen sei n. Wir wollen uns überlegen, daß I genau aus den ganzzahligen Vielfachen von n besteht.

Dazu sei $m \in I$ ein beliebiges Element von I. Wir dividieren m mit Rest durch q; das Ergebnis sei

$$m : n = q \quad \text{Rest } r \quad \text{mit} \quad 0 \leq r < n \,.$$

Dann liegt mit m und n auch $r = m - qn$ in I und ist echt kleiner als n. Da n die kleinste positive Zahl in I ist, muß daher $r = 0$ sein, d.h. $m = qn$ ist ein ganzzahliges Vielfaches von n.

■

Definition: *a)* Ist R ein Ring und $f \in R$ so bezeichnen wir

$$(f) \underset{\text{def}}{=} \{rf \mid r \in R\}$$

als das von f erzeugte *Hauptideal.*
b) R heißt *Hauptidealring,* wenn jedes Ideal von R ein Hauptideal ist.

Das gerade bewiesene Lemma zeigt also, daß $\mathbb{Z}$ ein Hauptidealring ist.

Allgemeiner definieren wir

Definition: Ist R ein Ring und ist $M \subset R$ eine Teilmenge von R, so ist das *von M erzeugte Ideal* (M) das kleinste Ideal von R, das M enthält, d.h. den Durchschnitt aller Ideale, die M enthalten. Für eine endliche Menge $M = \{f_1, \ldots, f_m\}$ schreiben wir (M) kurz als $(f_1, \ldots, f_m)$. Die Menge M bezeichnen wir als ein *Erzeugendensystem* des Ideals I.

Diese Definition macht nicht wirklich klar, wie das von M erzeugte Ideal aussieht. Da uns in der Computeralgebra nur endlich erzeugte Ideale interessieren, möchte ich mich auf diesen Fall beschränken; die Verallgemeinerung auf beliebige Mengen M sollte für jeden, der den folgenden Beweis verstanden hat, offensichtlich sein.

Lemma: $\left(f_1, \ldots, f_m\right) = \left\{ \sum_{i=1}^{m} r_i f_i \;\middle|\; r_i \in R \right\}$

Beweis: Da jedes Ideal, das $f_1, \ldots, f_m$ enthält, auch für $r_1, \ldots, r_m \in R$ die Elemente $r_i f_i$ enthält und damit auch deren Summe, ist klar, daß die rechte Seite in jedem Ideal enthalten ist, das die f_i enthält. Außerdem ist die rechtsstehende Menge selbst ein Ideal: Da sie die f_i enthält, ist sie nicht leer; die Summe zweier Elemente ist offensichtlich wieder ein Element, da wir einfach die Koeffizienten addieren müssen, und wenn wir ein Element mit einem beliebigen Element $r \in R$ multiplizieren,

werden einfach alle Koeffizienten mit r multipliziert. Somit ist die rechte Seite in der Tat das kleinste Ideal, das alle f_i enthält. ∎

Sei nun $R = k[X_1, \ldots, X_n]$ der Polynomring in n Variablen über einem Körper k,und seien $f_1, \ldots, f_m \in R$ Polynome. Wir interessieren uns für die Lösungsmenge des durch die f_i gegebenen Gleichungssystems, also die Menge aller $(x_1, \ldots, x_n) \in k^n$, für die alle f_i verschwinden. Wir definieren gleich allgemein

Definition: Die Nullstellenmenge einer Teilmenge $M \subseteq k[X_1, \ldots, X_n]$ ist

$$V(M) \underset{\text{def}}{=} \left\{(x_1, \ldots, x_n) \in k^n \;\middle|\; f(x_1, \ldots, x_n) = 0 \text{ für alle } f \in M\right\}.$$

Im Falle eine endlichen Menge $M = \{f_1, \ldots, f_m\}$ schreiben wir kurz $V(f_1, \ldots, f_m)$.

(In der algebraischen Geometrie bezeichnet man Mengen dieser Art als Varietäten; daher der Buchstabe V.)

Lemma: Ist $I = (f_1, \ldots, f_m)$ das von den f_i erzeugte Ideal, so ist

$$V(I) = V(f_1, \ldots, f_m)\,.$$

Beweis: Da alle f_i in I liegen, ist natürlich $V(I) \subseteq V(f_1, \ldots, f_m)$. Umgekehrt sei $(x_1, \ldots, x_n)$ ein Element von $V(f_1, \ldots, f_m)$ und g irgendein Element von I. Nach dem vorigen Lemma gibt es Polynome $r_i \in R$; so daß $g = \sum_{i=1}^m r_i f_i$ ist. Damit ist auch

$$g(x_1, \ldots, x_n) = \sum_{i=1}^{m} r_i(x_1, \ldots, x_m) f_i(x_1, \ldots, x_m) = 0\,,$$

so daß $(x_1, \ldots, x_n)$ in $V(I)$ liegt. Damit ist das Lemma bewiesen. ∎

Dieses Lemma zeigt, daß zwei Gleichungssysteme

$$f_1(x_1, \ldots, x_n) = 0, \quad \ldots, \quad f_m(x_1, \ldots, x_n) = 0$$

und

$$g_1(x_1, \ldots, x_n) = 0, \quad \ldots, \quad g_r(x_1, \ldots, x_n) = 0$$

die gleiche Lösungsmenge haben, wenn die Ideale $(f_1, \ldots, f_m)$ und $(g_1, \ldots, g_r)$ übereinstimmen.

Die Umkehrung dieser Aussage ist allerdings falsch. Ein einfaches Gegenbeispiel haben wir bereits bei nur einer Gleichung in einer Variablen: Die Gleichungen

$$x = 0, \qquad x^2 = 0, \qquad x^3 = 0, \qquad \ldots$$

haben allesamt nur die Null als Lösung, aber natürlich sind die Ideale $\left(x^d\right) \lhd k[X]$ für verschiedene Werte von d verschieden. Später werden wir diese Frage, wann so etwas vorkommt, genauer untersuchen.

Zum Abschluß dieses Paragraphen soll nur noch kurz festgehalten werden, wie sich Ideale und Nullstellenmengen zueinander verhalten. Dazu müssen wir zunächst die Summe und das Produkt zweier Ideale definieren:

Definition: *a)* Die Summe $I + J$ zweier Ideale I, J eines Rings R ist das kleinste Ideal, das sowohl I als auch J enthält.
b) Das Produkt IJ dieser Ideale ist das kleinste Ideal, das alle Produkte fg mit $f \in I$ und $g \in J$ enthält.

Man überlegt sich leicht (mit dem gleichen Argument, mit dem wir das Ideal $(f_1, \ldots, f_m)$ oben explizit bestimmt haben), daß $I + J$ gerade die Menge aller $f + g$ mit $f \in I$ und $g \in J$ ist; IJ dagegen enthält im allgemeinen auch Elemente, die sich *nicht* in der Form fg mit $f \in I$ und $g \in J$ darstellen lassen: Ist etwa $I = J = (X, Y) \lhd \mathbb{R}[X, Y]$, so enthält IJ mit $X^2 = X \cdot X$ und $Y^2 = Y \cdot Y$ auch deren Summe $X^2 + Y^2$, die sich nicht als Produkt zweier Polynome aus $\mathbb{R}[X, Y]$ schreiben läßt. Wenn wir $\mathbb{R}$ durch $\mathbb{C}$ ersetzen, läßt sich $X^2 + Y^2$ zwar zerlegen als $(X + iY)(X - iY)$, aber auch in $\mathbb{C}[X, Y]$ gibt es irreduzible Polynome in $(X, Y) \cdot (X, Y)$, die sich somit nicht als Produkt darstellen lassen. In IJ liegen daher auch alle (endlichen) Summen der Form $\sum f_i g_i$ mit $f_i \in I$ und $g_i \in J$; da diese (analog zum obigen Argument) ein Ideal bilden, besteht IJ genau aus diesen Summen.

Satz: Für zwei Ideale I, J im Polynomring $R = k[X_1, \ldots, X_n]$ gilt
a) Ist $I \subseteq J$, so ist $V(J) \subseteq V(I)$

b) $V(I + J) = V(I) \cap V(J)$
c) $V(IJ) = V(I) \cup V(J)$

Beweis: a) Sei $(x_1, \ldots, x_n) \in V(J)$. Dann verschwindet $f(x_1, \ldots, x_n)$ für alle $f \in J$, erst recht also für alle $f \in I$, d.h. $(x_1, \ldots, x_n) \in V(I)$.

b) Da $I + J$ das kleinste Ideal ist, das sowohl I als auch J enthält, liegt $V(I + J)$ nach *a)* sowohl in $V(I)$ als auch in $V(J)$, also auch in deren Durchschnitt. Liegt umgekehrt ein Punkt $(x_1, \ldots, x_n)$ sowohl in $V(I)$ als auch in $V(J)$, so liegt er auch in $V(I + J)$, denn wie wir gerade gesehen haben, läßt sich jedes Element von $I + J$ schreiben als $f + g$ mit $f \in I$ und $g \in J$, und sowohl f als auch g verschwinden im Punkt $(x_1, \ldots, x_n)$.

c) Da IJ erzeugt wird von den Produkten fg mit $f \in I$ und $g \in J$ und jedes dieser Produkte sowohl in I als auch in J liegt, ist IJ eine Teilmenge sowohl von I als auch von J; somit liegt $V(I) \cup V(J)$ nach *a)* in $V(IJ)$. Umgekehrt sei $(x_1, \ldots, x_n) \in V(IJ)$, liege aber nicht in $V(I)$. Dann gibt es ein $f \in I$ mit $f(x_1, \ldots, x_n) \neq 0$. Für jedes $g \in J$ liegt aber fg in IJ, so daß das Produkt $f(x_1, \ldots, x_n)g(x_1, \ldots, x_n)$ verschwinden muß. Da die Funktionswerte im Körper k liegen und der Faktor $f(x_1, \ldots, x_n)$ nicht verschwindet, muß $g(x_1, \ldots, x_n) = 0$ sein für alle $g \in J$; der Punkt liegt also in $V(J)$. Somit liegt er in jedem Fall in $V(I) \cup V(J)$.

■

§2: Gauß und Euklid

Zur (exakten) Lösung eines linearen Gleichungssystems in mehreren Veränderlichen verwenden wir üblicherweise den GAUSS-Algorithmus. Für die Lösung eines System von Polynomgleichungen höheren Grades in nur einer Veränderlichen können wir den EUKLIDischen Algorithmus verwenden, denn die gemeinsamen Nullstellen zweier Polynome in einer Veränderlichen sind gerade die Nullstellen ihres größten gemeinsamen Teilers, so daß wir das System durch mehrfache Anwendung des EUKLIDischen Algorithmus reduzieren können auf eine einzige Polynomgleichung.

Der um 1966 von BRUNO BUCHBERGER vorgestellte Ansatz zur Lösung nichtlinearer Gleichungssysteme in mehreren Veränderlichen kann als

eine Kombination von Ideen hinter dem GAUSSschen Eliminationsverfahren und dem EUKLIDischen Algorithmus aufgefaßt werden; er hat Anwendungen, die weit über das Problem der Lösung nichtlinearer Gleichungssysteme hinausgehen. In der Tat wurde die Grundidee des Verfahrens bereits knapp vor BUCHBERGER, und ohne daß dieser davon wußte, von dem japanischen Mathematiker HEISUKE HIRONAKA entdeckt, der es für ein klassisches Problem der algebraischen Geometrie entwickelte: Für die damit bewiesene sogenannte Auflösung der Singularitäten einer algebraischen Varietät über einem Körper der Charakteristik Null erhielt HIRONAKA 1970 die Fields-Medaille, die höchste Auszeichnung der Mathematik.

Wenn wir ein lineares Gleichungssystem durch GAUSS-Elimination lösen, bringen wir es zunächst auf eine Treppengestalt, indem wir die erste vorkommende Variable aus allen Gleichungen außer der ersten eliminieren, die zweite aus allen Gleichungen außer den ersten beiden, uns so weiter, bis wir schließlich Gleichungen haben, deren letzte entweder nur eine Variable enthält oder aber eine Relation zwischen Variablen, für die es sonst keine weiteren Bedingungen mehr gibt. Konkret sieht ein Eliminationsschritt folgendermaßen aus: Wenn wir im Falle der beiden Gleichungen

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = u \quad \text{mit} \quad a_1 \neq 0 \tag{1}$$

$$b_1x_1 + b_2x_2 + \cdots + b_nx_n = v \tag{2}$$

die Variable X_1 mit Hilfe von (1) aus (2) eliminieren wollen, ersetzen wir die zweite Gleichung durch ihre Summe mit $-b_1/a_1$ mal der ersten. Die theoretische Rechtfertigung für diese Umformung besteht darin, daß das Gleichungssystem bestehend aus (1) und (2) sowie das neue Gleichungssystem dieselbe Lösungsmenge haben, und daran ändert sich auch dann nichts, wenn noch weitere Gleichungen dazukommen.

Ähnlich können wir vorgehen, wenn wir ein nichtlineares Gleichungssystem in nur einer Variablen betrachten: Am schwersten sind natürlich die Gleichungen vom höchsten Grad, also versuchen wir, die zu reduzieren auf Polynome niedrigeren Grades. Das kanonische Verfahren

dazu ist die Polynomdivision: Haben wir zwei Polynome

$$f = a_d X^d + a_{d-1} X^{d-1} + \cdots + a_1 X + a_0 \quad \text{und}$$
$$g = b_e X^e + b_{e-1} X^{e-1} + \cdots + b_1 X + b_0$$

mit $e \leq d$, so dividieren wir f durch g, d.h. wir berechnen einen Quotienten q und einen Rest r derart, daß $f = qg + r$ ist und r entweder verschwindet oder kleineren Grad als g hat. Konkret: Bei jedem Divisionsschritt haben wir ein Polynom

$$f = c_\delta X^\delta + c_{\delta-1} X^{\delta-1} + \cdots + c_1 X + c_0 \quad \text{mit} \quad c_\delta \neq 0\,,$$

das wir für $\delta \geq e$ mit Hilfe des Divisors

$$g = b_e X^e + b_{e-1} X^{e-1} + \cdots + b_1 X + b_0$$

reduzieren, indem wir es ersetzen durch

$$f - \frac{b_e}{c_\delta} X^{\delta-e} g\,.$$

Das führen wir so lange fort, bis f auf Null oder ein Polynom von kleinerem Grad als e reduziert ist: Das ist dann der Divisionsrest r. Auch hier ist klar, daß sich nichts an der Lösungsmenge ändert, wenn man die beiden Gleichungen f, g ersetzt durch g, r, denn

$$f = qg + r \quad \text{und} \quad r = f - qg\,,$$

d.h. f und g verschwinden genau dann für einen Wert x, wenn g und r an der Stelle x verschwinden.

In beiden Fällen ist die Vorgehensweise sehr ähnlich: Wir vereinfachen das Gleichungssystem schrittweise, indem wir eine Gleichung ersetzen durch ihre Summe mit einem geeigneter Vielfachen einer anderen Gleichung.

Dieselbe Strategie wollen wir auch anwenden Systeme von Polynomgleichungen in mehreren Veränderlichen. Erstes Problem dabei ist, daß wir nicht wissen, wie wir die Monome eines Polynoms anordnen sollen und damit, was der führende Term ist. Dazu gibt es eine ganze Reihe verschiedener Strategien, von denen je nach Anwendung mal die eine, mal die andere vorteilhaft ist.

§3: Monomordnungen und der Divisionsalgorithmus

Wir betrachten Polynome in n Variablen $X_1, \ldots, X_n$ über einem Körper k und setzen zur Abkürzung

$$X^\alpha = X_1^{\alpha_1} \cdots X_n^{\alpha_n} \quad \text{mit} \quad \alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}_0 \,.$$

Terme der Form X^α haben wir in §1 als Monome bezeichnet und ihnen die Summe der α_i als Grad zugeordnet.

Eine Anordnung der Monome ist offensichtlich äquivalent zu einer Anordnung auf $\mathbb{N}_0^n$, und es gibt sehr viele Möglichkeiten, diese Menge anzuordnen. Für uns sind allerdings nur Anordnungen interessant, die einigermaßen kompatibel sind mit der algebraischen Struktur des Polynomrings $k[X_1, \ldots, X_n]$; beispielsweise wollen wir sicherstellen, daß der führende Term des Produkts zweier Polynome das Produkt der führenden Terme der Faktoren ist – wie wir es auch vom Eindimensionalen her gewohnt sind. Daher definieren wir

Definition: *a)* Eine Monomordnung ist eine Ordnungsrelation „$<$“ auf $\mathbb{N}_0^n$, für die gilt

1. „$<$“ ist eine Linear- oder Totalordnung, d.h. für zwei Elemente $\alpha, \beta \in \mathbb{N}_0^n$ ist entweder $\alpha < \beta$ oder $\beta < \alpha$ oder $\alpha = \beta$.
2. Für $\alpha, \beta, \gamma \in \mathbb{N}_0^n$ gilt $\alpha < \beta \Longrightarrow \alpha + \gamma < \beta + \gamma$.
3. „$<$“ ist eine Wohlordnung, d.h. jede Teilmenge $I \subseteq \mathbb{N}_0^n$ hat ein kleinstes Element.

b) Für ein Polynom $f = \sum_{\alpha \in I} c_\alpha X^\alpha \in k[X_1, \ldots, X_n]$ mit $c_\alpha \neq 0$ für alle $\alpha \in I \subset \mathbb{N}_0$ sei γ das größte Element von I bezüglich einer fest gewählten Monomordnung. Dann bezeichnen wir bezüglich dieser Monomordnung

- γ = multideg f als Multigrad von f
- X^γ = FM(f) als führendes Monom von f
- c_γ = FK(f) als führenden Koeffizienten von f
- $c_\gamma X^\gamma$ = FT(f) als führenden Term von f

Der Grad deg f von f ist, wie in der Algebra üblich, der höchste Grad eines Monoms von f; je nach gewählter Monomordnung muß das nicht unbedingt der Grad des führenden Monoms sein.

Beispiele von Monomordnungen sind

a) Die lexikographische Ordnung: Hier ist $\alpha < \beta$ genau dann, wenn für den ersten Index i, in dem sich α und β unterscheiden, $\alpha_i < \beta_i$ ist. Betrachtet man Monome X^α als Worte über dem (geordneten) Alphabet $\{X_1, \ldots, X_n\}$, kommt hier ein Monom X^α genau dann vor X^β, wenn die entsprechenden Worte im Lexikon in dieser Reihenfolge gelistet werden. Die ersten beiden Forderungen an eine Monomordnung sind klar, und auch die Wohlordnung macht keine großen Probleme: Man betrachtet zunächst die Teilmenge aller Exponenten $\alpha \in I$ mit kleinstmöglichem α_1, unter diesen die Teilmenge mit kleinstmöglichem α_2, *usw.,* bis man bei α_n angelangt ist. Spätestens hier ist die verbleibende Teilmenge einelementig, und ihr einziges Element ist das gesuchte kleinste Element von I.

b) Die graduierte lexikographische Ordnung: Hier ist der Grad eines Monoms erstes Ordnungskriterium: Ist $\deg X^\alpha < \deg X^\beta$, so definieren wir $\alpha < \beta$. Falls beide Monome gleichen Grad haben, soll $\alpha < \beta$ genau dann gelten, wenn α im lexikographischen Sinne kleiner als β ist. Auch hier sind offensichtlich alle drei Forderungen erfüllt.

c) Die inverse lexikographische Ordnung: Hier ist $\alpha < \beta$ genau dann, wenn $\alpha_i < \beta_i$ für den *letzten* Index i, in dem sich α und β unterscheiden. Das entspricht offensichtlich gerade der lexikographischen Anordnung bezüglich des rückwärts gelesenen Alphabets $X_n, \ldots, X_1$. Entsprechend läßt sich natürlich auch bezüglich jeder anderen Permutation des Alphabets eine Monomordnung definieren, so daß diese Ordnung nicht sonderlich interessant ist – außer als Bestandteil der im folgenden definierten Monomordnung:

d) Die graduierte inverse lexikographische Ordnung: Wie bei der graduierten lexikographischen Ordnung ist hier der Grad eines Monoms erstes Ordnungskriterium: Falls $\deg X^\alpha < \deg X^\beta$, ist $\alpha < \beta$, und nur falls beide Monome gleichen Grad haben, soll $\alpha < \beta$ genau dann gelten, wenn α im Sinne der inversen lexikographischen Ordnung *größer* ist als β. Man beachte, daß wir hier also nicht nur die Reihenfolge der Variablen invertieren, sondern auch die Ordnungsrelation im Fall gleicher Grade. Es ist nicht schwer zu sehen, daß auch damit

eine Monomordnung definiert wird: Mit den ersten beiden Forderungen gibt es wie üblich keine Probleme, und wenn wir eine Menge M von Monomen haben, gibt es darin eine Teilmenge bestehend aus den Monomen kleinsten Grades. Da es für jeden Grad nur endlich viele Monome gibt, ist diese Menge endlich, hat also bezüglich der inversen lexikographischen Ordnung nicht nur ein kleinstes, sondern auch ein größtes Element. Dieses ist das kleinste Element von M bezüglich der graduierten invers lexikographischen Ordnung.

Für das folgende werden wir noch einige Eigenschaften einer Monomordnung benötigen, die in der Definition nicht erwähnt sind.

Als erstes wollen wir uns überlegen, daß bezüglich jeder Monomordnung auf $\mathbb{N}_0^n$ kein Element kleiner sein kann als $(0, \ldots, 0)$: Wäre nämlich $\alpha < (0, \ldots, 0)$, so wäre wegen der zweiten Eigenschaft auch

$$2\alpha = \alpha + \alpha < \alpha + (0, \ldots, 0) = \alpha$$

und so weiter, so daß wir eine unendliche Folge

$$\alpha > 2\alpha > 3\alpha > \cdots$$

hätten, im Widerspruch zur dritten Forderung.

Daraus folgt nun sofort, daß das Produkt zweier Monome größer ist als jeder der beiden Faktoren und damit auch, daß ein echter Teiler eines Monoms immer kleiner ist als dieses. Außerdem folgt, daß für ein Produkt von Polynomen stets $\mathrm{FM}(fg) = \mathrm{FM}(f) \cdot \mathrm{FM}(g)$ ist.

Die Eliminationsschritte beim GAUSS-Algorithmus können auch als Divisionen mit Rest verstanden werden, und beim EUKLIDischen Algorithmus ist ohnehin alles Division mit Rest. Für ein Verallgemeinerung der beiden Algorithmen auf Systeme nichtlinearer Gleichungssysteme brauchen wir also auch einen Divisionsalgorithmus für Polynome in mehreren Veränderlichen, der die eindimensionale Polynomdivision mit Rest und die Eliminationsschritte beim GAUSS-Algorithmus verallgemeinert.

Beim GAUSS-Algorithmus brauchen wir im allgemeinen mehr als nur einen Eliminationsschritt, bis wir eine Gleichung auf eine Variable reduziert haben; entsprechend wollen wir auch hier einen Divisionsalgorithmus betrachten, der gegebenenfalls auch mehrere Divisoren gleichzeitig behandeln kann.

Wir gehen also aus von einem Polynom $R = f \in k[X_1, \ldots, X_n]$, wobei k irgendein Körper ist, in dem wir rechnen können, meistens also $k = \mathbb{Q}$ oder $k = \mathbb{F}_p$ oder eine endliche Erweiterung davon. Dieses Polynom wollen wir dividieren durch die Polynome $f_1, \ldots, f_m \in R$, d.h. wir suchen Polynome $a_1, \ldots, a_m, r \in R$, so daß

$$f = a_1 f_1 + \cdots + a_m f_m + r$$

ist, wobei r in irgendeinem noch zu präzisierenden Weise kleiner als die f_i sein soll.

Da es sowohl bei GAUSS als auch bei EUKLID auf die Anordnung der Terme ankommt, legen wir als erstes eine Monomordnung fest; wenn im folgenden von führenden Termen *etc.* die Rede ist, soll es sich stets um die führenden Terme *etc.* bezüglich dieser Ordnung handeln.

Mit dieser Konvention geht der Algorithmus dann folgendermaßen:

Gegeben sind $f, f_1, \ldots, f_m \in R$
Berechnet werden $a_1, \ldots, a_m, r \in R$ mit $f = a_1 f_1 + \cdots + a_m f_m + r$, wobei r kein Monom enthält, das durch das führende Monom eines der f_i teilbar ist.

1. Schritt (Initialisierung): Setze $a_1 = \cdots = a_m = r = 0$ und $p = f$.

2. Schritt (Endebedingung): Im Falle $p = 0$ endet der Algorithmus.

3. Schritt (Divisionsschritt): Falls keiner der führenden Terme FT f_i den führenden Term FT p teilt, wird p ersetzt durch $p - \mathrm{FT}\, p$ und r durch $r + \mathrm{FT}\, p$. Andernfalls sei i der kleinste Index, für den FT f_i Teiler von FT p ist; der Quotient sei q. Dann wird a_i ersetzt durch $a_i + q$ und p durch $p - q f_i$. Weiter geht es mit dem 2. Schritt.

Offensichtlich ist die Bedingung $f - p = a_1 f_1 + \cdots + a_m f_m + r$ nach der Initialisierung im ersten Schritt erfüllt, und sie bleibt auch bei jeder Anwendung des Divisionsschritts erfüllt. Außerdem endet der Algorithmus nach endlich vielen Schritten: Bei jedem Divisionsschritt wird der führende Term von p eliminiert, und alle Monome, die eventuell neu dazukommen, sind kleiner oder gleich dem führenden Monom von f_i. Da letzteres das (alte) führende Monom von p teilt, kann es nicht größer

sein als dieses, d.h. der führende Term des neuen p ist kleiner als der des alten. Wegen der Wohlordnungseigenschaft einer Monomordnung kann es keine unendliche absteigende Kette von Monomen geben; daher muß der Algorithmus nach endlich vielen Schritten abbrechen.

Bei der klassischen Polynomdivision für Polynome in einer Variablen über einem Körper wissen wir, daß der Rest kleineren Grad hat als der Divisor. Das muß hier nicht der Fall sein; wir können nur sagen, daß der Rest keine Monome enthält, die durch den führenden Term eines der Divisoren f_i teilbar sind.

Um den Algorithmus besser zu verstehen, betrachten wir zunächst zwei Beispiele:

Als erstes dividieren wir $f = X^2Y + XY^2 + Y^2$ durch $f_1 = XY - 1$ und $f_2 = Y^2 - 1$.

Zur Initialisierung setzen wir $a_1 = a_2 = r = 0$ und $p = f$. Wir verwenden die lexikographische Ordnung; bezüglich derer ist der führende Term von p gleich X^2Y und der von f_1 gleich XY. Letzteres teilt X^2Y, wir setzen also

$$p \leftarrow p - Xf_1 = XY^2 + X + Y^2 \quad \text{und} \quad a_1 \leftarrow a_1 + X = X\,.$$

Neuer führender Term von p ist XY^2; auch das ist ein Vielfaches von XY, also setzen wir

$$p \leftarrow p - Yf_1 = X + Y^2 + Y \quad \text{und} \quad a_1 \leftarrow a_1 + Y = X + Y\,.$$

Nun ist X der führende Term von p, und der ist weder durch XY noch durch Y^2 teilbar, also kommt er in den Rest:

$$p \leftarrow p - X = Y^2 + Y \quad \text{und} \quad r \leftarrow r + X = X\,.$$

Der nun führende Term Y^2 von p ist gleichzeitig der führende Term von f_2 und nicht teilbar durch XY, also wird

$$p \leftarrow p - f_2 = Y + 1 \quad \text{und} \quad a_2 \leftarrow a_2 + 1 = 1\,.$$

Die verbleibenden Terme von p sind weder durch XY noch durch Y^2 teilbar, kommen also in den Rest, so daß wir als Ergebnis erhalten

$$f = a_1f_1 + a_2f_2 + r \quad \text{mit} \quad a_1 = X + Y, \quad a_2 = 1 \quad \text{und} \quad r = X + Y + 1\,.$$

Wenn wir statt durch das Paar (f_1, f_2) durch (f_2, f_1) dividiert hätten, hätten wir im ersten Schritt zwar ebenfalls X^2Y durch XY dividiert, denn durch Y^2 ist es nicht teilbar. Der neue führende Term XY^2 ist aber durch beides teilbar, und wenn f_2 an erster Stelle steht, nehmen wir im Zweifelsfall dessen führenden Term. Man rechnet leicht nach, daß man hier mit folgendem Ergebnis endet:

$$f = a_1 f_1 + a_2 f_2 + r \quad \text{mit} \quad a_1 = X + 1, \quad a_2 = X \quad \text{und} \quad r = X + 1 .$$

Wie wir sehen, sind also sowohl die „Quotienten“ a_i als auch der „Rest“ r von der Reihenfolge der f_i abhängig. Sie hängen natürlich im allgemeinen auch ab von der verwendeten Monomordnung; deshalb haben wir die schließlich eingeführt.

Als zweites Beispiel wollen wir $f = XY^2 - X$ durch die beiden Polynome $f_1 = XY + 1$ und $f_2 = Y^2 - 1$ dividieren. Im ersten Schritt dividieren wir XY^2 durch XY mit Ergebnis Y, ersetzen also f durch $-X - Y$. Diese beiden Terme sind weder durch XY noch durch Y^2 teilbar, also ist unser Endergebnis

$$f = a_1 f_1 + a_2 f_2 + r \quad \text{mit} \quad a_1 = Y, \quad a_2 = 0 \quad \text{und} \quad r = -X - Y .$$

Hätten wir stattdessen durch (f_2, f_1) dividiert, hätten wir als erstes XY^2 durch Y^2 dividiert mit Ergebnis X; da $f = X f_2$ ist, geht die Division hier ohne Rest auf. Der Divisionsalgorithmus erlaubt uns also nicht einmal die sichere Feststellung, ob f als Linearkombination der f_i darstellbar ist oder nicht; als alleiniges Hilfsmittel zur Lösung nichtlinearer Gleichungssysteme reicht er offenbar nicht aus. Daher müssen wir in den folgenden Paragraphen noch weitere Werkzeuge betrachten.

§4: Der Hilbertsche Basissatz

Die Grundidee des Algorithmus von BUCHBERGER besteht darin, das Gleichungssystem so abzuändern, daß möglichst viele seiner Eigenschaften bereits an den führenden Termen der Gleichungen ablesbar sind.

Angenommen, wir haben ein nichtlineares Gleichungssystem

$$f_1(X_1, \ldots, X_n) = \cdots = f_m(X_1, \ldots, X_n) = 0$$

mit $f_i \in R = k[X_1, \ldots, X_n]$; seine Lösungsmenge sei $\mathcal{L} \subseteq k^n$.

Wie wir aus §1 wissen, hängt $\mathcal{L}$ nur ab von dem Ideal $I = (f_1, \ldots, f_m)$; zur Lösung des Systems sollten wir daher versuchen, ein möglichst „einfaches“ Erzeugendensystem für dieses Ideal zu finden.

Ganz besonders einfach (wenn auch selten ausreichend) sind Ideale, die von Monomen erzeugt werden:

Definition: Ein Ideal $I \lhd R = k[X_1, \ldots, X_n]$ heißt *monomial,* wenn es von (nicht notwendigerweise endlich vielen) Monomen erzeugt wird.

Nehmen wir an, I werde erzeugt von den Monomen X^α mit α aus einer Indexmenge A. Ist dann X^β irgendein Monom aus I, kann es als endliche Linearkombination

$$X^\beta = \sum_{i=1}^{r} f_i X^{\alpha_i} \quad \text{mit} \quad \alpha_i \in A$$

geschrieben werden, wobei die f_i irgendwelche Polynome aus R sind. Da sich jedes Polynom als Summe von Monomen schreiben läßt, können wir f_i als k-Linearkombination von Monomen X^γ schreiben und bekommen damit eine neue Darstellung von X^β als Summe von Termen der Form $cX^\gamma X^\alpha$ mit $\alpha \in A$, $\beta \in \mathbb{N}_0^n$ und $c \in k$. Sortieren wir diese Summanden nach den resultierenden Monomen $X^{\gamma+\alpha}$ und fassen alle Summanden mit gleichem Monom zusammen, so entsteht eine k-Linearkombination verschiedener Monome, die insgesamt gleich X^β ist. Das ist aber nur möglich, wenn diese Summe aus dem einen Summanden X^β besteht, d.h. β läßt sich schreiben in der Form $\beta = \alpha + \gamma$ mit einem $\alpha \in A$ und einem $\gamma \in \mathbb{N}_0^n$.

Dies zeigt, daß ein Monom X^β genau dann in I liegt, wenn $\beta = \alpha + \gamma$ ist mit einem $\alpha \in A$ und einem $\gamma \in \mathbb{N}_0^n$, d.h. X^β ist das Produkt eines der erzeugenden Monome mit *irgendeinem* Monom. Das Ideal I besteht genau aus den Polynomen f, die sich als k-Linearkombinationen solcher Monome schreiben lassen.

Damit folgt insbesondere, daß ein Polynom f genau dann in einem monomialen Ideal I liegt, wenn jedes seiner Monome dort liegt.

Lemma von Dickson: Jedes monomiale Ideal in $R = k[X_1, \ldots, X_n]$ kann von endlich vielen Monomen erzeugt werden.

Der *Beweis* wird durch vollständige Induktion nach n geführt. Im Fall $n = 1$ ist alles klar, denn da sind die Monome gerade die Potenzen der einzigen Variable, und natürlich erzeugt jede Menge von Potenzen genau dasselbe Ideal wie die Potenz mit dem kleinsten Exponenten aus dieser Menge. Hier kommt man also sogar mit einem einzigen Monom aus.

Im Fall $n > 1$ und $\alpha \in \mathbb{N}_0^n$ setzen wir $X'^{\alpha} = X_1^{\alpha_1} \cdots X_{n-1}^{\alpha_{n-1}}$ und betrachten das Ideal

$$J = \left(X'^{\alpha} \;\middle|\; X^{\alpha} \in I \right) \lhd k[X_1, \ldots, X_{n-1}] \,.$$

Nach Induktionsvoraussetzung wird J erzeugt von endlich vielen Monomen X'^{α}

Jedes Monom aus aus dem endlichen Erzeugendensystem von J läßt sich in der Form X'^{α} schreiben mit einem $\alpha \in \mathbb{N}_0^n$, für das X^{α} in I liegt. Unter den Indizes α_n, die wir dabei jeweils an das $(n-1)$-Tupel $(\alpha_1, \ldots, \alpha_{n-1})$ anhängen, sei r der größte. Dann liegt $X'^{\alpha'} X_n^r$ für jedes Monom aus dem Erzeugendensystem von J in I und damit für jedes Monom aus J. Die endlich vielen Monome $X^{\alpha'} X_n^r$ erzeugen also zumindest ein Teilideal von I.

Es gibt aber natürlich auch noch Monome in I, in denen X_n mit einem kleineren Exponenten als r auftritt. Um auch diese Elemente zu erfassen, betrachten wir für jedes $s < r$ das Ideal $J_s \lhd k[X_1, \ldots, X_{n-1}]$, das von allen jeden Monomen X'^{α} erzeugt wird, für die $X'^{\alpha} X_n^s$ in I liegt. Auch jedes der J_s wird nach Induktionsannahme erzeugt von endlich vielen Monomen X'^{α}, und wenn wir die sämtlichen Monome $X'^{\alpha} X_n^s$ zu unserem Erzeugendensystem hinzunehmen (für alle $s = 0, 1, \cdots, r-1$), haben wir offensichtlich ein Erzeugendensystem von I aus endlich vielen Monomen gefunden.

■

LEONARD EUGENE DICKSON (1874–1954) wurde in Iowa geboren, wuchs aber in Texas auf. Seinen Bachelor- und Mastergrad bekam er von der University of Texas, danach ging er an die Universität von Chicago. Mit seiner 1896 dort eingereichte Dissertation *Analytic Representation of Substitutions on a Power of a Prime Number of Letters with a Discussion of the Linear Group* wurde er der erste dort promovierte Mathematiker. Auch die weiteren seiner 275 wissenschaftlichen Arbeiten, darunter acht Bücher, beschäftigen sich vor allem mit der Algebra und Zahlentheorie. Den größten Teil seines Berufslebens verbrachte er als Professor an der Universität von Chicago, dazu kommen regelmäßige Besuche in Berkeley.

Beliebige Ideale sind im allgemeinen nicht monomial; schon das von $X + 1$ erzeugte Ideal in $k[X]$ ist ein Gegenbeispiel, denn es enthält weder das Monom X noch das Monom 1, im Widerspruch zu der oben gezeigten Eigenschaft eines monomialen Ideals, zu jedem seiner Elemente auch dessen sämtliche Monome zu enthalten.

Um monomiale Ideale auch für die Untersuchung solcher Ideale nützlich zu machen, wählen wir eine Monomordnung auf R und definieren für ein beliebiges Ideal $I \lhd R \underset{\text{def}}{=} k[X_1, \ldots, X_n]$ das monomiale Ideal

$$\mathrm{FM}(I) = \Big(\mathrm{FM}(f) \;\Big|\; f \in I \setminus \{0\}\Big)\,,$$

das von den führenden Monomen *aller* Elemente von I erzeugt wird – außer natürlich dem nicht existierenden führenden Monom der Null.

Nach dem Lemma von DICKSON ist $\mathrm{FM}(I)$ erzeugt von endlich vielen Monomen. Jedes dieser Monome ist, wie wir eingangs gesehen haben, ein Vielfaches eines der erzeugenden Monome, also eines führenden Monoms eines Elements von I. Ein Vielfaches des führenden Monoms ist aber das führende Monom des entsprechenden Vielfachen des Elements von I, denn $\mathrm{FM}(X^\gamma f) = X^\gamma \,\mathrm{FM}(f)$, da für jede Monomordnung gilt $\alpha < \beta \Longrightarrow \alpha + \beta < \alpha + \gamma$. Somit wird $\mathrm{FM}(I)$ erzeugt von endlich vielen Monomen der Form $\mathrm{FM}(f_i)$, wobei die f_i Elemente von I sind. Wir wollen sehen, daß die Elemente f_i das Ideal I erzeugen; damit folgt insbesondere

Hilbertscher Basissatz: Jedes Ideal $I \lhd R = k[X_1, \ldots, X_n]$ hat ein endliches Erzeugendensystem.

Beweis: Wie wir bereits wissen, gibt es Elemente $f_1, \ldots f_m \in I$, so daß FM(I) von den Monomen FM(f_i) erzeugt wird. Um zu zeigen, daß die Elemente f_i das Ideal I erzeugen, betrachten wir ein beliebiges Element $f \in I$ und versuchen, es als R-Linearkombination der f_i zu schreiben. Division von f durch $f_1, \ldots, f_m$ zeigt, daß es Polynome $a_1, \ldots, a_m$ und r in R gibt derart, daß

$$f = a_1 f_1 + \cdots + a_m f_m + r \, .$$

Wir sind fertig, wenn wir zeigen können, daß der Divisionsrest r verschwindet.

Falls r *nicht* verschwindet, zeigt der Divisionsalgorithmus, daß das führende Monom FM(r) von r durch kein führendes Monom FM(f_i) eines der Divisoren f_i teilbar ist. Andererseits ist aber

$$r = f - (a_1 f_1 + \cdots + a_m f_m)$$

ein Element von I, und damit liegt FM(r) im von den FM(f_i) erzeugten Ideal FM(I). Somit muß FM(r) Vielfaches eines FM(f_i) sein, ein Widerspruch. Also ist $r = 0$.

■

DAVID HILBERT (1862–1943) wurde in Königsberg geboren, wo er auch zur Schule und zur Universität ging. Er promovierte dort 1885 mit einem Thema aus der Invariantentheorie, habilitierte sich 1886 und bekam 1893 einen Lehrstuhl. 1895 wechselte er an das damalige Zentrum der deutschen wie auch internationalen Mathematik, die Universität Göttingen, wo er bis zu seiner Emeritierung im Jahre 1930 lehrte. Seine Arbeiten umfassen ein riesiges Spektrum aus unter anderem Invariantentheorie, Zahlentheorie, Geometrie, Funktionalanalysis, Logik und Grundlagen der Mathematik sowie auch zur Relativitätstheorie. Er gilt als einer der Väter der modernen Algebra.

§5: Gröbner-Basen und der Buchberger-Algorithmus

Angesichts der Rolle der führenden Monome im obigen Beweis bietet sich folgende Definition an für eine Idealbasis, bezüglich derer möglichst viele Eigenschaften bereits an den führenden Monomen abgelesen werden können:

Definition: Eine endliche Teilmenge $G = \{g_1, \ldots, g_m\} \subset I$ eines Ideals $I \lhd R = k[X_1, \ldots, X_n]$ heißt Standardbasis oder GRÖBNER-Basis von I, falls die Monome FM(g_i) das Ideal FM(I) erzeugen.

WOLFGANG GRÖBNER wurde 1899 im damals noch österreichischen Südtirol geboren. Nach Ende des ersten Weltkriegs, in dem er an der italienischen Front kämpfte, studierte er zunächst an der TU Graz Maschinenbau, beendete dieses Studium aber nicht, sondern begann 1929 an der Universität Wien ein Mathematikstudium. Nach seiner Promotion ging er zu EMMY NOETHER nach Göttingen, um dort Algebra zu lernen. Aus materiellen Gründen mußte er schon bald nach Österreich zurück, konnte aber auch dort zunächst keine Anstellung finden, so daß er Kleinkraftwerke baute und im Hotel seines Vaters aushalf. Ein italienischen Mathematiker, der dort seinen Urlaub verbrachte, vermittelte ihm eine Stelle an der Universität Rom, die er 1939 wieder verlassen mußte, nachdem er sich beim Anschluß Südtirols an Italien für die deutsche Staatsbürgerschaft entschieden hatte. Während des zweiten Weltkriegs arbeitete er größtenteils an einem Forschungsinstitut der Luftwaffe, nach Kriegsende als Extraordinarius in Wien, dann als Ordinarius in Innsbruck, wo er 1980 starb. Seine Arbeiten beschäftigen sich mit der Algebra und algebraischen Geometrie sowie mit Methoden der Computeralgebra zur Lösung von Differentialgleichungen.

Die Theorie der GRÖBNER-Basen wurde von seinem Studenten BRUNO BUCHBERGER in dessen Dissertation entwickelt. BUCHBERGER wurde 1942 in Innsbruck geboren, wo er auch Mathematik studierte und 1966 bei GRÖBNER promovierte mit der Arbeit *Ein Algorithmus zum Auffinden der Basiselemente des Restklassenrings nach einem nulldimensionalen Polynomideal.* Er arbeitete zunächst als Assistent, nach seiner Habilitation als Dozent an der Universität Innsbruck, bis er 1974 einen Ruf auf den Lehrstuhl für Computermathematik an der Universität Linz erhielt. Dort gründete er 1987 das Research Institute for Symbolic Computation (RISC), dessen Direktor er bis 1999 war. 1989 initiierte er in Hagenberg (etwa 20 km nordöstlich von Linz) die Gründung eines Softwareparks mit angeschlossener Fachhochschule; er hat mittlerweile fast Tausend Mitarbeiter. Außer mit Computeralgebra beschäftigt er sich auch im Rahmen des Theorema-Projekts mit dem automatischen Beweisen mathematischer Aussagen.

Wie der obige Beweis des HILBERTschen Basissatzes zeigt, erzeugt eine GRÖBNER-Basis das Ideals. Außerdem hat jedes Ideal I im Polynomring eine GRÖBNER-Basis, denn nach dem Lemma von DICKSON hat das Ideal

der führenden Monome ein endliches Erzeugendensystem, und jedes Monom aus diesem Erzeugendensystem ist führendes Monom eines Polynoms $f_i \in I$. Die Menge der Polynome f_i ist offensichtlich eine GRÖBNER-Basis im Sinne der obigen Definition.

Bevor wir uns damit beschäftigen, wie man diese berechnen kann, wollen wir zunächst eine wichtige Eigenschaft betrachten.

$\{g_1, \ldots, g_m\}$ sei eine GRÖBNER-Basis eines Ideals $I \lhd R$. Wir wollen ein beliebiges Element $f \in R$ durch $g_1, \ldots, g_m$ dividieren. Dies liefert als Ergebnis

$$f = a_1 g_1 + \cdots + a_m g_m + r\,,$$

wobei kein Monom von r durch eines der Monome FM(g_i) teilbar ist. Wie wir wissen, sind allerdings bei der Polynomdivision im allgemeinen weder der Divisionsrest r noch die Koeffizienten a_i auch nur im entferntesten eindeutig. Wir wollen untersuchen, wie sich das hier verhält.

Angenommen. wir haben zwei Darstellungen

$$f = a_1 g_1 + \cdots + a_m g_m + r = b_1 g_1 + \cdots + b_m g_m + s$$

der obigen Form. Dann ist

$$(a_1 - b_1) g_1 + \cdots + (a_m - b_m) g_m = s - r\,.$$

Links steht ein Element von I, also auch rechts. Andererseits enthält aber weder r noch s ein Monom, das durch eines der Monome FM(g_i) teilbar ist, d.h. $r - s = 0$, da die FM(g_i) ja das Ideal FM(I) erzeugen. Somit ist bei der Division durch die Elemente einer GRÖBNER-Basis der Divisionsrest eindeutig bestimmt. Insbesondere ist f genau dann ein Element von I, wenn der Divisionsrest verschwindet. Wenn wir eine GRÖBNER-Basis haben, können wir also leicht entscheiden, ob ein gegebenes Element $f \in R$ im Ideal I liegt.

Nachdem im Fall einer GRÖBNER-Basis der Divisionsrest nicht von der Reihenfolge der Basiselemente abhängt, können wir ihn durch ein Symbol bezeichnen, das nur von der *Menge* $G = \{g_1, \ldots, g_m\}$ abhängt; wir schreiben $\overline{f}^G$.

Als nächstes wollen wir uns mit der Frage beschäftigen, wie wir für ein vorgegebenes Ideal I eine GRÖBNER-Basis bestimmen können.

Dazu müssen wir uns als erstes überlegen, *wie* das Ideal vorgegeben sein soll. Wenn wir damit rechnen wollen, müssen wir irgendeine Art von endlicher Information haben; was sich anbietet ist natürlich ein endliches Erzeugendensystem.

Wir gehen also aus von einem Ideal $I = (f_1, \ldots, f_m)$ und suchen eine GRÖBNER-Basis. Das Problem ist, daß die Monome $\mathrm{FM}(f_i)$ im allgemeinen nicht ausreichen, um das monomiale Ideal $\mathrm{FM}(I)$ zu erzeugen, denn dieses enthält ja *jedes* Monom eines jeden Elements von I und nicht nur das führende. Wir müssen daher neue Elemente produzieren, deren führende Monome in den gegebenen Elementen f_i oder auch anderen Elementen von I erst weiter hinten vorkommen.

BUCHBERGERs Idee dazu war die Konstruktion sogenannter S-Polynome: Seien $f, g \in R$ zwei Polynome; $\mathrm{FM}(f) = X^\alpha$ und $\mathrm{FM}(g) = X^\beta$ seien ihre führenden Monome, und X^γ sei das kgV von X^α und X^β, d.h. $\gamma_i = \max(\alpha_i, \beta_i)$ für $i = 1, \ldots, n$. Das S-Polynom von f und g ist

$$S(f,g) = \frac{X^\gamma}{\mathrm{FT}(f)} \cdot f - \frac{X^\gamma}{\mathrm{FT}(g)} \cdot g\,.$$

Da $\frac{X^\gamma}{\mathrm{FT}(f)} \cdot f$ und $\frac{X^\gamma}{\mathrm{FT}(g)} \cdot g$ beide nicht nur dasselbe führende Monom X^γ haben, sondern es wegen der Division durch den führenden *Term* statt nur das führende Monom auch beide mit Koeffizient eins enthalten, fällt es bei der Bildung von $S(f,g)$ weg. Daher ist das führende Monom von $S(f,g)$ kleiner als X^γ. Das folgende Lemma ist der Kern des Beweises, daß S-Polynome alles sind, was wir brauchen, um GRÖBNER-Basen zu berechnen.

Lemma: Für die Polynome $f_1, \ldots, f_m \in R$ sei

$$S = \sum_{i=1}^{m} \lambda_i X^{\alpha_i} f_i \quad \text{mit} \quad \lambda_i \in k \quad \text{und} \quad \alpha_i \in \mathbb{N}_0^n$$

eine Linearkombination, zu der es ein $\delta \in \mathbb{N}_0^n$ gebe, so daß alle Summanden X^δ als führendes Monom haben, d.h. $\alpha_i + \mathrm{multideg}\, f_i = \delta_i$ für $i = 1, \ldots, m$. Falls $\mathrm{multideg}\, S < \delta$ ist, gibt es Elemente $\lambda_{ij} \in k$, so daß

$$S = \sum_{i=1}^{m} \sum_{j=1}^{m} \lambda_{ij} X^{\gamma_{ij}} S(f_i, f_j)$$

ist mit $X^{\gamma_{ij}} = \mathrm{kgV}\big(\mathrm{FM}(f_i), \mathrm{FM}(f_j)\big)$.

Beweis: Der führende Koeffizient von f_i sei μ_i; dann ist $\lambda_i \mu_i$ der führende Koeffizient von $\lambda_i X^{\alpha_i} f_i$. Somit ist multideg S genau dann kleiner als δ, wenn $\sum_{i=1}^{m} \lambda_i \mu_i$ verschwindet. Wir normieren alle $X^{\alpha_i} f_i$ auf führenden Koeffizienten eins, indem wir $p_i = X^{\alpha_i} f_i / \mu_i$ betrachten; dann ist

$$\begin{aligned} S = \sum_{i=1}^{m} \lambda_i \mu_i p_i = {} & \lambda_1 \mu_1 (p_1 - p_2) + (\lambda_1 \mu_1 + \lambda_2 \mu_2)(p_2 - p_3) + \cdots \\ & + (\lambda_1 \mu_1 + \cdots + \lambda_{m-1} \mu_{m-1})(p_{m-1} - p_m) \\ & + (\lambda_1 \mu_1 + \cdots + \lambda_m \mu_m) p_m \,, \end{aligned}$$

wobei der Summand in der letzten Zeile genau dann verschwindet, wenn multideg $S < \delta$.

Da alle p_i denselben Multigrad δ und denselben führenden Koeffizienten eins haben, kürzen sich in den Differenzen $p_i - p_j$ die führenden Terme weg, genau wie in den S-Polynomen. In der Tat: Bezeichnen wir den Multigrad von $\mathrm{kgV}\big(\mathrm{FM}(f_i), \mathrm{FM}(f_j)\big)$ mit γ_{ij}, so ist

$$p_i - p_j = X^{\delta - \gamma_{ij}} S(f_i, f_j) \,.$$

Damit hat die obige Summendarstellung von S die gewünschte Form. ∎

Daraus folgt ziemlich unmittelbar

Satz: Ein Erzeugendensystem $f_1, \ldots, f_m$ eines Ideals I im Polynomring $R = k[X_1, \ldots, X_n]$ ist genau dann eine GRÖBNER-Basis, wenn jedes S-Polynom $S(f_i, f_j)$ bei der Division durch $f_1, \ldots, f_m$ Rest Null hat.

Beweis: Als R-Linearkombination von f_i und f_j liegt das S-Polynom $S(f_i, f_j)$ im Ideal I; falls $f_1, \ldots, f_m$ eine GRÖBNER-Basis von I ist, hat es also Rest Null bei der Division durch $f_1, \ldots, f_m$.

Umgekehrt sei $f_1, \ldots, f_m$ ein Erzeugendensystem von $I \lhd R$ mit der Eigenschaft, daß alle $S(f_i, f_j)$ bei der Division durch $f_1, \ldots, f_m$ (in irgendeiner Reihenfolge) Divisionsrest Null haben. Wir wollen zeigen, daß $f_1, \ldots, f_m$ dann eine GRÖBNER-Basis ist, daß also die führenden Monome $\mathrm{FM}(f_1), \ldots, \mathrm{FM}(f_m)$ das Ideal $\mathrm{FM}(I)$ erzeugen.

Sei also $f \in I$ ein beliebiges Element; wir müssen zeigen, daß FM(f) im von den FM(f_i) erzeugten Ideal liegt.

Da f in I liegt, gibt es eine Darstellung

$$f = h_1 f_1 + \cdots + h_m f_m \quad \text{mit} \quad h_i \in R\,.$$

Falls sich hier bei den führenden Termen nichts wegkürzt, ist der führende Term von f die Summe der führenden Terme gewisser Produkte $h_i f_i$, die allesamt dasselbe führende Monom FM(f) haben. Wegen FM($h_i f_i$) = FM(h_i) FM(f_i) liegt FM(f) daher im von den FM(f_i) erzeugten Ideal.

Falls sich die maximalen unter den führenden Termen FT($h_i f_i$) gegenseitig wegkürzen, läßt sich die entsprechende Teilsumme der $h_i f_i$ nach dem vorigen Lemma auch als eine Summe von S-Polynomen schreiben. Diese wiederum lassen sich nach Voraussetzung durch den Divisionsalgorithmus als Linearkombinationen der f_i darstellen. Damit erhalten wir eine neue Darstellung

$$f = \widetilde{h}_1 f_1 + \cdots + \widetilde{h}_m f_m \quad \text{mit} \quad \widetilde{h}_i \in R\,,$$

in der der maximale Multigrad eines Summanden echt kleiner ist als in der obigen Darstellung, denn in der Darstellung als Summe von S-Polynomen sind die Terme mit dem maximalem Multigrad verschwunden.

Mit dieser Darstellung können wir wie oben argumentieren: Falls sich bei den führenden Termen nichts wegkürzt, haben wir FM(f) als Element des von den FM(f_i) erzeugten Ideals dargestellt, andernfalls erhalten wir wieder via S-Polynome und deren Reduktion eine neue Darstellung von f als Linearkombination der f_i mit noch kleinerem maximalem Multigrad der Summanden, und so weiter. Das Verfahren muß schließlich mit einer Summe ohne Kürzungen bei den führenden Termen enden, da es nach der Wohlordnungseigenschaft einer Monomordnung keine unendliche absteigende Folge von Multigraden geben kann. ∎

Der BUCHBERGER-Algorithmus in seiner einfachsten Form macht aus diesem Satz ein Verfahren zur Berechnung einer GRÖBNER-Basis aus einem vorgegebenen Erzeugendensystem eines Ideals:

Gegeben sind m Elemente $f_1, \ldots, f_m \in R = k[X_1, \ldots, X_n]$.

Berechnet wird eine GRÖBNER-Basis $g_1, \ldots, g_r$ des davon erzeugten Ideals $I = (f_1, \ldots, f_m)$ mit $g_i = f_i$ für $i \leq m$.

1. Schritt (Initialisierung): Setze $g_i = f_i$ für $i = 1, \ldots, m$; die Menge $\{g_1, \ldots, g_m\}$ werde mit G bezeichnet.

2. Schritt: Setze $G' = G$ und teste für jedes Paar $(f, g) \in G' \times G'$ mit $f \neq g$, ob der Rest r bei der Division von $S(f, g)$ durch die Elemente von G' (in irgendeiner Reihenfolge angeordnet) verschwindet. Falls nicht, wird G ersetzt durch $G \cup \{r\}$.

3. Schritt: Ist $G = G'$, so endet der Algorithmus mit G als Ergebnis; andernfalls geht es zurück zum zweiten Schritt.

Wenn der Algorithmus im dritten Schritt endet, ist der Rest bei der Division von $S(f, g)$ durch die Elemente von G stets das Nullpolynom; nach dem gerade bewiesenen Satz ist G daher eine GRÖBNER-Basis. Da sowohl die S-Polynome als auch ihre Divisionsreste in I liegen und G ein Erzeugendensystem von I enthält, ist auch klar, daß es sich dabei um eine GRÖBNER-Basis von I handelt. Wir müssen uns daher nur noch überlegen, daß der Algorithmus nach endlich vielen Iterationen abbricht.

Wenn im zweiten Schritt ein nichtverschwindender Divisionsrest r auftaucht, ist dessen führendes Monom durch kein führendes Monom eines Polynoms $g \in G$ teilbar. Das von den führenden Monomen der $g \in G$ erzeugte Ideal von R wird daher größer, nachdem G um r erweitert wurde. Wenn dies unbeschränkt möglich wäre, erhielten wir daher eine unendliche aufsteigende Folge von monomialen Idealen J_i, von denen jedes echt größer ist als sein Vorgänger:

$$J_1 < J_2 < \cdots J_i < J_{i+1} < \cdots .$$

Natürlich ist auch die Vereinigung J aller J_i ein monomiales Ideal, hat also nach dem Lemma von DICKSON ein endliches Erzeugendensystem $\{M_1, \ldots, M_q\}$. Da jedes M_j in einem J_i und damit auch in allen folgenden liegen muß, gibt es ein m, so daß alle M_j in J_m liegen. Damit ist $J = (M_1, \ldots, M_q) \subseteq J_m$, im Widerspruch zur Annahme, daß J_{m+1} und damit auch J echt größer als M_j ist.

Der Algorithmus kann natürlich auf mehrere offensichtliche Weisen optimiert werden: Beispielsweise stößt man beim wiederholten Durchlaufen des zweiten Schritts immer wieder auf dieselben S-Polynome, die daher nicht jedes Mal neu berechnet werden müssen, und wenn eines dieser Polynome einmal Divisionsrest Null hatte, hat es auch bei jedem weiteren Durchgang Divisionsrest Null, denn dann wird ja wieder durch dieselben Polynome (plus einiger neuer) dividiert. Es gibt inzwischen auch zahlreiche nicht offensichtliche Verbesserungen und Optimierungen; wir wollen uns aber mit dem Prinzip begnügen und stattdessen später noch einige andere Themen behandeln.

Der BUCHBERGER-Algorithmus hat den Nachteil, daß er das vorgegebene Erzeugendensystem in jedem Schritt größer macht ohne je ein Element zu streichen. Dies ist weder beim GAUSS-Algorithmus noch beim EUKLIDischen Algorithmus der Fall, bei denen jeweils eine Gleichung durch eine andere *ersetzt* wird. Obwohl wir sowohl die Eliminationsschritte des GAUSS-Algorithmus als auch die einzelnen Schritte der Polynomdivisionen beim EUKLIDischen Algorithmus durch S-Polynome ausdrücken können, *müssen* wir im allgemeinen Fall zusätzlich zu g und $S(f,g)$ auch noch das Polynom f beibehalten; andernfalls kann sich die Lösungsmenge ändern:

Als Beispiel können wir das Gleichungssystem

$$f(X,Y) = X^2Y + XY^2 + 1 = 0 \quad \text{und} \quad g(X,Y) = X^3 - XY - Y = 0$$

betrachten. Wenn wir mit der lexikographischen Ordnung arbeiten, sind hier die einzelnen Monome bereits der Größe nach geordnet, insbesondere stehen also die führenden Monome an erster Stelle und

$$S(f,g) = Xf(X,Y) - Yg(X,Y) = X^2Y^2 + XY^2 + X + Y^2 .$$

Der führende Term X^2Y^2 ist durch den führenden Term X^2Y von f teilbar; subtrahieren wir Yf vom S-Polynom, erhalten wir das nicht weiter reduzierbare Polynom

$$h(X,Y) = -XY^3 + XY^2 + X + Y^2 - Y .$$

Sowohl $g(X,Y)$ als auch $h(X,Y)$ verschwinden im Punkt $(0,0)$; dieser ist aber keine Lösung des Ausgangssystems, da $f(0,0) = 1$ nicht verschwindet.

Aus diesem Grund werden die nach dem BUCHBERGER-Algorithmus berechneten GRÖBNER-Basen oft sehr groß und unhandlich. Betrachten wir dazu als Beispiel das System aus den beiden Gleichungen

$$f_1 = X^3 - 2XY \quad \text{und} \quad f_2 = X^2Y - 2Y^2 + X$$

und berechnen eine GRÖBNER-Basis bezüglich der graduiert lexikographischen Ordnung.

$$S(f_1, f_2) = Yf_1 - Xf_2 = -X^2$$

ist weder durch den führenden Term von f_1 noch den von f_2 teilbar, muß also als neues Element f_3 in die Basis aufgenommen werden.

$$S(f_1, f_3) = f_1 + Xf_3 = -2XY$$

kann wieder mit keinem der f_i reduziert werden, muß also als neues Element f_4 in die Basis. Genauso ist es mit

$$f_5 = S(f_2, f_3) = f_2 + Yf_3 = -2Y^2 + X\,.$$

Für das so erweiterte Erzeugendensystem, bestehend aus den Polynomen

$$f_1 = X^3 - 2XY, \quad f_2 = X^2Y - 2Y^2 + X, \quad f_3 = -X^2,$$
$$f_4 = -2XY \quad \text{und} \quad f_5 = -2Y^2 + X\,,$$

sind die S-Polynome

$$S(f_1, f_2) = f_3, \quad S(f_1, f_3) = f_4 \quad \text{und} \quad S(f_2, f_3) = f_5$$

trivialerweise auf Null reduzierbar, die anderen Kombinationen müssen wir nachrechnen:

$$S(f_1, f_4) = Yf_1 + \frac{X^2}{2}f_4 = -2XY^2 = Yf_4$$

$$S(f_1, f_5) = Y^2f_1 + \frac{X^3}{2}f_5 = -2XY^3 + \frac{X^4}{2} = \frac{X}{2}f_1 + f_2 + Y^2f_4 - f_5$$

$$S(f_2, f_4) = f_2 + \frac{X}{2}f_4 = -2Y^2 + X = f_5$$

$$S(f_2, f_5) = Yf_2 + \frac{X^2}{2}f_5 = \frac{X^3}{2} + XY - 2Y^3 = \frac{1}{2}f_1 - \frac{1}{2}f_4 + Yf_5$$

$$S(f_3, f_4) = -Yf_3 - \frac{X}{2}f_4 = 0$$

$$S(f_3, f_5) = -Y^2 f_3 - \frac{X^2}{2}f_5 = \frac{1}{2}f_1 - \frac{1}{2}f_4$$

$$S(f_4, f_5) = -\frac{Y}{2}f_4 - \frac{X}{2}f_5 = \frac{X^2}{2} = -\frac{1}{2}f_3$$

Somit bilden diese fünf Polynome eine GRÖBNER-Basis des von f_1 und f_2 erzeugten Ideals.

Zum Glück brauchen wir aber nicht alle fünf Polynome. Das folgende Lemma gibt ein Kriterium, wann man auf ein Erzeugendes verzichten kann, und illustriert gleichzeitig das allgemeine Prinzip, wonach bei einer GRÖBNER-Basis alle wichtigen Eigenschaften anhand der führenden Termen ablesbar sein sollten:

Lemma: G sei eine GRÖBNER-Basis des Ideals $I \lhd k[X_1, \ldots, X_n]$, und $g \in G$ sei ein Polynom, dessen führendes Monom im von den führenden Monomen der restlichen Basiselemente erzeugten monomialen Ideal liegt. Dann ist auch $G \setminus \{g\}$ eine GRÖBNER-Basis von I.

Beweis: $G \setminus \{g\}$ ist nach Definition genau dann eine GRÖBNER-Basis von I, wenn die führenden Terme der Basiselemente das Ideal FM(I) erzeugen. Da G eine GRÖBNER-Basis von I ist und die führenden Terme egal ob mit oder ohne FT(g) dasselbe monomiale Ideal erzeugen, ist das klar. ∎

Man beachte, daß sich dieses Lemma nur anwenden läßt, wenn G eine GRÖBNER-Basis von I ist; wir können nicht schon während des Rechengangs im BUCHBERGER-Algorithmus Elemente streichen. Im obigen Beispiel etwa wird das Ideal $I = (f_1, f_2)$ natürlich auch erzeugt von f_1, f_2 und f_3; dabei ist FM(f_1) = X^3, FM(f_2) = X^2Y, und FM(f_3) = X^2 teilt beide dieser Monome. Wenn das Lemma auf die Basis f_1, f_2, f_3 anwendbar wäre, könnten wir also f_1 und f_2 streichen und f_3 wäre für sich allein eine GRÖBNER-Basis von I. Natürlich ist aber $I \neq (-X^2)$, denn weder f_1 noch f_2 sind Vielfache von X^2.

Von der Menge $\{f_1, f_2, f_3, f_4, f_5\}$ haben wir mit Hilfe des Kriteriums von BUCHBERGER verifiziert, daß sie eine GRÖBNER-Basis von I ist; deshalb können wir das Lemma darauf anwenden und f_1, f_2 streichen. Wir können das aber erst jetzt tun, denn im Verlauf der Berechnungen wurden f_1 und f_2 noch gebraucht um $f_4 = S(f_1, f_3)$ und $f_5 = S(f_2, f_3)$ zu konstruieren. Somit ist $I = (f_3, f_4, f_5)$, und darauf können wir das Lemma nicht weiter anwenden, denn

$$\mathrm{FM}(f_3) = X^2, \quad \mathrm{FM}(f_4) = XY \quad \text{und} \quad \mathrm{FM}(f_5) = Y^2,$$

und keines dieser drei Monome ist Vielfaches eines der anderen.

Zur weiteren Normierung können wir noch durch die führenden Koeffizienten teilen und erhalten dann die *minimale* GRÖBNER-Basis

$$\widetilde{f_3} = X^2, \quad \widetilde{f_4} = XY \quad \text{und} \quad \widetilde{f_5} = Y^2 - \frac{X}{2}.$$

Definition: Eine *minimale* GRÖBNER-Basis von I ist eine GRÖBNER-Basis von I mit folgenden Eigenschaften:

1.) Alle $g \in G$ haben den führenden Koeffizienten eins

2.) Für kein $g \in G$ liegt $\mathrm{FM}(g)$ im von den führenden Monomen der übrigen Elemente erzeugten Ideal.

Da ein Monom X^α genau dann im von einer Menge M von Monomen erzeugten Ideal liegt, wenn es durch eines dieser Monome teilbar ist, können wir die zweite Bedingung auch so ausdrücken, daß es keine zwei Elemente $g \neq g'$ in G geben darf, für die $\mathrm{FM}(g)$ ein Teiler von $\mathrm{FM}(g')$ ist.

Es ist klar, daß jede GRÖBNER-Basis zu einer minimalen GRÖBNER-Basis verkleinert werden kann: Durch Division können wir alle führenden Koeffizienten zu eins machen ohne etwas an der Erzeugung zu ändern, und nach obigem Lemma können wir nacheinander alle Elemente eliminieren, die die zweite Bedingung verletzen.

Wir können aber noch mehr erreichen: Wenn nicht das führende, sondern einfach *irgendein* Monom eines Polynoms $g \in G$ im von den führenden Termen der übrigen Elemente erzeugten Ideal liegt, ist dieses Monom teilbar durch das führende Monom eines anderen Polynoms $h \in G$. Wir

können den Term mit diesem Monom daher zum Verschwinden bringen, indem wir g ersetzen durch g minus ein Vielfaches von h. Da sich dabei nichts an den führenden Termen der Elemente von G ändert, bleibt G eine GRÖBNER-Basis. Wir können somit aus den Elementen einer minimalen GRÖBNER-Basis Terme eliminieren, die durch den führenden Term eines anderen Elements teilbar sind. Was dabei schließlich entstehen sollte, ist eine *reduzierte* GRÖBNER-Basis:

Definition: Eine reduzierte GRÖBNER-Basis von I ist eine GRÖBNER-Basis von I mit folgenden Eigenschaften:
1.) Alle $g \in G$ haben den führenden Koeffizienten eins
2.) Für kein $g \in G$ liegt ein Monom von g im von den führenden Monomen der übrigen Elemente erzeugten Ideal.

Die minimale Basis im obigen Beispiel ist offenbar schon reduziert, denn außer $\widetilde{f_5}$ bestehen alle Basispolynome nur aus dem führendem Term, und bei $\widetilde{f_5}$ ist der zusätzliche Term linear, kann also nicht durch die quadratischen führenden Monome der anderen Polynome teilbar sein.

Reduzierte GRÖBNER-Basis haben eine für das praktische Rechnen mit Idealen sehr wichtige zusätzliche Eigenschaft:

Satz: Jedes Ideal $I \lhd k[X_1, \ldots, X_n]$ hat (bei vorgegebener Monomordnung) eine eindeutig bestimmte reduzierte GRÖBNER-Basis.

Beweis: Wir gehen aus von einer minimalen GRÖBNER-Basis G und ersetzen nacheinander jedes Element $g \in G$ durch seinen Rest bei der Polynomdivision durch $G \setminus \{g\}$. Da bei einer minimalen GRÖBNER-Basis kein führendes Monom eines Element das führende Monom eines anderen teilen kann, ändert sich dabei nichts an den führenden Termen, G ist also auch nach der Ersetzung eine minimale GRÖBNER-Basis. In der schließlich entstehenden Basis hat kein $g \in G$ mehr einen Term, der durch den führenden Term eines Elements von $G \setminus \{g\}$ teilbar wäre, denn auch wenn wir bei der Reduktion der einzelnen Elemente durch eine eventuell andere Menge geteilt haben, hat sich doch an den führenden Termen der Basiselemente nichts geändert. Also gibt es eine reduzierte GRÖBNER-Basis.

Nun seien G und G' zwei reduzierte GRÖBNER-Basen von I. Jedes Element $f \in G'$ liegt insbesondere in I, also ist $\overline{f}^G = 0$. Insbesondere muß der führende Term von f durch den führenden Term eines $g \in G$ teilbar sein. Umgekehrt ist aber auch $\overline{g}^{G'} = 0$, d.h. der führende Term von g muß durch den führenden Term eines Elements von $f' \in G'$ teilbar sein. Dieser führende Term teilt dann insbesondere den führenden Term von f, und da G' als reduzierte GRÖBNER-Basis minimal ist, muß $f' = f$ sein. Somit gibt es zu jedem $g \in G$ genau ein $f \in G'$ mit FM(f) = FM(g); insbesondere haben G und G' dieselbe Elementanzahl. Tatsächlich muß sogar $f = g$ sein, denn $f - g$ liegt in I, enthält aber keine Term, der durch den führenden Term irgendeines Elements von G teilbar wäre. Also ist $f - g = 0$.

■

Bemerkung: Die Forderung in den Definitionen von minimalen und reduzierten GRÖBNER-Basen, daß alle führenden Koeffizienten eins sein müssen, ist zwar nützlich für theoretische Diskussionen, führt aber im Falle von Polynomen mit rationalen Koeffizienten oft dazu, daß die Koeffizienten Nenner haben. Computeralgebrasysteme können zwar mit rationalen Zahlen rechnen, indem sie diese durch Paare teilerfremder ganzer Zahlen darstellen, aber diese Rechnungen sind erheblich aufwendiger als solche mit ganzen Zahlen. Daher liefern einige Computeralgebrasysteme beim Kommando zur Berechnung einer reduzierten GRÖBNER-Basis anstelle von Polynomen mit führendem Koeffizienten eins solche mit teilerfremden ganzzahligen Koeffizienten.

Kapitel 2
Systeme von nichtlinearen Polynomgleichungen

GRÖBNER-Basen haben eine Vielzahl von Anwendungen in der Algebra; wir wollen uns hier vor allem damit beschäftigen, wie sie direkt oder im Zusammenspiel mit anderen Methoden zur expliziten Lösung nichtlinearer Gleichungssysteme führen können. Explizit angebbar sind die Lösungen meist nur, wenn die Lösungsmenge endlich ist; daher werden wir uns meist auf solche Systeme beschränken und interessieren uns daher auch für Kriterien, wie wir einem Gleichungssystem die Endlichkeit seiner Lösungsmenge ansehen können.

§1: Gröbner-Basen für nichtlineare Gleichungssysteme

Wir gehen aus von m Polynomgleichungen

$$f_i(x_1, \ldots, x_n) = 0 \quad \text{mit} \quad f_i \in k[X_1, \ldots, X_n] \quad \text{für} \quad i = 1, \ldots, m$$

und suchen die Lösungsmenge

$$\left\{(x_1, \ldots, x_n) \in k^n \;\middle|\; f_i(x_1, \ldots, x_n) = 0 \text{ für } i = 1, \ldots, m\right\}.$$

Diese wird allerdings oft leer sein; für $f_1 = X^2 - 2$ und $f_2 = Y^2 - 3$ aus $\mathbb{Q}[X]$ etwa ist diese Menge leer, da die Lösungen $(\pm\sqrt{2}, \pm\sqrt{3})$ nicht in $\mathbb{Q}^2$ liegen. Wir betrachten daher meist noch einen zweiten Körper K, der k enthält, und interessieren uns allgemeiner für die Lösungsmenge in K^n:

Definition: *a)* Ist I ein Ideal in $k[X_1, \ldots, X_n]$, und ist K ein Körper, der k enthält, setzen wir

$$V_K(I) = \left\{(x_1, \ldots, x_n) \in K^n \;\middle|\; f(x_1, \ldots, x_n) = 0 \text{ für alle } f \in I\right\}.$$

b) Für $I = (f_1, \cdots, f_m)$ schreiben wir auch kurz $V_K(f_1, \ldots, f_m)$ an Stelle von $V_K(I)$.

Der Körper k sollte dabei möglichst klein sein, denn mit den Elementen dieses Körpers müssen wir rechnen, und je größer der Körper, desto aufwendiger sind seine Rechenoperationen. In konkreten Beispielen werden wir uns meist auf $k = \mathbb{Q}$ beschränken und – soweit möglich – sogar versuchen, unsere Konstruktionen in $\mathbb{Z}[X]$ durchzuführen.

Der Körper K hingegen sollte so groß sein, daß er für ein Gleichungssystem, daß in irgendeinem Körper eine nichtleere endliche Lösungsmenge hat, diese Lösungsmenge enthält. Wir werden meist $K = \mathbb{C}$ betrachten.

Wie wir bereits aus §1 des vorigen Kapitels wissen, hängt die Lösungsmenge des Gleichungssystems nur ab vom Ideal $I = (f_1, \ldots, f_m)$; wir suchen ein Erzeugendensystem $\{g_1, \ldots, g_r\}$ dieses Ideals, aus dem wir mehr über die Mengen

$$V_K(I) = V_K(f_1, \ldots, f_m) = V_K(g_1, \ldots, g_r)$$

ablesen können. Wir erwarten natürlich, daß wir hier vor allem im Falle einer geeigneten GRÖBNER-Basis $\{g_1, \ldots, g_r\}$ eventuell Erfolg haben.

Viele Lösungsansätze für Gleichungssysteme in mehreren Veränderlichen beruhen auf der Elimination von Variablen: Im ℓ-ten Schritt suchen wir nach Bedingungen, die ein $(n - \ell)$-Tupel $(x_{\ell+1}, \ldots, x_n)$ erfüllen muß, wenn es ein ℓ-Tupel $(x_1, \ldots, x_\ell)$ gibt, so daß $(x_1, \ldots, x_n)$ in $V(I)$ liegt. Eine solche Bedingung ist trivial: Für jedes Polynom $f \in I$, in dem die Variablen $X_1, \ldots, X_\ell$ nicht vorkommen, muß $f(x_{\ell+1}, \ldots, x_n) = 0$ sein.

Definition: *a)* Das ℓ-te *Eliminationsideal* eines Ideal $I \lhd k[X_1, \ldots, X_n]$ ist $I_\ell = I \cap k[X_{\ell+1}, \ldots, X_n]$.
b) Eine Monomordnung $<$ heißt *Eliminationsordnung* für $X_1, \ldots, X_\ell$, wenn jedes Monom, das mindestens eine der Variablen $X_1, \ldots, X_\ell$ enthält, größer ist als alle Monome, die nur $X_{\ell+1}, \ldots, X_n$ enthalten.

Die lexikographische Ordnung mit $X_1 > X_2 > \cdots > X_{n-1} > X_n$ ist offensichtlich für jedes ℓ eine Eliminationsordnung für $X_1, \ldots, X_\ell$, die

graduiert lexikographische aber nicht, da bezüglich dieser beispielsweise $X_1 < X_n^2$ ist.

Satz: Ist G eine GRÖBNER-Basis von I bezüglich einer Eliminationsordnung für $X_1, \ldots, X_\ell$, so ist $G \cap I_\ell$ eine GRÖBNER-Basis von I_ℓ.

Beweis: Die Elemente von $G = \{g_1, \ldots, g_m\}$ seien so angeordnet, daß $G \cap I_\ell = \{g_1, \ldots, g_r\}$ ist. Wir müssen zeigen, daß sich jedes $f \in I_\ell$ als Linearkombination von $g_1, \ldots, g_r$ mit Koeffizienten aus $k[X_{\ell+1}, \ldots, X_n]$ darstellen läßt.

Der Divisionsalgorithmus bezüglich der lexikographischen Ordnung gibt uns eine Darstellung $f = h_1 g_1 + \cdots + h_m g_m$ von f als Element von I. Die Polynome $g_{r+1}, \ldots, g_m$ enthalten jeweils mindestens eine der Variablen $X_1, \ldots, X_\ell$, und da wir eine Eliminationsordnung verwenden, muß auch das führende Monom eine dieser Variablen enthalten. Da kein Monom von f eine dieser Variablen enthält, kann im Divisionsalgorithmus das führende Monom eines dieser Polynome nie Teiler des führenden Monoms des jeweils betrachteten Polynoms p sein, Somit ist $h_{r+1} = \cdots = h_m = 0$, und in keinem der Polynome $h_1, \ldots, h_r$ kann eine der Variablen $X_1, \ldots, X_\ell$ auftreten. Dies zeigt, daß f im von $g_1, \ldots, g_r$ erzeugten Ideal von $k[X_{\ell+1}, \ldots, X_n]$ liegt, d.h. dieses Ideal wird von $g_1, \ldots, g_r$ erzeugt.

Um zu zeigen, daß es sich dabei sogar um eine GRÖBNER-Basis handelt, können wir zum Beispiel zeigen, daß alle $S(g_i, g_j)$ mit $i, j \leq r$ ohne Rest durch $g_1, \ldots, g_r$ teilbar sind. Da G nach Voraussetzung eine GRÖBNER-Basis ist, sind sie auf jeden Fall ohne Rest durch G teilbar, und wieder kann bei der Division nie der führende Term eines Dividenden durch den eines g_i mit $i > r$ teilbar sein, d.h. $S(g_i, g_j)$ ist als Linearkombination von $g_1, \ldots, g_r$ mit Koeffizienten aus $k[g_1, \ldots, g_r]$ darstellbar. ∎

Daraus ergibt sich eine Strategie zur Lösung nichtlinearer Gleichungssysteme nach Art des GAUSS-Algorithmus: Wir gehen aus von der lexikographischen Ordnung, die ja für jedes ℓ eine Eliminationsordnung für $X_1, \ldots, X_\ell$ ist, und bestimmen eine (reduzierte) GRÖBNER-Basis für das von den Gleichungen erzeugte Ideal des Polynomrings $k[X_1, \ldots, X_n]$.

Dann betrachten als erstes das Eliminationsideal I_{n-1}. Dieses besteht nur aus Polynomen in X_n; falls wir mit einer reduzierten GRÖBNER-Basis arbeiten, gibt es darin höchstens ein solches Polynom.

Falls es ein solches Polynom gibt, muß jede Lösung des Gleichungssystem als letzte Komponente eine von dessen Nullstellen haben. Wir bestimmen daher diese Nullstellen (in K) und setzen sie nacheinander in das restliche Gleichungssystem ein. Dadurch erhalten wir Gleichungssysteme in $n-1$ Unbekannten, wo wir nach Gleichungen nur in X_{n-1} suchen können. Diese erhalten wir, indem wir bei allen Erzeugenden des Eliminationsideals I_{n-2} für X_n nacheinander die Werte aus $V_K(I_{n-1}) \subset k$ einsetzen. Nachdem wir so $V_K(I_{n-2}) \subset K^2$ bestimmt haben, können wir analog die Mengen $V_K(I_{n-3}) \subset K^3$ und so weiter bis $V_K(I) \subset K^n$ bestimmen.

Betrachten wir noch einmal das Beispiel gegen Ende von §5 des vorigen Kapitels mit

$$f_1 = X^3 - 2XY \quad \text{und} \quad f_2 = X^2Y - 2Y^2 + X\,.$$

Dort hatten wir die reduzierte GRÖBNER-Basis bezüglich der graduiert lexikographischen Ordnung berechnet; sie besteht aus

$$g_1 = X^2, \quad g_2 = XY \quad \text{und} \quad g_3 = Y^2 - \frac{X}{2}\,.$$

Da die graduiert lexikographische Ordnung keine Eliminationsordnung für X ist, können wir nicht erwarten, daß $\{g_1, g_2, g_3\} \cap k[Y]$ ein Erzeugendensystem des Eliminationsideals $(f_1, f_2) \cap k[Y]$ liefert, und in der Tat liegt keines der g_i in $k[Y]$. Zufälligerweise liegt aber $g_1 = X^2$ in $k[X]$, wir wissen also, daß für jede Lösung (x, y) des Gleichungssystems $x = 0$ sein muß. $g_2 = XY$ verschwindet für alle solche Punkte automatisch, und $g_3 = Y^2 - X/2$ verschwindet genau dann, wenn auch $y = 0$ ist. Somit ist $V(f_1, f_2) = \{(0, 0)\}$.

Wenn wir das Gleichungssystem mit dem hier vorgestellten Verfahren lösen wollen. müssen wir mit der lexikographischen Ordnung arbeiten. Da die führenden Terme von f_1 und f_2 bei beiden Ordnungen gleich sind und viele der zu berechnenden S-Polynome nur aus einem Term

bestehen, ändert sich zunächst nichts: Wie bei der graduiert lexikographischen Ordnung kommen wir auf

$$f_3 = S(f_1, f_2) = -X^2, \quad f_4 = S(f_1, f_3) = -2XY \quad \text{und}$$
$$f_5 = S(f_2, f_3) = X - 2Y^2 .$$

Auch $S(f_1, f_4) = -2XY^2 = Y f_4$ kann wie dort auf Null reduziert werden, bei der Berechnung von $S(f_1, f_5)$ ist jetzt aber nicht mehr Y^2, sondern X das führende Monom. Somit ist

$$S(f_1, f_5) = f_1 - X^2 f_5 = 2X^2Y^2 - 2XY = 2Y f_2 + 2f_4 + 4Y^3 ,$$

das S-Polynom läßt sich also modulo $\{f_1, f_2, f_3, f_4, f_5\}$ nicht auf Null reduzieren und wir müssen $f_6 = 4Y^3$ als neues Element in die Basis aufnehmen. Erst jetzt zeigt eine mühsame Rechnung, die man am besten seinem Computer überläßt, daß $S(f_i, f_j)$ für alle $1 \leq i < j \leq 6$ modulo $\{f_1, f_2, f_3, f_4, f_5, f_6\}$ auf Null reduziert werden kann, womit wir eine GRÖBNER-Basis gefunden haben.

Die führenden Monome der sechs Basiselemente bezüglich der lexikographischen Ordnung sind

$$\begin{array}{lll} \mathrm{FM}(f_1) = X^3 , & \mathrm{FM}(f_2) = X^2Y , & \mathrm{FM}(f_3) = -X^2 , \\ \mathrm{FM}(f_4) = -2XY , & \mathrm{FM}(f_5) = X, , & \mathrm{FM}(f_6) = 4Y^3 ; \end{array}$$

wir können also f_1 bis f_4 eliminieren. Die reduzierte GRÖBNER-Basis bedeutet besteht somit aus $g_1 = X - 2Y^2$ und $g_2 = Y^3$.

Das Eliminationsideal I_1 wird daher erzeugt von $g_2 = Y^3$, d.h. für jede Lösung (x, y) muß y verschwinden. Setzen wir $y = 0$ in g_1 ein, so sehen wir, daß auch x verschwinden muß, der Nullpunkt ist also die einzige Lösung.

Es war ein Zufall, daß wir dieses Ergebnis auch der GRÖBNER-Basis bezüglich der graduiert lexikographischen Ordnung ansehen konnten; bei komplizierteren Systemen wird dort oft jedes Basiselement alle Variablen enthalten, so daß wir nichts sehen können. Trotzdem kann die graduiert lexikographische Ordnung zur Lösung nichtlinearer Gleichungssysteme nützlich sein: 1993 publizierten J.C. FAUGÈRE, P. GIANNI, D. LAZARD und T. MORA einen heute nach ihren Anfangsbuchstaben

als FGLM benannten Algorithmus, der für ein Ideal I mit endlicher Nullstellenmenge $V(I)$ effizient eine GRÖBNER-Basis bezüglich der lexikographischen Ordnung bestimmt auf dem Umweg über die graduiert lexikographische Ordnung. Wir werden später sehen, daß wir im Falle einer endlichen Lösungsmenge diese auch ausgehend von einer beliebigen GRÖBNER-Basis mit alternativen Techniken bestimmen können.

Nun kann es beim obigen Verfahren für nichtlineare Gleichungssysteme natürlich vorkommen, daß I_{n-1} das Nullideal ist; falls unter den Lösungen des Systems unendlich viele Werte für die letzte Variable vorkommen, muß das sogar so sein. Es kann sogar vorkommen, daß *alle* Eliminationsideale außer $I_0 = I$ das Nullideal sind. In diesem Fall führt die gerade skizzierte Vorgehensweise zu nichts.

Bevor wir uns darüber wundern, sollten wir uns überlegen, was wir überhaupt unter der Lösung eines nichtlinearen Gleichungssystems verstehen wollen. Im Falle einer endlichen Lösungsmenge ist das klar: Dann wollen wir eine Auflistung der sämtlichen Lösungstupel. Bei einer unendlichen Lösungsmenge ist das aber nicht mehr möglich. Im Falle eines linearen Gleichungssystems wissen wir, daß die Lösungsmenge ein affiner Raum ist; wir können sie daher auch wenn sie unendlich sein sollte durch endlich viele Daten eindeutig beschreiben, zum Beispiel durch eine spezielle Lösung und eine Basis des Lösungsraums des zugehörigen homogenen Gleichungssystems.

Bei nichtlinearen Gleichungssystemen gibt es im allgemeinen keine solche Beschreibung unendlicher Lösungsmengen: Die Lösungsmenge des Gleichungssystems

$$X^2 + 2Y^2 + 3Z^2 = 100 \quad \text{und} \quad 2X^2 + 3Y^2 - Z^2 = 0$$

etwa ist die Schnittmenge eines Ellipsoids mit einem elliptischen Kegel; sie besteht aus zwei ovalen Kurven höherer Ordnung. Die GRÖBNER-Basis besteht in diesem Fall aus den beiden Polynomen

$$X^2 - 11Z^2 + 300 \quad \text{und} \quad Y^2 + 7Z^2 - 200\,,$$

stellt uns dieselbe Menge also dar als Schnitt eines hyperbolischen und eines elliptischen Zylinders. Eine explizitere Beschreibung der Lösungsmenge ist schwer vorstellbar.

Auf der Basis von STURMschen Ketten, dem Lemma von THOM und Verallgemeinerungen davon hat die semialgebraische Geometrie Methoden entwickelt, wie man auch allgemeinere Lösungsmengen nichtlinearer Gleichungssysteme durch eine sogenannte zylindrische Zerlegung qualitativ beschreiben kann; dazu wird der $\mathbb{R}^n$ in Teilmengen zerlegt, in denen die Lösungsmenge entweder ein einfaches qualitatives Verhalten hat oder aber leeren Durchschnitt mit der Teilmenge. Dadurch kann man insbesondere feststellen, in welchen Regionen des $\mathbb{R}^n$ Lösungen zu finden sind; diese Methoden sind Gegenstand der reell-algebraischen Geometrie.

In manchen Fällen lassen sich Lösungsmengen parametrisieren; wie man mit Methoden der algebraischen Geometrie zeigen kann, ist das aber im allgemeinen nur bei Gleichungen kleinen Grades der Fall und kommt daher für allgemeine Lösungsalgorithmen nicht in Frage.

Stets möglich ist das umgekehrte Problem, d.h. die Beschreibung einer parametrisch gegebenen Menge in impliziter Form. Hier gehen wir aus von Gleichungen der Form

$$x_1 = \varphi_1(t_1, \ldots, t_m), \quad \ldots, \quad x_n = \varphi_n(t_1, \ldots, t_m)\,,$$

und wir suchen Polynome $f_1, \ldots, f_r$ aus $k[X_1, \ldots, X_n]$, die auf der Menge aller jener $(x_1, \ldots, x_n)$ verschwinden, für die es eine solche Darstellung gibt (und eventuell noch auf Grenzwerten davon).

Dazu wählen wir eine lexikographische Ordnung auf dem Polynomring $k[T_1, \ldots, T_m, X_1, \ldots, X_n]$, bei der alle T_i größer sind als die X_j, und bestimmen eine GRÖBNER-Basis für das von den Polynomen $X_i - \varphi_i(T_1, \ldots, T_m)$ erzeugte Ideal. Dessen Schnitt mit $k[X_1, \ldots, X_n]$ ist ein Eliminationsideal, hat also als Basis genau die Polynome aus der GRÖBNER-Basis, in denen keine T_i vorkommen.

Fast genauso können wir auch zu einer vorgegebenen endlichen Menge von Punkten ein Gleichungssystem konstruieren, das genau diese Menge als Lösungsmenge hat; dies spielt beispielsweise in der algebraischen Statistik eine Rolle, wenn zu einem vorgegebenen Design die damit schätzbaren Modelle identifiziert werden sollen.

Wir gehen aus von r Punkten

$$P_i = \left(x_1^{(i)}, \ldots, x_n^{(i)}\right) \in k^n, \quad i = 1, \ldots, r,$$

und suchen ein Ideal $I \lhd k[X_1, \ldots, X_n]$, dessen Elemente genau in den Punkten P_i verschwinden. Im Falle nur eines Punktes P_i können wir einfach das Ideal

$$I_i = \left(X_1 - x_1^{(i)}, \ldots, X_n - x_n^{(i)}\right)$$

nehmen; bei mehreren Punkten brauchen wir den Durchschnitt der Ideale I_1 bis I_r, für den wir kein offensichtliches Erzeugendensystem haben.

Betrachten wir stattdessen die Punkte

$$Q_i = \left(t_1^{(i)}, \ldots, t_r^{(i)}, x_1^{(i)}, \ldots, x_n^{(i)}\right) \in k^{r+n} \quad \text{mit} \quad t_j^{(i)} = \begin{cases} 1 & \text{falls } i = j \\ 0 & \text{sonst} \end{cases},$$

so erzeugen die Polynome

$$\left(X_j - x_j^{(i)}\right) T_i \in k[T_1, \ldots, T_r, X_1, \ldots, X_n]$$

für $i = 1, \ldots, n$ und $j = 1, \ldots, r$ zusammen mit dem Polynom $T_1 + \cdots + T_r - 1$ ein Ideal, das alle Punkte Q_i als Nullstellen hat: Die Polynome $\left(X_j - x_j^{(i)}\right) T_i$ verschwinden in Q_i, da $x_j^{(i)}$ die j-te Koordinate von Q_i ist, und für $\ell \neq i$ verschwindet $\left(X_j - x_j^{(i)}\right) T_\ell$, da $t_\ell^{(i)}$ verschwindet.

Ist umgekehrt $Q = (t_1, \ldots, t_r, x_1, \ldots, x_n) \in k^{r+n}$ keiner der Punkte Q_i, so gibt es für jedes i mindestens eine Koordinate, in der sich Q von Q_i unterscheidet. Ist dies etwa die j-te Koordinate, so ist $X_j - x_j^{(i)}$ in Q von Null verschieden; $\left(X_j - x_j^{(i)}\right) T_i$ kann daher nur verschwinden, wenn $t_i = 0$ ist. Dies kann aber nicht für alle i der Fall sein, denn die Summe der t_i ist eins, da $T_1 + \cdots + T_r - 1$ verschwindet. Somit liegt Q nicht in $V(J)$.

Damit haben wir ein Ideal $J \lhd k[T_1, \ldots, T_r, X_1, \ldots, X_n]$ gefunden, dessen Nullstellen genau die Punkte $Q_1, \ldots, Q_r \in k^{r+n}$ sind. Die Punkte $P_1, \ldots, P_r$ sind die Projektionen der Q_i von k^{n+r} nach k^n; deshalb ist klar, daß alle Polynome aus

$$I \underset{\text{def}}{=} J \cap k[X_1, \ldots, X_n]$$

in den Punkten P_i verschwinden. Wir erhalten ein Erzeugendensystem dieses Ideals, indem wir bezüglich einer Eliminationsordnung für $T_1, \ldots, T_r$ eine GRÖBNER-Basis von J berechnen und davon nur die Polynome betrachten, die keine der Variablen T_i enthalten.

§2: Der Hilbertsche Nullstellensatz

Wie wir wissen, stimmen die Lösungsmengen zweier Gleichungssysteme

$$f_1(x_1, \ldots, x_n) = \cdots = f_m(x_1, \ldots, x_n) = 0$$

und

$$g_1(x_1, \ldots, x_n) = \cdots = g_p(x_1, \ldots, x_n) = 0$$

überein, wenn die Ideale $(f_1, \ldots, f_m)$ und $(g_1, \ldots, g_p)$ übereinstimmen. Umgekehrt folgt aber nicht aus der Gleichheit der Lösungsmengen, daß auch die Ideale gleich sein müssen. In diesem Paragraphen wollen wir genauer untersuchen, was hier gilt.

Als erstes müssen wir uns überlegen, *wo* wir nach Lösungen suchen: Wie wir bereits in §1 gesehen haben, wird die Lösungsmenge über dem kleinsten Körper, der alle Koeffizienten der Polynome enthält, oft leer sein, obwohl es Lösungen in größeren Körpern gibt. In den meisten Beispielen betrachten wir $k = \mathbb{Q}$ und $K = \mathbb{C}$ sein; wie wir wissen hat in $\mathbb{C}$ zumindest jedes nichtkonstante Polynom in einer Veränderlichen eine Nullstelle. Körper mit dieser Eigenschaft bezeichnen wir als *algebraisch abgeschlossen:*

Definition: Ein Körper k heißt *algebraisch abgeschlossen,* wenn jedes nichtkonstante Polynom $f \in k[X]$ mindestens eine Nullstelle in k hat.

Durch Polynomdivision folgt leicht induktiv:

Lemma: Ist k algebraisch abgeschlossen, so läßt sich jedes Polynom vom Grad d aus $k[X]$ schreiben als

$$f = c(X - x_1) \cdots (X - x_d) \quad \text{mit} \quad c \in k \setminus \{0\} \ \text{ und } \ x_1, \ldots, x_d \in k\,.$$

Die x_i müssen dabei nicht notwendigerweise verschieden sein.

■

Der Körper K soll im folgenden stets algebraisch abgeschlossen sein; zur Vereinfachung der Beweise wollen wir zusätzlich annehmen, daß er überabzählbar viele Elemente enthält. Die Sätze aus diesem Paragraphen gelten zwar auch ohne diese Zusatzvoraussetzung, jedoch erfordern die Beweise dann einen größeren Aufwand.

Sei also für den Rest dieses Paragraphen k irgendein Körper, und K sei ein algebraisch abgeschlossener Körper mit überabzählbar vielen Elementen, der k enthält.

Als erstes wollen wir uns mit der Frage beschäftigen, für welche Ideale $I \lhd k[X_1, \ldots, X_n]$ die Lösungsmenge $V_K(I)$ in K^n leer ist. Ein Beispiel ist offensichtlich: Natürlich ist $I = k[X_1, \ldots, X_n]$ ein Ideal, und da es insbesondere die Konstante eins enthält, ist $V_K(I) = \emptyset$. Eine (schwache) Form des HILBERTschen Nullstellensatzes besagt, daß dies das einzige Beispiel ist. Zur Vorbereitung des Beweises definieren wir

Definition: R sei ein Ring.
a) $I \lhd R$ ist ein *echtes* Ideal, falls $I \neq R$.
b) Ein echtes Ideal $\mathfrak{m} \lhd R$ heißt *maximales* Ideal, wenn R das einzige Ideal ist, das $\mathfrak{m}$ als echte Teilmenge enthält.
c) Ein echtes Ideal $\mathfrak{p} \lhd R$ heißt *Primideal,* wenn gilt: Liegt für zwei Elemente $f, g \in R$ das Produkt fg in $\mathfrak{p}$, so liegt mindestens einer der Faktoren f, g in $\mathfrak{p}$.

Wie aus der Zahlentheorie bekannt, teilt eine Primzahl p genau dann das Produkt zweier Zahlen a, b, wenn sie mindestens einen der beiden Faktoren teilt; in $\mathbb{Z}$ sind also die von den Primzahlen erzeugten Hauptideale Primideale. Dazu kommt wegen der Nullteilerfreiheit auch noch das Nullideal.

Durch vollständige Induktion beweist man leicht

Lemma: Ist $\mathfrak{p}$ ein Primideal und liegt ein Produkt $f_1 \cdots f_n$ von Elementen $f_i \in R$ in $\mathfrak{p}$, so liegt mindestens einer der Faktoren f_i in $\mathfrak{p}$. ∎

Lemma: Jedes maximale Ideal $\mathfrak{m} \lhd R$ ist ein Primideal.

Beweis: Das Produkt fg zweier Elemente $f, g \in R$ liege in $\mathfrak{m}$. Falls $f \in \mathfrak{m}$ sind wir fertig; andernfalls ist $\mathfrak{m}+(f) = R$ wegen der Maximalität von $\mathfrak{m}$; es gibt also Elemente $m \in \mathfrak{m}$ und $h \in R$, so daß $m + hf = 1$ ist. Damit ist $g = mg + hfg \in \mathfrak{m}$, denn $m \in \mathfrak{m}$ und $fg \in \mathfrak{m}$. ∎

Lemma: Jedes echte Ideal $I \lhd k[X_1, \ldots, X_n]$ liegt in einem maximalen Ideal $\mathfrak{m} \lhd k[X_1, \ldots, X_n]$.

Beweis: Falls I selbst maximal ist, sind wir fertig; andernfalls gibt es ein echtes Ideal I_1, das I als echte Teilmenge enthält. Auch wenn I_2 ein maximales Ideal ist, sind wir fertig; andernfalls gibt es ein echtes Ideal I_3, das I_2 als echte Teilmenge enthält, und so weiter. Wenn dieses Verfahren nach endlich vielen Schritten abbricht, haben wir ein maximales Ideal gefunden, das I enthält; andernfalls gibt es eine unendliche aufsteigende Folge von Idealen $I \subset I_1 \subset I_2 \subset \cdots$. Die Vereinigung aller I_j ist selbst ein Ideal in $k[X_1, \ldots, X_n]$ und hat damit nach dem HILBERTschen Basissatz ein endliches Erzeugendensystem $\{f_1, \ldots, f_m\}$. Jedes f_i liegt in einem der Ideale I_j und damit auch in allen I_ℓ mit $\ell > j$. Wegen der Endlichkeit des Erzeugendensystems gibt es daher einen Index r derart, daß alle f_i in I_r liegen. Dann ist aber $I = I_r = I_{r+1} = \cdots$, im Widerspruch zu der Annahme, daß jedes I_j echte Teilmenge von I_{j+1} ist. Somit bricht das Verfahren nach endlich vielen Schritten ab und liefert ein maximales Ideal $\mathfrak{m}$, in dem I enthalten ist. ∎

(Tatsächlich gilt auch dieses Lemma für beliebige Ringe; da dort der HILBERTsche Basissatz nicht gelten muß, beweist man es im allgemeinen Fall mit Hilfe des ZORNschen Lemmas.)

Schwache Form des Hilbertschen Nullstellensatzes: Für ein echtes Ideal $I \lhd k[X_1, \ldots, X_n]$ ist $V_K(I) \neq \emptyset$.

Beweis: Nach dem HILBERTschen Basissatz hat jedes Ideal I ein endliches Erzeugendensystem $\{f_1, \ldots, f_m\}$. Wir betrachten das von den f_i erzeugte Ideal $\bar{I}$ in $K[X_1, \ldots, X_n]$. Da eine Basis des k-Vektorraums $k[X_1, \ldots, X_n]/I$ auch Basis des K-Vektorraums $K[X_1, \ldots, X_n]/\bar{I}$ ist, muß auch $\bar{I}$ ein echtes Ideal von $K[X_1, \ldots, X_n]$ sein und liegt somit in

einem maximalen Ideal $\mathfrak{m} \lhd K[X_1, \ldots, K_n]$. Der Satz folgt somit aus der folgenden alternativen Version des HILBERTschen Nullstellensatzes:

Satz: Die maximalen Ideale $\mathfrak{m} \lhd K[X_1, \ldots, X_n]$ sind genau die Ideale

$$\mathfrak{m} = (X_1 - x_1, \ldots, X_n - x_n) \quad \text{mit} \quad (x_1, \ldots, x_n) \in K^n \, .$$

Beweis: $\mathfrak{m} = (X_1 - x_1, \ldots, X_n - x_n)$ ist der Kern der Abbildung

$$\begin{cases} K[X_1, \ldots, X_n] \to K \\ \qquad\qquad f \mapsto f(x_1, \ldots, x_n) \end{cases} .$$

Ist daher I ein Ideal, das $\mathfrak{m}$ echt enthält, so muß der Vektorraum $K[X_1, \ldots, X_n]/I$ ein echter Untervektorraum von $K[X_1, \ldots, X_n]/\mathfrak{m}$ sein. Da letzterer nach dem Homomorphiesatz isomorph zum eindimensionalen Vektorraum K ist, muß dies der Nullraum sein. Somit ist $I = K[X_1, \ldots, X_m]$, d.h. $\mathfrak{m}$ ist ein maximales Ideal.

Umgekehrt sei $\mathfrak{m}$ ein maximales Ideal. Wenn wir zeigen können, daß es Elemente $x_1, \ldots, x_n$ gibt, für die $X_i - x_i$ in $\mathfrak{m}$ liegt, ist $(X_1 - x_1, \ldots, X_n - x_n) \subseteq \mathfrak{m}$, und da links ein maximales Ideal steht, müssen beide Seiten gleich sein.

Angenommen, es gibt ein $i \in \{1, \ldots, n\}$, für das $X_i - x$ für kein $x \in K$ im Ideal $\mathfrak{m}$ liegt. Wegen der Maximalität von $\mathfrak{m}$ ist dann

$$\mathfrak{m} + (X - x) = K[X_1, \ldots, X_n] \quad \text{für alle } x \in K \, .$$

Somit gibt es für jedes $x \in K$ ein Polynom $f_x \in \mathfrak{m}$ sowie ein Polynom $h_x \in K[X_1, \ldots, X_n]$ derart, daß

$$f_x + h_x \cdot (X_i - x) = 1$$

ist. Da $1 \notin \mathfrak{m}$, ist dabei $h_x \neq 0$. Wir wählen für jedes $x \in K$ ein festes Polynom h_x (und damit auch f_x), das obige Gleichung erfüllt, und setzen $K_d = \{x \in K \,|\, \deg h_x = d\}$ für jedes $d \in \mathbb{N}_0$. Da K nach Voraussetzung überabzählbar viele Elemente enthält und K die Vereinigung der K_d ist, muß mindestens eine der Mengen K_d unendlich viele Elemente enthalten. (Nur an dieser Stelle geht die Voraussetzung der Überabzählbarkeit ein.)

Wir wählen eine solche Menge K_d und betrachten den Vektorraum $K[X_1, \ldots, K_n]_d$ aller Polynome vom Grad höchstens d. Da es nur endlich viele Monome vom Grad höchstens d gibt, ist dies ein endlichdimensionaler K-Vektorraum. Wir wählen eine natürliche Zahl r, die größer ist als seine Dimension, und dazu r Elemente $x^{(1)}, \ldots, x^{(r)} \in K$ mit $h_{x^{(i)}} \in k[X_1, \ldots, X_n]_d$. Dann muß es Elemente $\lambda_1, \ldots, \lambda_r \in K$ geben, die nicht allesamt verschwinden, derart, daß

$$\lambda_1 h_{x^{(1)}} + \cdots + \lambda_r h_{x^{(r)}} = 0$$

ist.

Dazu definieren wir

$$g = \sum_{j=1}^{r} \lambda_j \prod_{\ell \neq j} \left(X_i - x^{(\ell)}\right) \in K[X_i]\,.$$

Dieses Polynom liegt auch in $\mathfrak{m}$, denn wegen

$$1 = f_{x^{(j)}} + h_{x^{(j)}}\left(X_i - x^{(j)}\right) \quad \text{für } j = 1, \ldots, r$$

ist

$$\begin{aligned} g &= \sum_{j=1}^{r} \lambda_j \left(f_{x^{(j)}} + h_{x^{(j)}}\left(X_i - x^{(j)}\right)\right) \prod_{\ell \neq j} \left(X_i - x^{(\ell)}\right) \in K[X_i] \\ &= \sum_{j=1}^{r} \lambda_j f_{x^{(j)}} \prod_{\ell \neq j} \left(X_i - x^{(\ell)}\right) + \left(\sum_{j=1}^{r} \lambda_j h_{x^{(j)}}\right) \prod_{\ell=1}^{n} \left(X_i - x^{(\ell)}\right) \\ &= \sum_{j=1}^{r} \lambda_j \prod_{\ell \neq j} \left(X_i - x^{(\ell)}\right) f_{x^{(j)}} \in \mathfrak{m}\,, \end{aligned}$$

da $\sum\limits_{j=1}^{r} \lambda_j h_{x^{(j)}}$ verschwindet und alle $f_{x^{(j)}}$ in $\mathfrak{m}$ liegen.

g ist nicht das Nullpolynom, denn für jeden Index ν ist

$$g\left(x^{(\nu)}\right) = \sum_{j=1}^{r} \lambda_j \prod_{\ell \neq j} \left(x^{(\nu)} - x^{(\ell)}\right) = \lambda_\nu \prod_{\ell \neq \nu} \left(x^{(\nu)} - x^{(\ell)}\right)\,.$$

Da die $x^{(\ell)}$ paarweise verschieden sind und mindestens ein λ_ν nicht verschwindet, muß mindestens einer dieser Werte von Null verschieden sein.

Da g in $\mathfrak{m}$ liegt, kann g auch keine von Null verschiedene Konstante sein, hat also einen positiven Grad e. Über dem algebraisch abgeschlossenen Körper K zerfällt g daher in Linearfaktoren:

$$g = c(X_i - z_1)\dots(X_i - z_e) \quad \text{mit} \quad c \in K \setminus \{0\},\ z_1, \dots z_e \in k\,.$$

g liegt in $\mathfrak{m}$, aber nach Voraussetzung liegt keiner der Faktoren $X_i - z_j$ in $\mathfrak{m}$, und die Konstante $c \neq 0$ natürlich auch nicht. Dies ist ein Widerspruch, denn als maximales Ideal ist $\mathfrak{m}$ insbesondere ein Primideal. ∎

Somit hat also jedes echte Ideal $I \lhd k[X_1, \dots, X_n]$ zumindest in einem Erweiterungskörper K von k mindestens eine Nullstelle. Damit folgt umgekehrt

Satz: Das Gleichungssystem

$$f_1(x_1, \dots, x_n) = \dots = f_m(x_1, \dots, x_n) = 0$$

mit $f_1, \dots, f_m \in k[X_1, \dots, X_n]$ ist genau dann in jedem Erweiterungskörper K von k unlösbar, wenn es Polynome $h_1, \dots, h_m$ in $X_1, \dots, X_n$ gibt, so daß $h_1 f_1 + \dots + h_m f_m = 1$ ist.

Beweis: Im Falle der Unlösbarkeit ist das von $f_1, \dots, f_m$ erzeugte Ideal der ganze Polynomring, enthält also insbesondere die Eins. Da

$$(f_1, \dots, f_m) = \left\{ h_1 f_1 + \dots + h_m f_m \;\middle|\; h_1, \dots, h_m \in k[X_1, \dots, X_n] \right\},$$

hat auch die Eins eine Darstellung der verlangten Form.

Ist umgekehrt $h_1 f_1 + \dots + h_m f_m = 1$ für irgendwelche Polynome $h_1, \dots, h_m$, so ist für jeden Erweiterungskörper K von k und jedes n-Tupel $(x_1, \dots, x_n) \in K^n$

$$h_1(x_1, \dots, x_n) f_1(x_1, \dots, x_n) + \dots + h_m(x_1, \dots, x_n) f_m(x_1, \dots, x_n) = 1\,,$$

so daß nicht alle $f_j(x_1, \dots, x_n)$ verschwinden können. ∎

Wenn wir eine GRÖBNER-Basis eines Ideals I kennen, ist es einfach zu entscheiden, ob $I = k[X_1, \dots, X_n]$ ist (oder äquivalent, ob $1 \in I$): Da

der führende Term eines jeden Polynoms aus I durch den führenden Term eines Elements der GRÖBNER-Basis teilbar sein muß, enthält diese im Falle eines Ideals, das die Eins enthält, ein Polynom, dessen führendes Monom die Eins ist. Da diese bezüglich jeder Monomordnung das kleinste Monom ist, muß somit die GRÖBNER-Basis eine Konstante enthalten. Die zugehörige minimale und erst recht die reduzierte GRÖBNER-Basis besteht in diesem Fall nur aus der Eins.

Aus dem gerade bewiesenen Satz folgt mit einem 1929 von J.L. RABINOWITSCH gefundenen Trick die

Starke Form des Hilbertschen Nullstellensatzes: k sei ein beliebiger Körper und K ein überabzählbarer algebraisch abgeschlossener Erweiterungskörper von k. Falls für ein Ideal $I \lhd k[X_1, \ldots, X_n]$ ein Polynom $f \in k[X_1, \ldots, X_n]$ auf ganz $V_K(I)$ verschwindet, gibt es ein $q \in \mathbb{N}$, so daß f^q in I liegt.

Beweis: Wir erweitern den Polynomring $k[X_1, \ldots, X_n]$ mit einer neuen Variablen X_{n+1} zu $k[X_1, \ldots, X_{n+1}]$ und betrachten dort für ein Erzeugendensystem $\{f_1, \ldots, f_m\}$ von I das Gleichungssystem

$$f_1(x_1, \ldots, x_n) = \cdots = f_m(x_1, \ldots, x_n) = 1 - x_{n+1} f(x_1, \ldots, x_n) = 0 \,.$$

Für jeden Punkt $(x_1, \ldots, x_n, x_{n+1}) \in K^{n+1}$, für den die $f_j(x_1, \ldots, x_n)$ verschwinden, verschwindet auch $f(x_1, \ldots, x_n)$, d.h.

$$1 - x_{n+1} f(x_1, \ldots, x_n) = 1 \,.$$

Somit haben diese $n + 1$ Gleichungen keine gemeinsame Nullstelle; es gibt also Polynome $h_1, \ldots, h_{m+1} \in k[X_1, \ldots, X_{n+1}]$ derart, daß

$$h_1 f_1 + \cdots + h_m f_m + h_{m+1}(1 - X_{n+1} f) = 1$$

ist. Diese Gleichung bleibt gültig, wenn wir überall für X_{n+1} ein Polynom oder eine rationale Funktion in $X_1, \ldots, X_n$ einsetzen; wir setzen $X_{n+1} = 1/f$. Die h_j werden dann zu rationalen Funktionen in $X_1, \ldots, X_n$, wobei alle Nenner Potenzen von f sind. Ist f^q die höchste dieser Potenzen, so erhalten wir nach Multiplikation mit f^q eine Gleichung der Form

$$\widetilde{h}_1 f_1 + \cdots + \widetilde{h}_m f_m = f^q$$

mit $\widetilde{h}_j = f^q h_j\big(X_1, \ldots, X_n, 1/f\big) \in k[X_1, \ldots, X_n]$. Dies zeigt, daß f^q in $I = (f_1, \ldots, f_m)$ liegt.

■

Definition: R sei ein Ring und $I \lhd R$ ein Ideal von R. Das *Radikal* von I ist die Menge

$$\sqrt{I} \underset{\text{def}}{=} \left\{ f \in R \;\middle|\; \exists q \in \mathbb{N} : f^q \in I \right\}.$$

Das Radikal besteht also aus allen Ringelementen, die eine Potenz in I haben. Es ist selbst ein Ideal, denn sind $f, g \in \sqrt{I}$ zwei Elemente mit $f^p \in I$ und $g^q \in I$, so sind in

$$(f+g)^{p+q} = \sum_{\ell=0}^{p+q} \binom{p+q}{\ell} f^{p+q-\ell} g^{\ell}$$

die ersten q Summanden Vielfache von f^p, und die restlichen p sind Vielfache von g^q. Somit liegt jeder Summand in I, also auch die Summe. Für ein beliebiges $r \in R$ liegt natürlich auch rf in $\sqrt{I}$, denn seine q-te Potenz $(rf)^q = r^q f^q$ liegt in I, sobald f^q in I liegt.

Mit diesem neuen Begriff können wir den obigen Satz umformulieren:

Satz: Ein Polynom $f \in k[X_1, \ldots, X_n]$ verschwindet genau dann auf $V_K(I)$, wenn $f \in \sqrt{I}$.

■

Anders ausgedrückt heißt dies

Satz: Für zwei Ideale $I, J \lhd k[X_1, \ldots, X_n]$ ist $V_K(I) = V_K(J)$ genau dann, wenn $\sqrt{I} = \sqrt{J}$ ist.

■

Falls ein Ideal mit seinem Radikal übereinstimmt, enthält es *alle* Polynome, die auf $V_K(I)$ verschwinden; zwei Polynome nehmen genau dann in jedem Punkt von $V_K(I)$ denselben Wert an, wenn ihre Differenz in I liegt, wenn sie also modulo I dieselbe Restklasse definieren.

Wenn das Ideal I nicht mit seinem Radikal übereinstimmt, gilt zwar nicht mehr *genau dann,* aber wir können trotzdem die Elemente des

Faktorvektorraums $A = k[X_1, \ldots, X_n]/I$ auffassen als Funktionen von $V_K(I)$ nach K: Für jede Restklasse und jeden Punkt aus $V_K(I)$ nehmen wir einfach irgendein Polynom aus der Restklasse und setzen die Koordinaten des Punkts ein. Da die Differenz zweier Polynome aus derselben Restklasse in I liegt, wird sie nach Einsetzen des Punktes zu Null, der Wert hängt also nicht ab von der Wahl des Polynoms. Auch Polynome aus $K[X_1, \ldots, X_n]$ definieren in dieser Weise Funktionen $V_K(I) \to K$; hinreichend (aber nicht notwendig) dafür, daß zwei Polynome dieselbe Funktion definieren ist, daß ihre Differenz im von I erzeugten Ideal $\bar{I} \lhd K[X_1, \ldots, X_n]$ liegt.

Im Falle von Polynomen einer Veränderlichen ist jedes Ideal von $k[X]$ ein Hauptideal, denn nach dem HILBERTschen Basissatz hat es ein endlich Erzeugendensystem $\{f_, \ldots, f_m\}$ und wird daher offensichtlich von $f = \text{ggT}(f_1, \ldots, f_m)$ erzeugt. Ist $I = (f)$ mit einem Polynom $f \neq 0$ vom Grad d, so können wir die Restklassen repräsentieren durch die Polynome vom Grad höchstens $d - 1$, denn jedes Polynom $g \in k[X]$ hat dieselbe Restklasse wie sein Divisionsrest bei der Polynomdivision durch f. Somit ist $A = k[X]/I$ in diesem Fall ein d-dimensionaler Vektorraum. Da $V_K(I)$ gerade aus den Nullstellen von f in K besteht, von denen es höchstens d verschiedene gibt, liefert die Dimension von A eine obere Schranke für die Elementanzahl von $V_K(I)$; wenn wir die Nullstellen mit ihrer Vielfachheit zählen, ist die Dimension von A sogar *gleich* der Gesamtzahl der Nullstellen. Im nächsten Paragraphen wollen wir uns überlegen, wie man ähnliche Ergebnisse auch für Systeme von Polynomgleichungen in mehreren Veränderlichen finden kann.

§3: Gleichungssysteme mit endlicher Lösungsmenge

Auch hier gehen wir wieder aus von einem beliebigen Körper k sowie einem algebraisch abgeschlossenen Erweiterungskörper K mit überabzählbar vielen Elementen. Letztere Bedingung ist nur notwendig, weil wir sie im Beweis des HILBERTschen Nullstellensatzes verwendet haben; wie bereits dort erwähnt, gibt es auch Beweise für den Fall, daß K ein beliebiger algebraisch abgeschlossener Körper ist, so daß alle Sätze dieses Paragraphen tatsächlich auch ohne die Voraussetzung der Überabzählbarkeit von K gelten.

Satz: I sei ein Ideal im Polynomring $k[X_1, \ldots, X_n]$ über dem Körper k, und K sei ein überabzählbarer algebraisch abgeschlossener Körper, in dem k enthalten sei. Dann gilt: $V_K(I)$ ist genau dann endlich, wenn der Faktorring $A = k[X_1, \ldots, X_n]/I$ ein endlichdimensionaler k-Vektorraum ist. In diesem Fall ist die Dimension von A eine obere Schranke für die Elementanzahl von $V_K(I)$.

Den recht umfangreichen *Beweis* führen wir in mehreren Schritten:

1. Schritt: Wenn der Vektorraum A endliche Dimension hat, ist $V_K(I)$ endlich.

Bezeichnet nämlich d die Dimension von A, so sind für jedes i die Potenzen $1, X_i, \ldots, X_i^d$ linear abhängig; es gibt also ein Polynom aus $k[X_i]$, das modulo I zur Null wird und somit in I liegt. Für jeden Punkt aus $V_K(I)$ muß daher die i-te Koordinate eine Nullstelle dieses Polynoms sein. Damit kann die i-te Koordinate nur endlich viele Werte annehmen, und da dies für alle i gilt, ist $V_K(I)$ endlich.

2. Schritt: $\bar{I}$ sei das von I in $K[X_1, \ldots, X_n]$ erzeugte Ideal. Wenn $V_K(I)$ endlich ist, hat der K-Vektorraum $\bar{A} = K[X_1, \ldots, X_n]/\bar{I}$ endliche Dimension.

Besteht $V_K(I)$ nur aus endlich vielen Punkten, so nimmt jede der Koordinatenfunktionen $X_1, \ldots, X_n$ auf $V_K(I)$ nur endlich viele Werte an; es gibt also für jedes i ein Polynom aus $K[X_i]$, das auf ganz $V_K(I)$ verschwindet. Nach dem HILBERTschen Nullstellensatz muß eine Potenz dieses Polynoms in $\bar{I}$ liegen, es gibt also auch in $\bar{I}$ für jedes i ein Polynom nur in X_i. Somit gibt es einen Grad d_i derart, daß sich X_i^e für $e \geq d_i$ modulo $\bar{I}$ durch die endlich vielen X_i-Potenzen $1, X_i, \ldots, X_i^{d_i - 1}$ ausdrücken läßt. Damit läßt sich auch jedes Monom aus $K[X_1, \ldots, X_n]$ modulo $\bar{I}$ durch jene Monome ausdrücken, bei denen jede Variable X_i höchstens mit Exponent $d_i - 1$ auftritt. Da es nur endlich viele solche Monome gibt, ist $K[X_1, \ldots, X_n]/\bar{I}$ ein endlichdimensionaler K-Vektorraum.

3. Schritt: A ist genau dann endlichdimensional, wenn $\bar{A}$ endlichdimensional ist; in diesem Fall haben beide dieselbe Dimension.

Ist A endlichdimensional, so wählen wir eine Basis $\{b_1, \ldots, b_r\}$ und zu jedem Basiselement b_i ein Polynom $B_i \in k[X_1, \ldots, X_n]$, das modulo I gleich b_i ist. Zusammen mit einer Basis von I als k-Vektorraum bilden die B_i dann eine k-Vektorraumbasis von $k[X_1, \ldots, X_n]$. Über K wird die Basis von I zu einer K-Vektorraumbasis von $\bar{I}$, da sich jedes Element von $\bar{I}$ als eine K-Linearkombination von Elementen aus I schreiben läßt. Zusammen mit den B_i, die wir auch als Elemente von $K[X_1, \ldots, X_n]$ aufâãlfassen können, erhalten wir sowohl über k als auch über K eine Basis des ganzen jeweiligen Polynomrings, und damit ist klar, daß die Restklassen der B_i modulo $\bar{I}$ den Faktorring $\bar{A}$ erzeugen. Somit ist dieser als K-Vektorraum endlichdimensional.

Die Gleichheit von $\dim_k A$ und $\dim_K \bar{A}$ folgt, falls wir zeigen können, daß die Restklassen der B_i modulo $\bar{I}$ linear unabhängig sind.

Dazu zeigen wir die folgende, etwas allgemeinere Aussage: Sind $B_1, \ldots, B_r$ Polynome aus $k[X_1, \ldots, X_n]$ mit Restklassen $b_1, \ldots, b_r$ modulo I und Restklassen $\bar{b}_1, \ldots, \bar{b}_r$ modulo $\bar{I}$, so sind die b_i genau dann linear abhängig, wenn es die $\bar{b}_i$ sind.

Die eine Richtung ist einfach: Falls die b_i linear abhängig sind, gibt es Skalare $\lambda_i \in k$, die nicht alle verschwinden, so daß $\lambda_1 b_1 + \cdots + \lambda_r b_r$ der Nullvektor aus A ist. $\lambda_1 B_1 + \cdots + \lambda_r B_r$ liegt daher in I, also erst recht in $\bar{I}$, so daß auch $\lambda_1 \bar{b}_1 + \cdots + \lambda_r \bar{b}_r$ der Nullvektor aus $\bar{A}$ ist.

Wenn die $\bar{b}_i$ linear abhängig sind, gibt es $\lambda_i \in K$, so daß $\lambda_1 \bar{b}_1 + \cdots + \lambda_r \bar{b}_r$ der Nullvektor aus $\bar{A}$ ist, d.h. $\lambda_1 B_1 + \cdots + \lambda_r B_r$ liegt in $\bar{I}$. Da die λ_i nicht in k liegen müssen, nützt und das noch nichts, um etwas über die b_i auszusagen.

Um trotzdem deren lineare Abhängigkeit zu beweisen, wählen wir ein endliches Erzeugendensystem $f_1, \ldots, f_m$ des Ideals I. Wir wissen dann, daß es Polynome $g_1, \ldots, g_m$ aus $K[X_1, \ldots, X_n]$ gibt mit

$$\lambda_1 B_1 + \cdots + \lambda_r B_r = g_1 f_1 + \cdots + g_m f_m \,.$$

Die Polynome g_j sind K-Linearkombinationen von Monomen $M_{j\ell}$ in den Variablen X_i. Die obige Gleichung ist also äquivalent zu einer

Gleichung der Form

$$\lambda_1 B_1 + \cdots + \lambda_r B_r - \sum_{j=1}^{m} \sum_{\ell=1}^{r_j} \mu_{j\ell} M_{j\ell} f_j = 0$$

mit Elementen $\mu_{j\ell} \in K$, die von den g_j abhängen. Sortieren wir diese Gleichung nach Monomen, können wir dies so interpretieren, daß ein (recht großes) lineares Gleichungssystem in den Variablen λ_i und $\mu_{j\ell}$ eine nichttriviale Lösung hat. Da die B_i und die f_j Polynome mit Koeffizienten aus k sind, ist dies ein homogenes lineares Gleichungssystem mit Koeffizienten aus k. Seine Lösungsmenge über k ist ein k-Vektorraum, für den uns der GAUSS-Algorithmus eine Basis liefert. Da der GAUSS-Algorithmus nirgends aus dem Körper hinausführt, in dem die Koeffizienten liegen, ist dies auch eine Basis des Lösungsraums über K; die beiden Vektorräume haben also dieselbe Dimension. Da wir wissen, daß es über K eine nichttriviale Lösung gibt, muß es daher auch über k eine geben,

Es gibt somit Elemente $\lambda_i' \in k$ und $\mu_{j\ell}' \in k$, die das Gleichungssystem lösen. Damit ist dann

$$\lambda_1' B_1 + \cdots + \lambda_r' B_r = g_1' f_1 + \cdots + g_m' f_m$$

mit Polynomen $g_j' \in k[X_1, \ldots, X_n]$, die linke Seite liegt also im Ideal I. Somit ist $\lambda_1' b_1 + \cdots + \lambda_r' b_r$ der Nullvektor in A. Die λ_i' können nicht allesamt verschwinden, denn ansonsten müßte mindestens ein $\mu_{j\ell} \neq 0$ sein, Null wäre also gleich einer nichttrivialen Linearkombination von Monomen, was absurd ist. Also sind auch die b_i linear abhängig.

Bleibt noch zu zeigen, daß A endlichdimensional ist, wenn $\bar{A}$ endlichdimensional ist. Das folgt sofort aus der gerade gezeigten Äquivalenz der linearen Abhängigkeit über k und über K: Hat $\bar{A}$ die endliche Dimension d, so ist jede Teilmenge von $\bar{A}$ mit mehr als d Elementen linear abhängig. Damit ist, wie wir gerade gesehen haben, auch jede Teilmenge von mehr als d Elementen aus A linear abhängig über k, also ist A endlichdimensional.

Im nächsten Schritt wollen wir das Zählen der Lösungen zurückführen auf das Zählen von Nullstellen eines Polynoms einer Veränderlichen.

Definition: Ein Polynom $u \in K[X_1, \ldots, X_n]$ heißt *separierend,* wenn es für keine zwei Elemente von $V_K(I)$ denselben Wert annimmt.

4. Schritt: Falls $V_K(I)$ endlich ist, gibt es ein separierendes homogenes lineares Polynom $u = c_1X_1 + \cdots c_nX_n$. Wir können dabei für u eines der speziellen Polynome

$$u_a = X_1 + aX_2 + a^2X_3 + \cdots + a^{n-1}X_n$$

wählen, wobei a in einer beliebig vorgebbaren Teilmenge von K mit mehr als $(n-1)\binom{s}{2} = \frac{1}{2}s(s-1)(n-1)$ Elementen liegt.

Für je zwei verschiedene Punkte $z, w \in V_K(I)$ ist $u_a(z) = u_a(w)$ genau dann, wenn

$$(z_1 - w_1) + (z_2 - w_2)a + (z_3 - w_3)a^2 + \cdots + (z_n - w_n)a^{n-1}$$

verschwindet. Die Koordinaten z_i, w_i von z und w sind Elemente von K; die $a \in K$, für die $u_a(z) = u_a(w)$ ist, sind also die Nullstellen eines Polynoms in einer Veränderlichen über K vom Grad höchstens $n-1$. Daher gibt es höchstens $n-1$ Werte $a \in K$, für die $u_a(z) = u_a(w)$ ist. Ist $s = \#V_K(I)$ endlich, so gibt es $\binom{s}{2}$ Paare aus voneinander verschiedenen Elementen; somit gibt es höchstens $(n-1)\binom{s}{2}$ Elemente $a \in K$, für die $u_a(z) = u_a(w)$ für *irgendwelche* voneinander verschiedene Elemente von $V_K(I)$.

(Hier haben wir benutzt, daß jeder algebraisch abgeschlossene Körper unendlich ist. Falls bereits k unendlich ist, etwa $k = \mathbb{Q}$, können wir sogar ein $a \in k$ finden gibt es somit Polynome u_a, die für je zwei verschiedene Elemente von $V_K(I)$ verschiedene Werte annehmen. Falls bereits k ein unendlicher Körper ist, können wir sogar entsprechende $a \in k$ finden; in diesem Fall gibt es also schon in $k[X_1, \ldots, X_n]$ solche Polynome. Im hier meistens betrachteten Fall $k = \mathbb{Q}$ können wir etwa eine ganze Zahl a mit $0 \leq a \leq (n-1)\binom{s}{2}$ wählen.

5. Schritt: Die Elementanzahl s von $V_K(I)$ ist höchstens gleich der Dimension von A.

Da wir im 3. Schritt gesehen haben, daß $\dim_k A = \dim_K \bar{A}$ ist, können wir auch mit dieser Dimension argumentieren. Aus dem 4. Schritt wissen wir, daß es ein Polynom $u \in K[X_1, \ldots, X_n]$ gibt, das für jedes

Element von $V_K(I)$ einen anderen Wert annimmt. Wir ersetzen u durch seine Restklasse $\tilde{u}$ modulo $\bar{I}$ in $\bar{A}$ und wollen uns überlegen, daß die Elemente $1, \tilde{u}, \ldots, \tilde{u}^{s-1} \in \bar{A}$ linear unabhängig sind: Angenommen, es gibt eine Relation der Form $\sum_{\ell=0}^{s-1} \lambda_\ell \tilde{u}^\ell = 0$ mit $\lambda_\ell \in K$. Das Polynom $\sum_{\ell=0}^{s-1} \lambda_\ell u^\ell \in K[X_1, \ldots, X_n]$ liegt dann in $\bar{I}$, verschwindet also für jedes der s Elemente von $V_K(I)$. Da u für jedes dieser Elemente einen anderen Wert annimmt, hat das Polynom $\sum_{\ell=0}^{r-1} \lambda_\ell U^\ell \in k[U]$ einerseits mindestens s verschiedene Nullstellen in K, andererseits ist sein Grad kleiner als s. Das ist nur für das Nullpolynom möglich; somit verschwinden alle Koeffizienten λ_ℓ, was die behauptete lineare Unabhängigkeit beweist. Damit enthält $\bar{A}$ mindestens s linear unabhängige Elemente, d.h. $r = \dim_K \bar{A} \geq s = \#V_K(I)$. Damit ist die Behauptung und auch der gesamte Satz bewiesen.

∎

Betrachten wir als Beispiel das von $f = X^2 + Y^2 - 1$ und $g = X - Y$ erzeugte Ideal $I \lhd \mathbb{Q}[X, Y]$. Seine Lösungsmenge ist, geometrisch gesehen, der Schnitt des Einheitskreises mit der ersten Winkelhalbierenden, besteht also aus den beiden Punkten $\left(\frac{1}{2}\sqrt{2}, \frac{1}{2}\sqrt{2}\right)$ und $\left(-\frac{1}{2}\sqrt{2}, -\frac{1}{2}\sqrt{2}\right)$.

Der Polynomring $\mathbb{Q}[X, Y]$ hat als $\mathbb{Q}$-Vektorraum eine Basis bestehend aus allen Monomen $X^a Y^b$ mit $a, b \in \mathbb{N}_0$. Modulo I sind X und Y äquivalent, und damit ist $X^a Y^b \sim X^{a+b}$. Außerdem ist $2X^2$ äquivalent zu $X^2 + Y^2$, und das wiederum ist wegen f äquivalent zu 1, d.h. $X^2 \sim \frac{1}{2}$. Daher ist jedes Monom äquivalent entweder zu einer Konstanten (falls $a+b$ gerade) oder einem skalaren Vielfachen von X. Da I kein Polynom der Form $\lambda X + \mu$ enthält, sind X und 1 modulo I linear unabhängig; somit bilden ihre Restklassen eine Basis des Vektorraums $\mathbb{Q}[X, Y]/I$.

Ersetzen wir in diesem Beispiel g durch $X^2 - Y^2 = (X + Y)(X - Y)$, so schneiden wir den Kreis mit beiden Winkelhalbierenden und haben nun eine vierelementige Lösungsmenge

$$V_{\mathbb{C}}(I) = \left\{ \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right), \left(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right), \left(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right), \left(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right) \right\}.$$

Modulo dem neuen Ideal I sind X und Y nicht mehr äquivalent, sondern nur noch X^2 und Y^2. Jedes Monom ist somit äquivalent entweder zu

einer X-Potenz oder zu einem Monom der Form X^aY. Da auch hier $X^2 \sim \frac{1}{2}$, ist es somit äquivalent zu einem skalaren Vielfachen eines der Monome $1, X, Y$ oder XY. Da keine Linearkombination dieser vier Monome in I liegt, bilden ihre Restklassen eine Basis von $\mathbb{Q}[X,Y]/I$.

In diesen beiden Beispielen waren sowohl die Lösungsmengen als auch Basen der Faktorringe einfach zu finden; im Allgemeinen ist das eher nicht der Fall. Wenn wir eine GRÖBNER-Basis des Ideals I kennen, können wir leicht eine Vektorraumbasis des Faktorrings konstruieren:

Definition: $I \lhd k[X_1, \ldots, X_n]$ sei ein Ideal und G sei eine GRÖBNER-Basis bezüglich irgendeiner Monomordnung auf $k[X_1, \ldots, X_n]$. Ein Monom in $X_1, \ldots, X_n$ heißt *Standardmonom* (bezüglich G), wenn es für kein $g \in G$ durch das führende Monom von g teilbar ist.

Satz: Für jede GRÖBNER-Basis G eines Ideals $I \lhd k[X_1, \ldots, X_n]$ bilden die Restklassen der Standardmonome eine Vektorraumbasis von $k[X_1, \ldots, X_n]/I$.

Beweis: Zunächst sind diese Restklassen linear unabhängig, denn jede nichttriviale Linearkombination der Null entspräche einem Polynom h aus I, dessen sämtliche Monome Standardmonome sind. Da die führenden Monome der Elemente von G das Ideal FM(I) erzeugen, müßte daher FM(h) Vielfaches eines FM(g) mit $g \in G$ sein, was der Definition eines Standardmonoms widerspricht.

Für ein beliebiges $f \in k[X_1, \ldots, X_n]$ liefert uns der Divisionsalgorithmus eine Darstellung

$$f = \sum_{g \in G} a_g g + r \quad \text{mit} \quad a_g, r \in k[X_1, \ldots, X_n]\,,$$

wobei r eine k-Linearkombination von Standardmonomen ist. Da die Summe der $a_g g$ in I liegt, ist f also äquivalent zu einer k-Linearkombination von Standardmonomen, so daß seine Restklasse die entsprechende Linearkombination von deren Restklassen ist. ∎

Dieser Satz gilt unabhängig davon, ob $k[X_1, \ldots, X_n]/I$ als Vektorraum endlichdimensional ist; er liefert uns auch ein einfaches Kriterium dafür,

wann er endliche Dimension hat und wann somit die Lösungsmenge $V_K(I)$ endlich ist:

Lemma: G sei eine GRÖBNER-Basis eines Ideals $I \lhd k[X_1, \ldots, X_n]$ bezüglich irgendeiner Monomordnung. $V_K(I)$ ist genau dann endlich, wenn G für jedes i ein Polynom enthält, dessen führendes Monom eine X_i-Potenz ist.

Beweis: Falls die GRÖBNER-Basis für jedes i ein Polynom mit führendem Monom $X_i^{d_i}$ enthält, ist jedes Monom, in dem ein X_i mit einem Exponenten größer oder gleich d_i vorkommt, durch das führende Monom eines Elements der GRÖBNER-Basis teilbar. Die Monome, für die das nicht der Fall ist, haben für jedes i einen Exponenten echt kleiner d_i; es gibt also nur endlich viele Standardmonome. Somit hat A endliche Dimension, und $V_K(I)$ ist endlich.

Ist umgekehrt $V_K(I)$ endlich, so enthält $\bar{I}$ für jedes i ein Polynom aus $K[X_i]$ – siehe Schritt 2 im Beweis des obigen Satzes. Da die GRÖBNER-Basis von I gleichzeitig eine GRÖBNER-Basis von $\bar{I}$ ist, muß das führende Monom eines ihrer Elemente die höchste X_i-Potenz in diesem Polynom teilen, muß also selbst eine Potenz von X_i sein. ∎

Für den Fall, daß $V_K(I)$ endlich ist, läßt der obige Satz noch wie folgt verschärfen:

Satz: Ist $D = V_K(I)$ endlich und τ eine Monomordnung, so gibt es zu jeder Funktion $\varphi\colon D \to K$ eine Linearkombination f von Standardmonomen bezüglich τ derart, daß $f(x) = \varphi(x)$ für alle $x \in D$. Insbesondere ist die Dimension von $k[X_1, \ldots, X_n]/I$ größer oder gleich der Elementanzahl von D. Die beiden Zahlen sind genau dann gleich, wenn I das Ideal $I(D)$ aller auf D verschwindender Polynome ist, was wiederum dazu äquivalent ist, daß I ein Radikalideal ist.

Beweis: Wie wir im vierten Schritt des Beweises in §3 gesehen haben, gibt es ein lineares Polynom $\ell \in k[X_1, \ldots, X_n]$, das auf den verschiedenen Punkten von D verschiedene Werte annimmt. Dazu können wir über die NEWTONsche oder LAGRANGEsche Interpolationsformel ein

Polynom aus $K[T]$ finden, das für jeden Punkt $x \in D$ an der Stelle $t = \ell(x)$ den Wert $\varphi(x)$ annimmt. Setzen wir ℓ in dieses Polynom ein, erhalten wir ein Polynom $\widetilde{f}$ aus $k[X_1, \ldots, X_n]$, das für jedes $x \in D$ an der Stelle x den Wert $\varphi(x)$ annimmt. Da die Restklassen der Standardmonome eine Basis des Restklassenrings bilden, gibt es dazu eine Linearkombination f der Standardmonome, die sich nur durch ein Polynom aus $\bar{I}$ von g unterscheidet, d.h. $f(x) = g(x) = \varphi(x)$ für alle $x \in D$.

Die Dimension des Vektorraums aller Funktionen $D \to K$ ist gleich der Elementanzahl von D, denn die Funktionen, die jeweils einem Punkt aus D den Wert eins zuordnen und alle anderen auf Null abbilden, bilden eine Basis. Damit muß auch der Restklassenring mindestens diese Dimension haben.

Die beiden Dimensionen stimmen genau dann überein, wenn die obige Linearkombination f durch φ eindeutig bestimmt ist. Sind f_1 und f_2 zwei verschiedene solche Linearkombinationen, so verschwindet $f_1 - f_2$ auf ganz D, liegt also im Ideal $I(D)$. Genau dann, wenn dieses mit I übereinstimmt, können wir daraus folgern, daß $f_1 = f_2$ ist, und das ist nach dem HILBERTschen Nullstellensatz genau dann der Fall, wenn I ein Radikalideal ist. ∎

In §1 haben wir gesehen, wie man zu jeder endlichen Teilmenge $D \subset k^n$ ein Ideal $I \lhd k[X_1, \ldots, X_n]$ finden kann, für das $D = V_K(I)$ ist. Mit dem gerade bewiesenen Satz können wir nun sehen, daß das dort konstruierte Ideal gleich $I(D)$ ist:

Für $D = \{x^{(1)}, \ldots, x^{(r)}\} \subset k^n$ mit $x^{(i)} = (x_1^{(i)}, \ldots, x_n^{(i)})$ hatten wir die Punkte

$$y^{(i)} = (0, \ldots, 0, 1, 0, \ldots, 0, x_1^{(i)}, \ldots, x_n^{(i)}) \in k^{r+n}$$

betrachtet, wobei die Eins bei $y^{(i)}$ an der i-ten Stelle steht; die Menge dieser Punkte sei $\widetilde{D}$. Wie wir gesehen hatten, ist $\widetilde{D}$ die Nullstellenmenge jenes Ideals $J \lhd k[T_1, \ldots, T_r, X_1, \ldots, X_n]$, das erzeugt wird von den Polynomen $f_{ij} = T_i(X_j - x_j^{(i)})$ und dem Polynom $g = T_1 + \cdots + T_r - 1$. Wir wollen uns als erstes überlegen, daß J das Ideal *aller* auf $\widetilde{D}$ verschwindenden Funktionen ist: Da $f_{ij} \in J$, ist jedes Monom $T_i X_j$

modulo J äquivalent zu einem skalaren Vielfachen von T_i. Induktiv folgt, daß für jedes nichtkonstante Monom M in den X_j das Monom T_iM äquivalent ist zu einem skalaren Vielfachen von T_i. Da g in J liegt, ist M selbst äquivalent zu $T_1M + T_2M + \cdots + T_rM$ und damit zu einem linearen Polynom in den T_i. Somit ist jedes Polynom aus $k[T_1, \ldots, T_r, X_1, \ldots, X_n]$ äquivalent zu einem Polynom nur in den T_i.

Für zwei verschiedene Punkte $x^{(i)}$ und $x^{(\ell)}$ aus D gibt es mindestens einen Index j, für den $x_j^{(i)} \neq x_j^{(\ell)}$ ist. Mit f_{ij} und $f_{\ell j}$ enthält J auch das Polynom

$$T_\ell f_{ij} - T_i f_{\ell j} = T_\ell T_i (x^{(\ell)} - x^{(i)})$$

und damit das Produkt T_iT_ℓ, so daß jedes Monom, das zwei verschiedene T_i enthält, modulo J verschwindet. Außerdem liegt für jedes T_i auch das Polynom $T_ig = T_iT_1 + \cdots + T_iT_r$ in J; da alle Produkte T_iT_ℓ mit $\ell \neq i$ in J liegen, muß auch $T_i^2 \in J$ sein. Somit ist jedes Polynom äquivalent zu einem linearen Polynom in den T_i, wobei wir dieses homogen wählen können, da 1 äquivalent ist zur Summe der T_i.

Dies zeigt, daß der Restklassenring modulo J als k-Vektorraum höchstens die Dimension r hat. Eine kleinere Dimension kann er nicht haben, da sich jede Funktion $\widetilde{D} \to K$ durch eine Linearkombination von Vertretern der Basisvektoren realisieren läßt und r die Mächtigkeit von $\widetilde{D}$ ist. Damit folgt aus dem gerade bewiesenen Satz, daß J ein Radikalideal sein muß. Dann ist aber auch $I = J \cap k[X_1, \ldots, X_n]$ ein Radikalideal, d.h. $I = I(D)$.

Kapitel 3
Anwendung auf Designs

Eine der Grundaufgaben der Statistik besteht darin, ausgehend von Stichproben Modelle zu entwickeln und deren Parameter zu schätzen. Wir gehen aus von einem Körper k, den wir meist als Teilkörper der reellen Zahlen auffassen werden. Da wir allerdings im Körper k exakt rechnen mußsen, sollte k so klein wie möglich sein, etwa $k = \mathbb{Q}$.

Definition: Ein *Design* D in k^n ist eine endliche Teilmenge von k^n.

Oftmals haben die n Koordinaten der Punkte aus D inhaltliche Interpretationen, indem sie die Ausprägungen verschiedener Faktoren kodieren; eine Stichprobe dient dann dazu, den Einfluß verschiedener der Faktoren auf ein Ergebnis abzuschätzen. Beispiele sind etwa der Ertrag eines landwirtschaftlichen Produkts in Abhänbgigkeit von Düngung, Bewässerung, Bodenbeschaffenheit *usw.,* oder der Preis, den ein Kunde für ein Auto zu zahlen bereit ist in Abhängigkeit von verschiedenen Ausstattungsmerkmalen. Oft werden für diese Faktoren nur endlich viele Stufen betrachtet. Diese können auf einer reinen Nominalskala liegen, etwa $\{$*rot*, *blau*, *gelb*$\}$ oder $\{$*vorhanden*, *nicht vorhanden*$\}$, weshalb man sich in der Literatur im Falle von ℓ Stufen oft mit der Kodierung durch die Zahlen von Null bis $\ell - 1$ begnügt. Für die folgende Theorie bringt diese Einschränkung allerdings keinerlei Vereinfachung; deshalb werden wir für jede der Stufen beliebige Elemente eines Körpers k zulassen:

Definition: *a)* Ein volles faktorielles Design für n Faktoren ist ein kartesisches Produkt von n endlichen Teilmengen $S_i \subset k$. Falls alle M_i jeweils ℓ Elemente haben, sprechen wir von einem ℓ^n-Design.

b) Ein (fraktionelles) faktorielles Design für n Faktoren ist eine Teilmenge eines vollen faktoriellen Designs für diese Faktoren.

Offensichtlich kann jedes Design $D \subset k^n$ als fraktionelles faktorielles Design angesehen werden, etwa für S_i gleich Menge aller möglicher Werte, die die Variable x_i auf D annehmen kann.

§1: Allgemeine lineare Modelle

Sei $D = \{x^{(1)}, \ldots, x^{(r)}\} \subset k^n$. Typischerweise ist für jeden Punkt $x^{(j)} \in D$ ein Wert $y_i \in k$ gegeben; gesucht ist eine Funktion $f\colon k^n \to k$ mit $f(x^{(j)}) = y_j$ oder zumindest $f(x^{(j)}) \approx y_j$ für alle j.

Für den allgemeinsten (linearen) Ansatz zum Auffinden solcher Funktionen wählen wir eine endliche Menge $\mathbb{F} = \{f_1, \ldots, f_s\}$ von Funktionen $f_i\colon k^n \to k$ und betrachten Funktionen f der Form $f = \sum_{i=1}^{s} \theta_i f_i$ mit $\theta_1, \ldots, \theta_s \in k$.

Definition: Die Z-Matrix zu $\mathbb{F}$ und D ist die $s \times r$-Matrix Z mit Einträgen $z_{ij} = f_i(x^{(j)})$.

Für $f = \sum_{i=1}^{s} \theta_i f_i$ soll dann idealerweise gelten

$$y_j = \sum_{i=1}^{s} \theta_i f_i(x^{(j)}) = \sum_{i=1}^{s} \theta_i z_{ij} \; ;$$

ausgedrückt mit den Vektoren

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_r \end{pmatrix} \quad \text{und} \quad \theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_s \end{pmatrix}$$

wird dies zu $y^T = \theta^T Z \quad$ oder $\quad Z^T \theta = y$.

Falls $r < s$ ist, kann dieses Gleichungssystem keine eindeutig bestimmte Lösung haben. Für $r = s$ gibt es genau dann eine, wenn die Z-Matrix invertierbar ist; dann ist $\theta = (Z^T)^{-1} y$.

In der Statistik ist meist $r > s$; in diesem Fall wird es im allgemeinen keine Lösung geben. Dann müssen wir uns begnügen mit einem Vektor θ derart, daß der Fehler $\varepsilon = y - Z^T \theta$ möglichst klein ist.

Wenn wir davon ausgehen, daß die Komponenten ε_i dieses Vektors von einer Vielzahl voneinander unabhängiger externer Störgrößen abhängen, sagt uns der zentrale Grenzwertsatz, daß die ε_i zumindest näherungsweise normalverteilt sein sollten. Ein *maximum likelihood* Ansatz führt dann zu der Forderung, daß die (EUKLIDische) Länge des Vektors ε minimal sein soll, d.h. ε ist der Abstand des Vektors y vom Untervektorraum $U = \{Z^T\theta \mid \theta \in k^s\}$.

Damit ist ε das Lot von y auf U, liegt also insbesondere im orthogonalen Komplement $U^\perp$ von U, d.h. $\left\langle \varepsilon, Z^T\theta \right\rangle = 0$ für alle $\theta \in k^s$.

Wie aus der linearen Algebra bekannt, ist

$$\left\langle \varepsilon, Z^T\theta \right\rangle = \langle Z\varepsilon, \theta \rangle = \left\langle Z(Z^T\theta - y), \theta \right\rangle = \left\langle ZZ^T\theta - Zy, \theta \right\rangle = 0$$

für alle $\theta \in k^s$. Der einzige Vektor aus k^s, der auf allen Vektoren aus k^s senkrecht steht, ist der Nullvektor; somit muß

$$ZZ^T\theta = Zy$$

sein. ZZ^T ist eine quadratische Matrix; es gibt daher genau dann eine eindeutig bestimmte Lösung, wenn diese Matrix nichtsingulär ist. Das wiederum ist äquivalent dazu, daß Z (und damit auch Z^T) den Rang s hat: In diesem Fall ist nämlich die Abbildung von k^r nach k^s, die einem Vektor x den Vektor Zx zuordnet, surjektiv, und die Abbildung von k^s nach k^r, die θ auf $Z^T\theta$ abbildet, hat ein s-dimensionales Bild. Somit hat die Abbildung von k^r nach k^r, die einem Vektor $x \in k^r$ den Vektor $ZZ^Tx = Z(Z^Tx)$ zuordnet, ein s-dimensionales Bild, d.h. der Rang von ZZ^T ist s. Ist umgekehrt der Rang von Z kleiner als s, so erst recht der von ZZ^T, so daß ZZ^T in der Tat genau dann nichtsingulär ist, wenn der Rang von Z gleich s ist.

Damit haben wir gezeigt

Satz: Falls der Rang von Z gleich s ist, gibt es genau einen Vektor $\theta \in k^s$, für den die Differenz $Z^t\theta - y$ minimale Länge hat, nämlich $\theta = (ZZ^T)^{-1}y$.

■

§2: Polynomiale lineare Modelle

An dieser Vorlesung soll es in erster Linie um Modelle gehen, bei denen die Funktionen aus $\mathbb{F}$ Monome sind. Schon lange vor dem Aufkommen der algebraischen Statistik betrachtete man im Sinne möglichst einfacher Modelle vorzugsweise Mengen $\mathbb{F}$, die mit jedem Monom auch dessen sämtliche Teiler enthalten:

Definition: k sei ein Körper, $R = k[X_1, \ldots, X_n]$ ein Polynomring über k, und $\mathbb{T}$ sei die Menge aller Monome $X_1^{e_1} \cdots X_n^{e_n}$ mit $e_i \in \mathbb{N}_0$.
a) Eine nichtleere Teilmenge $\mathcal{O}$ von $\mathbb{T}$ heißt *Ordnungsideal,* wenn für jedes Monom aus $\mathcal{O}$ auch dessen sämtliche Teiler in $\mathcal{O}$ liegen.
b) Ein lineares Modell mit $\mathbb{F} \subset \mathbb{T}$ heißt *monomiales lineares Modell;* die Menge $\mathbb{F}$ wird als der *Träger* des Modells bezeichnet.
c) Das Modell heißt *vollständig,* wenn $\mathbb{F}$ ein Ordnungsideal ist.

Das Ergebnis am Ende des letzten Kapitels zeigt, wie wir zu jedem Design D ein vollständiges lineares Modell finden können: Wir fassen D zunächst auf als Nullstellenmenge eines Ideals I des Polynomrings R. Wie man ein solches Ideal I finden kann, haben wir am Ende von §1 des vorigen Kapitels gesehen, und ganz am Ende des Kapitels haben wir uns überlegt, daß es das volle Designideal $I(D)$ ist. Danach wählen wir eine Monomordnung τ auf R und berechnen dazu eine GRÖBNER-Basis von I.

Definition: *a)* $\mathrm{Est}_\tau(D)$ ist die Menge aller Standardmonome zu einer GRÖBNER-Basis von $I(D)$ bezüglich der Monomordnung τ.
b) Die Menge aller Mengen $\mathrm{Est}_\tau(D)$, wobei τ die sämtlichen Monomordnungen von R durchläuft, heißt der (GRÖBNER-)*Fächer* von D.

Für praktische Zwecke ist diese Konstruktion meist recht aufwendig, da wir dazu GRÖBNER-Basen des Designideals kennen und damit auch berechnen müssen.

Im Falle eines vollen faktoriellen Design $D = S_1 \times \cdots \times S_n$ haben wir damit keinerlei Schwierigkeiten: Offensichtlich bilden die Polynome

$$g_i = \prod_{x \in S_i} (X_i - x)$$

ein Erzeugendensystem von $I(D)$, und dieses Erzeugendensystem ist nach dem Kriterium von BUCHBERGER bezüglich jeder Monomordnung eine GRÖBNER-Basis nach dem folgenden

Lemma: Sind $f, g \in k[X, Y]$ zwei Polynome, wobei f nur von X und g nur von Y abhängt, so ergibt $S(f, g)$ bezüglich jeder Monomordnung bei der Divison durch f, g den Rest Null.

Beweis: Sei $f = \sum_{i=1}^{d} a_i X^i$ und $g = \sum_{j=1}^{e} b_j X^j$. Wir können o.B.d.A. davon ausgehen, daß $a_d = b_e = 1$ ist; dann ist

$$S(f,g) = Y^e f - X^d g = \sum_{i=0}^{d-1} a_i X^i Y^e - \sum_{j=0}^{e-1} X^d Y^j \ .$$

Der führende Term hiervon ist je nach Monomordnung entweder von der Form $a_i X^i Y^e$ (mit $i = d-1$, falls $a_{d-1} \neq 0$) oder von der Form $b_j X^d Y^j$ (mit $j = e-1$ falls $b_{e-1} \neq 0$). Im ersten Fall ist der führende Term durch das führende Monom Y^e von g teilbar, im zweiten durch FM(f) = X^d. Durch Subtraktion von $a_i X^i g$ *bzw.* $b_j Y^j f$ können wir diesen führenden Term eliminieren, wobei etwa neu hinzukommende Terme von der gleichen Bauart mit kleinerem i *bzw.* j sind.

Der neue führende Term ist wieder entweder von der Form $aX^i Y^e$ oder $bX^d Y^j$, und wieder läßt er sich eliminieren durch Subtraktion von $aX^i g$ oder $bY^j f$. Auf diese Weise lassen sich nacheinander alle Terme von $S(f, g)$ eliminieren, so daß der Divisionsalgorithmus den Rest Null liefert. ∎

Damit ist klar, daß die Standardmonome für ein volles faktorielles Design $D = S_1 \times \cdots \times S_n$ für jede Monomordnung genau aus den Monomen $X_1^{e_1} \cdots X_n^{e_n}$ besteht, für die alle e_i kleiner sind als die Elementanzahl der entsprechenden Menge S_i.

§3: Designs mit minimalem Fächer

Der Fächer eines vollständigen faktoriellen Designs besteht aus genau einem Blatt; kleiner kann er nicht werden. Deshalb definieren wir

Definition: Ein Design D hat minimalen Fächer, wenn alle Monomordnungen τ auf die gleiche Menge $\mathrm{Est}_\tau(D)$ führen.

Das ist insbesondere dann der Fall, wenn alle Monomordnungen zur gleichen GRÖBNER-Basis führen; betrachten wir also diesen Fall etwas genauer:

Definition: *a)* Eine Teilmenge $G = \{g_1, \ldots, g_m\}$ eines Ideals I von $k[X_1, \ldots, X_n]$ heißt *universelle* GRÖBNER-Basis, wenn sie bezüglich jeder Monomordnung auf $k[X_1, \ldots, X_n]$ eine GRÖBNER-Basis ist.
b) G heißt *Super-G-Basis* bezüglich einer Monomordnung τ, wenn jede Teilmenge von G bezüglich τ eine GRÖBNER-Basis des von ihr erzeugten Ideals ist.

Die beiden hier definierten Konzepte haben eigentlich nichts miteinander zu tun; bei den Designs, die wir hier betrachten wollen, ist aber ihr Zusammenspiel nützlich. Wir beginnen mit einem Kriterium zur Charakterisierung von Super-G-Basen:

Lemma: $G \subset I$ ist genau dann eine Super-G-Basis, wenn für je zwei Elemente $f, g \in G$ gilt:

$$\mathrm{FM}_\tau\big(\mathrm{ggT}(f,g)\big) = \mathrm{ggT}\big(\mathrm{FM}_\tau(f), \mathrm{FM}_\tau(g)\big)\,. \qquad (*)$$

Beweis durch Induktion nach $m = \#G$:

Für $m = 2$ ist $G = \{f, g\}$, und es ist klar, daß $\{f\}$ und $\{g\}$ GRÖBNER-Basen der Hauptideale (f) und (g) sind. Wir müssen daher zeigen, daß $\{f, g\}$ genau dann eine GRÖBNER-Basis von I ist, wenn $\mathrm{FM}_\tau\big(\mathrm{ggT}(f,g)\big) = \mathrm{ggT}\big(\mathrm{FM}_\tau(f), \mathrm{FM}_\tau(g)\big)$ ist.

Mit $h = \mathrm{ggT}\big(\mathrm{FM}_\tau(f), \mathrm{FM}_\tau(g)\big)$ ist

$$S(f,g) = \frac{\mathrm{FM}_\tau(g)}{\mathrm{FK}_\tau(f)h} - \frac{\mathrm{FM}_\tau(f)}{\mathrm{FK}_\tau(g)h}g\,,$$

und nach dem Kriterium von BUCHBERGER ist G genau dann eine GRÖBNER-Basis, wenn sich dieses Polynom modulo $\{f, g\}$ reduzieren läßt, wenn es also Polynome p, q gibt, so daß $S(f,g) = pf + qg$ ist. Mit

$$\widetilde{g} = \frac{\mathrm{FM}_\tau(g)}{\mathrm{FK}_\tau(f)h} - p \quad \text{und} \quad \widetilde{f} = \frac{\mathrm{FM}_\tau(f)}{\mathrm{FK}_\tau(g)h} - q$$

ist also $\widetilde{g}f - \widetilde{f}g = 0$, und der jeweils erste Summand ist der führende Term von $\widetilde{g}$ *bzw.* $\widetilde{f}$. Da h der ggT der führenden Monome von f und g ist, sind die führenden Monome von $\widetilde{f}$ und $\widetilde{g}$ teilerfremd, und damit sind auch $\widetilde{f}$ und $\widetilde{g}$ teilerfremd. Wegen $\widetilde{g}f = \widetilde{f}g$ muß daher $\widetilde{f}$ ein Teiler von f sein und $\widetilde{g}$ einer von g. Da Polynomringe faktoriell sind, gibt es somit ein Polynom $\widetilde{h}$ derart, daß $f = \widetilde{h}\widetilde{g}$ und $g = \widetilde{h}\widetilde{f}$ ist. Wegen der Teilerfremdheit von $\widetilde{f}$ und $\widetilde{g}$ folgt daraus, daß $\widetilde{h} = \text{ggT}(f, g)$ ist, und wenn wir die führenden Monome betrachten, sehen wir, daß h das führende Monom von $\widetilde{h}$ ist. Somit gilt $(*)$, falls G eine GRÖBNER-Basis ist.

Ist umgekehrt $(*)$ erfüllt und etwa $\widetilde{h} = \text{ggT}(f, g)$, so gibt es teilerfremde Polynome $\widetilde{f}, \widetilde{g}$ derart, daß $f = \widetilde{h} \cdot \widetilde{f}$ und $g = \widetilde{h} \cdot \widetilde{g}$ ist. Nach $(*)$ ist das führende Monom h von $\widetilde{h}$ der größte gemeinsame Teiler von $\text{FM}_\tau(f)$ und $\text{FM}_\tau(g)$. Somit ist $\text{FM}_\tau\left(\widetilde{f}\right) = \text{FM}(f)/h$ und $\text{FM}_\tau\left(\widetilde{g}\right) = \text{FM}(g)/h$. Insgesamt sei

$$\widetilde{f} = \frac{\text{FM}_\tau(f)}{\text{FK}_\tau(f)} - q \quad \text{und} \quad \widetilde{g} = \frac{\text{FM}_\tau(g)}{\text{FK}_\tau(g)} - p$$

mit geeigneten Polynomen p und q; dann ist $S(f, g) = pf + qg$, so daß G nach dem Kriterium von BUCHBERGER eine GRÖBNER-Basis ist.

Damit ist der Induktionsanfang $m = 2$ gezeigt; sei nun $m > 2$ und $G = \{g_1, \ldots, g_m\}$. Falls G seine Super-G-Basis ist, muß für je zwei Elemente g_i und g_j auch $\{g_i, g_j\}$ eine GRÖBNER-Basis von (g_i, g_j) sein, d.h. nach dem gerade bewiesenen Fall $m = 2$ gilt obige Gleichung.

Gilt umgekehrt obige Gleichung für alle (i, j) und ist G' eine Teilmenge von G, die zwei feste Elemente g_i und g_j enthält, so ist zunächst $\{g_i, g_j\}$ nach dem Fall $m = 2$ eine GRÖBNER-Basis von (g_i, g_j), also läßt sich $S(g_i, g_j)$ nach dem Kriterium von BUCHBERGER modulo $\{g_i, g_j\}$ und damit erst recht modulo G' auf Null reduzieren, d.h. G' ist eine GRÖBNER-Basis des von G' erzeugten Ideals. Somit ist G eine Super-G-Basis.

■

Polynome, für die dieses Kriterium sehr einfach nachzuweisen ist, sind *Distraktionen* (Zerstreuungen) von Monomen:

Definition: $M = X_1^{e_1} \cdots X_n^{e_n}$ sei ein Monom, und für jedes $i = 1, \ldots, n$ sei ein Vektor $a^{(i)} \in k^{\ell_i}$ gegeben, wobei $\ell_i \geq e_i$. Die *Distraktion* von M bezüglich dieser Vektoren ist

$$D(M) = \prod_{i=1}^{n} \prod_{j=1}^{e_i} \left(X_i - a_j^{(i)}\right) .$$

Falls alle Vektoren $a^{(i)}$ Nullvektoren sind, ist $D(M) = M$. Für ein volles faktorielles Design $D = S_1 \times \cdots \times S_n$ mit $\ell_i = \#S_i$ Vektoren $a^{(i)}$, deren Komponenten (in irgendeiner Reihenfolge) die verschiedenen Elemente von S_i sind, ist

$$D(X_i^{\ell_i}) = \prod_{j=1}^{\ell_i} \left(X_i - a_j^{(i)}\right) = \prod_{x \in S_i} \left(X_i - x\right)$$

und

$$I(D) = \left(D(X_1^{\ell_1}), \ldots, D(X_n^{\ell_n})\right) .$$

Die Tatsache, daß dieses Erzeugendensystem eine universelle GRÖBNER-Basis ist, folgt nun auch aus dem folgenden

Satz: $M_1, \ldots, M_r$ seien Monome aus $k[X_1, \ldots, X_n]$, und $a^{(1)}, \ldots, a^{(n)}$ seinen Vektoren über k mit hinreichend vielen Komponenten. Dann bilden die Polynome $D(M_1), \ldots, D(M_r)$ bezüglich jeder Monomordnung eine universelle GRÖBNER-Basis des von ihnen erzeugten Ideals. Falls keines der Monome M_i ein anderes teilt, ist diese GRÖBNER-Basis reduziert.

Beweis: Aus der Definition von $D(M)$ folgt sofort, daß alle Monome in $D(M)$ Teiler von M sind; daher ist M bezüglich jeder Monomordnung das führende Monom von $D(M)$. Außerdem ist klar, daß zwei Distraktionspolynome $D(M)$ und $D(M')$ genau dann Teiler voneinander sind, wenn dasselbe für M und M' gilt. Damit ist insbesondere der ggT zweier Polynome $D(M)$ und $D(M')$ gleich $D\big(\mathrm{ggT}(M, M')\big)$, so daß das Kriterium aus obigem Lemma erfüllt ist. Somit bilden die $D(M_i)$ eine Super-G-Basis des von ihnen erzeugten Ideals, insbesondere also eine GRÖBNER-Basis.

Diese GRÖBNER-Basis ist minimal genau dann, wenn kein führendes Monom eines $D(M_i)$ das eines $D(M_j)$ mit $j \neq i$ teilt, wenn also kein M_i ein anderes teilt. In diesem Fall ist die Basis auch reduziert, denn würde M_i irgendein Monom von $D(M_j)$ teilen, so auch M_j selbst, da alle Monome in $D(M_j)$ Teiler von M_j sind.

■

Definition: Ein Design $D \subset \mathbb{N}_0^n$ heißt *Stufendesign,* wenn für jeden Punkt $(x_1, \ldots, x_n) \in D$ auch die sämtlichen Punkte $(u_1, \ldots, u_n) \in \mathbb{N}_0^n$ mit $u_i \leq x_i$ für alle i in D liegen.

Lemma: Jedes Stufendesign ist die Nullstellenmenge eines Ideals, das von Distraktionen von Monomen bezüglich hinreichend langer Vektoren $a^{(i)} = (0, 1, \ldots, \ell_i)$ erzeugt wird.

Beweis durch Induktion nach n: Für $n = 1$ gibt es ein $\ell_1 \in \mathbb{N}_0$, so daß

$$D = \{0, 1, \ldots, \ell_1\} = V\big(X_1(X_1 - 1)\ldots(X_1 - \ell_1)\big) = V\big(D(X_1^{\ell_1+1})\big)$$

ist.

Ist $n > 1$ und kann x_n Werte aus $\{0, \ldots, \ell_n\}$ annehmen, so beachten wir zunächst, daß für jeden dieser Werte a auch

$$\big\{(x_1, \ldots, x_{n-1}) \in \mathbb{N}^{n-1} \;\big|\; (x_1, \ldots, x_{n-1}, a) \in D\big\}$$

ein Stufendesign ist, also Nullstellenmenge einer Menge $\mathcal{M}_a$ von Distraktionen von Monomen in $X_1, \ldots, X_{n-1}$. Offensichtlich ist dann D die Nullstellenmenge der Polynome $D(MX_n^{a+1})$ mit $M \in \mathcal{M}_a$ für $a = 0, \ldots, \ell_n$ und von $D(X_n^{\ell_n+1})$.

■

Zusammen mit dem vorigen Satz folgt

Korollar: Jedes Stufendesign hat minimalen Fächer.

■

§4: Fraktionen eines vollen faktoriellen Designs

Nun sei $\mathcal{F}$ eine Teilmenge eines vollen faktoriellen Designs D, und G sei die reduzierte universelle GRÖBNER-Basis von $I(D)$. Dann ist $I(D)$

eine Teilmenge von $I(\mathcal{F})$; es gibt also ein Erzeugendensystem von $I(\mathcal{F})$, das G enthält. Wir können versuchen, das zur Berechnung von Modellen zu $\mathcal{F}$ zu verwenden. Ein erstes technisches Hilfsmittel dazu ist das folgende

Lemma: $\mathcal{O} \subset \mathbb{T}$ sei ein Ordnungsideal. Dann gibt es eine eindeutig bestimmte minimale Teilmenge Min($\mathcal{O}$) von $\mathbb{T}$ derart, daß $\mathbb{T} \setminus \mathcal{O}$ genau aus den Vielfachen der Monome aus Min($\mathcal{O}$) besteht.

Beweis: Das von $\mathbb{T} \setminus \mathcal{O}$ erzeugte monomiale Ideal $I \lhd k[X_1, \ldots, X_n]$ enthält kein Monom aus $\mathcal{O}$, denn wie wir wissen, enthält ein monomiales Ideal genau die Monome, die durch eines der erzeugenden Monome teilbar sind. Da ein Ordnungsideal mit jedem Monom auch dessen sämtliche Teiler enthält, müste im Falle eines Monoms aus $\mathcal{O}$ in I das Erzeugendensystem $\mathbb{T} \setminus \mathcal{O}$ ein Element von $\mathcal{O}$ enthalten, was natürlich absurd ist.

Nach dem Lemma von DICKSON hat I ein Erzeugendensystem aus endlich vielen Monomen. Ein solches Erzeugendensystem ist minimal, wenn keines der Monome dort durch ein anderes teilbar ist. Ein gegebenes endliches Erzeugendensystem läßt sich problemlos auf ein minimales reduzieren; also gibt es solche minimalen Erzeugendensysteme.

Angenommen, $\{M_1, \ldots, M_p\}$ und $\{N_1, \ldots, N_s\}$ sind zwei solche minimale Erzeugendensysteme. Da jedes N_i im von den M_j erzeugten monomialen Ideal I liegt, muß N_i durch (mindestens) ein M_j teilbar sein. Da I auch von $N_1, \ldots, N_s$ erzeugt wird, muß umgekehrt M_j durch ein N_ℓ teilbar sein, d.h. $N_\ell | M_j | N_i$. Da wir nur minimale Erzeugendensysteme betrachten, folgt $N_\ell = N_i$, also insbesondere $N_i = M_j$. Jedes N_i liegt also im Erzeugendensystem $\{M_1, \ldots, M_p\}$, und da dies ein minimales Erzeugendensystem ist, folgt $\{M_1, \ldots, M_p\} = \{N_1, \ldots, N_q\}$, d.h. es gibt genau ein minimales Erzeugendensystem von I.

Da die Monome in I genau die aus $\mathbb{T} \setminus \mathcal{O}$ sind, besteht diese Menge somit genau aus den Vielfachen der Monome aus $\{M_1, \ldots, M_p\}$, so daß wir Min($\mathcal{O}$) = $\{M_1, \ldots, M_p\}$ setzen können.

■

Falls $\mathcal{O}$ Teilmenge von $\mathrm{Est}_\tau(D)$ für ein volles faktorielles Design D ist, können wir das lineare Modell zu $\mathcal{O}$ natürlich auf Grund von D

bestimmen; falls $\mathcal{O}$ allerdings deutlich kleiner als $\text{Est}_\tau(D)$ ist, sollte das auch mit deutlich geringerem Aufwand möglich sein, nämlich mit jeder Fraktion $\mathcal{F}$ von D mit $\text{Est}_\tau(\mathcal{F}) \supseteq \mathcal{O}$. Es genügt, wenn wir die minimalen unter diesen Fraktionen bestimmen, denn jede andere enthält mindestens eine von diesen.

Via GRÖBNER-Basen und Monomordnungen lassen sich (wenn auch mit beträchtlichem Aufwand) alle diese Fraktionen bestimmen; wir wollen uns hier aber mit einem weniger ambitiösen Ziel begnügen und nur die Fraktionen bestimmen, die Nullstellenmengen von Idealen sind, die von Distraktionen von Polynomen erzeugt werden. Die Theorie dazu finden wir bei

> LORENZO ROBBIANO, MARIA PIERA ROGANTIN: Full factorial designs and distracted fractions *in* BRUNO BUCHBERGER, FRANZ WINKLER [HRSG]: Gröbner bases and applications Linz, *Cambridge University Press,* 1998, Seite 473–482

wo auch Aussagen über die Anzahl der so zu findenden und der insgesamt existierenden Fraktionen $\mathcal{F}$ zu finden sind. Ansätze zur allgemeinen Lösung des Problems findet man in

> MASSIMO CARBOARA, LORENZO ROBBIANO: Families of Ideals in Statistics *in* KÜCHLIN [HRSG.]: Proceedings of the 1997 ISSAC, *ACM Press,* 1997, Seite 404–409

$\text{Min}(\mathcal{O})$ wird im allgemeinen eine ganze Reihe von Monomen enthalten, die in $\text{FM}_\tau\big(I(D)\big)$ liegen, und die uns nicht interessieren müssen, wenn wir uns auf Teilmengen $\mathcal{F} \subset D$ beschränken; wir definieren daher

Definition: $\text{CutOut}(\mathcal{O}) = \text{Min}(\mathcal{O}) \setminus \text{FM}_\tau\big(I(D)\big)$

Damit gilt

Satz: $D = S_1 \times \cdots \times S_n$ sei ein volles faktorielles Design, G sei die universelle reduzierte GRÖBNER-Basis von $I(D)$, und $\mathcal{O}$ sei ein Ordnungsideal, das in $\text{Est}_\tau(D)$ liegt, mit $\text{CutOut}(\mathcal{O}) = \{M_1, \ldots, M_r\}$. Die Vektoren $a^{(i)}$ seien so definiert, daß ihre Komponenten gleich den verschiedenen Elementen von S_i in irgendeiner Reihenfolge sind. Dann

bildet G zusammen mit den Polynomen $D(M_1), \ldots, D(M_r)$ bezüglich jeder Monomordnung τ eine GRÖBNER-Basis eines Designideals einer Fraktion $\mathcal{F}$ von D mit $\mathrm{Est}_\tau(\mathcal{F}) = \mathcal{O}$.

Beweis: Wie wir bereits wissen, sind die Elemente der GRÖBNER-Basis G von $I(D)$ die Distraktionen der Monome $X_i^{\#S_i}$ bezüglich der $a^{(i)}$. Daher besteht die Vereinigung von G mit der Menge der M_j nur aus Distraktionen von Monomen bezüglich fester Vektoren und ist damit nach obigem Satz bezüglich jeder Monomordnung eine GRÖBNER-Basis des davon erzeugten Ideals I. Für $\mathcal{F} = V(I)$ wird $\mathrm{FM}_\tau(I)$ erzeugt von den führenden Monomen der Elemente von G und den M_i, also von $\mathrm{CutOut}(\mathcal{O})$ und $\mathrm{FM}_\tau\big(I(D)\big)$ und damit von $\mathrm{Min}_\tau(\mathcal{O})$. Somit ist $\mathrm{Est}_\tau(\mathcal{F}) = \mathcal{O}$. ∎

Betrachten wir als Beispiel $D = \{0,1,2\} \times \{0,1\}^{10}$, das volle faktorielle Design für elf Faktoren, deren erster drei Stufen hat; alle weiteren haben nur zwei. D enthält somit $3 \times 2^{10} = 3\,072$ Punkte aus $\mathbb{R}^{11}$.

Wir erwarten nicht, daß jede der 3 072 Faktorkombinationen relevant ist und beschränken und auf das Modell

$$\mathcal{O} = \{1, X_1, \ldots, X_{11}, X_1X_2, X_1X_3, X_1X_4, X_1X_2X_3, X_2X_3, X_1^2\}\,.$$

Es gibt also nur eine Dreierinteraktion, und nur ein Monom kommt als Quadrat vor. Die bezüglich der Teilbarkeitsrelation maximalen Elemente von $\mathcal{O}$ sind $X_5, \ldots, X_{11}, X_1X_2X_3, X_1X_4$ und X_1^2. Die Menge $\mathrm{Min}(\mathcal{O})$ enthält somit alle Monome der Form MX_i, wobei M irgendeines dieser Monome ist.

$\mathrm{FM}_\tau\big(I(D)\big)$ wird erzeugt von X_1^3 und $X_2^2, \ldots, X_{10}^2$; $\mathrm{CutOut}(\mathcal{O})$ besteht also aus den davon verschiedenen Monomen aus $\mathrm{Min}(\mathcal{O})$.

Um kurze Polynome zu bekommen, wählen wir $a^{(1)} = (0,1,2)$ und $a^{(i)} = (0,1)$ für $i \geq 2$. Die Distraktionen der Monome sind also für alle Monome, die kein Quadrat enthalten, einfach die Monome selbst, und für die $X_1^2X_i$ sind es die Polynome $X_1(X_1 - 1)X_i$.

Der Nullpunkt und alle Punkte, bei denen genau eine Koordinate von Null verschieden ist, sind Nullstellen aller dieser Polynome; da x_1 auch

den Wert zwei annehmen kann, sind dies schon einmal dreizehn Lösungen.

Falls ein x_i mit $i \geq 5$ von Null verschieden ist, müssen alle anderen x_j verschwinden, da $X_i X_j$ eines der Erzeugenden des Ideals ist. Wegen der Polynome $D(X_1^2 X_j) = X_1(X_1 - 1)X_j$ müssen auch im Falle $x_1 = 2$ alle x_j mit $j > 1$ verschwinden.

Ist $x_4 = 1$, so kann wegen der Polynome $X_1 X_4 X_j$ auch $x_1 = 1$ sein, falls alle anderen x_j verschwinden; es gibt also außer dem Einheitspunkt noch den Punkt $(1, 0, 0, 1, 0, \ldots, 0)$.

Wenn alle x_i mit $i \geq 4$ verschwinden, ist für x_1 bis x_3 jede Kombination aus Nullen und Einsen möglich; dies ergibt acht Lösungen, von denen vier neu sind. Insgesamt haben wir also 18 Punkte, was erwartungsgemäß gleich der Elementanzahl von $\mathcal{O}$ ist.

Kapitel 4
Markov-Basen für Kontingenztests

Zu den Grundaufgaben der Statistik gehört auch die Frage nach dem Zusammenhang zwischen zwei oder mehreren Zufallsvariablen. Um beispielsweise zu testen, ob ein Medikament besser ist als ein anderes, teilt man die Probanden zufallsgesteuert ein in zwei Kontrollgruppen, die mit je einem der beiden Medikamente behandelt werden, und mißt den Erfolg. Falls die beiden Zufallsvariablen *Medikament* und *Erfolg* voneinander unabhängig sind, haben beide Medikamente gleich viel (oder wenig) Erfolg, andernfalls läßt sich mit einer gewissen Wahrscheinlichkeit eines der beiden als das bessere identifizieren.

§1: Kontingenztafeln

Für zwei unabhängige Zufallsvariablen X und Y mit Werten in $\{1,\dots,r\}$ *bzw.* $\{1,\dots,c\}$ ist $p(X=x,Y=y)=p(X=x)p(Y=y)$ für alle $(x,y)\in\{1,\dots,r\}\times\{1,\dots,c\}$. Wegen der Beziehungen

$$p(X=x)=\sum_{y=1}^{c}p(X=x,Y=y)$$

und

$$p(Y=y)=\sum_{x=1}^{r}p(X=x,Y=y)$$

können wir das auch so ausdrücken, daß $p(X=x,Y=y)$ nur abhängt von den beiden Summen

$$\sum_{y=1}^{c}p(X=x,Y=y)\quad\text{und}\quad\sum_{x=1}^{r}p(X=x,Y=y)\,.$$

Wir wählen eine natürliche Zahl n und betrachten n Werte (x_k, y_k) der Zufallsvariablen $X \times Y$. Dazu definieren wir eine neue Zufallsvariable $U = (U_{ij})_{\substack{i=1\dots r \\ j=1\dots c}}$, wobei U_{ij} die Anzahl von Paaren (x_k, y_k) mit $x_k = i$ und $y_k = j$ bezeichnet. Außerdem betrachten wir noch die Zufallsvariablen

$$U_{i+} = \sum_{j=1}^{c} U_{ij} \quad \text{und} \quad U_{+j} = \sum_{i=1}^{r} U_{ij}$$

für $i = 1, \dots, r$ und $j = 1, \dots, c$. Die Erwartungswerte der Zufallsvariablen U_{ij} sind dann

$$\mathbb{E}(U_{ij}) = np(X = i, Y = j) = n \cdot \sum_{j=1}^{c} p(X = i, Y = j) \cdot \sum_{i=1}^{r} p(X = i, Y = j)$$

Entsprechend sind die Erwartungswerte von U_{i+} und U_{+j} gleich

$$\mathbb{E}(U_{i+}) = n \cdot \sum_{j=1}^{c} p(X = i, Y = j) \quad \text{und} \quad \mathbb{E}(U_{+j}) = n \cdot \sum_{i=1}^{r} p(X = i, Y = j)\,.$$

Durch Vergleich der Erwartungswerte folgen die Beziehungen

$$\mathbb{E}(U_{ij}) = \frac{\mathbb{E}(U_{i+})\mathbb{E}(U_{+j})}{n}$$

$$\mathbb{E}(U_{i+}) = \sum_{j=1}^{c} \mathbb{E}(U_{ij}) \quad \text{und} \quad \mathbb{E}(U_{+j}) = \sum_{i=1}^{r} \mathbb{E}(U_{ij})\,.$$

Natürlich können wir schon wegen der Ganzzahligkeit der u_{ij} nicht erwarten, daß für einen beobachteten Wert u von U auch $u_{ij} = u_{i+}u_{+j}/n$ ist, aber die Abweichung zwischen den beiden Seiten sollte mit großer Wahrscheinlichkeit klein sein. Als Maß der Abweichung wählen wir die Zahl

$$\chi^2(u) = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(u_{ij} - \hat{u}_{ij})^2}{\hat{u}_{ij}} \quad \text{mit} \quad \hat{u}_{ij} = \frac{u_{i+}u_{+j}}{n}\,.$$

In der Statistik zeigt man, daß die Verteilung der Zufallsvariablen $\chi^2(U)$ asymptotisch gegen eine χ^2-Verteilung mit $(r-1)(c-1)$ Freiheitsgraden konvergiert, falls alle u_{ij} gegen unendlich gehen.

Leider können aus praktischen Gründen oft nur Experimente realisiert werden, bei denen zumindest einige der u_{ij} recht klein sind. Eine Faustregel besagt, daß man definitiv nicht mit der χ^2-Verteilung arbeiten sollte, wenn nicht alle u_{ij} mindestens gleich fünf sind.

§2: Fishers exakter Test

Das erste Verfahren, die Wahrscheinlichkeit für die Zufälligkeit der Abweichung der Werte von U_{ij} und $U_{i+}U_{+j}/n$ auch im Falle kleiner Werte einiger u_{ij} zu schätzen, war FISHERs exakter Test. Er betrachtet zu einer gegebenen Kontingenztafel (u_{ij}) alle Tafeln (v_{ij}) mit $v_{ij} \in \mathbb{N}_0$ und $v_{i+} = u_{i+}$ für alle i sowie $v_{+j} = u_{+j}$ für alle j. Für jede dieser Tafeln berechnet er das zugehörige χ^2 und schätzt die Wahrscheinlichkeit für die Zufälligkeit der Abweichung als den Anteil aller Tafeln, die zu keinem größeren χ^2 führen als die gegebene Tabelle (u_{ij}). Im Falle von Vierfeldertests ist das auch noch bei moderat großen Zeilen- und Spaltensummen praktikabel, denn offensichtlich sind durch diese Summen alle v_{ij} eindeutig festgelegt, sobald man einen dieser vier Werte kennt. Mit wachsender Zahl der Freiheitsgrade steigt allerdings auch der Aufwand für FISHERs exakten Test dramatisch an, so daß die Betrachtung aller Tabellen mit denselben Randverteilungen wie die gegebene Tabelle nicht mehr mit realistischen Aufwand möglich ist.

§3: Log-lineare Modelle

Bevor wir uns überlegen, wie wir in solchen Fällen vorgehen können, wollen wir zunächst die betrachtete Situation etwas verallgemeinern. Bei der Untersuchung auf Unabhängigkeit sollten die Erwartungswerte der U_{ij} nur abhängen von denen der U_{i+} und der U_{+j}. Auch im Falle abhängiger Größen kann es sein, daß es eine begrenzte Zahl von Linearkombinationen der U_{ij} gibt mit der Eigenschaft, daß die Erwartungswerte aller U_{ij} aus denen dieser Linearkombinationen berechnet werden können. Solche Situationen formalisiert der Begriff eines log-linearen Modells:

Definition: $X_1, \ldots, X_m$ seien Zufallsvariablen, und X_ℓ nehme Werte

aus der Menge $\{1,\dots,r_\ell\}$ an. Für

$$i=(i_1,\dots.i_m)\in\mathcal{R}\underset{\text{def}}{=}\{1,\dots,r_1\}\times\cdots\times\{1,\dots,r_m\}$$

sei $p_i = p(X_1 = i_1,\dots,X_m = i_m)$. Weiter sei $A\in\mathbb{Z}^{d\times\#\mathcal{R}}$ eine Matrix, deren Spalten alle die gleiche Summe haben. Das log-lineare Modell $\mathcal{M}_A$ zur Matrix A ist die Menge aller Wahrscheinlichkeitstabellen $(p_i)_{i\in\mathcal{R}})$ mit der Eigenschaft, daß der Zeilenvektor $(\log p_i)_{i\in\mathcal{R})}$ im von den Zeilenvektoren von A aufgespannten Untervektorraum von $\mathbb{R}^{\#\mathcal{R}}$ liegt.

Als Beispiel betrachten wir das obige Unabhängigkeitsmodell. Hier ist $\mathcal{R}=\{1,\dots,r\}\times\{1,\dots,c\}$ und

$$p_{(i,j)}=p(X=i,Y=j)=p(X=i)\cdot p(Y=j)\,.$$

Somit ist $\log p_{(i,j)}=\log p(X=i)+\log p(Y=j)\,.$

Ist A die $(r+c)\times rc$-Matrix, deren Spalte mit Index (i,j) in den Komponenten i und $r+j$ eine Eins stehen hat und sonst lauter Nullen, so ist

$$(logp_{(1,1)},\dots,\log p_{(r,c)})=\sum_{i=1}^{r}\log p(X=i)a^{(1)}+\sum_{j=1}^{c}\log p(Y=j)a^{(r+j)}\,,$$

wobei $a^{(\ell)}$ für den ℓ-ten Zeilenvektor von A steht. Für $r=3$ und $c=2$ etwa ist

$$A=\begin{pmatrix}1&1&0&0&0&0\\0&0&1&1&0&0\\0&0&0&0&1&1\\1&0&1&0&1&0\\0&1&0&1&0&1\end{pmatrix}$$

und

$$\begin{array}{lcccccc}&&(p_{(1,1)}&p_{(1,2)}&p_{(2,1)}&p_{(2,2)}&p_{(3,1)}&p_{(3,2)})\\=\log p(X=1)&\times&(1&1&0&0&0&0)\\+\log p(X=2)&\times&(0&0&1&1&0&0)\\+\log p(X=3)&\times&(0&0&0&0&1&1)\\+\log p(Y=1)&\times&(1&0&1&0&1&0)\\+\log p(Y=2)&\times&(0&1&0&1&0&1)\,.\end{array}$$

Wenn wir eine Kontingenztafel wie gerade eben als einen Vektor u auffassen, gibt uns das Matrixprodukt Au die sämtlichen Zeilen- und Spaltensummen; beim Unabhängigkeitsmodell hängen die Erwartungswerte der Variablen U_{ij} nur von diesen ab. Entsprechend erwarten wir bei einem beliebigen log-linearen Modell, daß auch dort die Erwartungswerte der Zufallsvariablen U_i, $i \in \mathcal{R}$, nur vom Zufallvsvariablenvektor AU abhängen sollten. Sobald wir wissen, wie sich die Erwartungswerte der U_i aus denen von AU berechnen lassen, können wir FISHERs exakten Test auch auf diese Situation verallgemeinern, indem wir zu einer gegebenen Tafel u aller Tafeln $v = (v_i)_{i\in\mathcal{R}}$ betrachten, für die $Av = Au$ ist.

Definition: Die *Faser* $\mathcal{F}(u)$ zu einer gegebenen Tafel $(u_i)_{i\in\mathcal{R}}$ bezüglich des Modells $\mathcal{M}_A$ ist die Menge aller $v \in \mathbb{N}_0^{\#\mathcal{R}}$ mit $Av = Au$.

Auch hier wird die Faser oft zu groß sein als daß wir χ^2 für jedes einzelne Element ausrechnen können, so daß wir uns mit einer Stichprobe begnügen müssen, die wir uns über eine Irrfahrt verschaffen wollen. Die einzelnen Schritte der Irrfahrt sind durch Elemente einer sogenannten MARKOV-Basis gegeben:

Definition: Eine MARKOV-Basis des log-linearen Modells $\mathcal{M}_A$ ist eine endliche Menge $\mathcal{B}$ von Tafeln $b \in \mathbb{Z}^{\#\mathcal{R}}$ mit $Ab = 0$, für die gilt: Für jede Tafel $u \in \mathbb{N}_0^{\#\mathcal{R}}$ und je zwei Elemente $v, v' \in \mathcal{F}(u)$ gibt es eine Folge von Elementen $b_1, \ldots, b_L \in \mathcal{B}$, so daß gilt:
1.) Für $\ell = 1, \ldots, L$ ist $v + b_1 + \cdots + b_\ell \in \mathbb{N}_0^{\#\mathcal{R}}$
2.) $v + b_1 + \cdots + b_L = v'$

Im Falle des Unabhängigkeitsmodells für zwei Zufallsvariablen können wir leicht eine MARKOV-Basis finden: Wir nehmen einfach alle Tafeln $b^{(ijk\ell)}$, $i, k = 1, \ldots, r$ und $j, \ell = i, \ldots, c$, deren Einträge allesamt verschwinden mit Ausnahme von

$$b_{ij}^{(ijk\ell)} = b_{k\ell}^{(ijk\ell)} = 1 \quad \text{und} \quad b_{i\ell}^{(ijk\ell)} = b_{kj}^{(ijk\ell)} = -1\,.$$

Im nächsten Paragraphen wollen wir uns überlegen, daß es für jedes log-lineare Modell eine MARKOV-Basis gibt.

§4: Markov-Basen und Ideale

Wir führen für jedes $i \in \mathcal{R}$ eine Variable p_i ein und betrachten den Polynomring $k[p_i | i \in \mathcal{R}]$. Jeder Tabelle $u \in \mathbb{N}_0^{\#\mathcal{R}}$ ordnen wir das Monom $p^u \underset{\text{def}}{=} \prod_{i \in \mathcal{R}} p_i^{u_i} \in k[p_i | i \in \mathcal{R}]$ zu.

Definition: *a)* Eine nichtleere Teilmenge $\mathcal{L} \subset \mathbb{Z}^{\#\mathcal{R}}$ heißt *Gitter,* wenn für je zwei Elemente $u, v \in \mathcal{L}$ auch $u + v$ und $-u$ in $\mathcal{L}$ liegen.
b) Eine endliche Teilmenge $\mathcal{B}$ eines Gitters $\mathcal{L}$ heißt MARKOV-Basis, wenn es zu je zwei Elementen $v, v' \in \mathbb{N}_0^{\#\mathcal{R}}$ mit $v' - v \in \mathcal{L}$ eine Folge von Elementen $b_1, \ldots, b_L \in \mathcal{B}$ gibt, so daß gilt:
1.) Für $\ell = 1, \ldots, L$ ist $v + b_1 + \cdots + b_\ell \in \mathbb{N}_0^{\#\mathcal{R}}$
2.) $v + b_1 + \cdots + b_L = v'$
c) Das *Gitterideal* zu $\mathcal{L} \subset \mathbb{Z}^{\mathcal{R}}$ ist

$$I_{\mathcal{L}} \underset{\text{def}}{=} \left(p^u - p^v \mid u, v \in \mathbb{N}_0^{\mathcal{R}} \setminus \{0\}, u - v \in \mathcal{L}\right).$$

Offensichtlich ist für ein log-lineares Modell $\mathcal{M}_A$ die Menge $\mathcal{L}$ aller $u \in \mathbb{Z}^{\#\mathcal{R}}$ mit $Au = 0$ ein Gitter, und eine MARKOV-Basis davon ist genau das, was im vorigen Paragraphen definiert wurde.

Für $u \in \mathbb{Z}^{\mathcal{R}}$ definieren wir $u^+, u^- \in \mathbb{N}_0^{\mathcal{R}}$ durch

$$u_i^+ = \max(u_i, 0) \quad \text{und} \quad u_0^- = \max(-u_i, 0).$$

Dann ist $u = u^+ - u^-$, und dies ist die einzige Darstellung von u als Differenz zweier Elemente $v, w \in \mathbb{N}_0^{\mathcal{R}}$, bei der für kein $i \in \mathcal{R}$ sowohl u_i als auch v_i von Null verschieden sind.

Satz: Eine Menge $\mathcal{B} \subset \mathcal{L}$ ist genau dann eine MARKOV-Basis von $\mathcal{L}$, wenn die Polynome $p^{b^+} - p^{b^-}$ mit $b \in \mathcal{B}$ das Ideal $I_{\mathcal{L}}$ erzeugen.

Der *Beweis* wird in drei Schritten geführt:

1. Schritt: $I_{\mathcal{L}} = I' \underset{\text{def}}{=} \left(p^{u^+} - p^{u^-} \mid u \in \mathcal{L}\right)$

Es ist klar, daß I' in $I_{\mathcal{L}}$ liegt, denn $u^+ - u^- = u$ liegt in $\mathcal{L}$. Wir müssen zeigen, daß es kein Element $f \in I_{\mathcal{L}}$ gibt, das nicht in I' liegt.

Angenommen, es gibt solche Polynome. Wir führen auf $\mathbb{N}_0^{\mathcal{R}}$ eine Monomordnung ein und betrachten die Menge aller führenden Monome von Polynomen $f \in I_{\mathcal{L}} \setminus I'$. Wegen der Wohlordnungseigenschaft von Monomordnungen enthält diese Menge ein minimales Element; f sei ein Polynom aus $I_{\mathcal{L}}$ mit diesem führenden Monom.

...

■

(vgl. BERND STURMFELS: Gröbner Bases and Convex Polytopes, *AMS University Lecture Series* **8***, Kapitel 4)*

Da jedes Ideal eines Polynomrings eine endliche GRÖBNER-Basis hat, zeigt dieser Satz die Existenz von MARKOV-Basen.

§5: Markov-Ketten

Ein häufiger zitiertes Vorbild einer Irrfahrt ist der Heimweg eines Betrunkenen über einen Platz mit vielen Laternen. Jedesmal, wenn er sich den Kopf an einer Laterne anschlägt, geht er (mit gleicher Wahrscheinlichkeit) nach rechts oder nach links. Seine Entscheidungen an den verschiedenen Laternen sind unabhängig voneinander, hängen also insbesondere nicht davon ab, *wie* er an die aktuelle Laterne gekommen ist. MARKOV-Ketten sind stochastische Prozesse mit entsprechenden Eigenschaften:

Definition: *a)* Ein stochastischer Prozess $X^{(1)}, X^{(2)}, \ldots$ aus Zufallsvariablen mit Werten in einer Menge $A = \{a_1, \ldots, a_m\}$ heißt MARKOV-Prozess oder MARKOV-Kette, wenn für alle $n \in \mathbb{N}$ gilt

$$\begin{aligned} p(X^{(n+1)} = x^{(n+1)} \mid X^{(1)} = x^{(1)}, \ldots, X^{(n)} = x^{(n)}) \\ = p(X^{(n+1)} = x^{(n+1)} \mid X^{(n)} = x^{(n)}) \,. \end{aligned}$$

b) Eine MARKOV-Kette heißt *zeitinvariant,* wenn die bedingte Wahrscheinlichkeit $p(X^{(n+1)} = y \mid X^{(n)} = x)$ nicht von n abhängt. In diesem Fall setzen wir setzen wir $p_{ij} = p(X^{(n+1)} = a_j \mid X_n = a_i)$ und bezeichnen die $m \times m$-Matrix P mit Einträgen p_{ij} als die *Übergangsmatrix* des Prozesses.

Der russische Mathematiker ANDREĬ ANDREEVIČ MARKOV (Андрей Андреевич Марков, 1856–1922) studierte in Sankt Petersburg, wo er später auch Professor wurde. Er beschäftigte sich zunächst hauptsächlich mit Zahlentheorie und Analysis; erst später folgen die wahrscheinlichkeitstheoretischen Arbeiten, für die er heute vor allem bekannt ist. Der Name Марков wird in lateinischen Buchstaben verschieden transkribiert; MARKOVs französische Arbeiten erschienen mit der Schreibweise MARKOFF; nach den klassischen deutschen Transkriptionsregeln müßte man MARKOW schreiben. Die Schreibweise MARKOV entspricht den englischen Regeln und scheint sich mittlerweile in der Mathematik ziemlich durchgesetzt zu haben.

Wir betrachten im folgenden nur zeitinvariante MARKOV-Ketten. Zeitinvarianz muß selbstverständlich nicht bedeuten, daß alle Zufallsvariablen $X^{(n)}$ dieselbe Verteilung haben, sie sind allerdings auch nicht unabhängig voneinander: Bezeichnet $p^{(n)} \in \Delta_{m-1}$ die Verteilung von $X^{(n)}$, so ist

$$p_j^{(n+1)} = p(X^{(n+1)} = a_j) = \sum_{i=1}^{m} p(X^{(n+1)} = a_j \mid X^{(n)} = a_i) = \sum_{i=1}^{m} p_i^{(n)} p_{ij} \; ;$$

wenn wir die $p^{(n)}$ als Zeilenvektoren schreiben, ist also $p^{(n+1)} = p^{(n)} P$.

Definition: *a)* Eine MARKOV-Kette heißt *stationär,* wenn alle $X^{(n)}$ dieselbe Wahrscheinlichkeitsverteilung haben.
b) Eine MARKOV-Kette heißt *reversibel,* wenn für alle i, j, n gilt: $p_i^{(n)} p_{ij} = p_j^{(n)} p_{ji}$.
c) Eine MARKOV-Kette heißt *irreduzibel,* wenn es für je zwei mögliche Werte a_i, a_j stets ein $r \in \mathbb{N}$ gibt, so daß $p(X^{(r+1)} = a_j \mid X^{(1)} = a_i) > 0$ ist.
d) Der Wert a_i heißt *aperiodisch,* wenn der ggT aller natürlicher Zahlen r mit $p(X^{(r+1)} = a_i \mid X^{(1)} = a_i) > 0$ gleich eins ist.

Satz: Ist eine MARKOV-Kette reversibel, irreduzibel und aperiodisch, so ist sie stationär.

Zum *Beweis* sei auf Lehrbücher zur Theorie der MARKOV-Ketten verwiesen, zum Beispiel

ACHIM KLENKE: Wahrscheinlichkeitstheorie, *Springer,*^{2}2008, Satz 18.18

Wir werden im folgenden, soweit nicht explizit etwas anderes gesagt ist, stets annehmen, daß unsere MARKOV-Ketten zeitinvariant sind. In diesem Fall können wir die Wahrscheinlichkeitsverteilungen aller Zufallsvariablen aus der von X_1 und der Übergangsmatrix berechnen: Ist allgemein $p_i^{(n)}$ die Wahrscheinlichkeit, mit der X_n dem Wert a_i annimmt, so ist

$$p(X_n = a_{i_0}, X_{n+1} = a_{i_1}, \ldots, X_{n+r} = a_{i_r})$$
$$= p(X_n = a_{i_0}) \prod_{\ell=1}^{r} p(X_{n+\ell} = a_{i_\ell} \mid X_{n+\ell-1} = a_{\ell-1}) \,.$$

Für $r = 1$ wird das zu

$$p(X_n = a_i, X_{b+1} = a_j) = p(X_n = a_i) p(X_{n+1} = a_j \mid X_n = a_i) = p_i^{(n)} p_{ij} \,,$$

was wir auch einfacher mit Matrizen und Vektoren formulieren können: Ist $\mathfrak{p}^{(n)} = (p_1^{(n)}, \ldots, p_m^{(n)})^T$ der Spaltenvektor der Wahrscheinlichkeitsverteilung zu X_n, so ist $\mathfrak{p}^{(n+1)} = A^T \mathfrak{p}^{(n)}$ und damit $\mathfrak{p}^{(n)} = \left(A^T\right)^{n-1} \mathfrak{p}^{(1)}$.

Somit bestimmen $\mathfrak{p}^{(1)}$ und die Übergangsmatrix A die Wahrscheinlichkeitsverteilungen aller X_n und erlauben damit auch die Berechnung der Wahrscheinlichkeiten aller Teiltupel, die von der MARKOV-Kette produziert werden.

§6: Die Markovketten-Montecarlo Methode

Wir gehen aus einem log-linearen Modell $\mathcal{M}_A$ und einer MARKOV-Basis $\mathcal{B}$ von $\mathrm{Kern}_{\mathbb{Z}}(A)$. Zu einer beobachteten Tabelle u betrachten wir eine Zufallsvariable U mit Werten in $\mathcal{F}(u)$ und interessieren uns für die Wahrscheinlichkeit

$$p\left(\chi^2(U) \geq \chi^2(u)\right) \,.$$

Der folgende Algorithmus von METROPOLIS liefert eine Schätzung dieser Wahrscheinlichkeit:

Initialisierung: Wähle ein beliebiges Element $u^{(1)}$ aus $\mathcal{F}(u)$.

t-ter Iterationsschritt: Wähle zufällig ein Element $b_t \in \mathcal{B}$, wobei jedes Element mit gleicher Wahrscheinlichkeit auftreten kann, sowie ein Element $\varepsilon_t \in \{+1, -1\}$, wobei beide Möglichkeiten Wahrscheinlichkeit $\frac{1}{2}$ haben. Dann ist

$$u^{(t+1)} = \begin{cases} u^{(t)} + \varepsilon_t b_t & \text{mit Wahrscheinlichkeit } q \\ u^{(t)} & \text{mit Wahrscheinlichkeit } 1 - q \end{cases}$$

$$\text{mit } q = \min\left(1, \frac{p\big(U = u^{(t)} + \varepsilon_t b_t \;\big|\; U \in \mathcal{F}(u)\big)}{p\big(U = u^{(t)} \;\big|\; U \in \mathcal{F}(u)\big)}\right) .$$

Zu dieser neuen Tabelle wird χ^2 berechnet, und anhand einer hinreichend langen Folge dieser Werte wird die obige Wahrscheinlichkeit geschätzt. Um die Schätzun möglichst unabhängig von $u^{(1)}$ zu machen, werden dabei meist die ersten Glieder der Folge ignoriert.

Um zu sehen, daß wir so tatsächlich eine unverzerrte Schätzung der Wahrscheinlichkeit bekommen, müssen wir zeigen, daß die erzeugte Folge von χ^2-Werten als erwartete Häufigkeit den gesuchten Anteil aller Elemente der Faser mit einem Wert von χ^2, der den für die gegebene Tabelle von u nicht übersteigt. Siehe dazu das oben zitierte Buch von KLENKE, insbesondere die Paragraphen 18.2 und 18.3.