Wolfgang K. Seiler

Einführung in die Algebraische Statistik

Vorlesung im Herbstsemester 2021
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. In der online Version werde ich alle bekannten Fehler korrigieren.

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

Verglichen mit den klassischen Teilgebieten der Mathematik ist die Algebraische Statistik sehr jung: Die ersten Arbeiten dazu erschienen erst in den letzten Jahren des vorigen Jahrhunderts, die meisten erst in diesem. Von daher ist auch noch nicht so ganz klar, womit sich die Algebraische Statistik alles beschäftigt: Grundsätzlich geht es um die Anwendung algebraischer Methoden auf statistische Fragestellungen, aber man ist noch weit davon entfernt genau zu wissen, welche statistischen Fragestellungen einer algebraischen Behandlung zugänglich sind und welche nicht. In dieser Vorlesung wird es um zwei Themengebiete gehen, über die in den vergangenen zwei Jahrzehnten erfolgreich gearbeitet wurde, die aber bei weitem nicht das gesamte Gebiet der Algebraischen Statistik umfassen.

Als erstes geht es darum, zu einer vorgegebenen (oder gezielt gewählten) Stichprobe statistische Modelle zu finden, deren Parameter auf Grund der vorliegenden Daten bestimmt werden können.

Das zweite Thema sind Kontingenztafeln, wie sie etwa bei Vierfeldertests auftreten. Unter geeigneten Bedingungen wie hinreichender Größe der Stichprobe und Annahmen über die zu Grunde liegende Verteilungsfunktion kann man hier Hypothesen durch χ^2-Tests überprüfen. Oft sind aber wegen eines zu hohen Aufwands oder zu hoher Kosten nur kleine Stichproben möglich, und über die Verteilungsfunktionen weiß man auch nicht immer Bescheid. Schon lange, bevor irgend jemand von Algebraischer Statistik redete, gab es auch dazu bereits alternative Verfahren wie beispielsweise FISHERs exakten Test, jedoch sind diese sehr aufwendig. Mit den sogenannten MARKOV-Basen liefert

die Algebraische Statistik eine weitere Methode, die iterativ und mit geringerem Aufwand als exakte Verfahren in solchen Fällen deutlich zuverlässigere Ergebnisse liefert als χ^2-Tests.

§1: Statistische Modelle

Bei manchen Fragestellungen ist auf Grund von Naturgesetzen bekannt, welche Form der Zusammenhang zwischen zwei oder mehreren Größen hat; unbekannt sind nur die Parameter. Mißt man beispielsweise für einen festen Leiter Spannung und Stromstärke, so ist nach dem OHMschen Gesetz die Spannung U stets gleich dem Widerstand R des Leiters mal der Stromstärke I. Gesucht ist der Wert des Widerstands R.

Bei wirtschafts- und sozialwissenschaftlichen Fragestellungen ist oft kein Gesetz bekannt; hier muß man versuchen, eine möglichst einfache Funktion zu finden, die die beobachteten Daten möglichst gut beschreibt.

Allgemein betrachten wir als Modelle Funktionen

$$f\colon \Theta \times \mathbb{R}^n \to \mathbb{R}^m ,$$

die jedem n-tupel $x = (x_1, \ldots, x_n)$ von Eingangsgrößen ein m-tupel $y = (y_1, \ldots, y_m)$ von Ausgangsgrößen zuordnen in Abhängigkeit von r Parametern $(\theta_1, \ldots, \theta_r) \in \Theta \subseteq \mathbb{R}^r$. Wir beschränken uns hier auf den häufigsten Fall $m = 1$ einer einzigen Ausgangsgröße y. Außerdem betrachten wir nur Modelle, die in den *Parametern* $\theta_1, \ldots, \theta_r$ linear sind. In den Eingangsgrößen $x_1, \ldots, x_n$ sollen die Modelle in der Algebraischen Statistik polynomial sein; wir lassen also keine transzendenten Funktionen zu. (Tatsächlich kann man mit den Methoden der Algebraischen Statistik auch gewisse transzendente Funktionen wie Exponentialfunktionen und trigonometrische Funktionen behandeln, indem man Funktionen der Art $e^{\lambda x}$ als Variablen betrachtet.)

Zur Klärung der Begriffe seien zunächst einige grundlegende Definitionen aus der Algebra kurz wiederholt:

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

1.) R ist 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 $f, g, h \in R$, 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* in dieser Vorlesung stets für einen *kommutativen Ring* stehen.

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.i Oft schreiben wir kurz X^α mit $\alpha = (\alpha_1, \ldots, \alpha_n)$.
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 $\mathbb{Q}[X, Y]$ aufzufassen als den Polynomring in Y über dem Ring $R = \mathbb{Q}[X]$; daher die allgemeinere Definition.

In der Statistik arbeitet man meist über dem Körper $\mathbb{R}$ der reellen Zahlen. In der Algebraischen Statistik wollen wir allerdings (soweit möglich) exakt algebraisch rechnen, und da gibt es im Falle von $\mathbb{R}$ Probleme: Da dieser Körper überabzählbar ist, können wir nur die Elemente einer Teilmenge vom Maß Null überhaupt durch endliche Formelausdrücke oder sonstige Beschreibungen angeben, und selbst für diese hat D. RICHARDSON 1968 gezeigt, daß es keinen Algorithmus geben kann, der entscheidet, ob zwei solche Beschreibungen dieselbe Zahl definieren oder nicht – selbst wenn man für die Beschreibungen nur sehr einfache Formelausdrücke zuläßt.

Wenn wir es mit nur endlich vielen Daten zu tun haben (und das ist beim konkreten Rechnen natürlich immer der Fall), gibt es allerdings stets einen abzählbaren Teilkörper von $\mathbb{R}$, der alle diese Zahlen enthält; oft wird uns sogar der Körper $\mathbb{Q}$ der rationalen Zahlen genügen. Wir legen uns daher nicht auf einen Körper fest, sondern arbeiten über einem beliebigen Körper k, der in dieser Vorlesung freilich immer ein relativ „kleiner“ Teilkörper von $\mathbb{R}$ oder (in ganz seltenen Fällen) eventuell auch $\mathbb{C}$ sein wird.

Modelle, die in den Eingangsgrößen polynomial sind, können aufgefaßt werden als Polynome, deren Koeffizienten Funktionen der Paramter θ_ν sind. Für jedes unserer Modelle gibt es also eine endliche Menge M von Monomen $M_\nu = X^{\alpha^{(\nu)}} \in k[X_1, \ldots, X_n]$, $\nu = 1, \ldots, s$, wobei

$\alpha^{(\nu)} = \left(\alpha_1^{(\nu)}, \ldots, \alpha_n^{(\nu)}\right)$ in $\mathbb{N}_0^n$ liegt, und für jedes ν haben wir eine Funktion $a_\nu \colon \Theta \to k$ derart, daß

$$f(\theta_1, \ldots, \theta_r, x_1, \ldots, x_n) = \sum_{\nu=1}^{s} a_\nu(\theta_1, \ldots, \theta_r) x_1^{\alpha_1^{(\nu)}} \cdots x_n^{\alpha_n^{(\nu)}}$$

ist. Wir wollen nur Modelle betrachten, die in $\theta_1, \ldots, \theta_r$ linear sind, und das auch noch auf die einfachst mögliche Weise: Für uns ist $r = s$ und $\Theta = \mathbb{R}^r$, also

$$f(\theta_1, \ldots, \theta_r, x_1, \ldots, x_n) = \sum_{\nu=1}^{r} \theta_\nu x_1^{\alpha_1^{(\nu)}} \cdots x_n^{\alpha_n^{(\nu)}} .$$

Tatsächlich machen wir noch eine weitere Einschränkung: Um möglichst kleine Monome zu bekommen, verlangen wir, daß die Menge der auf der rechten Seite auftretenden Monome ein *Ordnungsideal* ist im Sinne der folgenden

Definition: Eine endliche Menge von Monomen heißt *Ordnungsideal,* falls sie mit jedem Element auch dessen sämtliche Teiler enthält.

Liegt also das Monom X^2Y in einem Ordnungsideal, so muß dieses auch die Monome $1, X, Y, X^2$ und XY enthalten.

Die Bezeichnung *Ordnungsideal* ist leider etwas unglücklich, denn Ordnungsideale habe mit den Idealen in Ringen, die wir in Kürze betrachten werden, nicht das Geringste zu tun. Da sich die Bezeichnung aber eingebürgert hat, müssen wir damit leben.

Unter einer *Stichprobe* verstehen wir eine endliche Teilmenge $\mathcal{S}$ von k^{n+1} mit Elementen der Form $(x_1, \ldots, x_n, y)$, wobei y die beobachtete Ausgangsgröße zu den Eingangsgrößen $x_1, \ldots, x_n$ ist. Die dazu gehörige Menge der n-tupel $(x_1, \ldots, x_n) \in \mathbb{R}^n$ bezeichnen wir als ein *Design,* da diese Teilmenge oft *a priori* so gewählt wird, daß die zugehörigen y-Werte möglichst viel Information liefern. Es gibt verschiedene Methoden solche Designs zu konstruieren, darunter insbesondere auch viele algebraische, allerdings wird die Designtheorie (oder Theorie der optimalen Versuchsplanung) üblicherweise nicht zur Algebraischen Statistik gezählt und ist auch deutlich älter als diese. Sie wird auch in dieser Vorlesung keine Rolle spielen.

Ausgehend von einem Design $D \subset \mathbb{R}^n$ können wir eine Stichprobe auch auffassen als eine Funktion $D \to \mathbb{R}$, die als Funktion auf einer endlichen Menge einfach durch die Tabelle der Funktionswerte gegeben ist und keine speziellen mathematischen Eigenschaften haben muß.

Wenn wir eine Stichprobe $\mathcal{S} \subset \mathbb{R}^{n+1}$ und ein Modell $f\colon \Theta \times \mathbb{R}^n \to \mathbb{R}$ haben, wollen wir die Parameter $(\theta_1, \ldots, \theta_r) \in \Theta$ so bestimmen, daß im Idealfall gilt

$$f(\theta_1, \ldots, \theta_r, x_1, \ldots, x_n) = y \quad \text{für alle } (x_1, \ldots, x_n, y) \in \mathcal{S}\,.$$

Das wird im Allgemeinen nicht möglich sein, denn selbst wenn der Zusammenhang zwischen Eingangs- und Ausgangsgrößen wie beim OHMschen Gesetz feststeht, wird es auf Grund von Meßfehlern und Zufallsschwankungen (etwa der Temperaturabhängigkeit des Widerstands) fast nie ein $\theta \in \Theta$ geben, für das diese Gleichung für jedes Element der Stichprobe gilt. Wir müssen uns daher begnügen, ein θ zu finden, so daß

$$f(\theta_1, \ldots, \theta_r, x_1, \ldots, x_n) \approx y \quad \text{für alle } (x_1, \ldots, x_n, y) \in \mathcal{S}\,.$$

Der seit GAUSS übliche Ansatz zur Definition der ungefähren Gleichheit in diesem Zusammenhang ist die *Methode der kleinsten Quadrate:* Wir suchen ein $\theta \in \Theta$ derart, daß die Summe der quadratischen Abweichungen zwischen linker und rechter Seite minimal wird.

§2: Bestimmung der Parameter durch ein lineares Gleichungssystem

Wir interessieren uns in dieser Vorlesung nur für Modelle, die linear in den Variablen $\theta_1, \ldots, \theta_r$ sind. Für jedes Element $(x_1^{(j)}, \ldots, x_n^{(j)}, y^{(j)})$ der Stichprobe $\mathcal{S}$ gibt es daher Elemente $a_{j1}, \ldots, a_{jr}$ aus einem geeigneten Körper $k \subset \mathbb{R}$ derart, daß

$$f(\theta_1, \ldots, \theta_r, x_1^{(j)}, \ldots, x_n^{(j)}) = \sum_{\nu=1}^{r} a_{j\nu}\theta_\nu$$

ist. Bezeichnet θ den Spaltenvektor mit Komponenten $\theta_1, \ldots, \theta_r$ und y den mit Komponenten $y_1, \ldots, y_s$, sollte mit der Matrix $A = (a_{j\nu})$ aus $k^{s\times r}$ also gelten

$$A\theta = y\,.$$

Tatsächlich wird dieses lineare Gleichungssystem für θ meist überbestimmt und unlösbar sein. Nach GAUSS suchen wir stattdessen nach einem Vektor θ derart, daß der EUKLIDische Abstand zwischen den beiden Vektoren $A\theta$ und y minimal wird.

Falls das Gleichungssystem lösbar ist, ist dieser Abstand für jeden Lösungsvektor gleich Null, und wir sind fertig – kürzer kann kein Abstand sein.

Andernfalls ist für jeden Vektor θ, und damit auch für den, den wir suchen, das Produkt $A\theta$ von y verschieden; für ein optimales θ sei es etwa gleich $\bar{y}$. Dann ist $\bar{y}$ ein Vektor, der sich in der Form $A\theta$ darstellen läßt, und unter allen solchen Vektoren ist er derjenige, für den die Länge des Differenzvektors $\bar{y} - y$ minimal ist.

Die Matrix $A \in k^{s\times r}$ definiert eine lineare Abbildung

$$\varphi\colon k^r \to k^s; \quad \theta \mapsto A\theta\,;$$

deren Bildraum sei U. Falls die rechte Seite y in U liegt, ist das Gleichungssystem lösbar; andernfalls suchen wir einen Vektor $\theta \in k^r$, für den die Länge des Vektors $A\theta - y$ minimal wird. Da die Vektoren, die sich in der Form $A\theta$ darstellen lassen, genau die Vektoren aus U sind, ist somit $A\theta = \pi_U(y)$ die orthogonale Projektion von y nach U. Diese könnten wir *im Prinzip* bestimmen, indem wir die QR-Zerlegung von A berechnen, denn dann sind die ersten Spalten von Q eine Basis von U, die durch die weiteren Spalten zu einer Basis von ganz k^s ergänzt wird; danach haben wir ein lösbares lineares Gleichungssystem.

Wir wollen uns überlegen, wie wir θ auch ohne die rechnerisch aufwendige QR-Zerlegung bestimmen können.

Für den gesuchten Vektor θ (oder für die gesuchten Vektoren θ) ist $A\theta = \varphi_U(y)$. Da $A\theta$ bereits in U liegt, ist $\pi_U(A\theta) = A\theta$, also ist die Gleichung $A\theta = \pi_U(y)$ äquivalent zu

$$\pi_U(A\theta) = \pi_U(y) \quad \text{oder} \quad A\theta - y \in \operatorname{Kern}\pi_U = U^\perp\,.$$

Das orthogonale Komplement $U^\perp$ von U besteht aus allen Vektoren $y \in k^n$, die senkrecht stehen auf U, für die also gilt

$$\langle A\theta, y\rangle = 0 \quad \text{für alle } \theta \in k^m\,.$$

Wie aus der Linearen Algebra bekannt, gilt für Skalarprodukte die Gleichung

$$\langle A\theta, y\rangle = \left\langle \theta, A^T y\right\rangle ,$$

wobei A^T die zu A transponierte Matrix bezeichnet. (Falls wir über einem nichtreellen Teilkörper der komplexen Zahlen arbeiten, gilt Entsprechendes, wenn wir an Stelle der transponierten Matrix deren komplex konjugierte Matrix nehmen, also die adjungierte Matrix $A^* = \overline{A^T}$.) y liegt also genau dann in $U^\perp$, wenn $A^T y$ senkrecht steht auf allen Vektoren $\theta \in k^r$. Ein solcher Vektor aus k^r ist insbesondere $A^T y$ selbst; wegen der positiven Definitheit des Skalarprodukts ist also $A^T y = 0$. Da aus $A^T y = 0$ für alle $\theta \in k^r$ folgt, daß $\left\langle \theta, A^T y\right\rangle$ verschwindet, ist damit

$$U^\perp = \left\{ y \in k^s \;\middle|\; A^T y = 0 \right\} .$$

$A\theta - y$ liegt daher genau dann im Kern von π_U, wenn $A^T(A\theta - y) = 0$ ist oder, anders ausgedrückt, wenn θ eine Lösung des linearen Gleichungssystems

$$(A^T A)\,\theta = A^T y$$

ist. Dieses Gleichungssystem läßt sich schnell aufstellen und dann, falls es Lösungen gibt, nach GAUSS lösen.

Als Beispiel betrachten wir den einfachsten Fall der linearen Regression, die Bestimmung einer Ausgleichsgeraden. Hier erwarten wir einen linearen Zusammenhangs $\theta_1 x + \theta_2 = y$ zu N Wertepaaren $(x_i, y_i) \in \mathbb{R}^2$, wobei N sinnvollerweise größer als zwei sein sollte. Wir haben dann N Gleichungen

$$\theta_1 x_i + \theta_2 = y_i$$

mit unbekannten Parametern θ_1 und θ_2 und bekannten Zahlen x_i und y_i. Wir haben also ein lineares Geichungssystem von N Gleichungen in den beiden Variablen θ_1 und θ_2.

Fassen wir die Werte x_i zusammen zu einem Vektor $x \in \mathbb{R}^N$ und die y_i zu einem Vektor $y \in \mathbb{R}^N$, so läßt sich dieses Gleichungssystem kurz schreiben als

$$\theta_1 x + \theta_2 1 = y ,$$

wobei $1 \in \mathbb{R}^N$ jenen Vektor bezeichnet, dessen sämtliche Komponenten gleich eins sind.

Die Matrix des Gleichungssystems ist somit die $N \times 2$-Matrix A mit Spalten x und 1. A^T ist somit die $2 \times N$-Matrix, in deren erster Zeile die x_i stehen, während in der zweiten lauter Einser stehen. Somit ist

$$A^T A = \begin{pmatrix} \langle x, x\rangle & \langle x, 1\rangle \\ \langle x, 1\rangle & \langle 1, 1\rangle \end{pmatrix} \quad \text{und} \quad A^T y = \begin{pmatrix} \langle x, y\rangle \\ \langle 1, y\rangle \end{pmatrix} .$$

Das Gleichungssystem wird also zu

$$\langle x, x\rangle\, \theta_1 + \langle x, 1\rangle\, \theta_2 = \langle x, y\rangle \quad \text{und} \quad \langle x, 1\rangle\, \theta_1 + N\theta_2 = \langle 1, y\rangle .$$

Seine Matrix ist genau dann singulär, wenn ihre Determinante verschwindet, wenn also $N \cdot \langle x, x\rangle = \langle x, 1\rangle^2$ ist. Nach der CAUCHY-SCHWARZschen Ungleichung ist

$$|\langle 1, x\rangle| \le |1| \cdot |x| = \sqrt{N}\,|x| , \quad \text{also} \quad |\langle 1, x\rangle|^2 \le N \cdot \langle x, x\rangle .$$

Ausgeschrieben ist das die Ungleichung

$$\left(\sum_{i=1}^{N} x_i\right)^2 \le N \sum_{i=1}^{N} x_i^2 ,$$

und bekanntlich wird die CAUCHY-SCHWAZsche Ungleichung genau dann zur Gleichung, wenn die beiden Vektoren linear abhängig sind. Im Falle von x und 1 ist das genau dann der Fall, wenn alle x_i denselben Wert haben, was in praktischen Anwendungen fast nie der Fall sein dürfte. In diesem Fall ist die erste Gleichung ein Vielfaches der zweiten, und es gibt unendlich viele Lösungen. In allen anderen Fällen ist die Matrix invertierbar, so daß es eine eindeutige Lösung gibt.

Führen wir die in der Ausgleichsrechnung traditionell benutzten Abkürzungen

$$[x^r] = \sum_{i=1}^{N} x_i^r , \quad [y^r] = \sum_{i=1}^{N} x_i^r \quad \text{und} \quad [x^r y^s] = \sum_{i=1}^{N} x_i^r y_i^s$$

ein, so erhält das Gleichungssystem die übersichtlichere Gestalt

$$[x^2]\theta_1 + [x]\theta_2 = [xy] \quad \text{und} \quad [x]\theta_1 + N\theta_2 = [y] .$$

Subtraktion von $[x]/[x^2]$ mal der ersten Gleichung von der zweiten führt auf

$$\left(N - \frac{[x]^2}{[x^2]}\right)\theta_2 = [y] - \frac{[x]}{[x^2]}[xy]$$

oder $(N[x^2] - [x]^2)\theta_2 = [y][x^2] - [x][xy]$, d.h.

$$\theta_2 = \frac{[y][x^2] - [x][xy]}{N[x^2] - [x]^2} \,.$$

(Man beachte, daß im Falle der eindeutigen Lösbarkeit sowohl $[x^2] > 0$ als auch $N[x^2] - [x]^2 > 0$ ist.)

Einsetzen von θ_2 in die erste Gleichung ergibt dann auch

$$\theta_1 = \frac{[xy] - [x]\theta_2}{[x^2]} \,.$$

Zurück zu unserem Ausgangsproblem: Wir haben eine Stichprobe

$$\mathcal{S} = \left\{(x_1^{(j)}, \ldots, x_n^{(j)}, y^{(j)}) \;\middle|\; j = 1, \ldots, m\right\} \subset \mathbb{R}^{n+1}$$

und suchen dazu Modelle, deren Parameter sich auf Grund der Stichprobe schätzen lassen. Unsere Strategie dabei wird folgende sein: Wir bestimmen zunächst alle durch Ordnungsideale gegebenen Modelle, deren Parameter sich an Hand der Stichprobe berechnen lassen. Diese Modelle werden, da wir exakte Gleichheit fordern, meist viel zu viele Parameter enthalten. Nach Berechnung dieser Parameter für jedes der erhaltenen Modellen sehen wir hoffentlich, welche Monome jeweils eine wichtige Rolle spielen, d.h. also Koeffizienten haben, die sich deutlich von der Null unterschreiden, und welche nicht. Je nach Anwendung liefert uns auch irgendeine Theorie Aussagen darüber, welche Monome wichtig sein sollten. Wenn wir uns auf die wichtigen Monome beschränken, erhalten wir jeweils ein Modell, für das wir die Parameter nicht mehr so bestimmen können, daß Gleichheit herrscht, aber darauf können wir die gerade betrachtete Methode der kleinsten Quadrate anwenden. Wenn wir das für verschiedene Modelle durchführen, können

wir die EUKLIDischen Abstände zwischen den linken und rechten Seiten vergleichen und erhalten auch daraus Hinweise, wie gut die einzelnen Modelle zur Beschreibung des Zusammenhangs zwischen den Eingangs- und den Ausgangsgrößen geeignet sind.

Für den ersten Schritt, das Auffinden aller geeigneter Ordnungsideale zu einer gegebenen Stichprobe, verwendet die Algebraische Statistik die 1966 von BRUNO BUCHBERGER in seiner Dissertation eingeführten und nach seinem Lehrer benannten GRÖBNER-Basen, mit denen wir uns daher als nächstes beschäftigen müssen.

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. GRÖBNER-Basen liefern zu einem gegebenen nichtlinearen Gleichungssystem ein einfacheres System mit der gleichen Lösungsmenge; es ist eine Art Verallgemeinerung der Treppengestalt, die der GAUSS-Algorithmus liefert. Zumindest bei endlichen Lösungsmengen lassen sich diese auch konkret angeben – sofern wir die Nullstellen von Polynomen einer Veränderlichen explizit berechnen 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

$$a_1X_1 + \cdots + a_nX_n = b$$

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

Bei der Lösung nichtlinearer Gleichungssysteme können wir versuchen, ähnlich vorzugehen. Angenommen, wir suchen die Menge der gemeinsamen Nullstellen der beiden Polyonome

$$f = X^2Y^2 + 2X^3 - 3X^2 - X \quad \text{und} \quad g = Y^2 + X - 3\,,$$

also die Menge

$$\mathcal{L} = \left\{(x,y) \in \mathbb{R} \;\middle|\; f(x,y) = g(x,y) = 0\right\}.$$

Wir könnten den Term X aus beiden Gleichungen eliminieren, indem wir die Summe $f + g$ betrachten, aber offensichtlich hilft uns das nicht wirklich weiter. Nützlicher wäre es wahrscheinlich, einen der „größeren“ Terme zu eliminieren. Linearkombinationen mit Skalaren als Koeffizienten helfen uns dabei nicht weiter. Wenn wir aber g mit X^2 multiplizieren erhalten wir das Polynom $X^2Y^2 + X^3 - 3X^2$, das genau wie f den Term X^2Y^2 enthält, und

$$f - X^2 g = X^3 - X = X(X+1)(X-1)$$

ist in der Tat einfacher als die Ausgangspolynome f und g: Die Differenz hängt nur noch von X ab und verschwindet bei $x = 0$ und $x = \pm 1$. Setzen wir diese drei Werte nacheinander in die beiden Polynome ein, erhalten wir für $x = 0$

$$f(0,y) = 0 \quad \text{und} \quad g(0,y) = y^2 - 3 \quad \text{mit Lösung} \quad y = \pm\sqrt{3}\,,$$

Für $x = 1$ erhalten wir

$$f(1,y) = y^2 - 2 \quad \text{und} \quad g(0,y) = y^2 - 2 \quad \text{mit Lösung} \quad y = \pm\sqrt{2}\,,$$

und für $x = -1$ schließlich erhalten wir

$$f(-1,y) = y^2 - 4 \quad \text{und} \quad g(0,y) = y^2 - 4 \quad \text{mit Lösung} \quad y = \pm 2\,.$$

Die Lösungsmenge ist somit

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

Im Gegensatz zum Fall der linearen Gleichungssysteme sollten wir uns im nichlinearen Fall also nicht darauf beschränken, Gleichungen mit Skalaren zu multiplizieren und entsprechende Linearkombinationen zu betrachten, sondern wir sollten die Multiplikation mit *beliebigen* Polynomen zulassen. Dies führt auf den Begriff des *Ideals* in einem Ring:

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 $I = \{nq \mid q \in \mathbb{Z}\}$.

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 für das praktische Rechnen nur endlich erzeugte Ideale interessieren, möchte ich mich auf diesen Fall beschränken; die Verallgemeinerung auf beliebige Mengen M sollte für jeden, der den nachfolgenden Beweis verstanden hat, offensichtlich sein.

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

Beweis: Da jedes Ideal, das $f_1, \ldots, f_m$ enthält, auch für beliebige Elemente $r_1, \ldots, r_m \in R$ die Produkte $r_i f_i$ 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 der einzelnen f_i addieren müssen. Wenn wir schließlich ein Element der rechten Seite 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 $V(f_1, \ldots, f_m)$. Falls wir uns für Lösungen aus einem größeren Körper $K > k$ interessieren, schreiben wir entsprechend $V_K(I)$ und $V_K(f_1, \ldots, f_m)$.

(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. einfache Gegenbeispiele können wir schon bei nur einer Gleichung in einer Variablen finden:

Die Polynome $X^2+1, X^2+2, X^2+3, \ldots \in \mathbb{R}[X]$ haben allesamt keine reellen Nullstellen, Trotzdem sind die Ideale

$$\left(X^2+1\right), \left(X^2+1\right), \left(X^2+1\right), \ldots$$

natürlich allesamt verschieden. Dieses Problem verschwindet allerdings, wenn wir uns nicht mehr nur auf reelle Nullstellen beschränken, sondern auch komplexe zulassen.

Anders ist das bei den Polynomen $X, X^2, X^3, \ldots$. Egal über welchem Körper wir arbeiten, hat jedes dieser Polynome die Null als einzige Nullstelle, aber trotzdem sind die Ideale $\left(X^d\right) \lhd k[X]$ für verschiedene Werte von d verschieden.

Für das Problem, zu einer gegebenen Stichprobe die damit identifizierbaren Modelle zu finden, betrachtet die Algebraische Statistik das zugehörige Design als Lösungsmenge eines nichtlinearen Gleichungssystems. Es ist daher wichtig, den Zusammenhang zwischen Gleichungssystemen (oder Idealen) und deren Lösungsmengen zu kennen. Damit beschäftigt sich der Rest dieses Paragraphen.

Als erstes definieren wir die Summe und das Produkt zweier Ideale:

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 im Komplexen gibt es Gegenbeispiele: So ist etwa das Polynom $X^2 + Y^2 - Z^2 \in \mathbb{C}[X, Y, Z]$ irreduzibel, also nicht als Produkt zweier nichtkonstanter Faktoren darstellbar, aber es liegt trotzdem im Produkt des Ideals $I = (X, Y, Z)$ mit sich selbst. Im Produkt 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 $f \in J$. Da I eine Teilmenge von J ist, verschwindet $f(x_1, \ldots, x_n)$ erst recht für alle $f \in I$. Somit liegt. $(x_1, \ldots, x_n)$ 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. Nach *a)* liegt $V(I) \cup V(J)$ daher 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, die zudem noch den praktischen Vorteil eines im Allgemeinen deutlich kleineren Grads aufweist.

BUCHBERGERs 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 damals höchste Auszeichnung der Mathematik. (Seit 2003 vergibt die Norwegische Akademie der Wissenschaften den wie einen Nobelpreis dotierten Abelpreis.)

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

$$f = a_dX^d + a_{d-1}X^{d-1} + \cdots + a_1X + a_0 \quad \text{und}$$
$$g = b_eX^e + b_{e-1}X^{e-1} + \cdots + b_1X + b_0$$

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

$$f = c_\delta X^\delta + c_{\delta-1}X^{\delta-1} + \cdots + c_1X + c_0 \quad \text{mit} \quad c_\delta \neq 0\,,$$

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

$$g = b_eX^e + b_{e-1}X^{e-1} + \cdots + b_1X + b_0$$

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

Den Grad deg f von f hatten wir in Kapitel 0 als den höchsten Grad eines Monoms von f definiert; 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 derer 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^β, 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 Anord-

nung 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 umkehren, 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.

Diese vier Beispiele spielen bei vielen klassischen Anwendungen von GRÖBNER-Basen eine Hauptrolle, es gibt aber noch unendlich viele weitere mögliche Monomordnungen. Für unsere Zwecke in der Algebraischen Statistik brauchen wir *alle* diese Monomordnungen. Zum Glück wird sich aber zeigen, daß stets eine endliche Teilmenge davon ausreicht, um alle identifizierbaren Modelle zu bestimmen.

Für das folgende werden wir noch einige Eigenschaften einer (beliebigen) 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 FM(fg) = FM(f) · 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 irgendeiner 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 $f, f_1, \ldots, f_m \in R$ und eine Monomordnung auf 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

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

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:

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

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 $\alpha \in A, \gamma \in \mathbb{N}_0^n$ 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 kamen 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

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

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 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 FM($X^\gamma f$) = X^γ FM(f), da für jede Monomordnung gilt $\alpha < \beta \Longrightarrow \alpha + \beta < \alpha + \gamma$. Somit wird FM($I$) erzeugt von endlich vielen Monomen der Form 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 der

Hilbertsche 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 damals auch international führende Zentrum der deutschen 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 Ideal. 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 FM(f_i) im allgemeinen nicht ausreichen, um das monomiale Ideal FM(I) zu erzeugen: Für $I = (f_1, f_2)$ mit $f_1 = X^2 + X$ und $f_2 = X^2 + Y$ etwa ist bezüglich der lexikographischen Ordnung FM(f_1) = FM(f_2) = X^2, da aber beispielsweise auch $f_1 - f_2 = X - Y$ in I liegt, enthält FM(I) zumindest auch noch dessen führenden Term X, so daß X^2 allein nicht zur Erzeugung von FM(I) ausreicht. Wir müssen daher Linearkombinationen der f_i finden, deren führende Monome durch kein führendes Monom eines f_i teilbar sind, und zumindest diese noch mit dazu nehmen.

BUCHBERGERs Idee dazu orientiere sich an den Eliminationsschritten im GAUSS-Algorithmus, die er mit den sogenannten S-Polynomen verallgemeinerte auf Linearkombinationen mit Polynomen als Koeffizienten: Seien $f, g \in R$ zwei Polynome; FM(f) = X^α und FM(g) = X^β 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. $\alpha^{(i)} + \text{multideg } f_i = \delta$ für $i = 1, \ldots, m$. Falls 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

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

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 $\gamma^{(ij)}$, 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 das Kriterium von BUCHBERGER, wonach ein Erzeugendensystem eines Ideals genau dann eine GRÖBNER-Basis ist, wenn jedes S-Polynom zweier Erzeugender bei der Division durch das Erzeugendensystem den Rest Null ergibt. Da wir gesehen haben, daß es durchaus von der Reihenfolge der Divisoren abhängen kann, ob wir einen verschwindenden oder nichtverschwindenden Rest erhalten, wollen wir die Bedingung „Rest Null“ etwas abschwächen:

Definition: Ein Polynom f reduziert auf Null bezüglich der Polynome $f_1, \ldots, f_m$, wenn es Polynome $h_1, \ldots, h_m$ gibt, für die gilt
1.) $f = h_1 f_1 + \cdots + h_m f_m$
2.) Bezüglich der betrachteten Monomordnung ist $\mathrm{FM}(h_i f_i)$ für jedes i kleiner oder gleich $\mathrm{FM}(f)$.

Wenn der Divisionsalgorithmus Rest Null liefert, ist diese Bedingung offensichtlich erfüllt, aber sie ist etwas allgemeiner und sie reicht aus für BUCHBERGERs Kriterium:

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 $S(f_i, f_j)$ bezüglich $f_1, \ldots, f_m$ auf Null reduziert.

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 $S(f_i, f_j)$ bezüglich $f_1, \ldots, f_m$ auf Null reduzieren. Wir wollen zeigen, daß $f_1, \ldots, f_m$ dann eine GRÖBNER-Basis ist, daß also die führenden Monome FM(f_1), …, FM(f_m) das Ideal 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 das führende Monom von f gleich dem führenden Monom mindestens eines Produkts $h_i f_i$ und somit ein Vielfaches des führenden Monoms von FM(f_i), so daß FM(f) im von den FM(f_i) erzeugten Ideal liegt.

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 als Linearkombinationen der f_i darstellen, wobei das führende Monom eines jeden Summanden in so einer Linearkombination höchstens so groß ist wie das führende Monom des jeweiligen S-Polynoms. 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 er-

halten 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 wäre 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 J_m 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, denn wenn eines dieser Polynome einmal Divisionsrest Null hatte, hat es auch bei jedem weiteren Durchgang Divisionsrest Null, da es ja wieder durch dieselben Polynome (plus einiger neuer) dividiert wird.

Der Hauptaufwand beim BUCHBERGER-Algorithmus besteht in der Berechnung und Reduktion der S-Polynome. Falls wir daher überflüssige Anwendungen des Divisionsalgorithmus im Voraus erkennen können, spart das viel Aufwand. Ein einfaches Kriterium dazu ist

Lemma: Sind FM(f) und FM(g) teilerfremd, so reduziert $S(f, g)$ modulo f, g auf Null.

Beweis: Aus der Definition eines S-Polynoms folgt sofort, daß sich an $S(f, g)$ nicht ändert, wenn wir f und/oder g mit einer Konstanten multiplizieren: Da wir in beiden Termen durch den führenden Köffizienten dividieren, ändert sich nichts am jeweiligen Produkt. Daher können wir o.B.d.A. davon ausgehen, daß sowohl f als auch g den führenden Koeffizienten eins haben.

Wenn die führenden Monome teilerfremd sind, ist ihr kleinstes gemeinsames Vielfaches gleich ihrem Produkt. Schreiben wir f = FM(f) + p

und $g = \mathrm{FM}(g) + q$, ist daher

$$\begin{aligned} S(f,g) &= \frac{\mathrm{FM}(f)\,\mathrm{FM}(g)}{\mathrm{FT}(f)} \cdot f - \frac{\mathrm{FM}(f)\,\mathrm{FM}(g)}{\mathrm{FT}(g)} \cdot g \\ &= \mathrm{FM}(g) \cdot f - \mathrm{FM}(f) \cdot g \\ &= (g-q) \cdot f - (f-p) \cdot g \\ &= gf - qf - fg + pg = -qf + pg\,. \end{aligned}$$

Damit ist $S(f,g)$ als Linearkombination von f und g dargestellt; um zu sehen, daß es modulo f und g auf Null reduziert, müssen wir noch zeigen, daß sich in der Summe $-qf + pg$ die führenden Terme nicht wegheben können. Dazu reicht es zu zeigen, daß die führenden Monome der beiden Summanden verschieden sind.

Wäre dies nicht der Fall, so wäre $\mathrm{FM}(q)\,\mathrm{FM}(f) = \mathrm{FM}(p)\,\mathrm{FM}(g)$. Da $\mathrm{FM}(f)$ und $\mathrm{FM}(q)$ teilerfremd sind, müßte also $\mathrm{FM}(q)$ ein Teiler von $\mathrm{FM}(g)$ sein und $\mathrm{FM}(p)$ einer von $\mathrm{FM}(f)$. Beides ist aber unmöglich, denn $\mathrm{FM}(f)$ ist echt gröser als $\mathrm{FM}(p)$ und $\mathrm{FM}(g)$ echt größer als $\mathrm{FM}(q)$, während ein Teiler eines Monoms bezüglich jeder Monomordnung kleiner ist als dieses. Damit ist das Lemma bewiesen. ∎

Teilerfremdheit der führenden Monome ist natürlich äquivalent dazu, daß es keine Variable gibt, die in *beiden* führenden Monomen vorkommt.

Es gibt inzwischen auch zahlreiche nicht offensichtliche Verbesserungen und Optimierungen des BUCHBERGER-Algorithmus. Wir wollen uns aber mit dem Prinzip begnügen um dafür mehr Zeit für Themen zu bekommen, die für die algebraische Statistik relevanter sind.

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:

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

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

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

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

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

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

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

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 $\mathrm{FM}(f_1) = X^3$, $\mathrm{FM}(f_2) = X^2Y$, und $\mathrm{FM}(f_3) = X^2$ 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 eine *minimale* GRÖBNER-Basis mit

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

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 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 FM(g) ein Teiler von 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 g_3 bestehen alle Basispolynome nur aus dem führendem Term, und bei g_3 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.