Nichtlineare Gleichungssysteme und ihre Lösungsmengen

Wenn ein lineares Gleichungssystem über einem Körper k unendlich viele Lösungen $(x_1, \ldots, x_n) \in k^n$ hat, lassen sich Konstanten a_i und b_{ij} aus k finden mit $i = 1, \ldots, n$ und $j = 1, \ldots, m$, so daß die Lösungen genau die n-tupel sind, die sich darstellen lassen in der Form

$$x_i = a_i + b_{i1}t_1 + \cdots + b_{im}t_m$$

mit m frei wählbaren Parametern $t_1, \ldots, t_m$ aus k. Die Lösungen lassen sich also als lineare Funktionen geeigneter Parameter darstellen.

Entsprechend könnte man bei nichtlinearen Gleichungssystemen mit unendlicher Lösungsmenge versuchen, die Lösungen als Polynomfunktionen geeigneter Parameter darzustellen. In manchen Fällen ist das tatsächlich möglich; für das Polynom $Y - X^2$ beispielsweise ist die Lösungsmenge eine Parabel, bestehend aus allen Punkten $(x, y) \in k^2$ mit $x = t$ und $y = t^2$ für ein geeignetes $t \in k$.

Auch in etwas komplizierteren Fällen kann das noch gelegentlich funktionieren. Bei der Nullstellenmenge des Polynoms $Y^2 - X^3$ etwa, der NEILschen Parabel, sind alle Paare (x, y) mit $x = t^2$ und $y = t^3$ Lösungen, und umgekehrt läßt sich auch für jede Lösung (x, y) ein entsprechender t-Wert finden: Ist $x = 0$, muß auch $y = 0$ sein, und $t = 0$ liefert das Paar $(0, 0)$. Für $x \neq 0$ können wir $t = y/x$ setzen; dann ist

$$t^2 = \frac{y^2}{x^2} = \frac{x^3}{x^2} = x \qquad \text{und} \qquad t^3 = \frac{y^3}{x^3} = \frac{y^3}{y^2} = y\,.$$

Beim Polynom $X^2 + Y^2 - 1 \in \mathbb{R}[X]$ wird die Sache schon schwieriger: Aus der Analysis wissen wir, daß sich alle Lösungen (x, y) in der Form $(\cos t, \sin t)$ schreiben lassen mit $t \in [0,\, 2\pi)$, aber das ist erstens keine algebraische Darstellung, und zweitens können wir die Werte von Sinus und Kosinus für die meisten Argumente nur näherungsweise bestimmen, während wir in der Computeralgebra *exakt* rechnen möchten.

Polynome $p, q \in \mathbb{R}[T]$ von positivem Grad mit $p(t)^2 + q(t)^2 = 1$ für alle t kann es nicht geben: Da jedes Polynom außer dem Nullpolynom höchstens endlich viele Nullstellen hat, müßte dann $p^2 + q^2 - 1$ das Nullpolynom sein, also $p^2 + q^2 = 1$ in $\mathbb{R}[T]$. Sei etwa der höchste

Term von p gleich aT^n und der von q gleich bT^m mit $a \neq 0$ und $b \neq 0$. Ist $n > m$, so hat $p^2 + q^2$ den höchsten Term a^2T^{2n}, kann also wegen $n > m \geq 0$ unmöglich gleich dem konstanten Polynom 1 sein. Entsprechend ist auch $m > n$ unmöglich. Im Falle $n = m$ ist der führende Term von $p^2 + q^2$ gleich $(a^2 + b^2)T^{2n}$, was für $n > 0$ verschwinden müßte. Das ist aber unmöglich, da $a^2 + b^2$ genau dann verschwindet, sowohl a als auch b verschwinden. Also müssen p und q konstant sein und liefern somit nur einen Punkt des Kreises.

Läßt man allerdings auch Quotienten von Polynomen, also rationale Funktionen zu, so läßt sich eine Parametrisierung finden: Mit

$$x = \frac{1 - t^2}{1 + t^2} \qquad \text{und} \qquad y = \frac{2t}{1 + t^2}$$

ist $x^2 + y^2 = 1$ für alle t, denn nach den binomischen Formeln ist

$$x^2 + y^2 = \frac{(1 - t^2)^2}{(1 + t^2)^2} + \frac{4t^2}{(1 + t^2)^2} = \frac{1 - 2t^2 + t^4 + 4t^2}{(1 + t^2)^2} = \frac{1 + 2t^2 + t^4}{(1 + t^2)^2} = 1 \,.$$

Mit einer Ausnahme lassen sich auch alle Punkte der Kreislinie so darstellen: Ist $y = 0$, so muß $t = 0$ sein, also $x = 1$, so daß es für den Punkt $(-1, 0)$ keine solche Darstellung gibt. (Er ist der Limes für $t \to \infty$.)

Für einen Punkt (x, y) mit $y \neq 0$ können wir den Quotienten $c = x/y$ betrachten; falls (x, y) zum Parameter t gehört, ist dann

$$c = \frac{1 - t^2}{2t} \quad \text{oder} \quad t^2 + 2ct = 1 \,.$$

Also ist $(t + c)^2 = 1 + c^2$, d.h. mit $t = -c \pm \sqrt{1 + c^2}$ haben wir zwei verschiedene Werte von t, für die das gilt, und die liefern die beiden Schnittpunkte der Geraden $x = cy$ mit der Kreislinie.

Wenn man mit sogenannten homogenen Koordinaten $(x : y : z)$ rechnet, bei denen es nur auf die Verhältnisse zwischen den einzelnen Koordinaten ankommt, nicht aber auf die Werte, kann man diese Parametrisierung sogar polynomial schreiben als $(1 - t^2 : 2t : 1 + t^2)$, und wenn man dann noch als Parameterraum die projektive Gerade mit homogenen Koordinaten $(t : s)$ nimmt und $(x : y : z) = (s^2 - t^2 : 2st : s^2 + t^2)$

schreibt, verschwindet auch die Sonderrolle des Punktes $(-1, 0)$, denn für $(t : s) = (1 : 0)$ erhalten wir $(-1 : 0 : 1)$, also die homogenen Koordinaten von $(-1, 0)$.

Selbst wenn man rationale Funktionen zur Parametrisierung zuläßt oder – wie gerade angedeutet – mit homogenen Koordinaten in den projektiven Raum geht, kann man aber mit Methoden der algebraischen Geometrie zeigen, daß so eine Parametrisierung nur in sehr wenigen Fällen existiert; im Allgemeinen müssen dazu die Grade sehr klein sein. Für eine ebene Kurve ohne Singularitäten (d.h. in jedem Punkt gibt es genau eine Gerade, die die Kurve mit Vielfachheit größer eins schneidet, die Tangente) etwa muß der Grad gleich eins oder zwei sein, d.h. nur Geraden und Kegelschnitte lassen sich so parametrisieren. Für allgemeine Gleichungssysteme ist es daher meist aussichtslos, nach einer Parameterdarstellung der Lösungsmenge zu suchen.

Falls allerdings eine Teilmenge von k^n parametrisch durch Polynome gegeben ist, läßt sich stets ein nichtlineares Gleichungssystem finden, das alle Elemente dieser Menge (und eventuell noch einige wenige weitere) als Lösungen hat. Angenommen, wir haben eine Parameterdarstellung

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

mit Polynomen $\varphi_1, \ldots, \varphi_n \in k[T_1, \ldots, T_m]$, 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, also eine Eliminationsordnung für $T_1, \ldots, T_m$, 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 kein T_i vorkommt.

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

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

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 $\left(X_j - x_j^{(\ell)}\right) T_\ell$ in Q_i, 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 $V(J)$. Dies gilt auch dann, wenn die Koordinaten von Q nicht in k, sondern in einem Erweiterungskörper K liegen.

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 und nur dort 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.

Betrachten wir als einfaches Beispiel die beiden Punkte $(1, 3)$ und $(2, 7)$ aus $\mathbb{Q}^2$. Dazu definieren wir zwei Punkte

$$Q_1 = (1, 0, 1, 3) \qquad \text{und} \qquad Q_2 = (0, 1, 2, 7)\,,$$

und wir wissen, daß $\{Q_1, Q_2\}$ die Nullstennmenge der fünf Polynome

$$f_1 = (X-1)T_1, \quad f_2 = (X-2)T_2, \quad f_3 = (Y-3)T_1, \quad f_4 = (Y-7)T_2$$

und $f_5 = T_1 + T_2 - 1$ ist. Wir arbeiten mit der lexikographischen Ordnung, für die $T_1 > T_2 > X > Y$ ist und suchen eine GRÖBNER-Basis des von den fünf Polynomen erzeugten Ideals. Deren Berechnung überlassen wir besser einem Computeralgebrasystem. Maple liefert als reduzierte Basis die aus den Polynomen

$$g_1 = Y^2 - 10Y + 21, \qquad g_2 = 4X - Y - 1,$$
$$g_3 = 4T_2 + Y - 3, \qquad \text{und} \qquad g_4 = 4T_1 + Y - 7\,,$$

woraus wir sehen, daß Maple nicht die in der Vorlesung verwendete Konvention benutzt, wonach in einer reduzierten GRÖBNER-Basis alle führenden Koeffizienten eins sein müssen, sondern stattdessen primitive Polynome mit ganzzahligen Koeffizienten liefert. g_3 und g_4 enthalten jeweils eine der Variablen T_i; unser gesuchtes Ideal wird also erzeugt von den beiden restlichen Basiselementen g_1 und g_2, die beide in $Q[X, Y]$ liegen. Also ist $V(g_1, g_2) = \{Q_1, Q_2\}$, worauf wir eigentlich auch ohne die Berechnung von GRÖBNER-Basen gekommen sein könnten: $g_2 = (Y-3)(Y-7)$ hat als Nullstellen genau die y-Koordinaten der beiden Punkte, und das Interpolationspolynom $(Y-1)/4$, das den y-Koordinaten die zugehörigen x-Koordinaten zuordnet, führt auf g_2.

Mit dieser Strategie können wir nun auch leicht eine GRÖBNER-Basis bezüglich der lexikographischen Ordnung mit $Y > X$ konstruieren:

Hier hat $\widetilde{g}_1 = (X-1)(X-2) = X^2 - 3X + 2$ die beiden x-Koordinaten als Nullstellen, und $4X - 1$ ist das Interpolationspolynom, das dazu die jeweilige y-Koordinate liefert, d.h. $\widetilde{g}_2 = Y - 4X + 1$. Es ist klar, daß diese beiden Polynome eine GRÖBNER-Basis bilden, denn ihre führenden Terme X^2 und Y sind teilerfremd, so daß sich das S-Polynom auf Null reduzieren läßt.

Kehren wir zurück zur Frage, was wir tun können, wenn ein nichtlineares Gleichungssystem eine unendliche Lösungsmenge hat. Da es in diesem Fall meist keine Möglichkeit gibt, diese Menge explizit darzustellen, bleibt nur die Strategie, ein möglichst einfaches Gleichungssystem mit derselben Lösungsmenge zu finden, das uns hoffentlich etwas mehr über diese verrät. Damit stellt sich die Frage, wann zwei Gleichungssystem die gleiche Lösungsmenge haben.

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: Als triviales Beispiel können wir für irgendeinen Körper k in k^3 die Lösungsmege der beiden Gleichungssysteme

$$x = y = z = 0 \qquad \text{und} \qquad x^2 = y^3 = z^{10} = 0$$

betrachten. Offensichtlich haben beide nur den Nullpunkt als Lösung, aber die Ideale (X, Y, Z) und (X^2, Y^3, Z^{10}) in $k[X, Y, Z]$ sind definitiv verschieden.

Als erstes müssen wir uns überlegen, *wo* wir nach Lösungen suchen: Wie wir bereits bei Gleichungssystemen wie $x^2 = 2$ und $y^2 = 3$ in $\mathbb{Q}^2$ 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 bekanntlich als *algebraisch abgeschlossen,* und wir werden bei der Beantwortung obiger Frage immer von der Lösungsmenge in einem algebraisch abgeschlossenen Körper ausgehen.

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