Werk

Titel: Vermeidung von Divisionen.

Autor: Strassen, Volker; Ramachandra, K.

Jahr: 1973

PURL: https://resolver.sub.uni-goettingen.de/purl?243919689_0264|log16

Vermeidung von Divisionen

Von *Volker Strassen* in Zürich

Summary

The extent to which the use of divisions may speed up the evaluation of polynomials is estimated from above. In particular it is shown that for multiplying general matrices the use of divisions does not decrease the number of $(*,/)$-operations. The computational complexity of the multiplication of matrices from a linear algebraic group G is estimated from below by a simple algebraic invariant of the Liering of G. In particular, the multiplication of orthogonal matrices is treated.

§ 1. Einleitung

Seien k ein unendlicher Körper und K ein endlich erzeugter Oberkörper von k. Aus Gründen der Einfachheit nehmen wir an, daß bei Berechnungen in K die k-linearen Operationen keine Zeit kosten, während die übrigen Multiplikationen und Divisionen jeweils eine Zeiteinheit beanspruchen.

Der Grundgedanke dieser Arbeit ist die Ersetzung von Berechnungen in K durch Berechnungen in formalen Potenzreihenringen $k[[x_1, \ldots, x_n]]$, wobei n der Transzendenzgrad von K über k ist. In solchen Ringen kann man eine Division vermöge der Formel

$$(1-f)^{-1} = \sum_{r \geqq 0} f^r$$

(f eine Potenzreihe ohne konstanten Term) auf unendlich viele Multiplikationen und Additionen zurückführen. Rechnet man statt mit Potenzreihen mit deren homogenen Komponenten bis zu einem genügend hohen Grade, so reduziert sich eine Division auf endlich viele Ringoperationen.

Auf die so angedeutete Weise zeigen wir in § 2, daß sich bei der Berechnung von Mengen von quadratischen Formen (z. B. für die Matrizenmultiplikation) die Verwendung von Divisionen überhaupt nicht lohnt (nach einer Bemerkung bei Winograd [10] hat auch P. Ungar dies bewiesen, sein Beweis scheint allerdings nicht publiziert zu sein). Das Hauptergebnis von § 2 besagt, daß man bei der Berechnung einer Menge von Polynomen vom Grade $\leqq d$ in einem Körper $K = k(x_1, \ldots, x_n)$ durch Verzicht auf Divisionen eine Verlangsamung von höchstens einem Faktor $4d \lg_2 d$ in Kauf nimmt. (Nach Sieveking [12] kann man diesen Faktor sogar durch $7d$ ersetzen.)

In § 3 vergleichen wir den optimalen Zeitaufwand ζ_m für die Multiplikation zweier allgemeiner m-reihiger Matrizen mit der entsprechenden Größe η_m für die Inversion einer

Matrix und zeigen

$$\zeta_m \leqq 3\eta_{2m}$$

und

$$\eta_{2^p} \leqq 6 \sum_{i=0}^{p-1} 2^{p-1-i}\zeta_{2^i} + 2^p.$$

Ähnliche Abschätzungen findet man ohne Beweis bei Winograd [10]. Mit Hilfe einer plausiblen Vermutung über die Berechnung von zwei Mengen von quadratischen Formen in disjunkten Unbestimmtenmengen zeigen wir, daß ζ_m und η_m die gleiche Größenordnung haben.

In § 4 führen wir den zur Berechnung einer Menge von quadratischen Formen benötigten Zeitaufwand im wesentlichen zurück auf den Rang eines Tensors $\tau \in U \otimes V \otimes W$ (U, V, W endlichdimensionale k-Vektorräume). Der Rang ist in Verallgemeinerung einer der möglichen Definitionen des Rangs von Matrizen durch

$$\operatorname{Rg}\tau = \min\{N: \tau = \sum_{q=1}^{N} u_q \otimes v_q \otimes w_q \text{ mit geeigneten } u_q \in U, v_q \in V, w_q \in W\}$$

gegeben. Die Untersuchung des Rangs ersetzt das durch [7] motivierte Studium von Berechnungen quadratischer Formen in nichtkommutierenden Unbestimmten, und zwar mit einem Gewinn an Bequemlichkeit und Durchsichtigkeit (siehe § 4, Anwendungen 12, (ii)). Eine ähnliche symmetrische Formulierung der Berechnungskomplexität stammt im Falle der Matrizenmultiplikation von Gastinel [16]. Wir formulieren ein nichtkommutatives Analogon zur oben erwähnten Vermutung und beweisen ein paar einfache Sätze über den Rang, indem wir eine genauere Diskussion auf eine folgende Arbeit verschieben. Als einfache Anwendung bestimmen wir die Berechnungslänge einer einzelnen quadratischen Form über einem beliebigen unendlichen Körper k mit Char $k \neq 2$.

In § 5 verallgemeinern wir die Methode und einen Teil der Ergebnisse von § 2 auf Berechnungen im Funktionenkörper einer irreduziblen Varietät, indem wir die formale Potenzreihenentwicklung einer Funktion an einem einfachen Punkt heranziehen. Wir führen das allerdings nur für die Multiplikation von Matrizen einer irreduziblen linearen algebraischen Gruppe durch (hier ergibt sich durch die Homogenität der Gruppenvarietät eine Vereinfachung; im allgemeinen Fall arbeitet man am besten generisch, indem man in Potenzreihen über K entwickelt). Als Ergebnis erhalten wir z. B., daß der optimale Zeitaufwand zur Multiplikation zweier orthogonaler Matrizen (mit Division) nicht kleiner ist als der halbe Mindestaufwand zur Multiplikation schiefsymmetrischer Matrizen gleicher Größe (ohne Division). Allgemein gilt:

Der Mindestaufwand bei der Multiplikation zweier Matrizen einer irreduziblen linearen algebraischen Gruppe ist $\geqq \frac{1}{4}$ mal dem Rang im obigen Sinne des Strukturtensors des Lierings der Gruppe.

Wir benutzen die Bezeichnungen und Resultate von [8]. Besonders der Begriff der L-Schranke, der Transitivitätssatz und der Simulationssatz werden dauernd verwandt (wir zitieren den Simulationssatz auch, wenn wir eines seiner Korollare meinen). Ferner sagen wir, wie in [8], k-Ring statt k-Algebra und k-Körper statt kommutative k-Divisionsalgebra. $\sqcup$ bedeutet disjunkte Vereinigung, $\lfloor a \rfloor$ bzw. $\lceil a \rceil$ die größte ganze Zahl $\leqq a$ bzw. die kleinste ganze Zahl $\geqq a$.

Eine erste Version der Resultate von § 2 und § 3 habe ich 1967/68 am Statistics Department der University of California, Berkeley gefunden. Ich danke der National Science Foundation für ihre Unterstützung (NSF GP — 7454).

§ 2. Vermeidung von Divisionen

In dieser Arbeit ist k stets ein unendlicher Körper. Wir untersuchen Berechnungen in kommutativen k-Ringen und in kommutativen k-Ringen mit Division durch Einheiten. Letzteres sind Palgebren A vom Typ $\Omega' = \{0, 1, +, -, *, /\} \sqcup k$ mit einstelligen $\lambda \in k$, deren induzierte Ω-Algebren (mit $\Omega = \{0, 1, +, -, *\} \sqcup k$) kommutative k-Ringe sind und wo / Division durch Einheiten bedeutet (also $\mathrm{Def}(/) = A \times \{\text{Einheiten des } k\text{-Rings } A\}$). k-Körper sind offenbar spezielle kommutative k-Ringe mit Division durch Einheiten. Wir wählen $1_{\{*,/\}}$ als Operationszeit für Ω', und $1_{\{*\}}$ als Operationszeit für Ω. Arbeiten wir in einem k-Ring A (mit oder ohne Division durch Einheiten) über einer Teilmenge E von A, d. h. in der kanonischen E-Expansion von A, so wird stets $z(e) = 0$ für $e \in E$ angenommen. Statt $1_{\{*,/\}}$ könnten wir auch irgendeine Operationszeit z' mit $o < z'(*) \leqq z'(/)$ und $z'(\omega) = o$ für $\omega \notin \{*, /\}$ zugrundelegen. Alle Resultate würden mit leichten Modifikationen richtig bleiben.

Das folgende anschaulich evidente Lemma werden wir oft stillschweigend benutzen.

Lemma 1. *Seien A ein kommutativer k-Ring mit oder ohne Division durch Einheiten, E, E', F, F' endlich $< A$. Ist F k-linear abhängig von F' und E' k-linear abhängig von E, so gilt*

$$L(F \bmod E) \leqq L(F' \bmod E').$$

Beweis (ausführlich). Nach dem Transitivitätssatz ist

$$L(F \bmod E) \leqq L(F \bmod F') + L(F' \bmod E') + L(E' \bmod E).$$

Es genügt also etwa

$$L(F \bmod F') = 0$$

zu zeigen. Wiederum nach dem Transitivitätssatz können wir annehmen, F sei ein. punktig, etwa $= \{a\}$. Da unsere Operationszeit auf linearen Operationen verschwindet, ist $\{b \in A\colon L(b \bmod F') = 0\}$ ein k-Vektorraum, der F' umfaßt. Also $L(a \bmod F') = 0$.

Seien nun $x_1, \ldots, x_n$ Unbestimmte über k. Wir fassen $k[\boldsymbol{x}] = k[x_1, \ldots, x_n]$ als k-Ring über $\{x_1, \ldots, x_n\}$, $k(\boldsymbol{x})$ als k-Körper über $\{x_1, \ldots, x_n\}$ auf (dies gilt für die ganze Arbeit). Die Einbettung $k[\boldsymbol{x}] \to k(\boldsymbol{x})$ ist ein φ-Homomorphismus, wo φ die Einbettung von $\Omega \sqcup \{x_1, \ldots, x_n\}$ in $\Omega' \sqcup \{x_1, \ldots, x_n\}$ ist. Nach dem Simulationssatz gilt also für $F < k[\boldsymbol{x}]$

$$L_{k(\boldsymbol{x})}(F) \leqq L_{k[\boldsymbol{x}]}(F).$$

Satz 2. *Sei F eine endliche Menge von Polynomen $\in k[\boldsymbol{x}]$ vom Grad $\leqq d$. Dann ist*

$$L_{k[\boldsymbol{x}]}(F) \leqq \frac{d(d-1)}{2} L_{k(\boldsymbol{x})}(F).$$

Beweis. Sei β eine ausführbare Berechnung von F in $k(\boldsymbol{x})$ mit

$$L_{k(\boldsymbol{x})}(F) = L(\beta).$$

Da k unendlich ist, gibt es ein $\boldsymbol{\lambda} \in k^n$ so, daß alle von 0 verschiedenen Ergebnisse von β Einheiten im lokalen Ring $\mathcal{O}_{\boldsymbol{\lambda}}$ des Punktes $\boldsymbol{\lambda}$ sind (dieser besteht aus allen rationalen Funktionen, deren gekürzter Nenner an der Stelle $\boldsymbol{\lambda}$ nicht verschwindet). β ist also auch ausführbar im kommutativen k-Ring über $\{x_1, \ldots, x_n\}$ mit Division durch Einheiten $\mathcal{O}_{\boldsymbol{\lambda}}$ und hat dort die gleiche Ergebnisfolge. Also

$$L(\beta) \geqq L_{\mathcal{O}_{\boldsymbol{\lambda}}}(F).$$

Sei $k[[\boldsymbol{x} - \boldsymbol{\lambda}]] = k[[x_1 - \lambda_1, \ldots, x_n - \lambda_n]]$ der Ring der formalen Potenzreihen in den $x_i - \lambda_i$. Wir haben einen Homomorphismus von kommutativen k-Ringen über $\{x_1, \ldots, x_n\}$ mit Division durch Einheiten

$$\mathcal{O}_\lambda \to k[[\boldsymbol{x} - \boldsymbol{\lambda}]],$$

der $k[\boldsymbol{x}]$ identisch abbildet (siehe [2], S. 59). Nach dem Simulationssatz ist also

$$L_{\mathcal{O}_\lambda}(F) \geqq L_{k[[\boldsymbol{x}-\boldsymbol{\lambda}]]}(F).$$

Sei $f = f_d$ die vereinigungstreue Abbildung $2^{k[[\boldsymbol{x}-\boldsymbol{\lambda}]]} \to 2^{k[\boldsymbol{x}]}$, die jeder Potenzreihe die Menge ihrer homogenen Teile vom Grade $\leqq d$ zuordnet. Wir schreiben $f(a)$ für $f(\{a\})$ und $f(a, b)$ für $f(\{a, b\})$.

$$L_{k[\boldsymbol{x}]}(f(0)) = L_{k[\boldsymbol{x}]}(f(1)) = 0, \quad L_{k[\boldsymbol{x}]}(f(a \pm b) \bmod f(a, b)) = L_{k[\boldsymbol{x}]}(f(\alpha a) \bmod f(a)) = 0,$$

und wenn $a = \sum_{r \geqq 0} a_r$, $b = \sum_{r \geqq 0} b_r$ mit a_r, b_r homogen vom Grade r,

$$(1) \qquad L_{k[\boldsymbol{x}]}(f(a\,b) \bmod f(a, b))$$

$$= L_{k[\boldsymbol{x}]}(\{\sum_{r+s=t} a_r \cdot b_s : 0 \leqq t \leqq d\} \bmod a_0, \ldots, b_d) \leqq \frac{d(d-1)}{2},$$

da $a_0, b_0 \in k$. Falls b Einheit, also $b_0 \neq 0$, haben wir auch

$$(2) \qquad L_{k[\boldsymbol{x}]}(f(a/b) \bmod f(a, b)) \leqq \frac{d(d-1)}{2}$$

(dies sieht man leicht durch Induktion nach d unter Verwendung von

$$\begin{aligned} &L_{k[\boldsymbol{x}]}(f_d(a/b) \bmod f_{d-1}(a/b), f_d(a, b)) \\ &= L_{k[\boldsymbol{x}]}((a/b)_d \bmod f_{d-1}(a/b), f_d(a, b)) \\ &= L_{k[\boldsymbol{x}]}((a_d - \sum_{0 \leqq \delta \leqq d-1} (a/b)_\delta b_{d-\delta}) b_0^{-1} \bmod f_{d-1}(a/b), f_d(a, b)) \leqq d - 1). \end{aligned}$$

Also nach dem Simulationssatz

$$L_{k[\boldsymbol{x}]}(f(F)) \leqq \frac{d(d-1)}{2} L_{k[[\boldsymbol{x}-\boldsymbol{\lambda}]]}(F).$$

Da d eine obere Schranke für den Grad der Polynome F ist, haben wir

$$L_{k[\boldsymbol{x}]}(F \bmod f(F)) = 0.$$

Fassen wir die erhaltenen Resultate zusammen, so folgt die Behauptung des Satzes.

Der Beweis ist natürlich ganz konstruktiv.

Korollar 3. *Ist F eine endliche Menge von quadratischen Formen, so ist die Einbettung $k[\boldsymbol{x}] \to k(\boldsymbol{x})$ L-autark für F.*

Bemerkung 4. Ist E eine weitere endliche Menge von quadratischen Formen, so zeigt der Beweis des letzten Satzes, daß die Einbettung $k[\boldsymbol{x}] \to k(\boldsymbol{x})$ auch L-autark für F mod E ist.

Der Faktor $\frac{d(d-1)}{2}$ in Satz 2 ist keineswegs optimal, besonders für große d läßt er sich wesentlich verbessern.

Satz 5. *Unter den Voraussetzungen von Satz 2 gilt*

$$L_{k[x]}(F) \leqq 4d \lg_2 d L_{k(x)}(F)$$

($\lg_2$ bezeichnet den Logarithmus zur Basis 2).

Beweis. Wie der Beweis von Satz 2, indem man an Stelle von (1), (2) das folgende Lemma benutzt.

Lemma 6. *Seien* $a = \sum_{r \geqq 0} a_r$, $b = \sum_{r \geqq 0} b_r \in k[[x - \lambda]]$, $d \geqq 2$. *Dann ist*

$$L_{k[x]}(f_d(a\,b) \bmod f_d(a, b)) \leqq 2d - 3$$

und falls $b_0 \neq 0$,

$$L_{k[x]}(f_d(a/b) \bmod f_d(a, b)) \leqq 4d \lg_2 d.$$

Beweis. Seien $u = \sum_{r \geqq r_0} u_r$, $v = \sum_{s \geqq s_0} v_s \in k[[x - \lambda]]$, u_r, v_s homogen vom Grade r bzw. s, $r_0 + s_0 \leqq d$. Seien $\lambda_1, \ldots, \lambda_{2(d-r_0-s_0)+1}$ paarweise und von 0 verschiedene Elemente von k.

$$\begin{aligned} L(f(u \cdot v) \bmod f(u, v)) &= L\left(f\left(\left(\sum_{r=r_0}^{d-s_0} u_r\right) \cdot \left(\sum_{s=s_0}^{d-r_0} v_s\right)\right) \bmod f(u, v)\right) \\ &\leqq L\left(\left\{ \sum_{\substack{r+s=t \\ r_0 \leqq r \leqq d-s_0 \\ s_0 \leqq s \leqq d-r_0}} u_r v_s : r_0 + s_0 \leqq t \leqq 2d - r_0 - s_0 \right\} \bmod f(u, v)\right) \qquad (3) \\ &= L\left(\left\{ \sum_{t=r_0+s_0}^{2d-r_0-s_0} \left(\sum_{\substack{r+s=t \\ r_0 \leqq r \leqq d-s_0 \\ s_0 \leqq s \leqq d-r_0}} u_r v_s \right) \lambda_i^t : i = 1, \ldots, 2(d - r_0 - s_0) + 1 \right\} \bmod f(u, v)\right) \end{aligned}$$

(wegen Lemma 1, weil $\mathrm{Det}(\lambda_i^t) \neq 0$ nach Vandermonde)

$$\begin{aligned} &\leqq L\left(\left\{\left(\sum_{r=r_0}^{d-s_0} u_r \lambda_i^r\right)\left(\sum_{s=s_0}^{d-r_0} v_s \lambda_i^s\right) : i = 1, \ldots, 2(d - r_0 - s_0) + 1\right\} \bmod u_{r_0}, \ldots, v_{d-r_0}\right) \\ &\leqq 2(d - r_0 - s_0) + 1 \end{aligned}$$

nach Lemma 1. Hieraus folgt die erste zu beweisende Ungleichung:

$$L(f(ab) \bmod f(a, b)) = L(f(a - a_0)\,(b - b_0)) \bmod f(a - a_0, b - b_0)) \leqq 2(d - 2) + 1.$$

Zum Beweis der zweiten können wir $d > 4$ annehmen (denn $\frac{d(d-1)}{2} \leqq 4d \lg_2 d$ für $2 \leqq d \leqq 4$), ferner $b_0 = 1$. Wir setzen $\tilde{b} = 1 - b$. Es ist

$$a/b = a \sum_{i \geqq 0} \tilde{b}^i = a_0 + c \sum_{i \geqq 0} \tilde{b}^i,$$

wobei $c = a_0 \tilde{b} + a - a_0$ verschwindenden konstanten Koeffizienten hat. Nach Lemma 1, dem Transitivitätssatz und der schon bewiesenen Ungleichung ergibt sich

$$\begin{aligned} L(f(a/b) \bmod f(a, b)) &\leqq L\left(f\left(c \sum_{i=0}^{d-1} \tilde{b}^i\right) \bmod f(c, \tilde{b})\right) \qquad (4) \\ &\leqq 2d - 3 + L\left(f\left(\sum_{i=0}^{d-1} \tilde{b}^i\right) \bmod f(\tilde{b})\right). \end{aligned}$$

Wir zeigen nun für $2^{p-1} \leqq d$ durch Induktion nach p

$$(5) \qquad L\left(f\left(\sum_{i=0}^{2^p-1} \tilde{b}^i, \tilde{b}^{2^p-1}\right) \bmod f(\tilde{b})\right) \leqq (4d-2)(p-1) - 2^{p+1} + 4.$$

Dies ist klar für $p = 1$. Allgemein gilt

$$L\left(f\left(\sum_{i=0}^{2^{p+1}-1} \tilde{b}^i, \tilde{b}^{2^p}\right) \bmod f(\tilde{b})\right) \leqq L\left(f\left(\sum_{0}^{2^{p+1}-1} \tilde{b}^i, \tilde{b}^{2^p}\right) \bmod f\left(\sum_{0}^{2^p-1} \tilde{b}^i, \tilde{b}^{2^p}\right)\right)$$

$$+ L\left(f\left(\sum_{0}^{2^p-1} \tilde{b}^i, \tilde{b}^{2^p}\right) \bmod f\left(\sum_{0}^{2^p-1} \tilde{b}^i, \tilde{b}^{2^{p-1}}\right)\right) + L\left(f\left(\sum_{0}^{2^p-1} \tilde{b}^i, \tilde{b}^{2^{p-1}}\right) \bmod f(\tilde{b})\right) =: \mathrm{I} + \mathrm{II} + \mathrm{III}.$$

Wegen

$$\sum_{0}^{2^{p+1}-1} b^i = \left(\sum_{0}^{2^p-1} \tilde{b}^i\right)(1 + \tilde{b}^{2^p})$$

ist

$$\mathrm{I} \leqq 2d - 3,$$

wegen (3)

$$\mathrm{II} \leqq 2(d - 2^p) + 1$$

und nach Induktionsvoraussetzung

$$\mathrm{III} \leqq (4d-2)(p-1) - 2^{p+1} + 4.$$

Damit ist (5) gezeigt. Nun sei $2^{p-1} < d \leqq 2^p$. Dann folgt aus (4) und (5)

$$L(f(a/b) \bmod f(a, b)) \leqq 2d - 3 + (4d-2)(p-1) - 2^{p+1} + 4$$

$$\leqq 4d(p-1) - (2p-3) \leqq 4d(p-1) \leqq 4d \lg_2 d.$$

Ein zu Satz 5 ähnliches Resultat erhält man für eine überall positive beschränkte (etwa auf k konstante) Operationszeit z', wenn k algebraisch abgeschlossen ist (oder wenigstens die nötigen Einheitswurzeln enthält):

$$L_{k[x]}(F) \leqq \text{const}\,(d \lg d L_{k(x)}(F) + \#(F) d (\lg d)^2).$$

Hier kann man sich nämlich auf Berechnungen von F in $k(x)$ beschränken, in denen höchstens $\#(F)$ Divisionen vorkommen, indem man ausgehend von einer optimalen Berechnung von F in $k(x)$ mit (ungekürzten) Zählern und Nennern separat rechnet und die Divisionen nur in den Endergebnissen ausführt. Die dadurch in Kauf genommene Verlangsamung beträgt nur einen konstanten Faktor. Ferner zeigt man nach dem Muster des Beweises von Lemma 6 leicht

$$L_{k[x]}(f(ab) \bmod f(a, b)) \leqq \text{const}\, d \lg d$$

$$L_{k[x]}(f(a/b) \bmod f(a, b)) \leqq \text{const}\, d (\lg d)^2,$$

indem man die schnelle Fouriertransformation verwendet. Speziell (und auf besonders einfache Weise) erhält man, daß bei der Berechnung einer Menge von Linearformen der Verzicht auf die Division nur um einen kleinen Faktor verlangsamt (der von der Operationszeit abhängt).

§ 3. Inversion von Matrizen

Ist $C \in M_m(k(x))$, d. h. ist C eine m-reihige quadratische Matrix mit Koeffizienten $c_{ij} \in k(x)$, so setzen wir

$$L(C) = L_{k(x)}(\{c_{ij} : 1 \leqq i, j \leqq m\}).$$

Analog ist $L(C_1, \ldots, C_r \bmod D_1, \ldots, D_s)$ definiert. Ist $C \in M_m(k[x])$, so setzen wir

$$L_{k[x]}(C) = L_{k[x]}(\{c_{ij}: ij \leqq m\}).$$

Seien A, B m-reihige Matrizen, deren Koeffizienten a_{ij}, b_{rs} paarweise verschiedene Unbestimmte x_p sind (das setzt natürlich $2m^2 \leqq n$ voraus). Wir setzen

$$\zeta_m = L(AB) = L_{k[x]}(AB), \ \eta_m = L(A^{-1})$$

(mit Hilfe des Simulationssatzes sieht man leicht, daß die rechten Seiten nicht von $n \geqq 2m^2$ und der Wahl von A, B abhängen).

Lemma 1. *Für beliebiges $C, D \in M_m(k(x))$ ist*

$$L(CD \bmod C, D) \leqq \zeta_m.$$

Es gibt ein von 0 verschiedenes Polynom $h \in k[a_{11}, \ldots, a_{mm}]$ so, daß für beliebiges $C \in M_m(k(x))$ aus $h(c_{11}, \ldots, c_{mm}) \neq 0$ folgt

C hat eine Inverse

und

$$L(C^{-1} \bmod C) \leqq \eta_m.$$

Insbesondere gilt dies also, wenn die c_{ij} algebraisch unabhängig über k sind.

Beweis. Zum Beweis der ersten Ungleichung wählen wir $n = 2m^2$, so daß die a_{ij}, b_{ij} gerade die Unbestimmten x_p durchlaufen, und wenden den Simulationssatz auf den durch $a_{ij} \mapsto c_{ij}$, $b_{ij} \mapsto d_{ij}$ bestimmten φ-Homomorphismus des von $k[x]$ induzierten k-Rings in den von $k(x)$ induzierten k-Körper an (φ ist die Einbettung von Ω in Ω').

Zum Beweis der zweiten Ungleichung wählen wir $n = m^2$ und eine optimale ausführbare Berechnung β von A^{-1}. Sei $h(a_{11}, \ldots, a_{mm}) = h(x_1, \ldots, x_n)$ das Produkt der in den Ergebnissen von β auftretenden von 0 verschiedenen Zähler und Nenner. Ist $h(c_{11}, \ldots, c_{mm}) \neq 0$, so sei $\mathcal{O}_C$ der lokale Ring aller $g \in k(x)$, für die $g(c_{11}, \ldots, c_{mm})$ definiert ist. Alle von 0 verschiedenen Ergebnisse von β sind Einheiten von $\mathcal{O}_C$; β berechnet also A^{-1} auch im kommutativen Ring mit Division durch Einheiten $\mathcal{O}_C$, d. h.

$$L_{\mathcal{O}_C}(A^{-1} \bmod A) \leqq \eta_m.$$

Nun wende man den Simulationssatz auf den durch $a_{ij} \mapsto c_{ij}$ definierten Homomorphismus $\mathcal{O}_C \to k(x)$ an.

Lemma 2. *$D \in M_m(k(x))$ habe paarweise verschiedene Unbestimmte x_p als Koeffizienten. Dann sind die Koeffizienten von $D(D+1)$ algebraisch unabhängig über k.*

Beweis. O. B. d. A. sei k algebraisch abgeschlossen. Die Abbildung $\Delta \mapsto \Delta(\Delta + 1)$ für $\Delta \in M_m(k)$ ist ein Endomorphismus ψ der algebraischen Varietät $M_m(k) \approx k^{m^2}$. Es genügt offenbar zu zeigen, daß ψ dominant ist. Nach dem Satz über die Dimension der Fasern ([4], S. 92) genügt es also z. B. nachzuweisen, daß für paarweise verschiedene $\theta_1, \ldots, \theta_m \in k$ die Gleichung

$$\Delta(\Delta + 1) = (\delta_{ij}\theta_i)$$

nur endlich viele Lösungen Δ besitzt. Sei Δ eine Lösung und seien $f(x) = \Pi(x - \delta_i)$ das charakteristische Polynom von Δ, $g(y) = \Pi(y - \theta_i)$ das charakteristische Polynom von

$\Delta(\Delta+1)$. Dann ist

$$\begin{aligned} g(x(x+1)) &= \operatorname{Det}(x(x+1)-\Delta(\Delta+1)) = \operatorname{Det}((x-\Delta)(x+1+\Delta)) \\ &= (-1)^m \operatorname{Det}(x-\Delta)\operatorname{Det}(-x-1-\Delta) = (-1)^m f(x) f(-x-1) \\ &= \prod_i (x(x+1)-\delta_i(\delta_i+1)), \end{aligned}$$

also

$$g(y) = \prod_i (y-\delta_i(\delta_i+1)),$$

also o. B. d. A. $\theta_i = \delta_i(\delta_i+1)$. Die δ_i sind paarweise verschieden, weil die θ_i es sind. Also läßt sich Δ auf Diagonalgestalt bringen: Sei $\Lambda \in M_m(k)$ regulär mit

$$\Lambda^{-1}\Delta\Lambda = (\delta_{ij}\delta_i).$$

Es ergibt sich

$$\Lambda^{-1}(\delta_{ij}\theta_i)\Lambda = \Lambda^{-1}\Delta(\Delta+1)\Lambda = (\delta_{ij}\delta_i(\delta_i+1)) = (\delta_{ij}\theta_i).$$

Wegen der Verschiedenheit der θ_i sind Λ und damit auch Δ diagonal. Nun folgt die Behauptung unmittelbar.

Lemma 3. $\zeta_m \geqq m^2$, $\eta_m \geqq m^2$.

Beweis. Es ist

$\lambda(F) :=$ (Dimension des von $F \cup \{1, x_1, \ldots, x_n\}$ erzeugten k-Vektorraums) $-(n+1)$

eine L-Schranke.

Satz 4. *ζ_m und η_m sind monoton. Ferner*

$$\zeta_m \leqq 3\eta_{2m}, \quad \eta_{2^p} \leqq 6 \sum_{i=0}^{p-1} 2^{p-1-i}\zeta_{2^i} + 2^p.$$

Beweis. Monotonie von ζ_m:

$$\zeta_m = L(AB) = L\left(\begin{bmatrix} 0 & \ldots & 0 \\ \cdot & & \\ \cdot & & A \\ \cdot & & \\ 0 & & \end{bmatrix}\begin{bmatrix} 0 & \ldots & \\ \cdot & & \\ \cdot & & B \\ \cdot & & \\ 0 & & \end{bmatrix}\right) \leqq \zeta_{m+1}$$

nach Lemma 1.

Monotonie von η_m: Nach Lemma 1 gibt es $\alpha, \Gamma = \begin{pmatrix} \gamma_1 \\ \cdot \\ \cdot \\ \gamma_m \end{pmatrix}$ und $\Lambda = (\lambda_1, \ldots, \lambda_m)$ mit

$\alpha, \gamma_i, \lambda_j \in k$ und $\alpha \neq 0$ so, daß

$$L\left(\begin{pmatrix} \alpha & \Lambda \\ \Gamma & A \end{pmatrix}^{-1}\right) \leqq \eta_{m+1}.$$

Von der Identität

$$\begin{pmatrix} 1 & 0 \\ -\Gamma/\alpha & 1 \end{pmatrix}\begin{pmatrix} \alpha & \Lambda \\ \Gamma & A \end{pmatrix}\begin{pmatrix} 1 & -\Lambda/\alpha \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \alpha & 0 \\ 0 & A-(\Gamma\Lambda/\alpha) \end{pmatrix}$$

geht man zur Inversion über. Dann erhält man nach Lemma 2. 1

$$L\left(\begin{pmatrix} \alpha & \Lambda \\ \Gamma & A \end{pmatrix}^{-1}\right) = L\left(\begin{pmatrix} \alpha & 0 \\ 0 & A-(\Gamma\Lambda/\alpha) \end{pmatrix}^{-1}\right) = L((A-(\Gamma\Lambda/\alpha))^{-1}) = L(A^{-1})$$

(letzteres nach dem Simulationssatz). Also

$$\eta_m = L(A^{-1}) \leqq \eta_{m+1}.$$

Erste Ungleichung: Sei D eine $2m$-reihige Matrix mit paarweise verschiedenen Unbestimmten als Koeffizienten. Wie man sofort nachrechnet, hat $D(D+1)$ die Inverse $D^{-1} - (D+1)^{-1}$. Da nach Lemma 2 die Koeffizientenmenge von $D(D+1)$ den Transzendenzgrad $(2m)^2$ über k hat, gilt dies auch für ihre Inverse, also hat $D^{-1} - (D+1)^{-1}$ über k algebraisch unabhängige Koeffizienten. Der Transitivitätssatz und Lemma 1 liefern nun

$$\begin{aligned} L(D(D+1)) &= L\big((D^{-1} - (D+1)^{-1})^{-1}\big) \\ &\leqq L\big((D^{-1} - (D+1)^{-1})^{-1} \bmod D^{-1} - (D+1)^{-1}\big) \\ &\quad + L(D^{-1}) + L\big((D+1)^{-1}\big) \leqq 3\eta_{2m}. \end{aligned} \tag{1}$$

Wir schreiben

$$D = \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix}$$

mit m-reihigen D_{ij}. Aus dem Simulationssatz folgt

$$\begin{aligned} \zeta_m = L_{k[x]}(D_{12}D_{21}) &\leqq L_{k[x]}\left(\begin{pmatrix} 0 & D_{12} \\ D_{21}D_{22} & \end{pmatrix}\left(\begin{pmatrix} 0 & D_{12} \\ D_{21}D_{22} & \end{pmatrix} + 1\right)\right) \\ &\leqq L_{k[x]}(D(D+1)) = L(D(D+1)). \end{aligned}$$

Zusammen mit (1) ergibt das die Behauptung.

Zweite Ungleichung: Diese folgt durch Induktion aus der Ungleichung

$$\eta_{2m} \leqq 6\zeta_m + 2\eta_m,$$

welche sich aus [7], S. 356 oben, mit Hilfe von Lemma 1 ergibt (es ist klar, daß die beiden zu invertierenden Matrizen jeweils algebraisch unabhängige Koeffizienten haben).

Korollar 5. $\zeta_m \leqq 25\eta_m$

für alle m. Falls für ein α

$$\zeta_m = O(m^\alpha),$$

so auch

$$\eta_m = O(m^\alpha).$$

Beweis. Nach [7], S. 354f. ist

$$\zeta_{2m} \leqq 7\zeta_m. \tag{2}$$

Also nach Satz 4

$$\zeta_{2q} \leqq 7\zeta_q \leqq 21\eta_{2q}. \tag{3}$$

Auf elementare Weise zeigt man

$$\zeta_{2q+1} \leqq \zeta_{2q} + (12q^2 + 6q + 1),$$

also für $q \geqq 2$

$$\zeta_{2q+1} \leqq \zeta_{2q} + 4(2q)^2 \leqq 21\eta_{2q} + 4\eta_{2q}$$

wegen Lemma 3 und (3), also

$$\zeta_{2q+1} \leqq 25\eta_{2q+1}.$$

Für $m \leqq 3$ folgt $\zeta_m \leqq 25\eta_m$ aus Lemma 3.

Ist $\zeta_m = O(m^\alpha)$, so ist $\alpha \geqq 2$ nach Lemma 3. Sei $2^{p-1} < m \leqq 2^p$. Dann ist

$$\eta_m \leqq \eta_{2^p} \leqq \text{const} \sum_{i=0}^{p-1} 2^{p-1+i(\alpha-1)} = O(2^{p\alpha}) = O(m^\alpha).$$

Aus dem Satz folgt auch ziemlich leicht, daß ζ_m und η_m auf einer unendlichen Teilmenge von $\mathbb{N}$ die gleiche Größenordnung haben. Mit Hilfe einer einleuchtenden Vermutung kann man $\eta_m = O(\zeta_m)$ allgemein zeigen.

Vermutung 6. Seien X_1, X_2 disjunkte Teilmengen von $\{x_1, \ldots, x_n\}$, $F_1 < k(X_1)$, $F_2 < k(X_2)$ Mengen von quadratischen Formen. Dann ist[1])

$$L(F_1 \cup F_2) = L(F_1) + L(F_2).$$

Korollar 7. *Ist Vermutung* 6 *richtig, so haben* ζ_m *und* η_m *die gleiche Größenordnung.*

Beweis. Seien

$$C = \begin{pmatrix} C_{11} & 0 & 0 & C_{14} \\ 0 & C_{22} & C_{23} & 0 \\ 0 & C_{32} & C_{33} & 0 \\ C_{41} & 0 & 0 & C_{44} \end{pmatrix}, D = \begin{pmatrix} D_{11} & 0 & D_{13} & 0 \\ 0 & D_{22} & 0 & D_{24} \\ D_{31} & 0 & D_{33} & 0 \\ 0 & D_{42} & 0 & D_{44} \end{pmatrix}$$

$4m$-reihige Matrizen, wobei die C_{ij}, D_{rs} m-reihig sind. Die Koeffizienten aller C_{ij}, D_{rs} seien paarweise verschiedene Unbestimmte. Aus Lemma 1 folgt

$$L(CD) \leqq \zeta_{4m}$$

und durch Ausrechnen unter Verwendung von Vermutung 6

$$8\zeta_m \leqq L(CD),$$

also

$$8\zeta_m \leqq \zeta_{4m}.$$

Aus dieser Ungleichung folgt durch Induktion nach q

$$\sum_{i=0}^{2q} 2^{2q-i}\zeta_{2^i} \leqq 6\zeta_{2^{2q}}.$$

Aus Satz 4 ergibt sich dann

$$\eta_{2^{2q}} = O(\zeta_{2^{2q}})$$

und daraus die Behauptung mittels (2).

§ 4. Der Rang eines Tensors

Definition 1. Seien U, V, W k-Vektorräume, $\tau \in U \otimes V \otimes W$. Dann heißt

$$\mathrm{Rg}(\tau) := \min\{N : \tau = \sum_{q=1}^{N} u_q \otimes v_q \otimes w_q \text{ mit geeigneten } u_q \in U,\ v_q \in V,\ w_q \in W\}$$

der Rang von τ.

Für den Fall $U = k^n$, $V = k^m$, $W = k^s$ ergibt sich mit $(\tau^{ijl}) \in k^n \otimes k^m \otimes k^s$

$$\mathrm{Rg}(\tau^{ijl}) = \min\{N : \tau^{ijl} = \sum_{q=1}^{N} u_q^i v_q^i w_q^l \text{ mit geeigneten } u_q^i, v_q^j, w_q^l \in k\}.$$

Man beweist leicht den folgenden

Satz 2. (i) *Sei*

$$\varphi: U \otimes V \otimes W \to V \otimes U \otimes W$$

[1]) Die Vermutung wird freilich durch das drastische Resultat von Moenck-Borodin [13] in Zweifel gezogen (siehe auch [12] und [14]).

der kanonische Isomorphismus, $\tau \in U \otimes V \otimes W$. *Dann ist*

$$\operatorname{Rg} \varphi(\tau) = \operatorname{Rg} \tau.$$

Entsprechend bei anderen Permutationen der Räume[2]).

(ii) *Seien* $\varphi\colon U \to U'$, $\psi\colon V \to V'$ *und* $\chi\colon W \to W'$ *lineare Abbildungen,* $\tau \in U \otimes V \otimes W$. *Dann ist*

$$\operatorname{Rg}((\varphi \otimes \psi \otimes \chi)\tau) \leqq \operatorname{Rg} \tau.$$

(iii) $$\operatorname{Rg}(\lambda\sigma + \mu\tau) \leqq \operatorname{Rg} \sigma + \operatorname{Rg} \tau \quad \text{für} \quad \lambda, \mu \in k.$$

Von besonderem Interesse ist der folgende Spezialfall:

Seien $U = U' \oplus U''$, $V = V' \oplus V''$, $W = W' \oplus W''$, *ferner* $\sigma \in U' \otimes V' \otimes W'$, $\tau \in U'' \otimes V'' \otimes W''$. *Dann ist* $\sigma \oplus \tau \in U \otimes V \otimes W$ *in natürlicher Weise definiert und es gilt*

$$\operatorname{Rg}(\sigma \oplus \tau) \leqq \operatorname{Rg} \sigma + \operatorname{Rg} \tau.$$

(iv) *Seien* $U = U' \otimes U''$, $V = V' \otimes V''$, $W = W' \otimes W''$, *ferner* $\sigma \in U' \otimes V' \otimes W'$, $\tau \in U'' \otimes V'' \otimes W''$. *Dann ist* $\sigma \otimes \tau \in U \otimes V \otimes W$ *in natürlicher Weise definiert und es gilt*

$$\operatorname{Rg}(\sigma \otimes \tau) \leqq \operatorname{Rg} \sigma \operatorname{Rg} \tau.$$

Vermutung 3. Mit den Bezeichnungen des Satzes 2, (iii) gilt stets

$$\operatorname{Rg}(\sigma \oplus \tau) = \operatorname{Rg} \sigma + \operatorname{Rg} \tau.$$

Daß in Satz 2, (iv) nicht immer das Gleichheitszeichen steht, ist leicht zu sehen. Wir stellen nun eine Beziehung zwischen Rang und Berechnungslänge her.

Satz 4. *Sei*

$$F = \{ \sum_{1 \leq i,j \leq n} \tau^{ijl} x_i x_j \colon 1 \leqq l \leqq s \}$$

eine Menge von quadratischen Formen aus $k[x]$. *Dann ist*

$$L_{k(x)}(F) = \min \{ \operatorname{Rg}(\tau^{ijl} + \alpha^{ijl})_{\substack{1 \leq i,j \leq n \\ 1 \leq l \leq s}} : \alpha^{ijl} = -\alpha^{jil} \text{ und } \alpha^{iil} = 0 \text{ für alle } i, j, l \},$$

wobei $(\tau^{ijl} + \alpha^{ijl})_{\substack{1 \leq i,j \leq n \\ 1 \leq l \leq s}}$ *als Element von* $k^n \otimes k^n \otimes k^s$ *aufgefaßt wird.*

Beweis. Nach § 2 können wir $L_{k(x)}(F)$ durch $L_{k[x]}(F)$ ersetzen.

$$\operatorname{Rg}(\tau^{ijl} + \alpha^{ijl}) \leqq N$$

bedeutet

$$\tau^{ijl} + \alpha^{ijl} = \sum_{q=1}^{N} u_q^i v_q^j w_q^l$$

mit geeigneten u_q^i, v_q^j, $w_q^l \in k$. Multiplikation mit $x_i x_j$ und Summation über i, j liefert

$$\sum_{i,j} \tau^{ijl} x_i x_j = \sum_{q=1}^{N} w_q^l \Big(\sum_i u_q^i x_i\Big) \cdot \Big(\sum_j v_q^j x_j\Big)$$

für alle l. Daraus folgt

$$L_{k[x]}(F) \leqq N,$$

[2]) Dies ist nichts anderes als die inzwischen von Hopcroft-Musinski [15] eingeführte Dualität.

also $\leqq$ in der Behauptung des Satzes. Zum Beweis von $\geqq$ sei $Q(E)$ für $E \subset k[x]$ die Menge der homogenen Teile vom Grade 2 aller $a \in E$. Man sieht leicht, daß

$$\lambda(E) := \min \{N\colon Q(E) \text{ in der linearen Hülle von}$$

$$\{(\sum_i u_q^i x_i)(\sum_j v_q^j x_j)\colon q = 1, \ldots, N\} \text{ mit geeigneten } u_q^i, v_q^i \in k\}$$

eine L-Schranke ist. Sei

$$N := L_{k[x]}(F).$$

Dann ist $N \geqq \lambda(F)$, also gibt es $u_q^i, v_q^j, w_q^l \in k$ mit

$$\sum_{i,j} \tau^{ijl} x_i x_j = \sum_{q=1}^{N} w_q^l (\sum_i u_q^i x_i)(\sum_j v_q^j x_j)$$

für alle l, d. h.

$$\sum_{i,j} (\tau^{ijl} - \sum_{q=1}^{N} u_q^i v_q^j w_q^l) x_i x_j = 0$$

für alle l. Setzen wir also

$$\alpha^{ijl} := \sum_{q=1}^{N} u_q^i v_q^j w_q^l - \tau^{ijl},$$

dann ist α^{ijl} schiefsymmetrisch in i, j und es gilt

$$\mathrm{Rg}(\tau^{ijl} + \alpha^{ijl}) = \mathrm{Rg}\left(\sum_{q=1}^{N} u_q^i v_q^j w_q^l\right) \leqq N.$$

Korollar 5. *Sei*

$$F = \{\sum_{1 \leqq i,j \leqq n} \tau^{ijl} x_i x_j \colon 1 \leqq l \leqq s\}$$

eine Menge von quadratischen Formen. Dann ist

$$\frac{1}{2} \mathrm{Rg}\,(\tau^{ijl} + \tau^{jil}) \leqq L_{k(x)}(F) \leqq \mathrm{Rg}\,(\tau^{il}).$$

Speziell, falls Char $k \neq 2$ *und* $\tau^{ijl} = \tau^{jil}$,

$$\frac{1}{2} \mathrm{Rg}\,(\tau^{ijl}) \leqq L_{k(x)}(F) \leqq \mathrm{Rg}\,(\tau^{ijl}).$$

Beweis. Sei (α^{ijl}) schiefsymmetrisch in i, j. Dann ist nach Satz 2

$$\mathrm{Rg}\,(\tau^{ijl} + \tau^{jil}) \leqq \mathrm{Rg}\,(\tau^{ijl} + \alpha^{ijl}) + \mathrm{Rg}\,(\tau^{jil} + \alpha^{jil}) = 2\,\mathrm{Rg}\,(\tau^{ijl} + \alpha^{ijl}).$$

Also nach Satz 4

$$\frac{1}{2} \mathrm{Rg}\,(\tau^{ijl} + \tau^{jil}) \leqq L_{k(x)}(F).$$

Der Rest ist klar.

Für Char $k = 2$ ist die zweite Aussage des Korollars falsch:

$$s = 1,\ \tau^{ij1} = \delta_{ij},\ L(\Sigma x_i^2) = 1,\ \mathrm{Rg}\,(\delta_{ij}) = n.$$

Später brauchen wir noch das folgende

Lemma 6. *Seien* $x_1, \ldots, x_p, y_1, \ldots, y_q$ *Unbestimmte über* k,

$$F = \left\{\sum_{\substack{1 \leqq i \leqq p \\ 1 \leqq j \leqq q}} \tau^{ijl} x_i y_j \colon 1 \leqq l \leqq s\right\}.$$

Dann ist

$$\frac{1}{2}\operatorname{Rg}(\tau^{ijl}) \leqq L_{k[x,y]}(F \bmod k[x] \cup k[y]) \leqq L_{[x,y]}(F) \leqq \operatorname{Rg}(\tau^{ijl}).$$

Beweis. Ist $E = \{a^{(1)}, \ldots, a^{(r)}\} < k[x, y]$, so schreiben wir

$$a^{(l)} = \sum_{i,j} \sigma^{ijl} x_i y_j + b^{(l)},$$

wo $b^{(l)}$ keine Monome der Gestalt $\sigma x_i y_j$ enthält, und setzen

$$\lambda(E) := \frac{1}{2}\operatorname{Rg}(\sigma^{ijl}).$$

Mittels Satz 2 sieht man leicht, daß λ wohldefiniert und eine L-Schranke mod $k[x] \cup k[y]$ ist.

Anwendung 7. Sei Char $k \neq 2$, $Q \in k(x)$ eine beliebige quadratische Form. Es ist

$$Q \sim x_1 x_2 + \cdots + x_{2m-1} x_{2m} + \hat{Q}(x_{2m+1}, \ldots, x_p)$$

mit definitem $\hat{Q}$. Wir behaupten

$$L_{k(x)}(Q) = p - m.$$

Zunächst folgt aus dem Simulationssatz, daß L konstant ist auf Äquivalenzklassen quadratischer Formen. Wegen $\hat{Q}(x_{2m+1}, \ldots, x_p) \sim \sum_{i=2m+1}^{p} \lambda_i x_i^2$ ist also

$$L(Q) = L(x_1 x_2 + \cdots + x_{2m-1} x_{2m} + \sum_{i=2m+1}^{p} \lambda_i x_i^2) \leqq p - m.$$

Andererseits können wir o. B. d. A. $p = n$ annehmen (nach Korollar 2. 3 und weil $k[x_1, \ldots, x_p]$ autark ist in $k[x]$). Dann ist Q nichtausgeartet und m ist die Dimension der maximalen Nullräume von Q. Wir nehmen an, $L(Q)$ sei $< n - m$. Ist (τ^{ij}) die Matrix von Q und ist (α^{ij}) schiefsymmetrisch mit

$$\operatorname{Rg}(\tau^{ij} + \alpha^{ij}) = L(Q) < n - m,$$

so gibt es $m + 1$ Vektoren, die von der Matrix $(\tau^{ij} + \alpha^{ij})$ annulliert werden (denn $\operatorname{Rg}(\tau^{ij} + \alpha^{ij})$ ist nichts anderes als der gewöhnliche Rang der Matrix $(\tau^{ij} + \alpha^{ij})$). Diese bilden einen Nullraum von Q, also $m + 1 \leqq m$, Widerspruch.

Zum Beispiel ist $L_{k(x)}(x_1^2 + \cdots + x_n^2) = n$ für $k = \mathbb{R}$ und $= \left\lceil \frac{n}{2} \right\rceil$ für $k = \mathbb{C}$. Das obige Resultat kann auch mit der Pan'schen Methode bewiesen werden.

Für den Rest dieses Abschnittes wollen wir unter einem k-Ring immer einen endlichdimensionalen nicht-assoziativen k-Ring verstehen (das ist ein endlichdimensionaler k-Vektorraum zusammen mit einer bilinearen Multiplikation, also eine Algebra vom Typ $k \sqcup \{0, +, -, *\}$ mit einstelligen $\lambda \in k$), vor allem um auch Lieringe miteinzubeziehen. Ist V ein endlichdimensionaler k-Vektorraum, so läßt sich eine bilineare Multiplikation in V durch einen Tensor $\sigma \in V^* \otimes V^* \otimes V$ beschreiben ([1], III. 66). Die Beschreibung ist bijektiv und folgendermaßen definiert: Für $u, v \in V$, $w \in V^*$ fasse man $u \otimes v \otimes w$ als Element von $(V^* \otimes V^* \otimes V)^*$ auf. Dann ist

$$w(u \cdot v) = (u \otimes v \otimes w)\sigma. \tag{1}$$

Definition 8. Sei V ein k-Ring, $\sigma \in V^* \otimes V^* \otimes V$ der die Multiplikation in V definierende Tensor. Dann setzen wir

$$\operatorname{Rg} V := \operatorname{Rg} \sigma.$$

Seien $e_1, \ldots, e_n$ eine Basis von V (als k-Vektorraum) und $e_1^*, \ldots, e_n^*$ die duale Basis von V^*. Sind σ^{ijl} die Koordinaten von σ bezüglich dieser Basen (aus technischen Gründen schreiben wir auch i, j als obere Indizes), d. h. gilt

$$\sigma = \sum_{i,j,l} \sigma^{ijl} e_i^* \otimes e_j^* \otimes e_l,$$

so wird (1) zu

$$e_i e_j = \sum_l \sigma^{ijl} e_l.$$

Kennt man also die Koordinaten u^i und v^j zweier Vektoren $u, v \in V$, so berechnen sich die Koordinaten des Produkts $u \cdot v$ wie folgt

$$(u \cdot v)^l = \sum_{i,j} \sigma^{ijl} u^i v^j. \tag{2}$$

Seien $x_1, \ldots, x_n, y_1, \ldots, y_n$ Unbestimmte über k. Dann ist

$$L_{k(x)}\left(\left\{\sum_{i,j} \sigma^{ijl} x_i y_j : 1 \leqq l \leqq n\right\}\right) = L_{k[x]}\left(\sum_{i,j} \sigma^{ijl} x_i y_j : 1 \leqq l \leqq n\right)$$

der kürzeste Zeitaufwand (bzgl. der durch die Operationszeit $1_{\{*,/\}}$ gegebenen Zählung) zur Berechnung des Produkts zweier Vektoren von V (vgl. auch [9], § 8).

Der folgende Satz ergibt sich aus Lemma 6.

Satz 9.

$$\frac{1}{2} \operatorname{Rg} V \leqq L_{k(x)}\left(\left\{\sum_{i,j} \sigma^{ijl} x_i y_j : 1 \leqq l \leqq n\right\}\right) \leqq \operatorname{Rg} V.$$

Satz 10. *Seien V, W k-Ringe.* (i) *Ist $f: V \to W$ ein injektiver Homomorphismus, so gilt*

$$\operatorname{Rg} V \leqq \operatorname{Rg} W.$$

(ii) *Ist $f: V \to W$ ein surjektiver Homomorphismus, so gilt*

$$\operatorname{Rg} W \leqq \operatorname{Rg} V.$$

(iii) $$\operatorname{Rg}(V \oplus W) \leqq \operatorname{Rg} V + \operatorname{Rg} W$$

(iv) $$\operatorname{Rg}(V \otimes W) \leqq \operatorname{Rg} V \cdot \operatorname{Rg} W.$$

Beweis. Seien σ, τ die die Multiplikationen in V, W beschreibenden Tensoren. Eine k-lineare Abbildung $f: V \to W$ ist genau dann ein Homomorphismus, wenn gilt

$$(\mathrm{id} \otimes \mathrm{id} \otimes f)\sigma = (f^* \otimes f^* \otimes \mathrm{id})\tau.$$

(i) Es gibt eine lineare Abbildung $g: W \to V$ mit $g \circ f = \mathrm{id}_V$, also

$$(f^* \otimes f^* \otimes g)\tau = (\mathrm{id} \otimes \mathrm{id} \otimes g) \circ (f^* \otimes f^* \otimes \mathrm{id})\tau = (\mathrm{id} \otimes \mathrm{id} \otimes g) \circ (\mathrm{id} \otimes \mathrm{id} \otimes f)\sigma = \sigma.$$

Die Behauptung folgt nun aus Satz 2.

(ii) Es gibt eine lineare Abbildung $g: W \to V$ mit $f \circ g = \mathrm{id}_W$, also $g^* \circ f^* = \mathrm{id}_{W^*}$, also

$$(g^* \otimes g^* \otimes f)\sigma = (g^* \otimes g^* \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \mathrm{id} \otimes f)\sigma = (g^* \otimes g^* \otimes \mathrm{id}) \circ (f^* \otimes f^* \otimes \mathrm{id})\tau = \tau.$$

Die Behauptung folgt wieder aus Satz 2.

(iii) und (iv) folgen direkt aus Satz 2, da $\sigma \oplus \tau$ und $\sigma \otimes \tau$ die die Multiplikation in $V \oplus W$ und $V \otimes W$ definierenden Tensoren sind.

Satz 11. *Sei V ein k-Ring mit Eins der Dimension n. Es gilt*

(i) $$\mathrm{Rg}\,(V) \geqq n$$

(ii) $$\mathrm{Rg}\,(V) = n \Leftrightarrow V \approx k \times \cdots \times k$$

Beweis. O. B. d. A. sei $V = k^n$ als Vektorraum. Sei (τ^{ijl}) der V zugrundeliegende Tensor und

$$\tau^{ijl} = \sum_{q=1}^{N} u_q^i v_q^j w_q^l$$

mit $N = \mathrm{Rg}\,(V)$. Sei $(e_1, \ldots, e_n) = 1 \in V$. Dann ist

$$\delta_{ml} = \sum_{i,j} \tau^{ijl} e_i \delta_{mj} = \sum_i \tau^{iml} e_i = \sum_i \sum_q u_q^i v_q^m w_q^l e_i = \sum_{q=1}^{N} \lambda_q v_q^m w_q^l \tag{3}$$

mit $\lambda_q = \sum_i u_q^i e_i$. Daraus folgt schon $N \geqq n$. Ist $N = n$, so folgt aus (3), daß (w_q^l) eine reguläre Matrix ist. Durch Basiswechsel können wir

$$(w_q^l) = (\delta_{ql})$$

erreichen, also

$$\tau^{ijl} = u_l^i v_l^j.$$

Aus (3) wird dann

$$\delta_{ml} = \lambda_l v_l^m,$$

also $\lambda_l \neq 0$ und

$$v_l^m = \frac{1}{\lambda_l} \delta_{ml}.$$

Aus Symmetriegründen

$$u_l^m = \frac{1}{\mu_l} \delta_{ml},$$

also

$$\tau^{ijl} = \begin{cases} \dfrac{1}{\lambda_l \mu_l} \text{ falls} & i = j = l \\ 0 & \text{sonst.} \end{cases}$$

Also ist $V \approx k \times \cdots \times k$. Die Umkehrung ist trivial.

Anwendungen 12. (i) Ist G eine abelsche Gruppe der Ordnung n, Char $(k) \nmid n$, so ist $\mathrm{Rg}\, k[G] = n$. Ist K ein separabler k_0-Körper, $[K: k_0] = n$, k algebraischer Abschluß von k_0 und V der k-Ring $K \otimes_{k_0} k$, so ist $\mathrm{Rg}\, V = n$.

(ii) Ist Vermutung 3 richtig, steht also insbesondere in Satz 10, (iii) das Gleichheitszeichen, und ist k etwa algebraisch abgeschlossen, so kann man nach dem Satz von Wedderburn den Rang halbeinfacher (assoziativer) k-Ringe auf den Rang von vollen Matrizenringen $M_m = M_m(k)$ zurückführen. $\mathrm{Rg}\, M_m$ ist bis auf einen Faktor $\leqq 2$ der minimale zur Multiplikation zweier allgemeiner m-reihiger Matrizen benötigte Zeitaufwand (bzgl. der Operationszeit $1_{\{*,/\}}$). In [7] wird

$$\mathrm{Rg}\, M_2 \leqq 7$$

gezeigt. Wegen $M_{2^p} \approx M_2 \otimes \cdots \otimes M_2$ folgt nach Satz 10

$$\mathrm{Rg}\, M_{2^p} \leqq 7^p,$$

und daraus leicht

$$\mathrm{Rg}\, M_m \leqq 3 m^{\lg 7}.$$

Tatsächlich erhält man in analoger Weise eine größenordnungsgleiche obere Abschätzung auch für die zur Multiplikation zweier Matrizen benötigte Gesamtanzahl von Operationen ([7], siehe auch [10], Corollary 7). Untere Schranken für Rg M_m sind nicht bekannt, abgesehen von der trivialen

$$\text{Rg } M_m \geqq m^2$$

(die z. B. aus Satz 11 folgt), und von

$$\text{Rg } M_2 \geqq 7$$

(Hopcroft-Kerr [3], Winograd [11]).

§ 5. Multiplikation spezieller Matrizen

Sei k algebraisch abgeschlossen, $G < Gl_n(k) < k^{n^2}$ eine irreduzible, lineare algebraische Gruppe (siehe z. B. [6], Chapter III, § 3), t_{ij} die von den Koordinatenprojektionen in k^{n^2} induzierten Elemente des Koordinatenringes von G. Seien π_1, π_2 die Projektionen von $G \times G$, $a_{ij} = t_{ij} \circ \pi_1$, $b_{ij} = t_{ij} \circ \pi_2$. Bekanntlich ist $G \times G$ eine irreduzible Varietät. Sei K der Funktionenkörper von $G \times G$, aufgefaßt als k-Körper. Wir interessieren uns für

$$L_K\left(\left\{\sum_{j=1}^{n} a_{ij}b_{jl} \colon i, l \leqq n\right\} \bmod a_{11}, \ldots, b_{nn}\right).$$

Nach [9], § 8 ist dies der minimale Zeitaufwand (bzgl. der Operationszeit $1_{\{*,/\}}$), den ein Programm zur Multiplikation von Matrizen $\in G$ für fast alle Inputmatrizenpaare beansprucht. Sei $\Theta < k^{n^2}$ der Tangentialraum an G im Punkte $1 \in G$. Da G nichtsingulär ist, ist $\dim \Theta = \dim G =: m$. Seien $\varepsilon^{(1)}, \ldots, \varepsilon^{(m)} \in k^{n^2}$ eine Basis von Θ, $\varepsilon^{(p)} = (\varepsilon_{ij}^{(p)})$. Seien ferner $x_1, \ldots, x_m, y_1, \ldots, y_m$ Unbestimmte über k. Wir fassen $k[\boldsymbol{x}, \boldsymbol{y}]$ auf als k-Ring über $\{x_1, \ldots, y_m\}$.

Satz 1.

$$L_K\left(\left\{\sum_{j=1}^{n} a_{ij}b_{jl} \colon i, l \leqq n\right\} \bmod a_{11}, \ldots, b_{nn}\right)$$

$$\geqq L_{k[\boldsymbol{x},\boldsymbol{y}]}\left(\left\{\sum_{j=1}^{n}\left(\sum_{p=1}^{m} \varepsilon_{ij}^{(p)} x_p\right)\left(\sum_{q=1}^{m} \varepsilon_{jl}^{(q)} y_q\right);\ i, l \leqq n\right\} \bmod k[\boldsymbol{x}] \cup k[\boldsymbol{y}]\right).$$

Beweis. Sei β eine optimale ausführbare Berechnung von

$$\left\{\sum_{j} a_{ij}b_{jl} \colon i, l \leqq n\right\} \bmod a_{11}, \ldots, b_{nn} \text{ in } K.$$

Es gibt $(\Gamma, \Delta) \in G \times G$ so, daß alle von 0 verschiedenen Ergebnisse a_i von β im Punkte (Γ, Δ) definiert und $\neq 0$ sind, daß also a_i Einheiten im lokalen Ring $\mathcal{O}_{(\Gamma,\Delta)} = \mathcal{O}_{G\times G,(\Gamma,\Delta)}$ sind. Wir fassen die lokalen Ringe von Punkten von $G \times G$ als kommutative k-Ringe mit Division durch Einheiten auf. Dann erhalten wir also

$$L_K\left(\left\{\sum_{j} a_{ij}b_{jl} \colon i, l \leqq n\right\} \bmod a_{11}, \ldots, b_{nn}\right) = L(\beta)$$

$$\geqq L_{\mathcal{O}_{(\Gamma,\Delta)}}\left(\left\{\sum_{j} a_{ij}b_{jl} \colon i, l \leqq n\right\} \bmod a_{11}, \ldots, b_{nn}\right).$$

Nun sei $h \colon \mathcal{O}_{(1,1)} \to \mathcal{O}_{(\Gamma,\Delta)}$ der durch

$$h(u)(M, N) := u(\Gamma^{-1}M, N\Delta^{-1})$$

für $M, N \in G$ definierte Isomorphismus. Wir haben

$$h(a_{ij}) = \sum_{r=1}^{n} t_{ir}(\Gamma^{-1}) a_{rj},\ h(b_{jl}) = \sum_{s=1}^{n} b_{js} t_{sl}(\Delta^{-1}).$$

Nach dem Simulationssatz gilt

$$\begin{aligned} & L_{\mathcal{O}_{(\Gamma,\Delta)}}(\{\sum_j a_{ij}b_{jl} : i, l \leqq n\} \bmod a_{11}, \ldots, b_{nn}) \\ = & L_{\mathcal{O}_{(1,1)}}(\{\sum_{r,s} t_{ir}(\Gamma^{-1})\,(\sum_j a_{rj}b_{js}) t_{sl}(\Delta^{-1}) : i, l \leqq n\} \\ & \bmod \{\sum_r t_{ir}(\Gamma^{-1}) a_{rj} : i, j \leqq n\} \cup \{\sum_s b_{js} t_{sl}(\Delta^{-1}) : j, l \leqq n\}) \\ = & L_{\mathcal{O}_{(1,1)}}(\{\sum_j a_{rj}b_{js} : r, s \leqq n\} \bmod a_{11}, \ldots, b_{nn}) \end{aligned}$$

nach Lemma 2. 1. Zusammen ergibt sich

$$\begin{aligned} (1) \qquad & L_K(\{\sum_j a_{ij}b_{jl} : i, l \leqq n\} \bmod a_{11}, \ldots, a_{nn}) \\ & \geqq L_{\mathcal{O}_{(1,1)}}(\{\sum_j a_{ij}b_{jl} : i, l \leqq n\} \bmod a_{11}, \ldots, b_{nn}). \end{aligned}$$

$\Theta \times \Theta < k^{n^2} \times k^{n^2}$ ist der Tangentialraum von $G \times G$ an der Stelle $(1, 1)$, also bilden $(\varepsilon^{(1)}, 0), \ldots, (\varepsilon^{(m)}, 0), (0, \varepsilon^{(1)}), \ldots, (0, \varepsilon^{(m)})$ eine Basis dieses Tangentialraums. Seien $u_1, \ldots, u_m$ lokale Parameter von G an der Stelle 1, und zwar so, daß $d_1 u_1, \ldots, d_1 u_m$ die zu $\varepsilon^{(1)}, \ldots, \varepsilon^{(m)}$ duale Basis ist. Dann sind $u_1 \circ \pi_1, \ldots, u_m \circ \pi_1, u_1 \circ \pi_2, \ldots, u_m \circ \pi_2$ lokale Parameter von $G \times G$ an der Stelle $(1, 1)$ und $d_{(1,1)}\, u_1 \circ \pi_1, \ldots, d_{(1,1)}\, u_m \circ \pi_2$ bilden die zu $(\varepsilon^{(1)}, 0), \ldots, (0, \varepsilon^{(m)})$ duale Basis. Es gibt einen Monomorphismus

$$\varphi : \mathcal{O}_{(1,1)} \to k[[x_1, \ldots, x_m, y_1, \ldots, y_m]]$$

von k-Ringen (mit Division durch Einheiten) so, daß

$$u_p \circ \pi_1 \mapsto x_p,\ u_p \circ \pi_2 \mapsto y_p$$

und so, daß $k[[x_1, \ldots, y_m]]$ vermöge φ die Vervollständigung von $\mathcal{O}_{(1,1)}$ nach seinem maximalen Ideal ist (für die Konstruktion siehe [6], Chapter II, § 2).

Um die linearen Komponenten von $\varphi(a_{ij})$ und $\varphi(b_{jl})$ zu bestimmen, bemerken wir

$$d_{(1,1)}\, a_{ij} = \sum_p \varepsilon_{ij}^{(p)} d_{(1,1)}\,(u_p \circ \pi_1)$$

$$d_{(1,1)}\, b_{jl} = \sum_q \varepsilon_{jl}^{(q)} d_{(1,1)}\,(u_q \circ \pi_2),$$

wie man leicht durch Auswerten an den Basisvektoren $(\varepsilon^{(1)}, 0)$ bis $(0, \varepsilon^{(m)})$ von $\Theta \times \Theta$ nachprüft. Daraus folgt (nach Konstruktion von φ)

$$\varphi(a_{ij}) = \delta_{ij} + \sum_p \varepsilon_{ij}^{(p)}\, x_p + \cdots$$

$$\varphi(b_{jl}) = \delta_{jl} + \sum_q \varepsilon_{jl}^{(q)}\, y_q + \cdots,$$

wobei $\varphi(a_{ij})$ bzw. $\varphi(b_{jl})$ Potenzreihen in den x_p bzw. y_q allein sind. Daher hat der quadratische Bestandteil von $\varphi(\sum_j a_{ij}b_{jl})$ die Gestalt

$$\sum_j (\sum_p \varepsilon_{ij}^{(p)} x_p)(\sum_q \varepsilon_{jl}^{(q)} y_q) + c_{il}(\boldsymbol{x}) + d_{il}(\boldsymbol{y}).$$

Wie im Beweis von Satz 2. 2 ergibt sich nun

$$L_{\mathcal{O}_{(1,1)}}(\{\sum_j a_{ij}b_{jl} : i, l \leqq n\} \bmod a_{11}, \ldots, b_{nn})$$
$$\geqq L_{k[[x,y]]}(\{\sum_j \varphi(a_{ij})\varphi(b_{jl}) : i, l \leqq n\} \bmod \varphi(a_{11}), \ldots, \varphi(b_{nn}))$$
$$\geqq L_{k[x,y]}(\{\sum_j (\sum_p \varepsilon_{ij}^{(p)} x_p)(\sum_q \varepsilon_{jl}^{(q)} y_q) : i, l \leqq n\} \bmod k[x] \cup k[y]).$$

Hieraus und aus (1) folgt der Satz.

Korollar 2.

$$L_K(\{\sum_j a_{ij}b_{jl} : i, l \leqq n\} \bmod a_{11}, \ldots, b_{nn})$$
$$\geqq \frac{1}{2} L_{k[x,y]}(\{\sum_j (\sum_p \varepsilon_{ij}^{(p)} x_p)(\sum_q \varepsilon_{jl}^{(q)} y_q) : i, l \leqq n\}).$$

Beweis. Dies folgt aus Satz 1 nach zweimaliger Anwendung von Lemma 4. 6.

Beispiel 3. Ist G die spezielle orthogonale Gruppe, so besteht ihr Tangentialraum an 1 aus den schiefsymmetrischen Matrizen. Korollar 2 besagt dann, daß der Mindestzeitaufwand von allgemeinen Verfahren (mit Division) zur Multiplikation m-reihiger orthogonaler Matrizen der Determinante 1 nicht kleiner ist als der halbe Mindestzeitaufwand allgemeiner Verfahren (ohne Division) zur Multiplikation m-reihiger schiefsymmetrischer Matrizen. Dieser läßt sich elementar sofort durch $\zeta_{\lfloor \frac{m}{2} \rfloor}$ abschätzen.

Korollar 4. *Sei* $\mathfrak{G}$ *der Liering von* G. *Dann gilt*

$$L_K(\{\sum_j a_{ij}b_{jl} : i, l \leqq n\} \bmod a_{11}, \ldots, b_{nn}) \geqq \frac{1}{4} \operatorname{Rg} \mathfrak{G}.$$

Beweis. Man verwende Lemma 4. 6 und die Tatsache, daß $\mathfrak{G}$ von Θ zusammen mit der Klammermultiplikation realisiert wird (siehe [5], LA 1. 3).

Dieses Korollar ist weniger genau als Korollar 2. Dafür eröffnet es die Möglichkeit, die Strukturtheorie der Lieringe auszunutzen (vgl. § 4, besonders Vermutung 4. 3).

Literatur

[1] *N. Bourbaki*, Eléments de mathématique, Algèbre I, Chapitres 1 à 3, Paris 1970.

[2] *N. Bourbaki*, Eléments de mathématique, Livre II, Algèbre, Chap. 4, Polynomes et fractions rationnelles, Paris 1959.

[3] *J. E. Hopcroft* and *L. R. Kerr*, On minimizing the number of multiplications necessary for matrix multiplication, TR 69—44, Dept. of Computer Science, Cornell University, Ithaca, N. Y. 1969.

[4] *D. Mumford*, Introduction to algebraic geometry, Harvard Lecture Notes.

[5] *J.-P. Serre*, Lie algebras and Lie groups, 1964 Lectures given at Harvard University, New York 1965.

[6] *I. R. Shafarevich*, Foundations of algebraic geometry, Russian Math. Surveys **24** (1969), 1—178.

[7] *V. Strassen*, Gaussian elimination is not optimal, Numerische Mathematik **13** (1969), 354—356.

[8] *V. Strassen*, Berechnung und Programm. I, Acta Informatica **1** (1972), 320—335.

[9] *V. Strassen*, Berechnung und Programm. II, Acta Informatica **2** (1973), 64—79.

[10] *S. Winograd*, On the algebraic complexity of functions, Actes Congrès Intern. Math. **3** (1970), 283—288.

[11] *S. Winograd*, On the multiplication of 2 by 2 matrices, IBM Research Report RC 2767, January 1970.