Algorithmische Beweise für Nichtnegativ- und Positivstellensätze

Markus Schweighofer

Diplomarbeit an der Universität Passau
bei Prof. Dr. Volker Weispfenning

Eingereicht im März 1999

Markus Schweighofer
Allersdorf 10
D-94262 Kollnburg

schweigh@fmi.uni-passau.de

Inhaltsverzeichnis

Kapitel 1

Einleitung

„Stellensätze“ sind Theoreme, die zeigen, daß aus einem geometrischen Sachverhalt eine algebraische Beziehung folgt, die den geometrischen Sachverhalt offensichtlich werden läßt. Genauer gesagt handelt es sich bei dem geometrischen Sachverhalt darum, daß ein Polynom f ein gewisses Vorzeichenverhalten zeigt auf einer Menge S von Punkten, die durch das Vorzeichenverhalten anderer Polynome definiert ist. Man unterscheidet die Stellensätze in Nullstellensätze, Nichtnegativstellensätze und Positivstellensätze, je nachdem ob der geometrische Sachverhalt $f = 0$ auf S, $f \geq 0$ auf S oder $f > 0$ auf S ist. Der bekannteste Stellensatz dürfte wohl folgender Nullstellensatz sein (hier ist $S = \{a\}$ durch das Polynom $X - a$ definiert):

Sei R ein kommutativer Ring mit Eins, $f \in R[X]$ und $a \in R$. Genau dann ist $f(a) = 0$, wenn $X - a$ in $R[x]$ ein Teiler von f ist.

Wie jeder andere Stellensatz behauptet er eine Äquivalenz einer geometrischen und einer algebraischen Bedingung, wobei der Schluß vom Geometrischen auf das Algebraische nichttrivial und der Schluß vom Algebraischen auf das Geometrische trivial ist. Namensgeber für alle Stellensätze ist der im Jahre 1893 von Hilbert bewiesene Hilbertsche Nullstellensatz:

Sei C ein algebraisch abgeschlossener Körper. Seien $f, q_1, \ldots, q_m \in C[X_1, \ldots, X_d]$. Genau dann gilt für alle $x \in C^d$

$$q_1(x) = \cdots = q_m(x) = 0 \implies f(x) = 0,$$

wenn es ein $N \in \mathbb{N}$ gibt, sodaß f^N in dem von $q_1, \ldots, q_m$ erzeugten Ideal in $C[X_1, \ldots, X_d]$ liegt.

In dieser Arbeit geht es allerdings ausschließlich um Nichtnegativ- und Positivstellensätze. Der wohl bekannteste Nichtnegativstellensatz ist die Lösung des im Jahre 1900 von Hilbert gestellten 17. Problems durch Artin im Jahre 1927 (siehe etwa [BCR]):

Sei R ein reell abgeschlossener Körper und $f \in R[X_1, \ldots, X_d]$. Genau dann ist $f \geq 0$ auf R^d, wenn f eine Darstellung der Form

$$f = \sum_i \left(\frac{f_i}{g_i}\right)^2$$

mit $f_i, g_i \in R[X_1, \ldots, X_d]$ besitzt.

Die Lösung des 17. Hilbertschen Problems, welches die Frage nach der Gültigkeit des obigen Satzes für $R = \mathbb{R}$ war, wurde nur möglich durch die Entwicklung der gänzlich neuen

Theorie der reell abgeschlossenen Körper, deren Prototyp $\mathbb{R}$ ist, durch Artin und Schreier. Lange Zeit schien es so, als wären die reell abgeschlossenen Körper der einzig richtige Rahmen, um Nichtnegativ- und Positivstellensätze zu beweisen. Es gelangen weitreichende Verallgemeinerungen und Varianten der Lösung des 17. Hilbertschen Problems. Einen Meilenstein setzte dabei Stengle im Jahre 1974 (siehe [St1]), in dessen Gefolgschaft zum Beispiel folgender Positivstellensatz steht (siehe [BCR]):

Sei R ein reell abgeschlossener Körper. Durch $p_1, \ldots, p_n \in R[X_1, \ldots, X_d]$ sei die Menge

$$S := \{x \in R^d \mid p_1(x) \geq 0, \ldots, p_n(x) \geq 0\}$$

definiert. Sei $f \in R[X_1, \ldots, X_d]$. Genau dann gilt $f > 0$ auf S, wenn f von der Form

$$f = \frac{1 + \sum_{e \in \{0,1\}^n} f_e p_1^{e_1} \cdots p_n^{e_n}}{\sum_{e \in \{0,1\}^n} g_e p_1^{e_1} \cdots p_n^{e_n}}$$

ist, wobei f_e und g_e für jedes $e \in \{0,1\}^n$ eine Summe von Quadraten in $R[X_1, \ldots, X_d]$ ist.

Diesen Satz und Funktionalanalysis benutzend, bewies im Jahre 1990 Schmüdgen in [Sch] mit der Absicht, Momentsequenzen positiver Borel-Maße auf $\mathbb{R}^d$ mit kompaktem Träger zu charakterisieren, das erstaunliche Resultat, daß man in obigem Satz unter der zusätzlichen Voraussetzung, daß S kompakt und $R = \mathbb{R}$ ist, in der Darstellung von f auf den Nenner verzichten kann, wenn man im Zähler statt 1 jede beliebige positive reelle Zahl zuläßt. Spätestens seit der 1998 veröffentlichten Dissertation von Wörmann [Wör] erkennt man, daß dieser Positivstellensatz ein Hybrid-Resultat aus Artin-Schreier-Theorie und einer bis dahin immer im Hintergrund gebliebenen Serie von Positivstellensätzen eines anderen Typs ist, den wir Kadison-Dubois nennen wollen. Ein reiner Vertreter des Typs Kadison-Dubois ist der folgende Positivstellensatz, den unseres Wissens zum ersten Mal Handelman 1988 in [Ha2] (für ein S, das zusätzlich nichtleeres Inneres hat) bewiesen hat:

Der durch die linearen Polynome (d.h. Polynome vom Grad ≤ 1) $p_1, \ldots, p_n \in \mathbb{R}[X_1, \ldots, X_d]$ definierte konvexe Polyeder

$$S := \{x \in \mathbb{R}^d \mid p_1(x) \geq 0, \ldots, p_n(x) \geq 0\}$$

sei kompakt und nichtleer. Sei $f \in \mathbb{R}[X_1, \ldots, X_d]$. Genau dann ist $f > 0$ auf S, wenn f von der Form

$$f = a + \sum_e a_e p_1^{e_1} \cdots p_n^{e_n}$$

ist mit $a \in \mathbb{R}^{>0}$ und $a_e \in \mathbb{R}^{\geq 0}$ für alle $e \in \mathbb{N}^n$, über die summiert wird.

Wie wir in dieser Arbeit sehen werden, zeigt dieser Satz die typischen Vor- und Nachteile seiner Gattung gegenüber den Artin-Schreier-Resultaten. Die Nachteile sind:

- Als Grundkörper können nur die reellen Zahlen verwendet werden.
- Die Menge S nimmt sehr bestimmte Formen an.
- Die Darstellung von f wird unter Umständen beliebig „groß“, wenn f auf S genügend kleine positive Werte annimmt.

Die Vorteile sind:

- Es tauchen keine Quadrate in der Darstellung von f auf.
- Es tauchen keine oder nur sehr bestimmte Nenner in der Darstellung von f auf.

- Die Darstellung von f scheint einer algorithmischen Berechnung zugänglich.

Der zuletzt genannte Vorteil ist die hauptsächliche Erkenntnis dieser Arbeit. Erstaunlich ist, daß dieser Vorteil etabliert werden kann, obwohl ihm der zuletzt genannte Nachteil zuwider zu laufen scheint.

Dieser Schein trügt!

Der erste Anhaltspunkt für diese Tatsache war, daß es einen Vertreter des Typs Kadison-Dubois gibt, der ganz offensichtlich eine algorithmische Berechnung der Darstellung von f erlaubt, obwohl er eben auch die schlechte Eigenschaft einer „explodierenden“ Darstellung von f bei kleinen Werten von f auf S zeigt. Es ist dies der folgende 1927 von Pólya ebenso trickreich wie kurz und elementar bewiesene Satz:

Sei $f \in \mathbb{R}[X_1, \ldots, X_d]$ eine Form (d.h. ein homogenes Polynom). Genau dann ist $f > 0$ auf $(\mathbb{R}^{\geq 0})^d \setminus \{0\}$, wenn f eine Darstellung

$$f = \frac{\sum_{e_1+\cdots+e_d=N+\deg f} a_e X_1^{e_1} \cdots X_d^{e_d}}{(X_1 + \cdots + X_d)^N}$$

mit $a_e \in \mathbb{R}^{>0}$ hat.

Der Algorithmus zur Berechnung einer Darstellung von f ist hier trivial: Wenn eine Form $f \in \mathbb{R}[X_1, \ldots, X_d]$ mit $f > 0$ auf $(\mathbb{R}^{\geq 0})^d \setminus \{0\}$ vorliegt, so berechne man für $N = 0, 1, 2, 3, \ldots$ das Produkt $f(X_1 + \cdots + X_d)^N$. Irgendwann erhält man eine Form, welche ein passender Zähler für die Darstellung von f ist. Trotzdem erfüllt der Satz den letzten der drei genannten Nachteile (in dem Sinne, daß wir mit $S = \{x \in (\mathbb{R}^{\geq 0})^d \mid x_1 + \cdots + x_d = 1\}$ anstelle von $(\mathbb{R}^{\geq 0})^d \setminus \{0\}$ arbeiten, was wegen der Homogenität von f dasselbe ist): In einer Darstellung der von $a \in \mathbb{R}$ abhängigen Form $X_1^2 - aX_1X_2 + X_2^2 = (X_1 - X_2)^2 + (2 - a)X_1X_2$ strebt dann zum Beispiel der kleinstmögliche Exponent N im Nenner gegen ∞, wenn a von unten gegen 2 strebt (siehe Abschnitt 3.1).

Wir zeigen in dieser Arbeit, daß man die Positivstellensätze vom Typ Kadison-Dubois auf den Pólyaschen Satz zurückführen kann, und zwar in gewisser algorithmischer Weise. Wir gewinnen so zum Beispiel einen schönen Algorithmus für den erwähnten Satz von Handelman, also zur Berechnung einer Darstellung eines Polynoms, welche offensichtlich macht, daß dieses Polynom auf einem nichtleeren kompakten konvexen Polyeder positiv ist. Beim Positivstellensatz von Schmüdgen, der ja nur zur Hälfte vom Typ Kadison-Dubois ist, können wir immerhin noch das Problem, eine Darstellung für jedes f mit $f > 0$ auf S zu finden, darauf reduzieren, für irgendein $s \in \mathbb{R}$ eine Darstellung von $s - \sum_{i=1}^d X_i^2$ zu finden. Da dort S als kompakt vorausgesetzt ist, gibt es für genügend großes s die letztere Darstellung. Allerdings muß dazugesagt werden, daß unser Algorithmus die Quadrate, die er für die Darstellung von f beim Schmüdgen-Positivstellensatz benötigt, im Wesentlichen aus dieser bis auf Weiteres vom Menschen zu findenden Darstellung von $s - \sum_{i=1}^d X_i^2$ bezieht.

An den Artin-Schreier-Anteil beim Schmüdgenschen Satz scheint man also algorithmisch nicht gut heranzukommen. Man sollte hier keine zu großen Hoffnungen auf den Satz von Pólya legen. Dieser wurde zwar schon einmal 1940 von Habicht (siehe [Hab]) herangezogen, um einen Positivstellensatz zu algorithmisieren, nämlich eine Abschwächung ausgerechnet des Ur-Artin-Schreier-Resultats, nämlich des 17. Hilbertschen Problems. In einer neueren Variante dieses Algorithmus von Loera und Santos (siehe [LS]) wird allerdings ganz deutlich, daß in dieser Abschwächung zur Darstellung von f nur noch die Quadrate $X_1^2, \ldots, X_d^2$ und ein einziges von f abhängiges Quadrat notwendig ist, ähnlich wie wir für den Schmüdgenschen Satz nur noch ganz bestimmte Quadrate benötigen, wenn eines der S definierenden Polynome p_i von der Form $s - \sum_{i=1}^d X_i^2$ mit $s \in \mathbb{R}$ ist.

Endlich sollten wir sagen, wodurch sich der Typ Kadison-Dubois unserer Vorstellung nach eigentlich konstituiert. Wir haben diesen Typ so bezeichnet nach dem Darstellungssatz von Kadison-Dubois (siehe Kapitel 4). Grob gesagt ist dieser eine weitreichende Verallgemeinerung des Satzes, daß jeder archimedisch angeordnete Körper (bis auf Isomorphie) ein Unterkörper von $\mathbb{R}$ ist. Statt archimedisch angeordneter Körper betrachtet man in einem sehr allgemeinen Sinn „archimedisch angeordnete“ Ringe. Bisher gab es zwei Beweise für (etwas unterschiedliche Versionen) dieses Satzes, den funktionalanalytischen Originalbeweis, und einen sehr viel kürzeren und (so weit wie möglich) algebraischen Beweis von Becker und Schwartz. Mit Hilfe des Satzes von Pólya geben wir in dieser Arbeit einen dritten und sehr kurzen Beweis zumindest für die Version des Satzes, die für die meisten der vielfältigen Anwendungen ausreicht. Zu diesen Anwendungen gehört etwa die Untersuchung $2k$-ter Potenzen in Körpern (siehe etwa [Be2]).

Angewandt auf eine gewisse Situation nimmt der Satz von Kadison-Dubois die Form eines Positivstellensatzes an mit einer etwas undurchsichtigen Voraussetzung. Diesen Positivstellensatz nennen wir „archimedischen Positivstellensatz“. Unter einem Positivstellensatz des Typs Kadison-Dubois verstehen wir einen Positivstellensatz, der ohne große Mühe aus diesem „archimedischen Positivstellensatz“ abgeleitet werden kann. Daß der Satz von Pólya von diesem Typ ist, ist erst seit der Dissertation von Wörmann [Wör] bekannt. Hier bestreiten wir nun den umgekehrten Weg und zeigen, daß der archimedische Positivstellensatz aus dem Satz von Pólya folgt.

Außerhalb dieses Programms steht ein zusätzliches Kapitel über univariate Polynome mit Koeffizienten aus einem Unterkörper K von $\mathbb{R}$, die auf der reellen Achse nirgends einen negativen Wert annehmen. Ein Algorithmus zur Darstellung solcher Polynome als gewichtete Summe (mit nichtnegativen Gewichten aus K) von Quadraten von Polynomen mit Koeffizienten aus dem selben Körper K wird entwickelt und implementiert. Es handelt sich dabei um eine geradlinige, aber mühsame Verallgemeinerung bekannter Resultate.

1.1 Leseanleitung

Diese Arbeit besteht aus zwei voneinander völlig unabhängigen Teilen: Der eine Teil ist Kapitel 2. Der andere Teil wird durch die Kapitel 3 und 4 gebildet. Beim Lesen des zweiten Teils kann man nach dem Abschnitt 3.2 guten Gewissens zu Kapitel 4 springen. Ebenso kann man nach Abschnitt 3.3 direkt zu Abschnitt 3.5 oder 3.6 springen.

In dieser Arbeit ist stets $0 \in \mathbb{N}$. Unter einem Ring verstehen wir einen kommutativen Ring mit 1. Unter einem Ringhomomorphismus verstehen wir konsequenterweise einen Homomorphismus von Ringen in diesem Sinne. Jeder Ringhomomorphismus muß also 1 auf 1 abbilden. Ein kompakter Raum muß nicht notwendig Hausdorffsch sein. Ausdrücke wie „$f = 0$ auf S“ oder „$f < g$ auf S“ sind zu lesen als „$f(x) = 0$ für alle $x \in S$“ bzw. „$f(x) < g(x)$ für alle $x \in S$“.

Ein letzter Hinweis betrifft die Verwendung des Begriffs „Algorithmus“ in dieser Arbeit. Um den hiesigen Sprachgebrauch zu motivieren, betrachten wir als Beispiel den bekannten Euklidischen „Algorithmus“ in $K[X]$ (K ein Körper). Er ist gleichzeitig ein „Algorithmus“ zur Berechnung und ein Beweis der Existenz des größten gemeinsamen Teilers zweier Polynome aus $K[X]$. Ist er wirklich beides? Nach der üblichen Formalisierung des Begriffs Algorithmus müßten wir für $K = \mathbb{R}$ effektiv mit reellen Zahlen rechnen können, um zurecht von einem Algorithmus sprechen zu können. Es gibt aber überabzählbar viele reelle Zahlen. Wir können also reelle Zahlen in den Maschinenmodellen, durch die der Begriff Algorithmus üblicherweise formalisiert wird, nicht einmal repräsentieren, geschweige denn damit rechnen. Häufig wird deswegen in irgendeiner Weise vorausgesetzt, daß in K das

Rechnen effektiv möglich ist. Wollen wir den „Algorithmus“ aber nur als Existenzbeweis lesen, so ist diese Voraussetzung unnötig.

Diesem Dilemma entfliehen wir, indem wir für jeden Körper K den Begriff eines Algorithmus *modulo Rechnen im Körper* K einführen. Gegenüber einem Algorithmus im üblichen Sinne soll ein Algorithmus modulo Rechnen im Körper K die folgenden zusätzlichen Möglichkeiten haben: Er kann Körperelemente aus K repräsentieren und auf Gleichheit testen. Er kann sich ein Element von K beschaffen. Er kann aus zwei Zahlen $a, b \in K$ die Zahlen $a + b$, $-a$, ab und falls $a \neq 0$ auch a^{-1} effektiv generieren. (Zur Übung überlege man sich einen Algorithmus modulo K, der das Nullelement und das Einselement von K generiert.) Sprechen wir von einem Algorithmus *modulo Rechnen im angeordneten Körper* K, so erlauben wir zusätzlich noch, daß der Algorithmus zwei Zahlen $a, b \in K$ einem „Orakel“ geben darf, welches dann auf die Frage, ob $a \leq b$ gilt, die richtige Antwort zurückliefert.

1.2 Vermischtes

Die Wendung „so daß“ wird in dieser Arbeit zusammengeschrieben. Obwohl dies in Deutschland unüblich ist, verzeihe der Leser das angesichts der Tatsache, daß diese Arbeit in unmittelbarer Nähe zu Österreich geschrieben wurde, wo die Zusammenschreibung „sodaß“ verbreitet ist.

Mein Dank gilt Prof. Weispfenning, der diese Arbeit betreute, und den übrigen hiesigen Professoren im Bereich Mathematik, vor allem Prof. Schwartz und Prof. Volger.

Kapitel 2

Nichtnegative univariate Polynome

Sei $f \in \mathbb{R}[X]$ ein univariates Polynom, das nirgends einen negativen Wert annimmt, d.h. $f(x) \geq 0$ für alle $x \in \mathbb{R}$. Dann ist f bekanntlich eine Summe von zwei Quadraten in $\mathbb{R}[X]$.

Dies zeigt man wie folgt: O.B.d.A. sei f normiert. Wir schreiben

$$f = \prod_j (X - a_j)^{r_j} \prod_k ((X - (b_k + ic_k))(X - (b_k - ic_k)))^{s_k}$$

in $\mathbb{C}[X]$ mit $a_j, b_k, c_k \in \mathbb{R}, r_j, s_k > 0$ und paarweise verschiedenen $a_j, (b_k + ic_k), (b_k - ic_k)$. Da f nirgends einen negativen Wert annimmt, müssen alle r_j gerade sein, also gibt es $g, p, q \in \mathbb{R}[X]$ mit

$$g^2 = \prod_j (X - a_j)^{r_j} \text{ und } q \pm ir = \prod (X - (b_k \pm ic_k))^{s_k}.$$

Dann ist $f = g^2(q + ir)(q - ir) = g^2(q^2 + r^2) = (gq)^2 + (gr)^2$.

Sei nun K ein Unterkörper von $\mathbb{R}$ und $f \in K[X]$, sodaß $f(x) \geq 0$ für alle $x \in \mathbb{R}$. Aus trivialen Gründen kann es passieren, daß f keine Summe von Quadraten in $K[X]$ ist, etwa wenn $K = \mathbb{Q}(\sqrt{2})$ und f das konstante Polynom $\sqrt{2}$ ist. Wäre nämlich $\sqrt{2}$ eine Summe von Quadraten in $K[X]$, so auch in K. In keiner Anordnung des Körpers $K = Q(\sqrt{2})$ dürfte dann $\sqrt{2}$ negativ sein. Man konstruiert aber leicht so eine Anordnung, indem man die natürliche Anordnung von $\mathbb{Q}(\sqrt{2})$ mit dem Körperautomorphismus $\mathbb{Q}(\sqrt{2}) \to \mathbb{Q}(\sqrt{2}) : \sqrt{2} \mapsto -\sqrt{2}$ transportiert.

Statt nach einer Darstellung als Summe von Quadraten in $K[X]$ fragen wir deshalb nach einer Darstellung als gewichtete Summe von Quadraten in $K[X]$ mit nichtnegativen Gewichten aus K. Genauer lautet die Frage:

Sei K ein Unterkörper von $\mathbb{R}$ und $f \in K[X]$, sodaß $f(x) \geq 0$ für alle $x \in \mathbb{R}$. Gibt es $a_i \in K^{\geq 0}$ und $g_i \in K[X]$, sodaß $f = \sum_i a_i g_i^2$?

Wir werden diese Frage positiv beantworten. Der Beweis liefert einen Algorithmus (modulo Rechnen im angeordneten Körper K) zur Berechnung geeigneter a_i, g_i. Für den Fall $K = \mathbb{Q}$ haben wir den Algorithmus implementiert. Man beachte übrigens, daß im Fall $K = \mathbb{Q}$ (genauso wie im Fall $K = \mathbb{R}$) die Gewichte a_i überflüssig sind, da für $p, q \in \mathbb{N}$ mit $q \neq 0$ gilt $\frac{p}{q} = \frac{pq}{q^2} = (\frac{1}{q})^2 + \cdots + (\frac{1}{q})^2$.

Im Abschnitt 2.5 befassen wir uns dann damit, inwiefern die Resultate verallgemeinerbar sind auf den Fall, daß wir anstelle von $(K, \mathbb{R})$ ein Paar (K, R) betrachten mit einem angeordneten Körper K und seinem reellen Abschluß R.

Der eingangs erwähnte Beweis der Existenz einer Darstellung als Summe von *zwei* Quadraten scheint in keiner Weise verallgemeinerbar auf den Fall eines zugrundeliegenden beliebigen Unterkörpers K von $\mathbb{R}$. Deswegen wäre es erstaunlich, wenn stets eine Darstellung als gewichtete Summe von *zwei* Quadraten existierte.

Stattdessen werden wir einen anderen einfachen Beweis verallgemeinern, der allerdings nur eine Darstellung als Summe von i.A. mehr als zwei Quadraten liefert.

2.1 Hilfsmittel

Wir brauchen einige Hilfsmittel, als erstes einige Tatsachen aus der Differentialrechnung, und zwar nur im Kontext von Polynomen. Obwohl diese Tatsachen durchaus bekannt sein dürften, haben wir sie nicht alle in der Literatur gefunden (das allgemeine Kriterium für lokale Extrema 2.13 haben wir zum Beispiel nur als Übungsaufgabe in [Brö] gefunden). Deshalb haben wir sie hier zusammengeschrieben und, da es keine größere Mühe bereitet, auch bewiesen.

Definition 2.1 (Ableitungen). *Wir definieren für $i \in \{1, \dots, d\}$ einen $\mathbb{R}$-Vektorraumendomorphismus $\frac{\partial}{\partial X_i}$ auf $\mathbb{R}[X_1, \dots, X_d]$ durch*

$$\frac{\partial}{\partial X_i} X_1^{e_1} \cdots X_d^{e_d} = \begin{cases} e_i X_1^{e_1} \cdots X_{i-1}^{e_{i-1}} X_i^{e_i - 1} X_{i+1}^{e_{i+1}} \cdots X_d^{e_d} & \text{falls } e_i \geq 1 \\ 0 & \text{falls } e_i = 0 \end{cases}$$

für alle $(e_1, \dots, e_d) \in \mathbb{N}^d$. Man überprüft sofort, daß $\frac{\partial}{\partial X_i} \circ \frac{\partial}{\partial X_j} = \frac{\partial}{\partial X_j} \circ \frac{\partial}{\partial X_i}$ für beliebige $i, j \in \{1, \dots, d\}$ gilt. Für $n \in \mathbb{N}$, $Y_1, \dots, Y_n \in \{X_1, \dots, X_d\}$ und $e_1, \dots, e_n \in \mathbb{N}$ bezeichne $\frac{\partial^{e_1 + \cdots + e_n}}{\partial Y_1^{e_1} \dots \partial Y_n^{e_n}}$ die Hintereinanderausführung (in beliebiger Reihenfolge) der $e_1 + \cdots + e_n$ Funktionen $\underbrace{\frac{\partial}{\partial Y_1}, \dots, \frac{\partial}{\partial Y_1}}_{e_1\text{-mal}}, \underbrace{\frac{\partial}{\partial Y_2}, \dots, \frac{\partial}{\partial Y_2}}_{e_2\text{-mal}}, \dots, \underbrace{\frac{\partial}{\partial Y_n}, \dots, \frac{\partial}{\partial Y_n}}_{e_n\text{-mal}}$. Ist $f \in \mathbb{R}[X]$ und $e \in \mathbb{N}$, so schreiben wir auch $f^{(e)}$ statt $\frac{\partial^e}{\partial X^e} f$. Statt $f^{(1)}, f^{(2)}, \dots$ schreiben wir auch $f', f'', \dots$

Lemma 2.2 (Produktregel). *Seien $p, q \in \mathbb{R}[X_1, \dots, X_d]$ und $i \in \{1, \dots, d\}$. Dann gilt:*

$$\frac{\partial pq}{\partial X_i} = \frac{\partial p}{\partial X_i} q + p \frac{\partial q}{\partial X_i}$$

Beweis: Die beiden Abbildungen

$$\mathbb{R}[X_1, \dots, X_d] \times \mathbb{R}[X_1, \dots, X_d] \to \mathbb{R}[X_1, \dots, X_d] : (p, q) \mapsto \frac{\partial pq}{\partial X_i}, (p, q) \mapsto \frac{\partial p}{\partial X_i} q + p \frac{\partial q}{\partial X_i}$$

sind bilinear. Daher genügt es, die Behauptung für den Fall $p = X_1^{e_1} \cdots X_d^{e_d}$ und $q = X_1^{e_1'} \cdots X_d^{e_d'}$ mit $(e_1, \dots, e_d), (e_1', \dots, e_d') \in \mathbb{N}^d$ nachzurechnen. □

Lemma 2.3. *Seien $p_1, \dots, p_n \in \mathbb{R}[X_1, \dots, X_d]$ und sei $i \in \{1, \dots, d\}$. Dann gilt:*

$$\frac{\partial p_1 \cdots p_n}{\partial X_i} = \frac{\partial p_1}{\partial X_1} p_2 \cdots p_n + p_1 \frac{\partial p_2}{\partial X_2} p_3 \cdots p_n + \cdots + p_1 \cdots p_{n-1} \frac{\partial p_n}{\partial X_n}$$

Beweis: Durch Induktion nach n mit Hilfe der Produktregel. □

Lemma 2.4 (Kettenregel). *Sei $f \in \mathbb{R}[X]$, $p \in \mathbb{R}[X_1, \ldots, X_d]$ und $i \in \{1, \ldots, d\}$. Dann gilt:*

$$\frac{\partial f(p)}{\partial X_i} = f'(p) \cdot \frac{\partial p}{\partial X_i}$$

Beweis: Es genügt, die Behauptung für $f = X^e$ mit $e \in \mathbb{N}$ zu zeigen. Dies macht man durch Induktion nach e mit Hilfe der Produktregel. □

Lemma 2.5 (Taylorentwicklung). *Sei $p \in \mathbb{R}[X_1, \ldots, X_d]$, $x \in \mathbb{R}^d$. Es gibt eine eindeutig bestimmte Familie $(a_e)_{e \in \mathbb{N}^d}$ mit endlichem Träger, sodaß*

$$p = \sum_{e \in \mathbb{N}^d} a_e (X_1 - x_1)^{e_1} \cdots (X_d - x_d)^{e_d},$$

und zwar

$$a_e = \frac{1}{e_1! \cdots e_d!} \cdot \frac{\partial^{e_1 + \cdots + e_d} p}{\partial X_1^{e_1} \ldots \partial X_d^{e_d}}(x) \quad \textit{für alle } e \in \mathbb{N}^d.$$

Beweis: Als Bild der Basis $(X_1^{e_1} \cdots X_d^{e_d})_{e \in \mathbb{N}^d}$ von $\mathbb{R}[X_1, \ldots, X_d]$ unter dem durch $X_i \mapsto X_i - x_i$ definierten $\mathbb{R}$-Algebrenisomorphismus (insbesondere $\mathbb{R}$-Vektorraumisomorphismus) von $\mathbb{R}[X_1, \ldots, X_d]$ ist auch $((X_1 - x_1)^{e_1} \cdots (X_d - x_d)^{e_d})_{e \in \mathbb{N}^d}$ eine Basis von $\mathbb{R}[X_1, \ldots, X_d]$. Damit ist Existenz und Eindeutigkeit von $(a_e)_{e \in \mathbb{N}^d}$ gezeigt. Die behauptete Gleichheit für a_e ergibt sich, wenn man $\frac{\partial^{e_1 + \cdots + e_d} p}{\partial X_1^{e_1} \ldots \partial X_d^{e_d}}(x)$ berechnet. □

Definition 2.6 (Form). *Sei $F \in \mathbb{R}[X_1, \ldots, X_d]$ und $k \in \mathbb{N}$. F heißt* k-Form*, wenn alle in F vorkommenden Monome (mit einem Koeffizienten $\neq 0$) den Grad k haben. F heißt* Form*, wenn F für ein $k \in \mathbb{N}$ eine k-Form ist.*

Beispiel 2.7. Das Nullpolynom ist eine k-Form für jedes $k \in \mathbb{N}$. Jede Form $F \neq 0$ ist genau für ein $k \in \mathbb{N}$ eine k-Form, nämlich für $k = \deg F$.

Definition 2.8 (Multinomialkoeffizient). *Für $k \in \mathbb{N}$ und $(e_1, \ldots, e_d) \in \mathbb{N}^d$ mit $e_1 + \cdots + e_d = k$ definieren wir den* Multinomialkoeffizienten

$$\binom{k}{e_1 \ldots e_d} := \frac{k!}{e_1! \cdots e_d!}.$$

Nach Lemma 2.5 gilt für jedes $p \in \mathbb{R}[X_1, \ldots, X_d]$ und für alle $x \in \mathbb{R}^d$

$$p = \sum_{k \in \mathbb{N}} \frac{1}{k!} \sum_{\substack{e \in \mathbb{N}^d \\ e_1 + \cdots + e_d = k}} \binom{k}{e_1 \ldots e_d} \frac{\partial^k p}{\partial X_1^{e_1} \ldots \partial X_d^{e_d}}(x) (X_1 - x_1)^{e_1} \cdots (X_d - x_d)^{e_d}.$$

Dies motiviert folgende Definition:

Definition 2.9 (Differential). *Sei $p \in \mathbb{R}[X_1, \ldots, X_d]$ und $x \in \mathbb{R}^d$. Für $k \in \mathbb{N}$ sei das* k-te Differential von p an der Stelle x *(in Zeichen $D^k p(x)$) die folgende k-Form:*

$$\sum_{\substack{e \in \mathbb{N}^d \\ e_1 + \cdots + e_d = k}} \binom{k}{e_1 \ldots e_d} \frac{\partial^k p}{\partial X_1^{e_1} \ldots \partial X_d^{e_d}}(x) X_1^{e_1} \cdots X_d^{e_d}$$

Beispiel 2.10. Sei $p \in \mathbb{R}[X_1, \ldots, X_d]$ und $x \in \mathbb{R}^d$. Dann ist $D^0p(x)$ die konstante 0-Form $p(x)$. Es ist $D^1p(x) = \frac{\partial p}{\partial X_1}(x)X_1 + \cdots + \frac{\partial p}{\partial X_d}(x)X_d$ die durch die Jordanmatrix $(\frac{\partial p}{\partial X_1}(x) \ldots \frac{\partial p}{\partial X_d}(x))$ von p an der Stelle x beschriebene Linearform (1-Form), und $D^2p(x)$ ist die durch die Hessematrix von p an der Stelle x beschriebene quadratische Form (2-Form).

Definition 2.11. *Eine* k-Form $F \in \mathbb{R}[X_1, \ldots, X_d]$ *heißt*

- positiv definit, *falls* $F(x) > 0$ *für alle* $x \in \mathbb{R}^d \setminus \{0\}$
- negativ definit, *falls* $F(x) < 0$ *für alle* $x \in \mathbb{R}^d \setminus \{0\}$
- indefinit, *falls es* $x, y \in \mathbb{R}^d \setminus \{0\}$ *mit* $F(x) < 0 < F(y)$ *gibt.*

Bemerkung 2.12. Sei $F \in \mathbb{R}[X_1, \ldots, X_d]$ eine k-Form. Dann gilt $F(\lambda x) = \lambda^k F(x)$ für alle $x \in \mathbb{R}^d$ und $\lambda \in \mathbb{R}$. Für $\lambda > 0$ ergibt sich, daß $F(\lambda x)$ dasselbe Vorzeichen hat wie $F(x)$. Das Vorzeichen von F ist also auf vom Nullpunkt ausgehenden Halbgeraden, die den Nullpunkt nicht enthalten, konstant. Deshalb kann man in obiger Definition $\mathbb{R}^d \setminus \{0\}$ ersetzen etwa durch die Einheitssphäre $S := \{x \in \mathbb{R}^d \mid \|x\| = 1\}$. Für ungerades k und $\lambda = -1$ erhält man $F(x) = -F(x)$. Für ungerades k kann also F weder positiv noch negativ definit sein.

Die folgenden Kriterien für lokale Extrema werden in der Analysis meist nur für den Fall $k = 1$ formuliert. Der Beweis für den allgemeinen Fall geht allerdings ganz analog.

Satz 2.13 (Kriterien für lokale Extrema). *Sei* $p \in \mathbb{R}[X_1, \ldots, X_d]$, $x \in \mathbb{R}^d$ *und* $k \in \mathbb{N}$, *sodaß* $D^1p(x) = \cdots = D^kp(x) = 0$. *Dann gilt:*

(i) Wenn $D^{k+1}p(x)$ *positiv definit ist, dann ist* x *eine isolierte lokale Minimalstelle von* p.

(ii) Wenn $D^{k+1}p(x)$ *negativ definit ist, dann ist* x *eine isolierte lokale Maximalstelle von* p.

(iii) Wenn $D^{k+1}p(x)$ *indefinit ist, dann ist* x *keine lokale Extremstelle von* p.

Beweis: O.B.d.A. sei $x = 0$. Wir können p in der Form $p = p(0) + F + r$ schreiben, wobei $F = D^{k+1}p(x)$ eine $(k+1)$-Form ist und r eine Summe von Monomen ist, von denen jedes einen Grad $> k+1$ hat. Wir behaupten $\lim_{x\to 0} \frac{r(x)}{\|x\|^{k+1}} = 0$. Hierfür reicht es zu zeigen, daß $\lim_{x\to 0} \frac{x_1^{e_1} \cdots x_d^{e_d}}{\|x\|^{k+1}} = 0$ für $(e_1, \ldots, e_d) \in \mathbb{N}^d$ mit $e_1 + \cdots + e_d > k+1$. Dies ist aber klar, denn der Zähler von $\frac{x_1^{e_1} \cdots x_d^{e_d}}{\|x\|^{k+1}}$ ist ein Produkt von mehr als $k+1$ Faktoren der Form x_i. Des gesamten Bruch kann man daher als Produkt von $k+1$ Faktoren der Form $\frac{x_i}{\|x\|}$ und wenigstens einem Faktor der Form x_i schreiben. Aus $\left|\left|\frac{x_i}{\|x\|}\right|\right| \leq 1$ und $\lim_{x\to 0} x_i = 0$ folgt daher die Behauptung.

Um nun (i) zu zeigen, sei F als positiv definit vorausgesetzt. Es ist dann nachzuweisen, daß es eine Umgebung U von 0 in $\mathbb{R}^d$ gibt, sodaß $F + r > 0$ auf $U \setminus \{0\}$ gilt. Da die Einheitssphäre $S := \{x \in \mathbb{R}^d \mid \|x\| = 1\}$ kompakt ist, nimmt F dort ein Minimum $c > 0$ an. Für alle $x \in \mathbb{R}^d \setminus \{0\}$ gilt

$$F(x) = \|x\|^{k+1} F\left(\frac{x}{\|x\|}\right) \geq \|x\|^{k+1} c.$$

Wegen $\lim_{x\to 0} \frac{r(x)}{\|x\|^{k+1}} = 0$ gibt es eine Umgebung U von 0 in $\mathbb{R}^d$, sodaß außerdem für alle $x \in U \setminus \{0\}$ gilt

$$\|x\|^{k+1} c > |r(x)|.$$

Insgesamt folgt für $x \in U \setminus \{0\}$, daß $F(x) + r(x) \geq F(x) - |r(x)| > 0$, wie gewünscht.

Analog zeigt man (ii). Um (iii) zu zeigen, setzen wir F als indefinit voraus. Wähle $x \in \mathbb{R}^d$ mit $\|x\| = 1$ und $F(x) > 0$. Für alle $\lambda \in \mathbb{R}$ gilt

$$F(\lambda x) = \lambda^{k+1} F(x) = \|\lambda x\|^{k+1} F(x).$$

Wegen $\lim_{\lambda \to 0} \frac{r(\lambda x)}{\|\lambda x\|^{k+1}} = 0$ gibt es ein $\varepsilon > 0$, sodaß außerdem für alle $\lambda \in]0, \varepsilon[$ gilt

$$\|\lambda x\|^{k+1} F(x) > |r(\lambda x)|.$$

Insgesamt folgt für $\lambda \in]0, \varepsilon[$, daß $F(\lambda x) + r(\lambda x) \geq F(\lambda x) - |r(\lambda x)| > 0$. Dies zeigt, daß der Nullpunkt keine lokale Minimalstelle von p ist. Analog zeigt man, daß er keine lokale Maximalstelle ist. □

Man beachte, daß durchaus auch der Fall $k = 0$ in obigem Satz zugelassen ist. In diesem Fall ist die Voraussetzung $D^1 p(x) = \cdots = D^k p(x) = 0$ leer. Wir erhalten dann:

Folgerung 2.14. *Sei $p \in \mathbb{R}[X_1, \ldots, X_d]$ und $x \in \mathbb{R}^d$ eine lokale Extremstelle von p. Dann ist $D^1 p(x) = 0$.*

Beweis: Angenommen $D^1 p(x) \neq 0$. Dann wäre $D^1 p(x)$ indefinit (vgl. Bemerkung 2.12). Also wäre nach (iii) aus Satz 2.13 der Punkt x keine lokale Extremstelle von p. Widerspruch. □

In Satz 2.13 können die Kriterien (i) und (ii) nur für gerades $k+1$ zum Zug kommen, und für ungerades $k+1$ mit $D^{k+1} p(x) \neq 0$ greift immer Kriterium (iii) (vgl. Bemerkung 2.12). Nicht isolierte lokale Extremstellen können mit den Kriterien aus dem Satz nicht entdeckt werden. Im Spezialfall $d = 1$ greift der Satz jedoch immer:

Satz 2.15 (lokale Extrema im Univariaten). *Sei $f \in \mathbb{R}[X]$ vom Grad ≥ 1, $x \in \mathbb{R}$. Bezeichne $k \in \mathbb{N}$ die kleinste Zahl mit $f^{(k+1)}(x) \neq 0$ (ein solches k existiert). Dann gilt:*

(i) Ist $k+1$ gerade und $f^{(k+1)}(x) > 0$, so ist x eine isolierte lokale Minimalstelle von f.

(i') Ist x eine lokale Minimalstelle von f, so ist $k+1$ gerade und $f^{(k+1)}(x) > 0$.

(ii) Ist $k+1$ gerade und $f^{(k+1)}(x) < 0$, so ist x eine isolierte lokale Maximalstelle von f.

(ii') Ist x eine lokale Maximalstelle von f, so ist $k+1$ gerade und $f^{(k+1)}(x) < 0$.

(iii) Ist $k+1$ ungerade, so ist x keine lokale Extremstelle von f.

(iii') Ist x keine lokale Extremstelle von f, so ist $k+1$ ungerade.

Beweis: Die Zahl k existiert nach Lemma 2.5 wegen $\deg f \geq 1$. Die Aussagen (i), (ii) und (iii) folgen direkt aus Satz 2.13 unter Beachtung von $D^e f(x) = f^{(e)}(x) \cdot X^e$ für alle $e \in \mathbb{N}$. Wir zeigen (i'): Sei x eine lokale Minimalstelle von f. Nach (ii) ist $k+1$ ungerade oder $f^{(k+1)}(x) > 0$. Nach (iii) ist $k+1$ gerade. Also ist $f^{(k+1)}(x) > 0$. Mit Hilfe von (i), (ii) und (iii) zeigt man analog (ii') und (iii'). □

Satz 2.16 (Satz von Rolle). *Sei $f \in \mathbb{R}[X]$ und $a, b \in \mathbb{R}$ mit $a < b$ und $f(a) = f(b) = 0$. Dann gibt es ein $x \in]a, b[$ mit $f'(x) = 0$.*

Beweis: O.B.d.A. gelte $f \neq 0$ und a und b seien benachbarte Nullstellen von f. Dann können wir f schreiben als $f = (X-a)^i (X-b)^j g$ mit $1 \leq i, j \in \mathbb{N}$ und $g \in \mathbb{R}[X]$, sodaß g keine

Nullstelle auf $[a,b]$ hat. Nach Lemma 2.3 und der Kettenregel 2.4 gilt dann

$$\begin{aligned} f' &= i(X-a)^{i-1}(X-b)^j g + j(X-a)^i(X-b)^{j-1}g + (X-a)^i(X-b)^j g' \\ &= (X-a)^{i-1}(X-b)^{j-1}\underbrace{(i(X-b)g + j(X-a)g + (X-a)(X-b)g')}_{h}. \end{aligned}$$

Es gilt $h(a) = i(a-b)g(a)$ und $h(b) = j(b-a)g(b)$. Da nach dem Zwischenwertsatz $g > 0$ auf $[a,b]$ oder $g < 0$ auf $[a,b]$ gilt, folgt $h(a) < 0 < h(b)$ oder $h(b) < 0 < h(a)$. Nach dem Zwischenwertsatz gibt es ein $x \in]a,b[$ mit $h(x) = 0$, also auch $f'(x) = 0$. □

Um Differentiale berechnen zu können, ist noch eine weitere Verallgemeinerung der Produktregel für Ableitungen hilfreich. Die bereits auf mehrere Faktoren verallgemeinerte Produktregel wird jetzt auf höhere Ableitungen nach mehreren Variablen verallgemeinert:

Lemma 2.17. *Seien $p_1, p_2, p_3 \in \mathbb{R}[X,T]$ und $e_X, e_T \in \mathbb{N}$. Dann gilt*

$$\frac{\partial^{e_X+e_T} p_1p_2p_3}{\partial X^{e_X}\partial T^{e_T}} = \sum_{\substack{e_{X1}+e_{X2}+e_{X3}=e_X \\ e_{T1}+e_{T2}+e_{T3}=e_T}} \binom{e_X}{e_{X1}e_{X2}e_{X3}}\binom{e_T}{e_{T1}e_{T2}e_{T3}} \frac{\partial^{e_{X1}+e_{T1}}p_1}{\partial X^{e_{X1}}\partial T^{e_{T1}}} \frac{\partial^{e_{X2}+e_{T2}}p_2}{\partial X^{e_{X2}}\partial T^{e_{T2}}} \frac{\partial^{e_{X3}+e_{T3}}p_3}{\partial X^{e_{X3}}\partial T^{e_{T3}}}.$$

Beweis: Durch Induktion nach $e_X + e_T$ zeigt man mit Hilfe von Lemma 2.3

$$\frac{\partial^{e_X+e_T} p_1p_2p_3}{\partial X^{e_X}\partial T^{e_T}} = \sum_{u\in\{1,2,3\}^{e_X}} \sum_{v\in\{1,2,3\}^{e_T}} \prod_{i\in\{1,2,3\}} \frac{\partial^{|\{j|u_j=i\}|+|\{j|v_j=i\}|}p_i}{\partial X^{|\{j|u_j=i\}|}\partial T^{|\{j|v_j=i\}|}}.$$

Für feste $e_{X1}, e_{X2}, e_{X3} \in \mathbb{N}$ mit $e_{X1} + e_{X2} + e_{X3} = e_X$ gibt es $\binom{e_X}{e_{X1}e_{X2}e_{X3}}$ Elemente u von $\{1,2,3\}^{e_X}$ mit $\{j \mid u_j = i\} = e_{Xi}$ für alle $i \in \{1,2,3\}$. Analoges gilt für T statt X, woraus die Behauptung folgt. □

Als nächstes Hilfsmittel brauchen wir die quadratfreie Zerlegung univariater Polynome.

Definition 2.18. *Sei K ein Körper und $g \in K[X]$. Man nennt g* quadratfrei *in $K[X]$, wenn es kein Primelement p von $K[X]$ gibt, sodaß p^2 ein Teiler von g ist*

Die quadratfreien Polynome aus $K[X]$ sind also genau diejenigen Polynome, in deren Primfaktorzerlegung jeder Primfaktor nur mit dem Exponenten 1 vorkommt. Anders ausgedrückt sind es genau die Polynome aus $K[X]$, die kein Quadrat eines Polynoms aus $K[X]$ vom Grad ≥ 1 als Teiler besitzen.

Definition 2.19 (quadratfreie Zerlegung). *Sei K ein Körper und $0 \neq f \in K[X]$. Eine Darstellung von f in der Form $f = ag_1^1g_2^2\cdots g_n^n$ mit $a \in K$ und normierten, quadratfreien und paarweise teilerfremden $g_i \in K[X]$ nennen wir* quadratfreie Zerlegung *von f in $K[X]$.*

Aus der Existenz und Eindeutigkeit der Primfaktorzerlegung in $K[X]$ folgert man sofort die Existenz und Eindeutigkeit der quadratfreien Zerlegung in $K[X]$:

Lemma 2.20. *Sei K ein Körper und $0 \neq f \in K[X]$.*

(i) Es gibt eine quadratfreie Zerlegung von f in $K[X]$.

(ii) Ist $f = ag_1^1g_2^2\cdots g_n^n$ eine solche, so ist g_i für jedes $i \in \{1,\ldots,n\}$ das Produkt aller normierten Primfaktoren p von f in $K[X]$ mit $\max\{e \in \mathbb{N} \mid p^e \text{ teilt } f\} = i$.

Im Gegensatz zur Primfaktorzerlegung ist die gröbere quadratfreie Zerlegung in $K[X]$ für Körper K der Charakteristik 0 sehr einfach zu berechnen (siehe [BW] für Körper anderer Charakteristik), wenn man annimmt, daß man im Körper K rechnen kann. Den Schlüssel dazu liefert das folgende Lemma.

Lemma 2.21. *Sei K ein Körper der Charakteristik 0. Sei $0 \neq f \in K[X]$ und p ein Primfaktor von f. Dann gilt:*

$$\max\{e \in \mathbb{N} \mid p^e \text{ teilt } f\} - 1 = \max\{e \in \mathbb{N} \mid p^e \text{ teilt } f'\}$$

Beweis: Schreibe $f = p^e g$ mit einem $g \in K[X]$, welches von p nicht geteilt wird. Dann gilt

$$f' = p^e g' + ep^{e-1}p'g = p^{e-1}(pg' + ep'g).$$

Zu zeigen ist, daß p kein Teiler von $pg' + ep'g$ ist. Angenommen p teilt $pg' + ep'g$. Dann teilt p auch $ep'g$. Da K Charakteristik 0 hat, ist $e \neq 0$ in K. Deshalb teilt p auch $p'g$. Weil p ein Primelement ist, aber kein Teiler von g, muß p ein Teiler von p' sein. Aus Gradgründen muß dann p' das Nullpolynom sein. Dies ist aber wiederum unmöglich, da K die Charakteristik 0 hat. □

Sei K ein Körper der Charakteristik 0. Wie berechnet man nun die quadratfreie Zerlegung in $K[X]$? Seien $g_1, \ldots, g_n$ normierte, quadratfreie und paarweise teilerfremde Polynome in $K[X]$. Es sei $g_1^1 g_2^2 \cdots g_n^n$ bekannt. Wir zeigen, wie man $g_1, \ldots, g_n$ daraus berechnen kann. Nach obigem Lemma und (ii) aus Lemma 2.20 ist der größte gemeinsame Teiler (den man etwa mit euklidischem Algorithmus leicht berechnen kann) von $g_1^1 g_2^2 \cdots g_n^n$ und $(g_1^1 g_2^2 \cdots g_n^n)'$ gleich $g_2^1 g_3^2 \cdots g_n^{n-1}$. Dividiert man $g_1^1 g_2^2 \cdots g_n^n$ durch $g_2^1 g_3^2 \cdots g_n^{n-1}$, so erhält man $g_1 g_2 \cdots g_n$. Indem man nun dasselbe ausgehend von $g_2^1 g_3^2 \cdots g_n^{n-1}$ wiederholt, erhält man $g_2 \cdots g_n$. Man macht induktiv so weiter und erhält also eine Liste $g_1 \cdots g_n,\ g_2 \cdots g_n,\ \ldots,\ g_{n-1} g_n,\ g_n,\ 1$. Durch weitere Divisionen erhält man daraus $g_1, \ldots, g_n$.

Für diese Berechnung der quadratfreien Zerlegung ist es offenbar egal, ob der zugrundeliegende Körper K kleiner oder größer ist. Dies liefert:

Lemma 2.22. *Sei K ein Körper der Charakteristik 0 und L ein Oberkörper von K. Jedes $0 \neq f \in K[X]$ hat dann in $K[X]$ dieselbe quadratfreie Zerlegung wie in $L[X]$.*

Offenbar ist ein Polynom $0 \neq f \in K[X]$ genau dann quadratfrei, wenn seine quadratfreie Zerlegung die triviale Gestalt $f = ag_1$ annimmt. Mit dem letzten Lemma folgt daher sofort:

Lemma 2.23. *Sei K ein Körper der Charakteristik 0 und L ein Oberkörper von K. Jedes $f \in K[X]$ ist genau dann in $K[X]$ quadratfrei, wenn es in $L[X]$ quadratfrei ist.*

Als letztes Hilfsmittel brauchen wir noch eine Schranke für die Nullstellen eines univariaten Polynoms in Abhängigkeit seiner Koeffizienten, eine sogenannte Cauchy-Schranke.

Lemma 2.24 (Cauchy-Schranke). *Sei K ein angeordneter Körper. Seien $a_0, \ldots, a_n \in K$ mit $a_n \neq 0$. Sei $x \in K$ mit $\sum_{i=0}^n a_i x^i = 0$. Dann gilt:*

$$|x| \leq \max\left\{1, \frac{|a_0|}{|a_n|} + \cdots + \frac{|a_{n-1}|}{|a_n|}\right\}$$

Beweis: Sei $|x| \geq 1$ Dann gilt $|a_n x^n| = |\sum_{i=0}^{n-1} a_i x^i| \leq \sum_{i=0}^{n-1} |a_i| \cdot |x|^i \leq \sum_{i=0}^{n-1} |a_i| \cdot |x|^{n-1}$. Man teile nun auf beiden Seiten durch $|x|^{n-1}$. □

2.2 Beweis der Existenz der Darstellung

Ausgangspunkt ist folgender bekannter Beweis dafür, daß jedes $f \in \mathbb{R}[X]$ mit $f \geq 0$ auf $\mathbb{R}$ eine Summe von Quadraten in $\mathbb{R}[X]$ ist:

O.B.d.A. sei $f \neq 0$. Wir führen den Beweis durch Induktion nach dem Grad von f. Der Induktionsanfang $\deg f = 0$ ist trivial. Für den Induktionsschritt sei nun $\deg f \geq 1$ ($\deg f = 1$ ist freilich unmöglich).

Falls f nicht quadratfrei ist, so können wir $f = gh^2$ mit Polynomen $g, h \in \mathbb{R}[X]$ schreiben, sodaß $\deg g < \deg f$. Offenbar nimmt $g = \frac{f}{h^2}$ außerhalb der endlich vielen Nullstellen von h keine und damit überhaupt keine negativen Werte an. Da also nach Induktionsvoraussetzung g eine Summe von Quadraten ist, gilt dasselbe für f.

Den Fall, daß f quadratfrei ist, führen wir auf den soeben behandelten Fall zurück, indem wir $f = g + h$ schreiben mit einem Polynom $g \in \mathbb{R}[X]$, von dem wir schon wissen, daß es die gewünschte Darstellung besitzt, und einem Polynom $h \in \mathbb{R}[X]$, das denselben Grad wie f hat, ebenfalls nirgends einen negativen Wert annimmt, aber nicht mehr quadratfrei ist. Daß h nicht quadratfrei ist, wollen wir dadurch sicherstellen, daß h eine Nullstelle $t \in \mathbb{R}$ hat (denn dann ist t eine lokale Minimalstelle von h, also $h(t) = h'(t) = 0$ und nach Taylorentwicklung 2.5 ist $(X - t)^2$ ein Teiler von h). Das Gewünschte klappt bereits, indem wir für g ein konstantes Polynom wählen, nämlich $g = \min\{f(x) \mid x \in \mathbb{R}\}$ (h ist damit auch festgelegt).

Sei nun K ein Unterkörper von $\mathbb{R}$ und $f \in K[X]$ mit $f \geq 0$ auf $\mathbb{R}$. Kann man obigen Beweis so abändern, daß er zeigt: Es gibt eine Darstellung $f = \sum_i a_i g_i^2$ mit $0 \leq a_i \in K$ und $g_i \in K[X]$, also eine Darstellung von f als *gewichtete* Summe von Quadraten von Polynomen *mit Koeffizienten aus dem Körper K*?

Der Induktionsanfang bereitet dank der zugelassenen Gewichte a_i keine Probleme. Im Induktionsschritt läßt sich der erste behandelte Fall ohne Probleme verallgemeinern. Im zweiten Fall, daß f quadratfrei ist, ist jedoch nicht mehr klar, wie man g wählen soll. Es stellt sich nämlich das Problem, daß $\min\{f(x) \mid x \in \mathbb{R}\} \notin K$ eintreten kann, daß also g eventuell kein Polynom aus $K[X]$ ist. Man kann g nicht als ein anderes konstantes Polynom wählen, wenn man erreichen möchte, daß h eine Nullstelle hat. Polynome vom Grad 1 kommen für g offenbar auch nicht in Frage. Das nächste, was wir versuchen können, ist für g ein Polynom vom Grad 2 zu wählen, also eine Parabel. Hier haben wir gleich das günstige Resultat, daß dann g schon die gewünschte Darstellung besitzt, falls $g \geq 0$ auf $\mathbb{R}$. Dies ist die altbekannte quadratische Ergänzung:

Lemma 2.25. *Sei K ein angeordneter Körper. Sei $g = aX^2 + bX + c \in K[X]$ mit $a, b, c \in K, a \neq 0$. Dann gilt $g = a(X + \frac{b}{2a})^2 + (c - \frac{b^2}{4a})$. Falls $g(x) \geq 0$ für alle $x \in K$, so gilt zusätzlich $a > 0$ und $c - \frac{b^2}{4a} \geq 0$.*

Beweis: Die behauptete Gleichung rechnet man sofort nach. Gelte $g(x) \geq 0$ für alle $x \in K$. Dann ist insbesondere $c - \frac{b^2}{4a} = g(-\frac{b}{2a}) \geq 0$. Angenommen $a < 0$. Dann folgt aus $g(x) \geq 0$ für alle $x \in K$, daß $(x + \frac{b}{2a})^2 \leq -\frac{1}{a}(c - \frac{b^2}{4a})$ für alle $x \in K$. Also gibt es ein $C \in K$, sodaß $x^2 \leq C$ für alle $x \in K$, insbesondere $4 \leq C$ und $C^2 \leq C$. Hieraus folgt $C \leq 1$ im Widerspruch zu $4 \leq C$. □

Wir sind für jedes f auf der Suche nach einem passenden g. Wir wollen noch einmal präzisieren, für welche f wir nach welchen g suchen. Eigentlich bräuchten wir nur $f \in K[X]$ betrachten, die quadratfrei sind und für die $f \geq 0$ auf $\mathbb{R}$ ist. Für diese f gilt insbesondere $f > 0$ auf $\mathbb{R}$, denn hätte f eine Nullstelle in $\mathbb{R}$, so wäre f nicht quadratfrei in $\mathbb{R}[X]$ und daher nach Lemma 2.23 auch nicht in $K[X]$. Da wir für die folgenden Betrachtungen nicht mehr brauchen, machen wir nur die zuletzt genannte Eigenschaft von f zur Voraussetzung:

Es sei $f \in K[X]$, sodaß $f(x) > 0$ für alle $x \in \mathbb{R}$. Wir suchen ein $g \in K[X]$ mit folgenden Eigenschaften:

(i) $\deg g \leq 2$

(ii) $0 \leq g(x)$ für alle $x \in \mathbb{R}$

(iii) $g(x) \leq f(x)$ für alle $x \in \mathbb{R}$

(iv) $f - g$ hat eine Nullstelle $t \in K$

Eigenschaft (i) haben wir uns dabei selbst auferlegt, um den Suchraum zu begrenzen und um die Forderung, daß g eine gewichtete Summe von Quadraten in $K[X]$ sein soll, durch die einfachere Forderung $0 \leq g$ auf $\mathbb{R}$ ausdrücken zu können, was wir in (ii) gemacht haben. Die Ungleichung (iii) drückt aus, daß $h = f - g$ keine negativen Werte annehmen darf. In (iv) haben wir statt $t \in \mathbb{R}$ sogar $t \in K$ gefordert, weil dies den Suchraum erheblich vereinfachen wird. Es ist noch zu beachten, daß $h = f - g$ eventuell nicht denselben Grad wie f hat, wenn $\deg f = \deg g = 2$ ist. Dies ist aber kein Problem, da wir den Fall $\deg f = 2$ im diskutierten Beweis nun gesondert mit Lemma 2.25 behandeln können.

Um mögliche Kandidaten für g zu ermitteln, setzen wir voraus, $g \in K[X]$ erfülle (i)-(iii) und (iv) mit $t \in K$. Wegen (i) und Taylorentwicklung 2.5 gilt $g = g(t) + g'(t)(X-t) + c(X-t)^2$ mit einem $c \in K$. Aus (iv) folgt $g(t) = f(t)$. Da nach (ii) $f - g \geq 0$ auf $\mathbb{R}$ ist, folgt zusammen mit (iv) $(f-g)'(t) = 0$, also $g'(t) = f'(t)$. Aus (ii) folgt, daß g und damit auch $g(X+t) = f(t) + f'(t)X + cX^2$ höchstens eine Nullstelle hat. Dann muß die Diskriminante $(f'(t))^2 - 4cf(t)$ des letzteren Polynoms ≤ 0 sein, also $c \geq \frac{(f'(t))^2}{4f(t)}$ (beachte $f(t) > 0$). Es gilt also: Jedes $g \in K[X]$, das (i)-(iii) und (iv) mit $t \in K$ erfüllt, ist von der Form $g = f_{t,c}$ mit $\frac{(f'(t))^2}{4f(t)} \leq c \in K$, wobei wir $f_{t,c} = f(t) + f'(t)(X-t) + c(X-t)^2$ setzen.

Als nächstes beobachten wir: Wenn es ein $c \in K$ mit $\frac{(f'(t))^2}{4f(t)} \leq c$ gibt, sodaß (i)-(iii) und (iv) für $t \in K$ erfüllt sind mit $g = f_{t,c}$, so sind (i)-(iii) und (iv) für t auch erfüllt mit $g = f_{t,c'}$ für $c' = \frac{(f'(t))^2}{4f(t)}$. Mit $g = f_{t,c'}$ sind dann nämlich sicher (i) und (iv) für t erfüllt. Da $f_{t,c'}$ entweder das konstante Polynom $f(t)$ ist oder genau eine Nullstelle hat, gilt auch (ii). Weil $f_{t,c'} \leq f_{t,c} \leq f$ auf $\mathbb{R}$ gilt, gilt schließlich (iii).

Wir schließen nun: Wenn es überhaupt ein $g \in K[X]$ gibt, welches (i)-(iv) erfüllt, dann gibt es ein $t \in K$, sodaß diese Bedingungen mit $f_t := f(t) + f'(t)(X-t) + \frac{(f'(t))^2}{4f(t)}(X-t)^2$ erfüllt sind. Wie oben gesehen, erfüllt jedes f_t die Bedingungen (i), (ii) und (iv). Kritisch ist nur (iii). Die neue Aufgabenstellung lautet also:

Es sei $f \in K[X]$, sodaß $f > 0$ auf $\mathbb{R}$. Wir suchen ein $t \in K$, sodaß für

$$f_t := f(t) + f'(t)(X-t) + \frac{(f'(t))^2}{4}(X-t)^2 \in K[X]$$

gilt $f_t \leq f$ auf $\mathbb{R}$.

Nach wie vor ist dabei das Problem, daß t aus K sein muß . Sonst könnten wir in jedem Fall t als eine globale Minimalstelle von f wählen, und es wäre f_t wieder unser konstantes Polynom $\min\{f(x) \mid x \in \mathbb{R}\}$. Die Idee ist jetzt, daß wir t in der Nähe einer geeigneten globalen Minimalstelle von f wählen. Das nächste Lemma (wo wir allgemeiner *lokale* Minimalstellen betrachten) zeigt, daß wir damit in einem auf gewisse Weise guten Sinn wenigstens lokal die Bedingung $f_t \leq f$ erfüllen können.

Lemma 2.26. *Sei $f \in \mathbb{R}[X]$, sodaß $f(x) > 0$ für alle $x \in \mathbb{R}$. Sei a eine lokale Minimalstelle von f. Für alle $t \in \mathbb{R}$ sei f_t definiert durch*

$$f_t = f(t) + f'(t)(X-t) + \frac{(f'(t))^2}{4f(t)}(X-t)^2 \in \mathbb{R}[X].$$

Dann gibt es eine Umgebung U von a in $\mathbb{R}$, sodaß für alle $t \in U$ und $x \in U$ gilt $f_t(x) \leq f(x)$.

Beweis: Wenn f den Grad 0 hat, ist $f_t = f(t)$ für alle $t \in \mathbb{R}$. Wir können dann $U = \mathbb{R}$ setzen. Habe nun f einen Grad ≥ 1 ($\deg f = 1$ ist freilich unmöglich). Nach Folgerung 2.14 und (i') aus Satz 2.15 gibt es dann eine gerade Zahl $n \geq 2$ mit

$$f^{(1)}(a) = \cdots = f^{(n-1)}(a) = 0 < f^{(n)}(a).$$

Wir behandeln zunächst den Fall $n = 2$. Hier gibt es ein $\varepsilon > 0$, sodaß

$$(\star) \quad f_t'' < f''(x) \quad \text{für alle} \quad t, x \in [a-\varepsilon, a+\varepsilon].$$

Das konstante Polynom $f_t'' = \frac{(f'(t))^2}{4f(t)} \in K$ strebt nämlich gegen 0, wenn t gegen a strebt. Daher gibt es ein $\varepsilon_1 > 0$ mit $f_t'' < \frac{f''(a)}{2}$ für alle $t \in [a-\varepsilon_1, a+\varepsilon_1]$. Außerdem gibt es wegen $f''(a) > 0$ ein $\varepsilon_2 > 0$ mit $\frac{f''(a)}{2} < f''(x)$ für alle $x \in [a-\varepsilon_2, a+\varepsilon_2]$. Man braucht nun nur $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$ zu setzen.

Sei $t \in [a-\varepsilon, a+\varepsilon]$ beliebig vorgegeben. Dann ist t die einzige Nullstelle von $f - f_t$ im Intervall $[a-\varepsilon, a+\varepsilon]$. Hätten wir nämlich noch eine weitere Nullstelle x im selben Intervall, so gäbe es nach dem Satz von Rolle 2.16 ein x' echt zwischen t und x mit $(f-f_t)'(x') = 0$. Da aber auch $(f-f_t)'(t) = 0$, gäbe es dann wieder nach dem Satz von Rolle ein x'' echt zwischen t und x' mit $(f-f_t)''(x'') = 0$. Wegen $x'' \in [a-\varepsilon, a+\varepsilon]$ ist das ein Widerspruch zu $(\star)$.

Wir setzen nun $U = [a-\varepsilon, a+\varepsilon]$. Zu zeigen ist $f_t \leq f$ auf U. Da $(f-f_t)'(t) = 0$ und nach $(\star)$ $(f-f_t)'' > 0$ gilt, folgt nach Satz 2.15, daß t eine lokale Minimalstelle von $f - f_t$ ist. Wie gesehen ist es außerdem die einzige Nullstelle von $f - f_t$ auf U. Daraus folgt $f - f_t \geq 0$ auf U.

Nun behandeln wir den verbleibenden Fall $n \geq 4$. Hier betrachten wir das Problem als zweidimensional: Da die ersten beiden Summanden der Taylorentwicklung an der Stelle t von f und f_t für beliebiges $t \in \mathbb{R}$ übereinstimmen, erhalten wir nach Bildung der Differenz der beiden Taylorentwicklungen und Durchmultiplizieren mit der positiven Zahl $4f(t)$ für alle $(x,t) \in \mathbb{R}^2$ die Äquivalenz

$$f_t(x) \leq f(x) \iff \sum_{k=2}^{\infty} \frac{4}{k!} f(t) f^{(k)}(t)(x-t)^k - (f'(t))^2(x-t)^2 \geq 0.$$

Entsprechend dem dabei vorkommenden Ausdruck definieren wir das Polynom

$$p = \sum_{k=2}^{\infty} \underbrace{\frac{4}{k!} f(T) \frac{\partial^k f(T)}{\partial T^k} (X-T)^k}_{g_k \in \mathbb{R}[X,T]} - \underbrace{\left(\frac{\partial f(T)}{\partial T}\right)^2 (X-T)^2}_{h \in \mathbb{R}[X,T]} \in \mathbb{R}[X,T].$$

Zu zeigen ist, daß p in einer Umgebung von (a,a) in $\mathbb{R}^2$ keine negativen Werte annimmt. Wegen $p(a,a) = 0$ ist dies dazu äquivalent, daß $(a,a) \in \mathbb{R}^2$ eine lokale Minimalstelle von p ist. Um dies zu zeigen, reichen leider unsere Kriterien von 2.13 aus der Differentialrechnung nicht aus, da hier keine *isolierte* lokale Minimalstelle vorliegt. Es verschwindet nämlich p auf der ganzen Diagonale, d.h. $p(x,x) = 0$ für alle $x \in \mathbb{R}$. Deshalb werden wir stattdessen zeigen, daß $q := \frac{p}{(X-T)^2} \in \mathbb{R}[X,T]$ ein (isoliertes) lokales Minimum an der Stelle (a,a) hat. Da dann sowohl q als auch $(X-T)^2$ an der Stelle (a,a) ein lokales Minimum haben und

dort verschwinden ($q(a,a) = 2f(a)f''(a) - (f'(a))^2 = 0$), folgt nämlich daraus, daß auch $p = q(X-T)^2$ an der Stelle (a,a) ein lokales Minimum hat. Wir werden nun zeigen:

$$D^1q(a,a) = \cdots = D^{n-3}q(a,a) = 0 \quad \text{und} \quad D^{n-2}q(a,a) \quad \text{ist positiv definit}$$

Nach (i) aus Satz 2.13 sind wir dann fertig. Dies ist die Stelle, wo unsere Voraussetzung $n \geq 4$ eingeht. Für $n = 2$ wäre Satz 2.13 nicht anwendbar. Nun besteht folgender Zusammenhang zwischen den Differentialen von p und von q: $(X-T)^2 \frac{1}{k!} D^k q(a,a) = \frac{1}{(k+2)!} D^{k+2} p(a,a)$ für alle $k \in \mathbb{N}$. Es gilt nämlich

$$\begin{aligned} \sum_{k\in\mathbb{N}} \frac{1}{k!} D^k p(a,a) &= p(X+a, T+a) = ((X+a)-(T+a))^2 q(X+a, T+a) \\ &= (X-T)^2 \sum_{k\in\mathbb{N}} \frac{1}{k!} D^k q(a,a) = \sum_{k\in\mathbb{N}} \frac{1}{k!} (X-T)^2 D^k q(a,a) \end{aligned}$$

und $\frac{1}{k!}(X-T)^2 D^k q(a,a)$ ist eine $(k+2)$-Form für alle $k \in \mathbb{N}$. Statt die zur Diskussion stehenden Differentiale von q an der Stelle (a,a) auszurechnen, werden wir nun doch die entsprechenden Differentiale von p an der Stelle (a,a) ausrechnen, da dies erstaunlicherweise viel einfacher ist. Zu zeigen ist dann:

(i) $D^3 p(a,a) = \cdots = D^{n-1} p(a,a) = 0$

(ii) $\frac{D^n p(a,a)}{(X-T)^2}$ ist positiv definit.

Um die betreffenden Differentiale von p an der Stelle (a,a) zu untersuchen, berechnen wir alle höheren partiellen Ableitungen $\frac{\partial^{e_X+e_T} h}{\partial X^{e_X} \partial T^{e_T}}(a,a)$, $\frac{\partial^{e_X+e_T} g_k}{\partial X^{e_X} \partial T^{e_T}}(a,a)$ für $e_X, e_T \in \mathbb{N}$, $e_X + e_T \leq n$ und $2 \leq k \in \mathbb{N}$. Dazu verwenden wir jeweils Lemma 2.17. Wir werden das abzuleitende Polynom jeweils geschickt in drei Faktoren p_1, p_2 und p_3 aufspalten, sodaß in der von diesem Lemma gelieferten Summe über die $e_{X1}, e_{X2}, e_{X3}, e_{T1}, e_{T2}, e_{T3}$ jeweils alle oder fast alle Summanden wegfallen, wenn man sie an der Stelle (a,a) auswertet. Seien also nun $e_{X1}, e_{X2}, e_{X3}, e_{T1}, e_{T2}, e_{T3}, e_X, e_T, k \in \mathbb{N}$ mit

(1) $e_{X1} + e_{X2} + e_{X3} = e_X$

(2) $e_{T1} + e_{T2} + e_{T3} = e_T$

(3) $e_X + e_T \leq n$

(4) $2 \leq k$

Wir beginnen mit h. Wir spalten h auf in $p_1 := \frac{\partial f(T)}{\partial T}$, $p_2 := \frac{\partial f(T)}{\partial T}$ und $p_3 := (X-T)^2$. Soll der zu $e_{X1}, e_{X2}, e_{X3}, e_{T1}, e_{T2}, e_{T3}$ gehörige Summand in Lemma 2.17 ausgewertet an der Stelle (a,a) nicht verschwinden, so muß offenbar gelten: $e_{X1} = e_{X2} = 0$, $e_{T1} \geq n-1$, $e_{T2} \geq n-1$, $e_{X3} + e_{T3} \geq 2$. Daraus folgt wegen (1) und (2), daß $e_X + e_T \geq 2(n-1) + 2 = 2n > n$ ist. Dies ist mit (3) unvereinbar. Also verschwinden alle höheren partiellen Ableitungen von h, für die wir uns interessieren.

Nun leiten wir g_k ab. Wir spalten g_k auf in $p_1 := \frac{4}{k!} f(T)$, $p_2 := \frac{\partial^k f(T)}{\partial T^k}$ und $p_3 := (X-T)^k$. Soll der zu $e_{X1}, e_{X2}, e_{X3}, e_{T1}, e_{T2}, e_{T3}$ gehörige Summand in Lemma 2.17 ausgewertet an der Stelle (a,a) jetzt nicht verschwinden, so muß gelten: $e_{X1} = e_{X2} = 0$, $e_{X3} + e_{T3} \geq k$. Mit (1), (2) und (3) folgt $k \leq n$. Dann muß aber $e_{T2} \geq n-k$ sein, damit der Summand nicht verschwindet. Die beiden bestehenden Abschätzungen für e_{T2} und $e_{X3} + e_{T3}$ liefern zusammengenommen bereits, daß die Summe der beiden Ausdrücke bereits $\geq n$ sein muß. Nach (1), (2) und (3) müssen diese beiden Abschätzungen dann bereits scharf gewesen sein, und es muß gelten: $e_{X1} = 0$, $e_{X_2} = 0$, $e_{X3} = e_X$, $e_{T1} = 0$, $e_{T2} = n-k$, $e_{T3} = k - e_X$. Dies ist überhaupt nur möglich, falls

$$(\diamond) \quad e_X + e_T = n \quad \text{und} \quad e_X \leq k \leq n.$$

Gilt ($\diamond$) nicht, so verschwinden alle Summanden. Gilt ($\diamond$), so lautet der einzige, der möglicherweise nicht verschwindet:

$$\begin{pmatrix} & e_X & \\ 0 & 0 & e_X \end{pmatrix}\begin{pmatrix} (n-k)+(k-e_X) \\ 0 \quad n-k \quad k-e_X \end{pmatrix}\left(\frac{4}{k!}f(T)\frac{\partial^n f(T)}{\partial T^n}\cdot\frac{\underbrace{\partial^k (X-T)^k}_{\partial X^{e_X}\partial T^{k-e_X}}}{k!(-1)^{k-e_X}}\right)(a,a).$$

Er vereinfacht sich zu

$$\binom{n-e_X}{n-k}4f(a)f^{(n)}(a)(-1)^{k-e_X}.$$

Insgesamt erhalten wir nun: (i) ist schon gezeigt. Um (ii) zu zeigen, rechnen wir:

$$\begin{aligned} D^n p(a,a) &= \sum_{e_X+e_T=n}\begin{pmatrix} n \\ e_X \quad e_T \end{pmatrix}\left(\sum_{k=\max\{e_X,2\}}^{n}\frac{\partial^{e_X+e_T}g_k}{\partial X^{e_X}\partial T^{e_T}}(a,a)\right)X^{e_X}T^{e_T} \\ &= \sum_{e_X+e_T=n}\begin{pmatrix} n \\ e_X \quad e_T \end{pmatrix}\left(\sum_{k=\max\{e_X,2\}}^{n}\binom{n-e_X}{n-k}4f(a)f^{(n)}(a)(-1)^{k-e_X}\right)X^{e_X}T^{e_T} \\ &= 4f(a)f^{(n)}(a)\sum_{i=0}^{n}\binom{n}{i}\left(\sum_{k=\max\{i,2\}}^{n}\binom{n-i}{n-k}(-1)^{k-i}\right)X^i T^{n-i} \end{aligned}$$

Die innere Summe können wir mit binomischem Lehrsatz berechnen, wenn wir den Summationsbereich geringfügig ändern:

$$\begin{aligned} \sum_{k=i}^{n}\binom{n-i}{n-k}(-1)^{k-i} &= \sum_{k=i}^{n}\binom{n-i}{(n-i)-(n-k)}(-1)^{k-i} = \sum_{k=i}^{n}\binom{n-i}{k-i}(-1)^{k-i} \\ &= \sum_{k=0}^{n-i}\binom{n-i}{k}(-1)^k = (1-1)^{n-i} = 0^{n-i} = \begin{cases} 0 \text{ falls } i<n \\ 1 \text{ falls } i=n \end{cases} \end{aligned}$$

Wir können also weiterrechnen:

$$\begin{aligned} D^n p(a,a) &= 4f(a)f^{(n)}(a)\left(-\sum_{k=0}^{1}\binom{n-0}{n-k}(-1)^{k-0}T^n\right. \\ &\qquad \left.-\binom{n}{1}\sum_{k=1}^{1}\binom{n-1}{n-k}(-1)^{k-1}XT^{n-1}+X^n\right) \\ &= 4f(a)f^{(n)}(a)((-1+n)T^n - nXT^{n-1}+X^n) \\ &= 4f(a)f^{(n)}(a)(X^n+(n-1)T^n - nXT^{n-1}) \end{aligned}$$

Wir behaupten nun:

$$(\square)\quad \frac{D^n p(a,a)}{(X-T)^2} = 4f(a)f^{(n)}(a)\sum_{i=0}^{n-2}(i+1)T^i X^{n-2-i}$$

Dazu rechnen wir nach:

$$\begin{aligned}
&\left(\sum_{i=0}^{n-2}(i+1)T^iX^{n-2-i}\right)(X^2-2XT+T^2)\\
&=\sum_{i=0}^{n-2}(i+1)T^iX^{n-i}-2\sum_{i=0}^{n-2}(i+1)T^{i+1}X^{n-1-i}+\sum_{i=0}^{n-2}(i+1)T^{i+2}X^{n-2-i}\\
&=\sum_{i=0}^{n-2}(i+1)T^iX^{n-i}-2\sum_{i=1}^{n-1}iT^iX^{n-i}+\sum_{i=2}^{n}(i-1)T^iX^{n-i}
\end{aligned}$$

Die zu $i \in \{2,\dots,n-2\}$ gehörigen Summanden der drei Summen löschen sich gegenseitig aus. Es bleiben die zu $i \in \{0,1,n-1,n\}$ gehörigen Summanden:

$$\begin{aligned}
&X^n+2TX^{n-1}-2TX^{n-1}-2(n-1)T^{n-1}X+(n-2)T^{n-1}X+(n-1)T^n\\
&=X^n-nT^{n-1}X+(n-1)T^n
\end{aligned}$$

Damit ist (□) gezeigt. Es bleibt zu zeigen, daß $4f(a)f^{(n)}(a)\sum_{i=0}^{n-2}(i+1)T^iX^{n-2-i}$ positiv definit ist. Da sowohl $f(a)$ als auch $f^{(n)}(a)$ positiv ist, reicht es zu zeigen, daß

$$\sum_{i=0}^{n-2}(i+1)T^iX^{n-2-i}$$

positiv definit ist. Seien hierzu $t,x \in \mathbb{R}$, nicht $t=x=0$. Falls $t=x$, so ist

$$\sum_{i=0}^{n-2}(i+1)t^ix^{n-2-i}=\sum_{i=0}^{n-2}(i+1)x^{n-2}=x^{n-2}\sum_{i=0}^{n-2}(i+1)\neq 0.$$

Sei also nun $t \neq x$. Angenommen $\sum_{i=0}^{n-2}(i+1)t^ix^{n-2-i}=0$. Durch Multiplikation mit $4f(a)f^{(n)}(a)(x-t)^2$ erhält man gemäß (□), daß $x^n-nt^{n-1}x+(n-1)t^n=0$. Wäre $t=0$, so wäre $x^n=0$ und somit $x=0=t\neq x$. Also $t\neq 0$, und es folgt $(\frac{x}{t})^n-n\frac{x}{t}+(n-1)=0$. Das Polynom $u:=X^n-nX+(n-1)\in\mathbb{R}[X]$ hat also eine Nullstelle. Da n gerade ist, besitzt dieses Polynom eine globale Minimalstelle ξ. Für jede globale Minimalstelle ξ von u gilt $u'(\xi)=0$, also $n\xi^{n-1}-n=0$, d.h. $\xi^{n-1}=1$. Da n gerade ist, folgt $\xi=1$. Es ist also 1 die einzige globale Minimalstelle von u, und es gilt $u(1)=0$, d.h. 1 ist die einzige Nullstelle von u. Also folgt $\frac{x}{t}=1$ und daher $x=t$. Wir behandeln aber gerade den Fall $x\neq t$. Widerspruch! □

Nach diesem mühsamen Beweis können wir nun darangehen, ein $t \in K$ zu finden, für das die Bedingung $f_t \leq f$ sogar auf ganz $\mathbb{R}$ erfüllt ist. Wie vor Lemma 2.26 bereits erwähnt, ist es naheliegend, ein geeignetes t in der Nähe einer globalen Minimalstelle von f zu suchen. Offenbar hat es aber keine Aussicht auf Erfolg, t zwischen zwei globalen Minimalstellen von f zu wählen, wenn t nicht selber eine solche ist. Denn es gilt $f_t(t)=f(t)$ und f_t kann nicht von t ausgehend nach beiden Seiten fallen. Es ist also klar, wo wir unser Glück versuchen können: etwas unterhalb des kleinsten globalen Minimalpunkts von f oder etwas überhalb des größten globalen Minimalpunkts von f. Wir formulieren den Satz nur in der ersten Variante. Es ist klar, daß dann auch die symmetrische Aussage gilt.

Satz 2.27. *Sei $f \in \mathbb{R}[X]$ vom Grad >0 mit $f(x)>0$ für alle $x\in\mathbb{R}$. Dann gibt es eine kleinste globale Minimalstelle a von f und ein $0<\varepsilon\in\mathbb{R}$, sodaß für alle $t\in\mathbb{R}$ mit $a-\varepsilon<t<a$*

$$f_t:=f(t)+f'(t)(X-t)+\frac{(f'(t))^2}{4f(t)}(X-t)^2\in\mathbb{R}[X]$$

ein Polynom vom Grad 2 ist mit $f_t\leq f$ auf $\mathbb{R}$.

Beweis: Die Existenz von a ist klar. Der Fall $\deg f = 1$ kann nicht eintreten. Im Fall $\deg f = 2$ gilt sogar $f_t \leq f$ auf $\mathbb{R}$ für *alle* $t \in \mathbb{R}$: Nach Taylorentwicklung 2.5 von f an der Stelle t gilt nämlich $f = f(t) + f'(t)(X-t) + \frac{f''(t)}{2}(X-t)^2$. Mit f hat dann auch $f(X+t) = f(t) + f'(t)X + \frac{f''(t)}{2}X^2$ keine Nullstelle, also negative Diskriminante, d.h. $(f'(t))^2 - 4f(t)\frac{f''(t)}{2} < 0$. Es folgt $\frac{(f'(t))^2}{4f(t)} < \frac{f''(t)}{2}$ und daher $f_t \leq f$ auf $\mathbb{R}$ (vergleiche die Definition von f_t mit der Taylorentwicklung von f an der Stelle t).

Für den Rest des Beweises sei $\deg f > 2$. Wir wählen U wie im Lemma 2.26, o.B.d.A. $U = [a-\varepsilon_0, a+\varepsilon_0]$ mit einem $\varepsilon_0 > 0$, sodaß f' keine Nullstelle auf $[a-\varepsilon_0, a[$ hat. Es gilt also $f_t \leq f$ auf U für alle $t \in U$. Da die Koeffizienten von f_t stetige Funktionen von t sind, ist auch die Funktion

$$U \to \mathbb{R} : t \mapsto C_t := \max\left\{1, \frac{|a_{0t}|}{|a_{nt}|} + \cdots + \frac{|a_{(n-1)t}|}{|a_{nt}|}\right\}$$

mit $f - f_t = \sum_{i=0}^{n} a_{it}X^i$, $a_{it} \in \mathbb{R}$, $n = \deg f > 2$ stetig (hierbei sind sogar $a_{3t}, \ldots, a_{nt}$ von t unabhängig). Nach Lemma 2.24 liegen für alle $t \in U$ alle Nullstellen von $f - f_t$ in $[-C_t, C_t]$. Mit U ist auch das Bild der stetigen Funktion $U \to \mathbb{R} : t \mapsto C_t$ kompakt, insbesondere beschränkt in $\mathbb{R}$. Wähle $C \in \mathbb{R}$ mit $C \geq C_t$ für alle $t \in U$. Dann liegen für alle $t \in U$ alle Nullstellen von $f - f_t$ im Intervall $[-C, C]$. Bei Vergrößerung von C bleibt diese Eigenschaft natürlich erhalten, weswegen wir o.B.d.A. $-C < a-\varepsilon_0 < a < a+\varepsilon_0 < C$ annehmen können. Betrachte nun $M := \min\{f(x) \mid x \in [-C, a-\varepsilon_0]\}$. Es gilt $f(a) < M$ nach Festlegung von a. Für jedes $t \in [a-\varepsilon_0, a[$ ist f_t ein Polynom vom Grad 2 mit genau der einen Nullstelle $N_t := \frac{-2f(t)}{f'(t)} + t$. Wenn $t \in [a-\varepsilon_0, a[$ gegen a strebt, so strebt $f'(t) < 0$ gegen 0 und $-2f(t)$ gegen $-2f(a) < 0$. Also strebt dann N_t gegen $+\infty$. Außerdem strebt dann $f_t(-C)$ gegen $f_a(-C) = f(a) < M$. Wir können also $\varepsilon \in]0, \varepsilon_0]$ wählen, sodaß $N_t \in [C, \infty[$ und $f_t(-C) < M$ für alle $t \in]a-\varepsilon, a[$. Schließlich zeigen wir für festes $t \in]a-\varepsilon, a[$ in fünf Schritten $f_t \leq f$ auf $\mathbb{R}$:

- Es gilt $f_t \leq f$ auf $]-\infty, -C]$: Es ist $f_t(-C) < M \leq f(-C)$ und auf $]-\infty, -C[$ hat $f - f_t$ keine Nullstelle.
- Es gilt $f_t \leq f$ auf $]-C, a-\varepsilon_0[$: f_t ist monoton fallend auf $]-\infty, N_t]$, also wegen $a-\varepsilon_0 < C \leq N_t$ auch auf $[-C, a-\varepsilon_0[$. Es folgt $f_t(x) \leq f_t(-C) < M \leq f(x)$ für alle $x \in]-C, a-\varepsilon_0[$.
- Es gilt $f_t \leq f$ auf $[a-\varepsilon_0, a[$: Dies folgt sofort aus $[a-\varepsilon_0, a[\subseteq U$ und der Wahl von U.
- Es gilt $f_t \leq f$ auf $[a, C[$: f_t ist monoton fallend auf $]-\infty, N_t]$, also wegen $C \leq N_t$ auch auf $[a, C[$. Da außerdem a ein globaler Minimalpunkt von f ist und $a \in U$ gilt, folgt $f_t(x) \leq f_t(a) \leq f(a) \leq f(x)$ für alle $x \in [a, C[$.
- Es gilt $f_t \leq f$ auf $[C, \infty[$: Es ist $f_t(N_t) = 0 < f(N_t)$ und $N_t \in [C, \infty[$. Da aber $f - f_t$ auf $]C, \infty[$ keine Nullstelle hat, folgt die Behauptung.

□

Wir können nun sofort den angepeilten Nichtnegativstellensatz beweisen:

Satz 2.28. *Sei K ein Unterkörper von $\mathbb{R}$ und $f \in K[X]$ vom Grad $n \geq 1$. Genau dann ist $f \geq 0$ auf $\mathbb{R}$, wenn f eine gewichtete Summe von n Quadraten in $K[X]$ ist, d.h. wenn es $a_1, \ldots, a_n \in K^{\geq 0}$ und $g_1, \ldots, g_n \in K[X]$ gibt mit $f = \sum_{i=1}^{n} a_i g_i^2$.*

Beweis: Eine Richtung ist trivial. Für die andere Richtung sei $f \geq 0$ auf $\mathbb{R}$. Offenbar muß dann n gerade sein. Wir führen den Beweis durch Induktion nach n. Den Induktionsanfang $n = 2$ erledigen wir mit Lemma 2.25. Für den Induktionsschritt sei nun $n \geq 4$.

Falls f nicht quadratfrei ist, so ist f eine gewichtete Summe von sogar nur $n-2$ Quadraten in $K[X]$. Wir können dann nämlich $f = gh^2$ mit Polynomen $g, h \in K[X]$ schreiben, sodaß $\deg g \leq (\deg f) - 2$. Offenbar gilt $g(x) = \frac{f(x)}{(h(x))^2} \geq 0$ für alle $x \in \mathbb{R}$ mit $h(x) \neq 0$. Da h nur endlich viele Nullstellen in $\mathbb{R}$ hat, folgt $g \geq 0$ auf $\mathbb{R}$. Nach Induktionsvoraussetzung ist also g eine gewichtete Summe von $n-2$ Quadraten in $K[X]$. Wegen $f = gh^2$ gilt dasselbe für f.

Sei nun f quadratfrei. Dann hat f keine Nullstelle auf $\mathbb{R}$, denn jede solche wäre eine mehrfache und würde implizieren, daß f nicht quadratfrei ist in $\mathbb{R}[X]$ und nach Lemma 2.23 auch nicht in $K[X]$. Also gilt $f > 0$ auf $\mathbb{R}$. Nach Satz 2.27 gibt es daher ein $t \in K$ (K liegt dicht in $\mathbb{R}$!) und eine Parabel $f_t \in K[X]$ mit $0 \leq f_t \leq f$ auf $\mathbb{R}$ und $f_t(t) = f(t)$. Dann ist $f - f_t$ ein Polynom vom Grad n, das nirgends einen negativen Wert annimmt und nicht quadratfrei ist, denn t ist mehrfache Nullstelle von $f - f_t$. Nach dem soeben behandelten Fall ist dann $f - f_t$ eine gewichtete Summe von $n-2$ Quadraten in $K[X]$. Da nach Lemma 2.25 f_t eine gewichtete Summe von 2 Quadraten in $K[X]$ ist, ist f eine solche von n Quadraten in $K[X]$. □

An dieser Stelle sollten wir erwähnen, daß für den Fall $K = \mathbb{Q}$ obiger Satz nicht besonders wertvoll erscheint. In [Pou] wird nämlich bewiesen, daß jedes $f \in \mathbb{Q}[X]$ mit $f \geq 0$ auf $\mathbb{R}$ eine (ungewichtete) Summe von fünf Quadraten in $\mathbb{Q}[X]$ ist. Für $K \neq \mathbb{Q}$ hat der Autor allerdings keine vergleichbaren Resultate gefunden.

2.3 Umsetzung in einen Algorithmus

Der Beweis von Satz 2.28 ist völlig algorithmisch. Bis auf das Rechnen im angeordneten Körper K sagt er uns, wie wir rekursiv die gewünschte Darstellung von f erhalten können. Es sind nur die Fragen zu klären, wie man im ersten Teil des Induktionsschritts für nicht quadratfreies f eine Zerlegung $f = gh^2$ berechnet und wie man im zweiten Teil des Beweises t berechnet.

Die erste Frage ist schnell geklärt: Man berechne, wie nach Lemma 2.21 diskutiert, die quadratfreie Zerlegung $f = ag_1^1 g_2^2 \cdots g_{2m+1}^{2m+1}$ von f (wir haben die Zerlegung mit einer ungeraden Anzahl von Faktoren hingeschrieben, was man natürlich immer machen kann). Da f nicht quadratfrei ist, gibt es ein Primelement p von $K[X]$, sodaß p^2 ein Teiler von f ist. Da die g_i paarweise teilerfremd sind, gibt es ein $i \in \{1, \ldots, m\}$, sodaß p^2 ein Teiler von g_i^i ist. Da g_1 quadratfrei ist, muß $i \geq 2$ sein. Es gibt also in dieser quadratfreien Zerlegung ein g_i mit $i \geq 2$ von positivem Grad. Man könnte nun $h = g_i$ setzen. Um den Algorithmus effizient zu machen und eine Darstellung mit wenig Summanden zu erhalten, scheint es eine sinnvolle Strategie, h von großem Grad zu wählen. Am besten setzt man also $h = g_2 g_3 g_4^2 g_5^2 g_6^3 g_7^3 \cdots g_{2m}^m g_{2m+1}^m$.

Zur zweiten Frage betrachten wir den Satz, der die Existenz von t gesichert hat, nämlich Satz 2.27. Er sagt, wir müßen $t \in K$ nur genügend nahe unterhalb der kleinsten globalen Minimalstelle $a \in \mathbb{R}$ von $f \in K[X]$ wählen. Nun ist a eine Nullstelle von $0 \neq f' \in K[X]$.

Man kennt eine Methode, die Anzahl der verschiedenen reellen Nullstellen von Polynomen $\neq 0$ aus $K[X]$ in einem Intervall mit Endpunkten aus K zu zählen. Man macht das mit Hilfe einer Sturmschen Kette dieses Polynoms (siehe dazu etwa die Lehrbücher [BW], [Mis] oder die Vorlesungsskripten [We1], [We2]). Mit Hilfe der Cauchy-Schranke 2.24 kann man nun zunächst ein Intervall mit Endpunkten aus K bestimmen, in dem sich alle Nullstellen des Polynoms befinden. Durch fortgesetztes Halbieren dieses Intervalls und jeweiliges Zählen der Nullstellen in dem neuen, kleineren Intervall kann man für jede der Nullstellen schließlich ein Intervall mit Endpunkten aus K finden, in dem nur noch eine Nullstelle liegt, ein *isolierendes* Intervall für diese Nullstelle. Durch weitere fortgesetzte Halbierung kann man diese isolierenden Intervalle beliebig verkleinern, d.h. die Differenz der oberen und unteren Intervallgrenze konvergiert gegen 0. Indem man die unteren Intervallgrenzen der isolierenden Intervalle (in irgendeiner Reihenfolge) aufzählt, die Intervalle halbiert, wieder

die unteren Intervallgrenzen aufzählt, und so fortfährt, erhält man eine Folge von Elementen von K, die für jede Nullstelle des Polynoms eine gegen diese Nullstelle konvergente Teilfolge enthält, deren Folgenglieder sich alle unterhalb der Nullstelle befinden.

Durch Anwendung dieser Methode auf $g = f'$ können wir eine Folge von Elementen von K aufzählen, die insbesondere eine Teilfolge enthält, die von unten gegen den kleinsten globalen Minimalpunkt von f konvergiert. Wir brauchen nur die Glieder dieser Folge nacheinander als mögliche Kandidaten für t ausprobieren. Irgendwann werden wir ein passendes t finden.

2.4 Implementierung

Für den Fall $K = \mathbb{Q}$ haben wir den Algorithmus im Computer-Algebra-System REDUCE (siehe [Hea] und [Mel]) implementiert. Die Datei `sos.red` (`sos` steht für „sum of squares“), die den Quelltext des Programms enthält, ist im Anhang ab Seite 85 aufgeführt. Der Quelltext ist lauffähig in Version 3.6 von REDUCE zusammen mit dem verbreiteten Zusatzpaket ROOTS (siehe [Kam]).

Das Zusatzpaket ROOTS wird dabei nur zur Berechnung Sturmscher Ketten verwendet. Eigentlich wäre es einfach, diese Sturmschen Ketten zu berechnen. Im wesentlichen ist dabei nur eine iterierte Polynomdivision mit Rest durchzuführen. Sind die Koeffizienten der vorkommenden Polynome dabei allerdings rationale Zahlen mit vom Betrag her großen Nennern und Zählern (solche Zahlen werden bei unserem Algorithmus ständig durch das t miteingeschleppt), so kommt es dabei eventuell zu Effizienzproblemen, je nachdem auf welchem LISP-System das REDUCE läuft. Nach einer persönlichen Mitteilung von Winfried Neun [Neu] liegt das vermutlich daran, daß in manchen LISP-Systemen die Berechnung des größten gemeinsamen Teilers für betragsmäßig große ganze Zahlen weniger effektiv implementiert ist. Offenbar umgeht das Zusatzpaket ROOTS diese Schwäche. Darauf deutet schon hin, daß dieses Paket ein eigenes Datenformat für univariate Polynome verwendet.

Der Quelltext in `sos.red` kann nach dem Starten von REDUCE durch Eingabe der Zeilen

```
faslout "sos"$
in "sos.red"$
faslend$
```

in ein Schnell-Lade-Modul `sos.b` übersetzt werden. Dieses kann bei späteren REDUCE-Aufrufen durch

```
load sos;
```

geladen werden. Es stehen dann zwei neue Funktionen `sos` und `sosprint` zur Verfügung. Angewandt auf Ausdrücke, die kein univariates Polynom mit rationalen Koeffizienten f mit $f \geq 0$ auf $\mathbb{R}$ darstellen, liefern beide Fehlermeldungen:

```
6: sos(x^7 + 1000);

       7
***** x  + 1000 invalid as polynomial having no negative values

7: sos(x + y);

***** x + y invalid as univariate polynomial

8: sos(x^x);

       x
***** x  invalid as polynomial
```

Dieselben Fehlermeldungen erscheinen bei jeweiliger Eingabe von `sosprint` statt `sos`. Werden die Funktionen jedoch mit einem Ausdruck aufgerufen, der ein univariates Polynom $f \in \mathbb{Q}[X]$ (für irgendeine Variable X) mit $f \geq 0$ auf $\mathbb{R}$ darstellt, so unterscheiden sie sich: Die Funktion `sosprint` druckt auf den Bildschirm dann eine Darstellung von f als gewichtete Summe (mit nichtnegativen rationalen Gewichten) von höchstens n Produkten von geradzahligen Potenzen von Polynomen aus $\mathbb{Q}[X]$ ($n = \deg f$) und gibt als Rückgabewert `nil` zurück:

```
13: r := sosprint(((x-1)^4 + x^2 - 2*x + 2) * (x + 1)^4);

       4   2       2            4        2
(x + 1) *(x  - 2*x)  + 3*(x + 1) *(x - 1)

14: r;

15:
```

Die ausgedruckte Darstellung wird nicht der Variable r zugewiesen, weil man sie auf diese Weise nicht zu Gesicht bekommen würde, denn der REDUCE-Evaluator würde sie umwandeln in die Normalform eines Polynoms. Man hat aber nun das Problem, daß man die durch `sosprint` gelieferte Darstellung nicht zur Weiterverarbeitung verwenden kann, da sie von `sosprint` nicht zurückgegeben wird. Zu diesem Zweck gibt es die zweite Prozedur `sos`. Sie liefert eine Liste zurück, deren erster Eintrag ein Polynom und deren restliche Einträge Gleichungen sind. Das Polynom ist ein Polynom mit rationalen Koeffizienten in neuen Unbestimmten (d.h. in der aktuellen REDUCE-Sitzung noch nicht vorgekommenen Variablen). Es ist eine Summe von n Monomen, von denen keines einen negativen Koeffizienten hat. Jedes Monom ist ein Produkt von geradzahligen Potenzen von Unbestimmten. Für jede der neuen Unbestimmten U enthält die zurückgegebene Liste genau eine Gleichung der Form $U = g$ mit einem Polynom $g \in \mathbb{Q}[X]$. Substituiert man die neuen Unbestimmten im Polynom durch die jeweilige rechte Seite der Gleichung für diese Unbestimmte, so erhält man wieder f.

```
3: r := sos(((x-1)^4 + x^2 - 2*x + 2) * (x + 1)^4);

          4       2          2
r := {g0005 *(g0007  + 3*g0008 ),

      g0005=x + 1,

            2
      g0007=x  - 2*x,

      g0008=x - 1}
```

Für $f = \frac{1}{16}X^6 + X^4 - \frac{1}{9}X^3 - \frac{11}{10}X^2 + \frac{2}{15}X + 2 \in \mathbb{Q}[X]$ liefert unsere Implementierung nach einer Rechenzeit von gut 9 Sekunden die Darstellung

$$\begin{aligned} f = {} & \frac{1}{16}\left(X + \frac{169}{192}\right)^2 X^2 \left(X - \frac{169}{192}\right)^2 \\ & + \frac{19493219788554285131886072535687}{331850852843102252537009591 5008}\left(X + \frac{169}{192}\right)^2 X^2 \\ & + \frac{116235169608485606412453930165625}{995552558529306757611028774 5024}\left(X + \frac{169}{192}\right)^2 \\ & \qquad \left(X - \frac{27006091539966003624431 1191}{193012660502929906718460000}\right)^2 \end{aligned}$$

$$+\frac{150310482233795647853003 29}{525114462807207287193600 0}\left(X-\frac{2494878809933707}{1116572999781024}\right)^2.$$

Alle Rechenzeiten sind gemessen auf einer SUN SPARC ULTRA 1 mit PSL als dem REDUCE zugrundeliegendem LISP. Für $f = X^8 - 3X^7 - 2X^6 - X^5 + X^4 - X + 2000$ erhalten wir nach einer Rechenzeit von 205 Sekunden eine Darstellung als gewichtete Summe von 5 Quadraten, in der ganze Zahlen mit einer Dezimaldarstellung von bis zu 237 Ziffern vorkommen. Dieses f hat etwa bei $3,14$ eine eine eindeutig bestimmte globale Minimalstelle und nimmt dort den Wert $293,9$ an.

Wir rücken nun diese Funktion näher an die Nullfunktion heran und setzen $f = X^8 - 3X^7 - 2X^6 - X^5 + X^4 - X + 1708,5$. An der unveränderten globalen Minimalstelle nimmt nun f etwa den Wert $1,4$ an. Der Algorithmus hat es nun schwieriger, ein passendes t zu finden. Er rechnet knappe 303 Sekunden und liefert eine Darstellung als gewichtete Summe von 5 Quadraten, in der ganze Zahlen mit einer Dezimaldarstellung von bis zu 319 Ziffern vorkommen.

Verkleinern wir den konstanten Koeffizienten von f, sodaß f nur noch knapp über der Nullfunktion liegt, setzen wir also $f = X^8 - 3X^7 - 2X^6 - X^5 + X^4 - X + 1707,11$ (der kleinste Wert, den f annimmt ist nun etwa $0,1$), so brauchen wir eine Rechenzeit von gut 505 Sekunden und die Darstellung enthält Zahlen mit bis zu 389 Ziffern.

2.5 Situation über reell abgeschlossenen Körpern

In diesem Abschnitt versuchen wir, die bisher erzielten Ergebnisse auf den Fall zu verallgemeinern, daß die Koeffizienten des darzustellenden Polynoms nun statt aus einem Unterkörper von $\mathbb{R}$ aus einem beliebigen angeordneten Körper K stammen.

Es stellt sich die Frage, wie man dann die geometrische Bedingung an ein $f \in K[X]$ formuliert, die äquivalent dazu sein soll, daß f eine gewichtete Summe von Quadraten in $K[X]$ ist. Bisher war diese Bedingung „$f \geq 0$ auf $\mathbb{R}$“. Da K nun kein Unterkörper von $\mathbb{R}$ mehr sein braucht, macht dies nicht länger Sinn.

Man stellt natürlich fest, daß K bisher dicht in $\mathbb{R}$ lag. Deswegen hätten wir bisher auch „$f \geq 0$ auf K“ statt „$f \geq 0$ auf $\mathbb{R}$“ schreiben können. Nun gibt es aber angeordnete Körper K und Polynome $f \in K[X]$ mit $f \geq 0$ auf K, die keine Summe von Quadraten in $K[X]$ sind aus dem einfachen Grund, weil es eine Erweiterung des angeordneten Körpers K gibt, auf der f negative Werte annimmt (siehe [Lor], Aufgabe 21.2).

Wir haben in den vorherigen Abschnitten nicht ohne Grund „$f \geq 0$ auf $\mathbb{R}$“ statt „$f \geq 0$ auf K“ geschrieben. Es wird sich nämlich herausstellen, daß die adäquate Bedingung nun „$f \geq 0$ auf Rlautet, wobei R der (bis auf K-Isomorphie) eindeutig bestimmte reelle Abschluß des angeordneten Körpers K ist (siehe die Lehrbücher [Lor], [Jac] oder das Skript [We2]).

Wenn wir nun versuchen, den Nichtnegativstellensatz 2.28 mit (K, R) statt $(K, \mathbb{R})$ zu beweisen, so kommen wir an mehreren Stellen in Probleme.

Zum Beispiel geht *der Beweis von* Satz 2.13 nicht ohne Probleme durch, denn dort bräuchte man, daß die Einheitssphäre $S := \{x \in R^d \mid x_1^2 + \cdots + x_d^2 = 1\}$ kompakt ist. Von einer „vernünftigen“ Topologie auf R^d würde man erwarten, daß in ihr die Mengen $C_\varepsilon(x) := \prod_{i=1}^d]x_1 - \varepsilon, x_1 + \varepsilon[^d$ mit $0 < \varepsilon \in R$ und $x \in R^d$ offen sind. Ist aber R nicht archimedisch, so gibt es ein $0 < \varepsilon \in R$ mit $\varepsilon < \frac{1}{n}$ für alle $n \in \mathbb{N}$. Dann ist $\{C_\varepsilon(x) \mid x \in S\}$ eine Überdeckung von S mit offenen Mengen ohne endliche Teilüberdeckung. Also ist dann S bezüglich keiner „vernünftigen“ Topologie auf R^d kompakt.

Es gibt allerdings das sogenannte Tarski-Prinzip, welches besagt, daß jede in Logik erster Stufe mit der Sprache der geordneten Ringe ausdrückbare Aussage, welche in $\mathbb{R}$ gilt, sogar in jedem beliebigen reell abgeschlossenen Körper R gilt (siehe etwa [Pr2], [Pr3], [We2], [rqe]). Nun scheint dies auf den Nichtnegativstellensatz 2.28 nicht direkt anwendbar zu sein, weil

dort noch der Unterkörper K in der Formulierung des Satzes auftaucht. Jedoch können wir Satz 2.27 auf diese Weise verallgemeinern, der die wichtigste Zutat des Beweises des Nichtnegativstellensatzes 2.28 bildet: Für jeden festen Grad von f ist dieser Satz in der Sprache der geordneten Ringe erster Stufe ausdrückbar. Also können wir ihn für jeden festen Grad von f und daher insgesamt verallgemeinern:

Satz 2.29. *Sei R ein reell abgeschlossener Körper. Sei $f \in R[X]$ vom Grad > 0 mit $f(x) > 0$ für alle $x \in R$. Dann gibt es eine kleinste globale Minimalstelle a von f und ein $0 < \varepsilon \in R$, sodaß für alle $t \in R$ mit $a - \varepsilon < t < a$*

$$f_t := f(t) + f'(t)(X-t) + \frac{(f'(t))^2}{4f(t)}(X-t)^2 \in R[X]$$

ein Polynom vom Grad 2 ist mit $f_t \leq f$ auf R.

Nun sieht man, daß der Beweis von 2.28 auch mit (K, R) statt $(K, \mathbb{R})$ durchgeht, *wenn K dicht in seinem reellen Abschluß R liegt*, d.h. wenn in jedem Intervall $]a, b[$ mit $a, b \in R$ und $a < b$ ein Element aus K liegt. Leider ist dies nicht immer der Fall (siehe [Lor], Aufgabe 21.2). Wir formulieren also die Verallgemeinerung des Nichtnegativstellensatzes 2.28 mit dieser zusätzlichen Voraussetzung:

Satz 2.30. *Es sei K ein angeordneter Körper, der dicht in seinem reellen Abschluß R liegt, und $f \in K[X]$ vom Grad $n \geq 1$. Genau dann ist $f \geq 0$ auf R, wenn f eine gewichtete Summe von n Quadraten in $K[X]$ ist, d.h. wenn es $a_1, \ldots, a_n \in K^{\geq 0}$ und $g_1, \ldots, g_n \in K[X]$ gibt mit $f = \sum_{i=1}^n a_i g_i^2$.*

Es stellt sich nun die Frage, wie die Situation ist, wenn der angeordnete Körper K nicht dicht in seinem reellen Abschluß R liegt. Die bekannte Lösung des 17. Hilbertschen Problems durch Artin bleibt auch in diesem Fall gültig. Sie besagt (siehe etwa [Lor] oder [We2]), daß jedes $f \in K[X_1, \ldots, X_d]$ mit $f \geq 0$ auf R^d eine Darstellung der Form $f = \sum_i a_i (\frac{f_i}{g_i})^2$ mit $0 \leq a_i \in K$ und $f_i, g_i \in K[X_1, \ldots, X_d]$ besitzt.

Hier scheint nur für den Fall $K = R$ etwas über die Anzahl der benötigten Summanden in der Darstellung bekannt zu sein (siehe [Pfi],[BCR]). Für diesen Fall interessieren wir uns jedoch nicht, da der eingangs des Kapitels erwähnte Beweis, daß jedes Polynom $f \in \mathbb{R}[X]$ mit $f \geq 0$ auf $\mathbb{R}$ eine Summe von zwei Quadraten in $\mathbb{R}[X]$ ist, auch mit R statt $\mathbb{R}$ durchgeht (man weiß , daß $C := R(\sqrt{-1})$ algebraisch abgeschlossen ist, wenn R reell abgeschlossen ist, siehe §20, Satz 7 in [Lor], Theorem 5.2 in [Jac] oder Seite 69 in [We2]).

Da wir uns hier aber nur für den Fall $d = 1$ interessieren, können wir wenigstens die Nenner h_i in der Darstellung beseitigen (bei gleichbleibender Anzahl von Summanden, über die jedoch nichts ausgesagt wird). Dies bewerkstelligen wir mit folgender Verallgemeinerung eines Satzes von Cassels [Cas] auf *gewichtete* Summen:

Satz 2.31. *Sei K ein angeordneter Körper. Für alle $f_1, \ldots, f_n, g_1, \ldots, g_n \in K[X]$, $0 \leq a_1, \ldots, a_n \in K$, für die $\sum_{i=1}^n a_i (\frac{f_i}{g_i})^2$ ein Element von $K[X]$ ist, gibt es $p_1, \ldots, p_n \in K[X]$ mit*

$$\sum_{i=1}^n a_i \left(\frac{f_i}{g_i}\right)^2 = \sum_{i=1}^n a_i p_i^2.$$

Beweis: O.B.d.A. ist $a_i > 0$ für alle $i \in \{1, \ldots, n\}$. Indem wir die Brüche $\frac{f_i}{g_i}$ auf einen gemeinsamen Hauptnenner bringen, können wir o.B.d.A. $g_i = g$ für alle $i \in \{1, \ldots, n\}$ schreiben. Daher reicht es folgendes zu zeigen:

Sei $h \in K[X]$ ein Polynom, für das es ein $g \in K[X]$ vom Grad ≥ 1 und $f_1, \ldots, f_n \in K[X]$ gibt mit $hg^2 = \sum_{i=1}^n a_i f_i^2$. Dann gibt es ein $G \neq 0$ aus $K[X]$ von kleinerem Grad als g und $F_1, \ldots, F_n \in K[X]$ mit $hG^2 = \sum_{i=1}^n a_i F_i^2$.

Wir beweisen dies: Nach Division mit Rest in $K[X]$ können wir schreiben

$$f_i = q_i g + r_i \qquad \text{mit} \qquad q_i, r_i \in K[X], \quad \deg r_i < \deg g \quad \text{oder} \quad r_i = 0$$

für $i \in \{1, \dots, n\}$.

Falls $r_i = 0$ für alle $i \in \{1, \dots, n\}$, so setzen wir $G = 1$ und $F_i = q_i$ für alle $i \in \{1, \dots, n\}$. Dann gilt

$$hG^2 = h = \frac{1}{g^2}(hg^2) = \frac{1}{g^2}\sum_{i=1}^{n} a_i f_i^2 = \sum_{i=1}^{n} a_i \left(\frac{f_i}{g}\right)^2 = \sum_{i=1}^{n} a_i q_i^2 = \sum_{i=1}^{n} a_i F_i^2.$$

Im Folgenden sei also die Menge $I := \{i \in \{1, \dots, n\} \mid r_i \neq 0\}$ nicht leer. Nun setzen wir

$$\sigma = \sum_{i=1}^{n} a_i q_i^2 - h, \qquad \tau = \sum_{i=1}^{n} a_i f_i q_i - gh,$$
$$F_i = \sigma f_i - 2\tau q_i \quad \text{für } i \in \{1, \dots, n\} \qquad \text{und} \qquad G = \sigma g - 2\tau.$$

Dann gilt

$$\begin{aligned} hG^2 &= \sigma^2 hg^2 - 4\sigma\tau gh + 4\tau^2 h \\ &= \sigma^2 \sum_{i=1}^{n} a_i f_i^2 - 4\sigma\tau(\tau + gh) + 4\tau^2(\sigma + h) \\ &= \sigma^2 \sum_{i=1}^{n} a_i f_i^2 - 4\sigma\tau \sum_{i=1}^{n} a_i f_i q_i + 4\tau^2 \sum_{i=1}^{n} a_i q_i^2 \\ &= \sum_{i=1}^{n} a_i (\sigma f_i - 2\tau q_i)^2 = \sum_{i=1}^{n} a_i F_i^2. \end{aligned}$$

Es ist noch $G \neq 0$ und $\deg G < \deg g$ zu zeigen. Hierzu rechnen wir nach:

$$\begin{aligned} G &= g \sum_{i=1}^{n} a_i q_i^2 - gh - 2\sum_{i=1}^{n} a_i f_i q_i + 2gh \\ &= \frac{1}{g}\left(g^2 \sum_{i=1}^{n} a_i q_i^2 + g^2 h - 2g \sum_{i=1}^{n} a_i f_i q_i\right) \\ &= \frac{1}{g}\left(g^2 \sum_{i=1}^{n} a_i q_i^2 + \sum_{i=1}^{n} a_i f_i^2 - 2g \sum_{i=1}^{n} a_i f_i q_i\right) \\ &= \frac{1}{g}\sum_{i=1}^{n} a_i (g^2 q_i^2 - 2(gq_i) f_i + f_i^2) \\ &= \frac{1}{g}\sum_{i=1}^{n} a_i (gq_i - f_i)^2 = \frac{1}{g}\sum_{i=1}^{n} a_i r_i^2 = \frac{1}{g}\sum_{i \in I} a_i r_i^2. \end{aligned}$$

Wäre $G = 0$, so wäre $\sum_{i \in I} a_i r_i^2 = 0$. Da der höchste Koeffizient von $a_i r_i^2$ für jedes $i \in I$ positiv ist, und I nicht leer ist, kann dies nicht sein. Also ist $G \neq 0$. Weil für alle $i \in I$ gilt $\deg r_i < \deg g$, haben wir $\deg G < 2\deg g - \deg g = \deg g$. □

Wie vorher angedeutet liefert dieser Satz zusammen mit der Lösung des 17. Hilbertschen Problems durch Artin den folgenden Nichtnegativstellensatz, in dem jedoch nichts mehr über die Anzahl der Summanden in der Darstellung ausgesagt wird:

Satz 2.32. *Es sei K ein angeordneter Körper und R sein reeller Abschluß . Sei $f \in K[X]$. Genau dann ist $f \geq 0$ auf R, wenn f eine gewichtete Summe von Quadraten in $K[X]$ ist, d.h. wenn es ein $n \in \mathbb{N}$, $a_1, \dots, a_n \in K^{\geq 0}$ und $g_1, \dots, g_n \in K[X]$ gibt mit $f = \sum_{i=1}^{n} a_i g_i^2$.*

Obwohl die Casselsche Elimination der Nenner 2.31 völlig algorithmisch ist (modulo Rechnen im Körper K), liefert sie uns keinen Algorithmus für obigen Satz. Dazu müßte man die Artinsche Lösung des 17. Hilbertschen Problems im univariaten Fall effektiv berechnen können. Nun ist dafür tatsächlich auch bisher schon ein Algorithmus bekannt, allerdings auch nur für den Fall, daß K ein Unterkörper von $\mathbb{R}$ ist (siehe Lemma 3.1 und die anschließende Diskussion in [LS], das darzustellende univariates Polynom habe o.B.d.A. keine reelle Nullstelle, vgl. auch [Hab]). In diesem Fall haben wir allerdings im Abschnitt 2.3 bereits einen Algorithmus für den stärkeren Nichtnegativstellensatz 2.28 angegeben.

Überhaupt haben wir in algorithmischer Hinsicht in diesem Abschnitt nichts neues erreicht: Auch für Satz 2.30 bekommen wir anders als für Satz 2.28 keinen Algorithmus. Im Beweis von Satz 2.30 gibt es zwar ein $t \in K$ mit $a - \varepsilon < t < a$, aber es ist nicht mehr klar, wie man ein solches berechnet. Im Abschnitt 2.3 hatten wir nämlich dazu ausgenutzt, daß bei fortlaufender Halbierung eines Intervalls, welches a enthält, die linke Intervallgrenze irgendwann größer als $a - \varepsilon$ wird. Dies ist aber nun i.A. nicht mehr der Fall, da $\varepsilon \in R$ infinitesimal positiv sein kann, d.h. $0 < \varepsilon < \frac{1}{n}$ für alle $n \in \mathbb{N}$.

Ob für den Fall, daß K nicht dicht in R liegt, statt der Aussage von Satz 2.32 vielleicht doch die stärkere Aussage von Satz 2.30 gilt, scheint nicht bekannt zu sein.

Kapitel 3

Auf kompakten Mengen positive Polynome

In diesem Kapitel geht es um folgende Situation: Es sei ein Unterkörper K von $\mathbb{R}$ gegeben. Es sei eine Teilmenge S des $\mathbb{R}^d$ definiert durch

$$S := \{x \in \mathbb{R}^d \mid q_1(x) = 0, \ldots, q_m(x) = 0, p_1(x) \geq 0, \ldots, p_n(x) \geq 0\}$$

mit $q_1, \ldots, q_m, p_1, \ldots, p_n \in K[X_1, \ldots, X_d]$. Es sei ein weiteres Polynom $f \in K[X_1, \ldots, X_d]$ gegeben mit $f > 0$ auf S. Wir suchen nach einer Darstellung von f, welche offensichtlich macht, daß $f > 0$ auf S gilt, und wollen eine solche Darstellung nach Möglichkeit auch berechnen. Genauer gesagt, soll die Darstellung am besten sogar von der Form

$$(\mathcal{R}_>) \qquad f = a + \sum_e a_e p_1^{e_1} \cdots p_n^{e_n} + \sum_{i=1}^{m} h_i q_i$$

sein mit $h_1, \ldots, h_m \in K[X_1, \ldots, X_d]$, $a \in K^{>0}$ und $a_e \in K^{\geq 0}$ für alle $e \in \mathbb{N}^n$, über die summiert wird. Wenn S nicht kompakt ist (dann ist S nicht beschränkt, denn S ist abgeschlossen), werden unsere Methoden nicht greifen.

Es wäre nur natürlich für diejenigen $f \in K[X_1, \ldots, X_d]$ mit $f \geq 0$ auf S nach der entsprechenden Darstellung $(\mathcal{R}_\geq)$ zu fragen, die aus $(\mathcal{R}_>)$ dadurch entstehe, daß man auf den ersten Summanden $a \in K^{>0}$ verzichtet. Hier schaut es jedoch schlecht aus: Wenn ein Polynom $f \in K[X_1, \ldots, X_d]$ eine Nullstelle $x \in S$ besitzt, für welche sogar die strikten Ungleichungen $p_1(x) > 0, \ldots, p_n(x) > 0$ gelten, so kann dieses Polynom offenbar entweder nur das Nullpolynom sein oder es besitzt keine Darstellung $(\mathcal{R}_\geq)$.

Wir werden uns also Gedanken um die Darstellung $(\mathcal{R}_>)$ machen, und zwar zunächst nur um deren Existenz, wobei sich ein weitgehend algorithmischer Zugang allerdings schon abzeichnen wird. In Abschnitt 3.1 beweisen wir die Existenz für den Fall, daß $q_1 = X_1 + \cdots + X_d = 1$ (zunächst $m = 1$, dann auch $m > 1$) und $n = d, p_1 = X_1, \ldots, p_n = X_d$.

In Abschnitt 3.2 befreien wir uns von diesen strengen Einschränkungen weitgehend. Wir zeigen: Schon wenn es Zahlen $s_1, \ldots, s_d \in K^{\geq 0}$ gibt, sodaß es etwa für die $2d$ Polynome $s_1 + X_1, s_1 - X_1, \ldots, s_d + X_d, s_d - X_d$ eine Darstellung $(\mathcal{R}_\geq)$ gibt (insbesondere ist dann $S \subseteq \prod_{i=1}^{d} [-s_i, s_i]$ kompakt), so gibt es für jedes $f \in K[X_1, \ldots, X_d]$ mit $f > 0$ auf S eine Darstellung $(\mathcal{R}_>)$.

Dies wird in Abschnitt 3.3 algorithmisch umgesetzt: Wenn man Darstellungen $(\mathcal{R}_\geq)$ etwa von $s_1 + X_1, s_1 - X_1, \ldots, s_d + X_d, s_d - X_d$ gefunden hat (wir sprechen davon, daß man

einen *Archimedizitätsnachweis* geführt hat), so kann man völlig algorithmisch für jedes $f \in K[X_1, \ldots, X_d]$ mit $f > 0$ auf S eine solche Darstellung bestimmen.

Im darauf folgenden Abschnitt 3.4 werden wir uns damit beschäftigen, wie man in verschiedenen Beispielen Archimedizitätsnachweise automatisch berechnen kann, wodurch das Problem der Berechnung der Darstellung $(\mathcal{R}_>)$ für diese Beispiele vollständig gelöst wird. Zu diesen Beispielen zählt der Fall, daß es lineare Polynome $q_{i_1}, \ldots, q_{i_k} \in \{q_1, \ldots, q_m\}$ und $p_{j_1}, \ldots, p_{j_l} \in \{p_1, \ldots, p_n\}$ gibt, sodaß die Obermenge

$$\{x \in \mathbb{R}^d \mid q_{i_1}(x) = 0, \ldots, q_{i_k}(x) = 0, p_{j_1}(x) \geq 0, \ldots, p_{j_l}(x) \geq 0\}$$

von S nichtleer und kompakt ist.

Im Abschnitt 3.5 geht es dann darum, was man macht, wenn eine Darstellung $(\mathcal{R}_>)$ nicht existiert. Für jede *ungerade* Zahl $k \in \mathbb{N}$ existiert allein unter der Voraussetzung, daß S kompakt ist, zu jedem $f \in K[X_1, \ldots, X_d]$ mit $f > 0$ auf S zumindest eine Darstellung der Form

$$(\mathcal{R}^k_>) \qquad f = a + \sum_e \left(\sum_j a_{ej} g_{ej}^{2k} \right) p_1^{e_1} \cdots p_n^{e_n} + \sum_{i=1}^m h_i q_i$$

mit $g_{ej}, h_i \in K[X_1, \ldots, X_d]$, $a \in K^{>0}$ und $a_{ej} \in K^{\geq 0}$. Hier braucht offenbar o.B.d.A. nur über solche e mit $e \in \{0, \ldots, 2k-1\}^n$ summiert zu werden. Auch hier wird der Leser fragen, wie es um die analoge Darstellung $(\mathcal{R}^k_\geq)$ bestellt ist, die aus $(\mathcal{R}^k_>)$ durch Weglassen des ersten Summanden $a \in K^{>0}$ entstehe, und wie es um die Darstellung $(\mathcal{R}^k_>)$ für *gerades* $k \geq 1$ steht. Mehr dazu in jenem Abschnitt.

Im letzten Abschnitt dieses Kapitels befassen wir uns mit einer übersichtlicheren Variante des entwickelten Verfahrens für bestimmte Spezialfälle.

Bemerkung 3.1. Die Existenz der Darstellung $(\mathcal{R}_>)$ zu beweisen, heißt nichts anderes als zu beweisen, daß es $a \in K^{>0}$ und $a_e \in K^{\geq 0}$ gibt mit

$$(\star) \qquad f \equiv a + \sum_e a_e p_1^{e_1} \cdots p_n^{e_n}$$

modulo dem von $q_1, \ldots, q_m$ in $K[X_1, \ldots, X_d]$ erzeugten Ideal. Aber auch in algorithmischer Hinsicht können wir die Summe $\sum_{i=1}^m h_i g_i$ in $(\mathcal{R}_>)$ vergessen. Wenn wir $a \in K^{>0}$ und $a_e \in K^{\geq 0}$ mit $(\star)$ schon kennen, so können wir mit Hilfe von Gröbnerbasen (siehe [BW], [Mis],[We1]) leicht $h_1, \ldots, h_m \in K[X_1, \ldots, X_d]$ berechnen, sodaß $(\mathcal{R}_>)$ gilt. Analoges gilt natürlich für $(\mathcal{R}_\geq)$, $(\mathcal{R}^k_>)$ und $(\mathcal{R}^k_\geq)$ anstelle von $(\mathcal{R}_>)$.

Man kann nämlich beim Berechnen einer Gröbnerbasis des von $q_1, \ldots, q_m$ erzeugten Ideals für jedes Element dieser Gröbnerbasis gleich eine Darstellung als Summe von Vielfachen der q_i mitberechnen (Algorithmus EXTGRÖBNER in [BW]). Daher reicht es aus, jedes Element des von $q_1, \ldots, q_m$ erzeugten Ideals als Summe von Vielfachen der Elemente einer Gröbnerbasis darzustellen, was selbstverständlich kein Problem ist.

3.1 Der Satz von Pólya als Ausgangspunkt

Definition 3.2. *Sei K ein angeordneter Körper. Sei $F \in K[X_1, \ldots, X_d]$ und $k \in \mathbb{N}$. F heißt k-*Form*, wenn alle in F vorkommenden Monome (mit einem Koeffizienten $\neq 0$) den Grad k haben. F heißt* Form*, wenn F für ein $k \in \mathbb{N}$ eine k-Form ist. Eine k-Form $F \in K[X_1, \ldots, X_d]$ nennen wir* Positivform *(bezüglich $K[X_1, \ldots, X_d]$), wenn sie die Gestalt*

$$F = \sum_{e_1+\cdots+e_d=k} a_e X_1^{e_1} \cdots X_d^{e_d} \quad \textit{mit} \quad a_e \in K^{>0} \quad \textit{für alle } e \in \mathbb{N}^d \textit{ mit } e_1 + \cdots + e_d = k$$

hat.

Beispiel 3.3. Das Nullpolynom ist eine k-Form für jedes $k \in \mathbb{N}$. Es ist keine Positivform. Jede Form $F \neq 0$ ist genau für ein $k \in \mathbb{N}$ eine k-Form, nämlich für $k = \deg F$. Das Polynom $X_1^2 + X_2$ ist eine Positivform (bezüglich $\mathbb{R}[X_1, X_2]$). Es ist allerdings keine Positivform bezüglich $\mathbb{R}[X_1, X_2, X_3]$.

Definition 3.4. *Wir definieren eine kompakte Teilmenge Δ_d von $\mathbb{R}^d$ durch*

$$\Delta_d = \{x \in (\mathbb{R}^{\geq 0})^d \mid x_1 + \cdots + x_d = 1\}.$$

Bemerkung 3.5. Sei $F \in \mathbb{R}[X_1, \ldots, X_d]$ eine k-Form. Dann gilt $F(\lambda x) = \lambda^k F(x)$ für alle $x \in \mathbb{R}^d$ und $\lambda \in \mathbb{R}$. Für $\lambda > 0$ ergibt sich, daß $F(\lambda x)$ dasselbe Vorzeichen hat wie $F(x)$. Das Vorzeichen von F ist also auf vom Nullpunkt ausgehenden Halbgeraden, die den Nullpunkt nicht enthalten, konstant. Es ergibt sich, daß für jede Form F folgende Äquivalenz gilt:

$$F > 0 \text{ auf } (\mathbb{R}^{\geq 0})^d \setminus \{0\} \iff F > 0 \text{ auf } \Delta_d.$$

Im Jahre 1927 bewies Pólya den folgenden Positivstellensatz:

Satz 3.6 (Pólya). *Sei $F \in \mathbb{R}[X_1, \ldots, X_d]$ eine Form. Genau dann gilt*

$$F > 0 \textit{ auf } (\mathbb{R}^{\geq 0})^d \setminus \{0\},$$

wenn es ein $N \in \mathbb{N}$ gibt, sodaß $F(X_1 + \cdots + X_d)^N$ eine Positivform ist.

Beweis: Die eine Richtung ist trivial. Für die andere Richtung sei vorausgesetzt, daß $F(x) > 0$ ist für alle $x \in (\mathbb{R}^{\geq 0})^d \setminus \{0\}$. Bezeichne k den Grad von F. Wir schreiben

$$F = \sum_{e_1+\cdots+e_d=k} a_e X_1^{e_1} \cdots X_d^{e_d} \quad \text{mit} \quad a_e \in \mathbb{R}$$

und definieren eine neue k-Form $G \in \mathbb{R}[X_1, \ldots, X_d, T]$ durch

$$G = \sum_{e_1+\cdots+e_d=k} a_e \prod_{i=1}^{d} \underbrace{X_i(X_i - T) \cdots (X_i - (e_i - 1)T)}_{e_i \text{ Faktoren}}.$$

Man beachte $G(X_1, \ldots, X_d, 0) = F$. Wir rechnen für beliebiges $N \in \mathbb{N}$:

$$\begin{aligned} F(X_1 + \cdots + X_d)^N &= (X_1 + \cdots + X_d)^N \sum_{e_1+\cdots+e_d=k} a_e X_1^{e_1} \cdots X_d^{e_d} \\ &= \left(\sum_{l_1+\cdots+l_d=N} \binom{N}{l_1 \ldots l_d} X_1^{l_1} \cdots X_d^{l_d} \right) \sum_{e_1+\cdots+e_d=k} a_e X_1^{e_1} \cdots X_d^{e_d} \\ &= \sum_{e_1+\cdots+e_d=k} \; \sum_{l_1+\cdots+l_d=N} a_e \binom{N}{l_1 \ldots l_d} X_1^{e_1+l_1} \cdots X_d^{e_d+l_d} \end{aligned}$$

$$= \sum_{e_1+\cdots+e_d=k} \sum_{\substack{r_1+\cdots+r_d=k+N \\ r_1\geq e_1,\ldots,r_d\geq e_d}} a_e \binom{N}{(r_1-e_1)\ldots(r_d-e_d)} X_1^{r_1}\cdots X_d^{r_d}$$

$$= \sum_{r_1+\cdots+r_d=k+N} \sum_{\substack{e_1+\cdots+e_d=k \\ e_1\leq r_1,\ldots,e_d\leq r_d}} a_e \frac{N!}{(r_1-e_1)!\cdots(r_d-e_d)!} X_1^{r_1}\cdots X_d^{r_d}$$

$$= \sum_{r_1+\cdots+r_d=k+N} \frac{N!}{r_1!\cdots r_d!} \left(\sum_{\substack{e_1+\cdots+e_d=k \\ e_1\leq r_1,\ldots,e_d\leq r_d}} a_e \frac{r_1!\cdots r_d!}{(r_1-e_1)!\cdots(r_d-e_d)!} \right) X_1^{r_1}\cdots X_d^{r_d}$$

$$= \sum_{r_1+\cdots+r_d=k+N} \frac{N!}{r_1!\cdots r_d!} \left(\sum_{\substack{e_1+\cdots+e_d=k \\ e_1\leq r_1,\ldots,e_d\leq r_d}} a_e \prod_{i=1}^{d} (r_i(r_i-1)\cdots(r_i-(e_i-1))) \right) X_1^{r_1}\cdots X_d^{r_d}$$

$$= \sum_{r_1+\cdots+r_d=k+N} \frac{N!}{r_1!\cdots r_d!} \left(\sum_{e_1+\cdots+e_d=k} a_e \prod_{i=1}^{d} (r_i(r_i-1)\cdots(r_i-(e_i-1))) \right) X_1^{r_1}\cdots X_d^{r_d}$$

$$= \sum_{r_1+\cdots+r_d=k+N} \frac{N!}{r_1!\cdots r_d!} G(r_1,\ldots,r_d,1) X_1^{r_1}\cdots X_d^{r_d}$$

$$= \sum_{r_1+\cdots+r_d=k+N} \frac{N!(k+N)^k}{r_1!\cdots r_d!} G\left(\frac{r_1}{k+N},\ldots,\frac{r_d}{k+N},\frac{1}{k+N}\right) X_1^{r_1}\cdots X_d^{r_d}$$

Weil für alle $r_1,\ldots,r_d \in \mathbb{N}$ mit $r_1+\cdots+r_d = k+N$ gilt $(\frac{r_1}{k+N},\ldots,\frac{r_d}{k+N}) \in \Delta_d$, und weil $\frac{1}{k+N}$ für $N\to\infty$ gegen 0 konvergiert, genügt es folgendes zu zeigen: Es gibt eine Umgebung V von 0 in $\mathbb{R}$, sodaß für alle $x\in\Delta_d$ und für alle $t\in V$ gilt $G(x,t)>0$.

Diese Umgebung V erhält man wie folgt: Für jedes $x\in\Delta_d$ ist $G(x,0)=F(x)>0$, also gibt es eine Umgebung von $(x,0)$ in $\mathbb{R}^{d+1}$, auf der $G>0$ gilt. Wir wählen zu jedem $x\in\Delta_d$ eine solche Umgebung von $(x,0)$, o.B.d.A. eine von der Form $U_x\times V_x$, wobei U_x eine offene Umgebung von x in $\mathbb{R}^d$ und V_x eine Umgebung von 0 in $\mathbb{R}$ ist. Die Menge $\{U_x \mid x\in\Delta_d\}$ bildet eine Überdeckung der kompakten Menge Δ_d durch offene Mengen. Daher gibt es eine Teilüberdeckung $\{U_x \mid x\in X\}$ mit einer endlichen Teilmenge X von Δ_d. Wir setzen $V=\bigcap\{V_x \mid x\in X\}$. Da X endlich ist, ist V wieder eine Umgebung von 0 in $\mathbb{R}$.

Außerdem leistet V das Gewünschte: Sei $x\in\Delta_d$ und $t\in V$. Wir zeigen $G(x,t)>0$. Wähle $y\in X$ mit $x\in U_y$. Dann gilt $(x,t)\in U_y\times V\subseteq U_y\times V_y$. Wegen $G>0$ auf $U_y\times V_y$ folgt $G(x,t)>0$. □

Bemerkung 3.7. Im Zusammenhang mit der Aussage des gerade bewiesenen Satzes sollte man beachten: Ist K ein Unterkörper von $\mathbb{R}$ und $F\in K[X_1,\ldots,X_d]$, so ist auch $F(X_1+\cdots+X_d)^N \in K[X_1,\ldots,X_d]$. Ist ferner $N\in\mathbb{N}$, sodaß $F(X_1+\cdots+X_d)^N$ eine Positivform ist, so ist für alle $N'\geq N$ das Polynom $F(X_1+\cdots+X_d)^{N'}$ ebenfalls eine Positivform.

Wir haben hier den Originalbeweis von Pólya (siehe [Pól] oder Problem 56 gegen Ende von Kapitel II in [HLP]) übernommen. Bemerkung 3.5 zeigt, daß der Satz durchaus etwas mit der Überschrift des Kapitels „Auf kompakten Mengen positive Polynome“ zu tun hat, da man in seiner Formulierung $(\mathbb{R}^{\geq 0})^d\setminus\{0\}$ durch die kompakte Menge Δ_d ersetzen kann. Der Satz von Pólya hat die außerordentlich angenehme Eigenschaft, daß es ein völlig offensichtliches Verfahren zur Berechnung der durch ihn garantierten Darstellung gibt: Man berechne für $N=0,1,2,\ldots$ das Produkt $F(X_1+\cdots+X_d)^N$ und überprüfe, ob es eine Positivform ist. Die folgende Bemerkung zeigt allerdings, daß unter gewissen Voraussetzungen bei genügend kleinen positiven Werten von F auf Δ_d dieser Algorithmus beliebig lange braucht, bis er eine Positivform findet:

Bemerkung 3.8. Für jede Form $F \in \mathbb{R}[X_1, \ldots, X_d]$, für die es ein $N \in \mathbb{N}$ gibt, sodaß $F(X_1 + \cdots + X_d)^N$ eine Positivform ist, bezeichnen wir das kleinste solche N als den Pólya-Exponenten von F. Gibt es kein solches N, so legen wir den Pólya-Exponenten von F als ∞ fest. Die Komplexität der durch den Satz von Pólya gegebenen Darstellung verhält sich im folgenden Sinn bösartig:

Sei $k \in \mathbb{N}$ und $E = \{(e_1, \ldots, e_d) \in \mathbb{N}^d \mid e_1 + \cdots + e_d = k\}$. Sei $F = \sum_{e \in E} c_e X_1^{e_1} \cdots X_d^{e_d}$ eine k-Form mit den Koeffizienten $c_e \in \mathbb{R}$ ($e \in E$), die an den Einheitsvektoren $\delta_1, \ldots, \delta_d$ des $\mathbb{R}^d$ ausgewertet positiv ist, aber eine Nullstelle $(\xi_1, \ldots, \xi_d) \in (\mathbb{R}^{\geq 0})^d \setminus \{0\}$ hat. Strebt dann $(a_e)_{e \in E}$ in $\mathbb{R}^E$ gegen $(c_e)_{e \in E}$, so strebt der Pólya-Exponent von $\sum_{e \in E} a_e X_1^{e_1} \cdots X_d^{e_d}$ gegen ∞.

Dies wurde in [Rob] bewiesen. Wir haben die Aussage leicht verschärft und den Beweis von der Terminologie der Nichtstandardanalysis befreit, sodaß er nun für einen Leser, der die Anfänge der Modelltheorie (siehe [We2], [Rot], [Pr1]) kennt, lesbar ist. Man beachte übrigens, daß in [Rob] der Wortlaut von Theorem 8.1 geändert werden muß. Statt „$\sum x_i \neq 0$“ muß dort „x_i not infinitesimal“ geschrieben werden. In der bisherigen Formulierung ist Theorem 8.1 trivialerweise falsch.

Nun zum Beweis mittels Modelltheorie: Ein Element a eines angeordneten Körpers heißt *endlich* (bzw. *infinitesimal*), wenn der Betrag von a kleiner als eine (bzw. jede) positive rationale Zahl ist. K heißt *archimedisch*, wenn K nur endliche Elemente enthält. Die Menge K_{fin} der endlichen Elemente von K bildet in jedem Fall einen Unterring von K, und die Menge K_{inf} der infinitesimalen Elemente von K bildet ein Primideal von K_{fin}.

Sei nun $N \in \mathbb{N}$. Zu zeigen ist, daß es eine Umgebung von $(c_e)_{e \in E}$ in $\mathbb{R}^E$ gibt, sodaß für alle $(a_e)_{e \in E}$ aus dieser Umgebung der Pólya-Exponent von $\sum_{e \in E} a_e X_1^{e_1} \cdots X_d^{e_d}$ größer als N ist. Es ist also zu zeigen, daß folgende Bedingung $(\star_K)$ für $K = \mathbb{R}$ gilt:

$(\star_K)$ Es gibt eine Umgebung U von $(c_e)_{e \in E}$ in K^E (bezüglich der Produkttopologie, die von der Ordnungstopologie auf K stammt), sodaß $(X_1 + \cdots + X_d)^N \sum_{e \in E} a_e X_1^{e_1} \cdots X_d^{e_d} \in K[X_1, \ldots, X_d]$ für kein $(a_e)_{e \in E} \in U$ eine Positivform ist.

Nun können wir $(\star_K)$ zumindest für alle nicht archimedischen Erweiterungen K geordneter Körper von $\mathbb{R}$ zeigen, indem wir U einfach „infinitesimal klein“ wählen, also etwa $U = \prod_{e \in E}]c_e - \varepsilon, c_e + \varepsilon[$ setzen mit einem $0 < \varepsilon \in K_{\text{inf}}$ (ein solches existiert). Sei dann nämlich $(a_e)_{e \in E} \in U$. Mit c_e ist auch a_e für jedes $e \in E$ endlich. Angenommen $G := (X_1 + \cdots + X_d)^N \sum_{e \in E} a_e X_1^{e_1} \cdots X_d^{e_d}$ wäre nun eine Positivform. Weil a_e für jedes $e \in E$ kongruent zu c_e modulo dem Ideal K_{inf} von K_{fin} ist, gilt dann in K_{fin}, daß $G(\xi_1, \ldots, \xi_d) \equiv (\xi_1 + \cdots + \xi_d)^N F(\xi_1, \ldots, \xi_d) = 0$ modulo K_{inf}. Also ist $G(\xi_1, \ldots, \xi_d) \in K_{\text{inf}}$. Dies ist widersprüchlich, denn andererseits können wir zeigen, daß $G(\xi_1, \ldots, \xi_d)$ nicht infinitesimal ist. Wählt man nämlich $i \in \{1, \ldots, d\}$ mit $\xi_i \in \mathbb{R}^{>0}$, so ist $G(\xi_1, \ldots, \xi_d) \geq G(\xi_i \delta_i) > 0$, und $G(\xi_i \delta_i) = \xi_i^{N+k} G(\delta_i)$ ist nicht infinitesimal. Denn K_{inf} ist ein Primideal in K_{fin}, und sowohl ξ_i^{N+k} als auch $G(\delta_i)$ ist ein Element von $K_{\text{fin}} \setminus K_{\text{inf}}$. Für $G(\delta_i)$ sieht man das wie folgt ein: Es ist $G(\delta_i)$ modulo K_{inf} kongruent zu $F(\delta_i)$, welches nach Voraussetzung eine positive reelle Zahl ist und damit nicht infinitesimal ist.

Die gerade bewiesene Tatsache, daß $(\star_K)$ für alle nicht archimedischen Erweiterungen K geordneter Körper von $\mathbb{R}$ gilt, drücken wir nun mit Hilfe des semantischen Folgerungsbegriffs $\models$ der Logik erster Stufe aus. Wir betrachten dazu eine Signatur mit Konstantensymbolen für jede reelle Zahl, Funktionssymbolen für $+$, $-$ und $\cdot$, sowie einem Relationssymbol für $<$. Offenbar läßt sich in dieser Signatur ein Satz φ erster Stufe formulieren, der in jeder Erweiterung K angeordneter Körper von $\mathbb{R}$ (bei natürlicher Interpretation aller Symbole) nichts anderes als $(\star_K)$ besagt. Dann gilt $\{n < c \mid n \in \mathbb{N}\} \cup D \cup \Sigma \models \varphi$, wobei c ein neues Konstantensymbol, D das Basisdiagramm (oft auch nur Diagramm genannt) von $\mathbb{R}$ und Σ ein Axiomensystem für angeordnete Körper sei. Nach dem Endlichkeitssatz der Logik erster Stufe gibt es eine endliche Teilmenge Φ von $\{n < c \mid n \in \mathbb{N}\} \cup D \cup \Sigma$ mit $\Phi \models \varphi$. Durch genügend große Interpretation von c kann man $\mathbb{R}$ zu einem Modell von Φ expandieren. Also gilt φ in $\mathbb{R}$. Das heißt, es gilt $(\star_K)$ für $K = \mathbb{R}$, und die Aussage ist bewiesen.

Wir illustrieren die Aussage mit einem Beispiel: Man betrachte die von $a \in \mathbb{R}$ abhängige

Form $X_1^2 - aX_1X_2 + X_2^2 \in \mathbb{R}[X_1, X_2]$. Für jedes $a < 2$ ist sie wegen

$$X_1^2 - aX_1X_2 + X_2^2 = (X_1 - X_2)^2 + (2-a)X_1X_2$$

positiv auf $(\mathbb{R}^{\geq 0})^2 \setminus \{0\}$ und besitzt damit nach dem Satz von Pólya einen endlichen Pólya-Exponenten. Strebt a gegen 2, so strebt dieser allerdings gegen ∞.

Wir beenden diese Betrachtungen über die Komplexität des Satzes von Pólya mit der Bemerkung, daß wir in der hergeleiteten Aussage für kein einziges $i \in \{1, \ldots, d\}$ auf die Voraussetzung $F(\delta_i) > 0$ verzichten können. Betrachte hierzu die von $a \in \mathbb{R}$ abhängige Form $aX_1 + X_2 \in \mathbb{R}[X_1, X_2]$. Für $a > 0$ hat sie den Pólya-Exponenten 0, obwohl sie für $a = 0$ im Punkt δ_2 positiv ist und eine Nullstelle auf $(\mathbb{R}^{\geq 0})^2 \setminus \{0\}$ hat.

Aus der eben gemachten Bemerkung folgt, daß der Satz von Pólya nicht für beliebige reell abgeschlossene Körper (siehe [Lor], [Jac], [We2]) anstelle von $\mathbb{R}$ gelten kann. Sonst bekäme man mit dem Endlichkeitssatz der Logik erster Stufe nämlich obere Schranken für den Pólya-Exponenten von k-Formen $F \in \mathbb{R}[X_1, \ldots, X_d]$ mit $F > 0$ auf $(\mathbb{R}^{\geq 0})^d \setminus \{0\}$, welche nur von k und d abhängten (siehe [Pr3], vgl. auch Satz 5.5.4 in [We2]). Dies widerspricht offenbar unserer Bemerkung. Im Jahre 1994 gaben Loera und Santos einen neuen Beweis für den Satz von Pólya, aus dem solche oberen Schranken, die allerdings zusätzlich noch von der Größe der Koeffizienten der Form und dem Minimum der Form auf Δ_d abhängen, extrahiert werden können (siehe [LS]). Unbefriedigend ist hier vor allem die Abhängigkeit vom Minimum der Form F auf Δ_d, da man dieses der Form F nicht unmittelbar ansieht. Wenn F nur ganzzahlige Koeffizienten hat, kann man wenigstens dieses Minimum abschätzen durch die Größe der Koeffizienten und einer unbekannten festen (also von allen Daten unabhängigen) Zahl, sodaß man für diesen Fall insgesamt eine Schranke in Abhängigkeit des Grades der Form, der Anzahl der Unbestimmten, der Größe der Koeffizienten und einer unbekannten von allen Daten unabhängigen Zahl erhält. Genaugenommen beweisen Loera und Santos:

Bemerkung 3.9. Sei $F \in \mathbb{R}[X_1, \ldots, X_d]$ eine k-Form mit $F > 0$ auf $(\mathbb{R}^{\geq 0})^d \setminus \{0\}$. Es sei b das Maximum der Beträge der Koeffizienten von F. Es sei μ das Minimum von F auf Δ_d. Dann gilt

(i) Für jedes $N \in \mathbb{N}$ mit $N \geq \frac{dk(bk+1)}{\mu}$ ist $F(X_1 + \cdots + X_d)^N$ eine Positivform.

(ii) Falls F ganzzahlige Koeffizienten hat, so gibt es eine von allen Daten unabhängige Zahl $\nu \in \mathbb{N}$ mit

$$\frac{1}{\mu} \leq b^{\left((1+\max\{k,d\})^{\nu(1+d)}\right)} 2^{\left((1+\max\{k,d\})^{\nu(1+d)}\right)}.$$

Man beachte, daß in einem kursierenden Vorabdruck von [LS] bei der Formulierung von Aussage 1 des dortigen Theorems 1.2 ein Leichtsinnsfehler unterlaufen ist: Es muß dort „F has integer coefficients“ statt „F has rational coefficients“ heißen. Sonst ist die Aussage falsch, wie man mit Hilfe von Bemerkung 3.8 leicht sieht, da die Abhängigkeit der Schranke von der Größe der Koeffizienten nur noch scheinbar gegeben wäre. In [LS] ist dieser Fehler bereits korrigiert.

Der Satz von Pólya läßt sich nicht in einen analogen Nichtnegativstellensatz verwandeln: Sei $F \in \mathbb{R}[X_1, \ldots, X_d]$ eine Form. Wenn es ein $N \in \mathbb{N}$ gibt, sodaß $F(X_1 + \cdots + X_d)^N$ keine negativen Koeffizienten hat, so ist $F \geq 0$ auf $(\mathbb{R}^{\geq 0})^d \setminus \{0\}$. Aber die Umkehrung ist i.A. falsch: Eine Form $F \in \mathbb{R}[X_1, \ldots, X_d]$ mit $F \geq 0$ auf $(\mathbb{R}^{\geq 0})^d \setminus \{0\}$ braucht nur eine Nullstelle im Inneren von $(\mathbb{R}^{\geq 0})^d \setminus \{0\}$ haben, dann ist $F = 0$ oder für jedes $N \in \mathbb{N}$ hat $F(X_1 + \cdots + X_d)^N$ einen negativen Koeffizienten.

Lemma 3.10. *Sei K ein Unterkörper von $\mathbb{R}$ und $f \in K[X_1, \ldots, X_d]$ mit $f > 0$ auf Δ_d. Dann ist f modulo dem Polynom $X_1 + \cdots + X_d - 1$ in $K[X_1, \ldots, X_d]$ äquivalent zu einer Positivform.*

Beweis: Indem wir in f jedes Monom von einem Grad $e < \deg f$ mit $(X_1 + \cdots + X_d)^{(\deg f)-e}$ multiplizieren, bekommen wir eine zu f modulo $X_1 + \cdots + X_d - 1$ äquivalente Form $F \in K[X_1, \ldots, X_d]$. Für alle $x \in \Delta_d$ gilt $F(x) = f(x) > 0$. Nach Bemerkung 3.5 und dem Satz von Pólya 3.6 gibt es ein $N \in \mathbb{N}$, sodaß $F(X_1 + \cdots + X_d)^N$ eine Positivform ist. Diese ist modulo $X_1 + \cdots + X_d - 1$ äquivalent zu F und damit zu f. □

Aus obigem Lemma folgt die Existenz der Darstellung $(\mathcal{R}_>)$ für alle $f \in K[X_1, \ldots, X_d]$ mit $f > 0$ auf S, falls $m = 1$, $q_1 = X_1 + \cdots + X_d - 1$ und $n = d, p_1 = X_1, \ldots, p_n = X_d$. Wenn nämlich eine Positivform $F \in K[X_1, \ldots, X_d]$ existiert mit $f \equiv F$ modulo q_1, so wähle man ein $a \in K^{>0}$ so klein, daß die Form $F - a(X_1 + \cdots + X_d)^{\deg F}$ keine negativen Koeffizienten besitzt. Dann gilt $f \equiv a + (F - a(X_1 + \cdots + X_d)^{\deg F})$ modulo q_1 und nach Bemerkung 3.1 folgt die Existenz von $(\mathcal{R}_>)$.

Erstaunlicherweise funktioniert das ganze auch noch, falls $m > 1$, falls man also noch zusätzliche Gleichungen hat, die S definieren. Im Allgemeinen ist dann S nur noch eine echte (eventuell sogar leere) Teilmenge von Δ_d, und obiger Beweis geht nicht mehr durch. Die Hinzunahme weiterer Gleichungen hat allerdings nicht nur den negativen Effekt, daß S eventuell kleiner wird, sondern auch den positiven, daß f nun modulo $q_1, \ldots, q_m$ i.A. zu mehr Polynomen äquivalent ist. Wir werden sogar sehen, daß f insbesondere zu einem Polynom äquivalent ist, welches positiv nicht nur auf S, sondern sogar wieder auf Δ_d ist. Dies ist etwa das Polynom $f + c(q_2^2 + \cdots + q_m^2)$ für genügend großes $c \in K$. Das Polynom $q_2^2 + \cdots + q_m^2$ ist nämlich positiv auf $\Delta_d \setminus S$ und es wird folgendes Lemma mit $U = S$, $V = \Delta_d$ und $r = q_2^2 + \cdots + q_m^2$ greifen:

Lemma 3.11. *Sei V ein kompakter Raum, $U \subseteq V$. Es seien f und r stetige Funktionen von V nach $\mathbb{R}$ mit folgenden Eigenschaften:*

$$f > 0 \text{ auf } U, \quad r \geq 0 \text{ auf } U \quad \text{und} \quad r > 0 \text{ auf } V \setminus U.$$

Dann gibt es ein $c_0 \in \mathbb{R}$, sodaß für alle $c \geq c_0$ gilt $f + cr > 0$ auf V.

Beweis: O.B.d.A. ist U offen in V, sonst gehen wir von U zu $f^{-1}(]0, \infty[)$ über. O.B.d.A. gelte auch $U \neq V$. Die Menge $V \setminus U$ ist abgeschlossen in einem kompakten Raum und daher selbst kompakt. Auf $V \setminus U$ nimmt daher r ein Minimum $\mu > 0$ und f ein Minimum $\mu' \in \mathbb{R}$ an. Für $c \in \mathbb{R}$ mit $c \geq 0$ gilt dann

$$\begin{array}{lll} f + cr \geq f > 0 & \text{auf} \quad U & \text{und} \\ f + cr \geq \mu' + c\mu & \text{auf} \quad V \setminus U. & \end{array}$$

Wegen $\mu > 0$ ist $\mu' + c\mu$ für genügend großes c größer als 0. □

Definition 3.12. *Sei K ein Unterkörper von $\mathbb{R}$ und I ein Ideal von $K[X_1, \ldots, X_d]$. Wir definieren dann die Nullstellenmenge $V_{\mathbb{R}}(I)$ von I durch*

$$V_{\mathbb{R}}(I) := \{x \in \mathbb{R}^d \mid q(x) = 0 \text{ für alle } q \in I\}.$$

Lemma 3.13. *Sei K ein Unterkörper von $\mathbb{R}$, I ein Ideal von $K[X_1, \ldots, X_d]$ mit $X_1 + \cdots + X_d - 1 \in I$. Es sei $f \in K[X_1, \ldots, X_d]$ mit $f > 0$ auf $V_{\mathbb{R}}(I) \cap (\mathbb{R}^{\geq 0})^d$. Dann ist f modulo dem Ideal I in $K[X_1, \ldots, X_d]$ äquivalent zu einer Positivform.*

Beweis: Das Ideal I ist nach dem Hilbertschen Basissatz endlich erzeugt, etwa

$$I = (X_1 + \cdots + X_d - 1, r_1, \ldots, r_t)$$

mit $r_1, \ldots, r_t \in K[X_1, \ldots, X_d]$. Wir wenden Lemma 3.11 an mit

$$U := V_{\mathbb{R}}(I) \cap (\mathbb{R}^{\geq 0})^d \subseteq \Delta_d =: V, \quad f|_V \quad \text{und} \quad r := (r_1^2 + \cdots + r_t^2)|_V.$$

Es gibt also ein $c \in K$, sodaß $f + c(r_1^2 + \cdots + r_t^2) > 0$ auf $V = \Delta_d$. Nach Lemma 3.10 ist $f + c(r_1^2 + \cdots + r_t^2)$ modulo $X_1 + \cdots + X_d - 1$, insbesondere modulo I äquivalent zu einer Positivform in $K[X_1, \ldots, X_d]$. Da f modulo I äquivalent ist zu $f + c(r_1^2 + \cdots + r_t^2)$, folgt die Behauptung. □

3.2 Der archimedische Positivstellensatz

Im letzten Abschnitt haben wir schon den algebraischen Begriff benutzt, der dem letzten Teil $\sum_{i=1}^m h_i q_i$ der Darstellung $(\mathcal{R}_>)$ (siehe Seite 33) entspricht. Es war dies das von $q_1, \ldots, q_m$ erzeugte Ideal I in $K[X_1, \ldots, X_d]$. Für den mittleren Teil der Darstellung $(\mathcal{R}_>)$ ist der angemessene Begriff dann der im Folgenden definierte von $K^{\geq 0} \cup \{\overline{p_1}, \ldots, \overline{p_n}\}$ erzeugte Semiring im Ring $K[X_1, \ldots, X_d]/I$. Hierbei haben wir schon folgende Konvention benutzt:

Die Restklasse $f + I$ nach einem Ideal I eines Elements von $K[X_1, \ldots, X_d]$ bezeichnen wir mit $\overline{f}$. Da der kanonische Ringhomomorphismus $K \to K[X_1, \ldots, X_d]/I$ eine Einbettung ist, außer wenn $I = K[X_1, \ldots, X_d]$ ist, notieren wir die Restklasse $\overline{a}$ eines Elements a von K oft wieder mit a.

Definition 3.14. *Sei R ein Ring. Eine Teilmenge P von R heißt* Semiring *in R, wenn gilt:*

$$0, 1 \in P, \quad P + P \subseteq P \quad \textit{und} \quad P \cdot P \subseteq P.$$

Für jedes $k \in \mathbb{N}$ heißt P Semiring k-ter Stufe *in R, wenn zusätzlich $r^{2k} \in P$ für jedes $r \in R$ (für $k = 0$ ist dies keine zusätzliche Forderung). Für $E \subseteq R$ und $k \in \mathbb{N}$ bezeichne $\langle E \rangle_k$ den* von E in R erzeugten Semiring k-ter Stufe*, also den kleinsten Semiring k-ter Stufe in R, der E enthält.*

Nun läßt sich elegant ausdrücken, wann ein Polynom $f \in K[X_1, \ldots, X_d]$ die uns interessierenden Darstellungen besitzt: f besitzt genau dann die Darstellung $(\mathcal{R}_{\geq}^k)$ (siehe Seite 34), wenn $\overline{f}$ in dem von $K^{\geq 0} \cup \{\overline{p_1}, \ldots, \overline{p_n}\}$ in $K[X_1, \ldots, X_d]/(q_1, \ldots, q_m)$ erzeugten Semiring k-ter Stufe liegt, den wir etwas schlampig einfach mit $\langle K^{\geq 0}, \overline{p_1}, \ldots, \overline{p_n} \rangle_k$ bezeichnen. Folglich besitzt f genau dann die Darstellung $(\mathcal{R}_>^k)$, wenn es ein $a \in K^{>0}$ gibt mit $\overline{f} - a \in \langle K^{\geq 0}, \overline{p_1}, \ldots, \overline{p_n} \rangle_k$. Damit sind auch die Darstellungen $(\mathcal{R}_>)$ bzw. $(\mathcal{R}_{\geq})$ erfaßt, denn sie entsprechen offensichtlich den Darstellungen $(\mathcal{R}_>^0)$ bzw. $(\mathcal{R}_{\geq}^0)$.

Um die aufgeworfene Frage nach der Existenz der Darstellung $(\mathcal{R}_>)$ zu untersuchen, müßten wir also eigentlich nur Semiringe in Faktorringen $K[X_1, \ldots, X_d]/I$ betrachten, die über $K^{\geq 0}$ endlich erzeugt sind, d.h. von der Form $P = \langle K^{\geq 0} \cup Q \rangle_0$ sind mit einer endlichen Teilmenge Q von $K[X_1, \ldots, X_d]/I$. Nicht alle Semiringe P mit $K^{\geq 0} \subseteq P$ haben diese Eigenschaft (siehe Beispiel 3.22). Allerdings wollen wir später angesichts der Fragen bezüglich der Darstellung $(\mathcal{R}_>^k)$ mit $1 \leq k \in \mathbb{N}$ über $K^{\geq 0}$ endlich erzeugte Semiringe k-ter Stufe als (über $K^{\geq 0}$ i.A. nicht endlich erzeugte) Semiringe (0-ter Stufe) auffassen. Außerdem werden wir uns im entscheidenden Moment sowieso wieder auf über $K^{\geq 0}$ endlich erzeugte Semiringe P beschränken können (siehe Lemma 3.21). Übrigens könnten wir auch für das Ideal I von $K[X_1, \ldots, X_d]$ voraussetzen, daß es endlich erzeugt ist. Diese Voraussetzung ist allerdings gehaltlos, da sie nach dem Hilbertschen Basissatz sowieso immer erfüllt ist.

Nachdem wir nun zu $q_1, \ldots, q_m, p_1, \ldots, p_n \in K[X_1, \ldots, X_d]$ und $k \in \mathbb{N}$ ein passendes algebraisches Objekt gebildet haben, welches die Darstellung $(\mathcal{R}_>^k)$ bzw. $(\mathcal{R}_{\geq}^k)$ beschreibt,

nämlich den Semiring k-ter Stufe $P := \langle K^{\geq 0}, \overline{p_1}, \ldots, \overline{p_n}\rangle_k$ in $K[X_1, \ldots, X_d]/(q_1, \ldots, q_m)$, können wir gottseidank aus diesem algebraischen Objekt das geometrische Objekt

$$S := \{x \in \mathbb{R}^d \mid q_1(x) = 0, \ldots, q_m(x) = 0, p_1(x) \geq 0, \ldots, p_n(x) \geq 0\}$$

zurückgewinnen, denn es ist $S = S(P)$, wobei $S(P)$ wie unten definiert sei. Zu $S(P)$ definieren wir dann gleich noch den etwas algebraisch anmutenderen Begriff $X(P)$ und zeigen anschließend, daß $X(P)$ eigentlich dasselbe wie $S(P)$ ist. Vorher brauchen wir noch eine andere kleine Definition:

Definition 3.15. *Sei K ein Unterkörper von $\mathbb{R}$, I ein Ideal von $K[X_1, \ldots, X_d]$ und $x \in V_{\mathbb{R}}(I)$. Den eindeutig bestimmten K-Algebrenhomomorphismus*

$$K[X_1, \ldots, X_d]/I \to \mathbb{R} : \overline{X_1} \mapsto x_1, \ldots, \overline{X_d} \mapsto x_d$$

bezeichnen wir mit ε_x. Für $p \in K[X_1, \ldots, X_d]/I$ schreiben wir statt $\varepsilon_x(p)$ auch $p(x)$.

Nach dem Homomorphiesatz existiert ε_x, denn $x \in V_{\mathbb{R}}(I)$ heißt nichts anderes als, daß I im Kern des K-Algebrenhomomorphismus $K[X_1, \ldots, X_d] \to \mathbb{R} : X_1 \mapsto x_1, \ldots, X_d \mapsto x_d$ enthalten ist. Die Eindeutigkeit von ε_x ist auch klar.

Definition 3.16. *Es sei K ein Unterkörper von $\mathbb{R}$ und I ein Ideal einer Polynomalgebra $K[X_1, \ldots, X_d]$ und $P \subseteq K[X_1, \ldots, X_d]/I$. Wir definieren dann*

$$S(P) := \{x \in V_{\mathbb{R}}(I) \mid \textit{für alle } p \in P \textit{ gilt } p(x) \geq 0\} \quad \textit{und}$$
$$X(P) := \{\varphi \mid \varphi : K[X_1, \ldots, X_d]/I \to \mathbb{R} \; K\textit{-Algebrenhomomorphismus}, \; \varphi(P) \subseteq \mathbb{R}^{\geq 0}\}.$$

Lemma 3.17 (Punkte als Homomorphismen). *Sei K ein Unterkörper von $\mathbb{R}$, I ein Ideal einer Polynomalgebra $K[X_1, \ldots, X_d]$ und $P \subseteq K[X_1, \ldots, X_d]/I$. Dann ist*

$$S(P) \to X(P) : x \mapsto \varepsilon_x$$

eine Bijektion.

Beweis: Offenbar ist die Abbildung wohldefiniert. Sie ist injektiv: Seien nämlich $x, y \in S(P)$ mit $\varepsilon_x = \varepsilon_y$. Für jedes $i \in \{1, \ldots, d\}$ folgt dann $x_i = \varepsilon_x(\overline{X_i}) = \varepsilon_y(\overline{X_i}) = y_i$, also $x = y$. Sie ist auch surjektiv: Sei hierzu $\varphi \in X(P)$, also $\varphi : K[X_1, \ldots, X_d]/I \to \mathbb{R}$ ein K-Algebrenhomomorphismus mit $\varphi(P) \subseteq \mathbb{R}^{\geq 0}$. Wir setzen $x := (\varphi(\overline{X_1}), \ldots, \varphi(\overline{X_d})) \in \mathbb{R}^d$. Dann ist $x \in V_{\mathbb{R}}(I)$, da für jedes $q \in I$ gilt $q(x) = q(\varphi(\overline{X_1}), \ldots, \varphi(\overline{X_d}) = \varphi(q(\overline{X_1}, \ldots, \overline{X_d})) = \varphi(\overline{q}) = \varphi(0) = 0$. Es ist sogar $x \in S(P)$, denn für jedes $p \in P$ gilt

$$p(x) = p(\varphi(\overline{X_1}), \ldots, \varphi(\overline{X_d})) = \varphi(p(\overline{X_1}, \ldots, \overline{X_d})) = \varphi(p) \geq 0.$$

Schließlich gilt $\varepsilon_x = \varphi$, denn für jedes $f \in K[X_1, \ldots, X_d]/I$ gilt

$$\varepsilon_x(f) = f(\varphi(\overline{X_1}), \ldots, \varphi(\overline{X_d})) = \varphi(f(\overline{X_1}, \ldots, \overline{X_d})) = \varphi(f).$$

□

Der einzige Grund, warum wir die Menge $X(P)$ eingeführt haben, war zu sehen, daß am Isomorphietyp der K-Algebra $K[X_1, \ldots, X_d]/I$ mit ausgezeichneter Teilmenge P und ausgezeichnetem Element f ablesbar ist, ob $f > 0$ auf $S(P)$ gilt. Nach dem letzten Lemma und der Definition von $X(P)$ ist dies eigentlich schon klar. Trotzdem wollen wir es im Folgenden noch formal darlegen:

Lemma 3.18. *Sei K ein Unterkörper von $\mathbb{R}$, I ein Ideal von $K[X_1,\dots,X_d]$, J ein Ideal von $K[Y_1,\dots,Y_n]$ und $\Psi : K[Y_1,\dots,Y_n]/J \to K[X_1,\dots,X_d]/I$ ein K-Algebrenisomorphismus. Sei $f \in K[X_1,\dots,X_d]/I$ und $P \subseteq K[X_1,\dots,X_d]/I$. Dann gilt:*

$$\{f(x) \mid x \in S(P)\} = \{(\Psi^{-1}(f))(y) \mid y \in S(\Psi^{-1}(P))\}$$

Insbesondere gilt:

$$f > 0 \text{ auf } S(P) \iff \Psi^{-1}(f) > 0 \text{ auf } S(\Psi^{-1}(P))$$

Beweis: Da Ψ ein K-Algebrenisomorphismus ist, ist

$$X(P) \to X(\Psi^{-1}(P)) : \varphi \mapsto \varphi \circ \Psi$$

eine Bijektion. Mit Lemma 3.17 folgt nun:

$$\begin{aligned}
\{f(x) \mid x \in S(P)\} &= \{\varepsilon_x(f) \mid x \in S(P)\} \\
&= \{\varphi(f) \mid \varphi \in X(P)\} \\
&= \{(\varphi \circ \Psi)(\Psi^{-1}(f)) \mid \varphi \in X(P)\} \\
&= \{\varphi(\Psi^{-1}(f)) \mid \varphi \in X(\Psi^{-1}(P))\} \\
&= \{\varepsilon_y(\Psi^{-1}(f)) \mid y \in S(\Psi^{-1}(P))\} \\
&= \{(\Psi^{-1}(f))(y) \mid y \in S(\Psi^{-1}(P))\}
\end{aligned}$$

□

Definition 3.19. *Sei P ein Semiring im Ring R. Ein Element $r \in R$ heißt* archimedisch *bzgl. P, wenn es ein $s \in \mathbb{N}$ gibt, sodaß $s+r \in P$ und $s-r \in P$. Die Menge aller bezüglich P archimedischen Elemente notieren wir (R unterdrückend) mit $A(P)$. Der Semiring P heißt archimedisch in R, wenn $A(P) = R$.*

Die gerade gemachte Definition ist eine weitreichende Verallgemeinerung des bekannten Begriffs eines archimedisch angeordneten Körpers: Ein angeordneter Körper K ist genau dann archimedisch angeordnet, wenn der Semiring $K^{\geq 0}$ in K archimedisch ist. Hier wenden wir obige Definition jedoch an auf die Situation $R = K[X_1,\dots,X_d]/I$ mit einem Unterkörper K von $\mathbb{R}$ und einem Ideal I von $K[X_1,\dots,X_d]$. Daß ein $f \in K[X_1,\dots,X_d]/I$ archimedisch bezüglich des Semirings P von $K[X_1,\dots,X_d]/I$ ist, bedeutet dann, daß f als Funktion auf der Menge $S(P)$ beschränkt ist *und dies im Semiring P auf die einfachste vorstellbare Weise dokumentiert wird*. Demnach ist P archimedisch in $K[X_1,\dots,X_d]/I$, wenn P die Beschränktheit aller Polynomfunktionen auf $S(P)$ dokumentiert. Insbesondere müßen dann alle Polynomfunktionen auf $S(P)$ auch beschränkt sein, was genau dann der Fall ist, wenn $S(P)$ kompakt ist. Die Menge $A(P)$ der auf $S(P)$ „dokumentiert beschränkten“ Elemente von $K[X_1,\dots,X_d]/I$ ist im Falle $K^{\geq 0} \subseteq P$ eine Unteralgebra von $K[X_1,\dots,X_d]/I$, ganz analog zur Menge der auf $S(P)$ beschränkten Elemente von $K[X_1,\dots,X_d]/I$:

Lemma 3.20. *Sei K ein Unterkörper von $\mathbb{R}$, I ein Ideal von $K[X_1,\dots,X_d]$ und P ein Semiring in $K[X_1,\dots,X_d]/I$ mit $K^{\geq 0} \subseteq P$. Dann ist $A(P)$ eine Unteralgebra der K-Algebra $K[X_1,\dots,X_d]/I$.*

Beweis: Da K ein Unterkörper von $\mathbb{R}$ ist, gibt es zu jedem $a \in K$ ein $s \in \mathbb{N}$ mit $-s \leq a \leq s$, also $s \pm a \in K^{\geq 0} \subseteq P$. Also gilt $K \subseteq A(P)$. Seien $f, g \in A(P)$. Wir müßen nur noch zeigen, daß dann auch $f+g$ und fg aus $A(P)$ sind. Dazu wählen wir $s, t \in \mathbb{N}$ mit $s \pm f \in P$ und $t \pm g \in P$. Dann gilt

$$(s+t) \pm (f+g) \in P + P \subseteq P \qquad \text{und}$$

$$(st \pm fg) = \frac{1}{2}((s \pm f)(t+g) + (s \mp f)(t-g)) \in K^{\geq 0}(P \cdot P + P \cdot P) \subseteq P.$$

□

In der Situation des gerade bewiesenen Lemmas kann man insbesondere die Archimedizität von P schon dadurch nachweisen, daß man für die Elemente eines Erzeugendensystems der K-Algebra $K[X_1,\dots,X_d]/I$ (etwa für $\overline{X_1},\dots,\overline{X_d}$) nachweist, daß sie archimedisch bezüglich P sind. Die Archimedizität von P manifestiert sich also schon in endlich vielen Elementen von P (in höchstens doppelt so vielen, wie das kleinste Erzeugendensystem der K-Algebra $K[X_1,\dots,X_d]/I$ Elemente hat). Ist daher P archimedisch, so gibt es einen über $K^{\geq 0}$ endlich erzeugten Untersemiring von P, der auch schon archimedisch ist. Das nächste Lemma sagt aber noch mehr aus, nämlich daß man diesen Untersemiring sogar groß genug wählen kann, daß die durch ihn definierte Teilmenge des $\mathbb{R}^d$ nicht zu groß wird.

Lemma 3.21. *Sei K ein Unterkörper von $\mathbb{R}$, I ein Ideal von $K[X_1,\dots,X_d]$ und P ein archimedischer Semiring in $K[X_1,\dots,X_d]/I$. Dann ist $S(P)$ kompakt. Ist ferner $f \in K[X_1,\dots,X_d]/I$ mit $f > 0$ auf $S(P)$, so gibt es eine endliche Teilmenge Q von P, sodaß $\langle K^{\geq 0} \cup Q\rangle_0$ archimedisch ist und $f > 0$ auf $S(Q)$ ist.*

Beweis: Zu jedem $i \in \{1,\dots,d\}$ gibt es wegen der Archimedizität von P ein $s_i \in \mathbb{N}$ mit $s_i + \overline{X_i} \in P$ und $s_i - \overline{X_i} \in P$. Es gilt also $S(P) \subseteq \prod_{i=1}^d [-s_i, s_i]$. Damit ist $S(P)$ beschränkt und wegen der Abgeschlossenheit kompakt. Sei nun $f \in K[X_1,\dots,X_d]/I$ mit $f > 0$ auf $S(P)$. Dann gilt

$$\prod_{i=1}^d [-s_i, s_i] \cap \{x \in \mathbb{R}^d \mid f(x) \leq 0\} \cap \bigcap \{\{x \in \mathbb{R}^d \mid p(x) \geq 0\} \mid p \in P\} = \emptyset.$$

Da in diesem Schnitt $\prod_{i=1}^d [-s_i, s_i]$ kompakt ist, und die übrigen Mengen abgeschlossen sind, ist schon ein endlicher Teilschnitt leer. Wir können also eine endliche Teilmenge P' von P wählen mit

$$\prod_{i=1}^d [-s_i, s_i] \cap \{x \in \mathbb{R}^d \mid f(x) \leq 0\} \cap \bigcap \{\{x \in \mathbb{R}^d \mid p(x) \geq 0\} \mid p \in P'\} = \emptyset.$$

Mit $Q := P' \cup \{s_i + \overline{X_i} \mid i \in \{1,\dots,d\}\} \cup \{s_i - \overline{X_i} \mid i \in \{1,\dots,d\}\} \subseteq P$ heißt dies $f > 0$ auf $S(Q)$. Da der Semiring $\langle K^{\geq 0} \cup Q\rangle_0$ die Menge $K^{\geq 0}$ enthält, ist er nach Lemma 3.20 genau dann archimedisch, wenn $\overline{X_i} \in A(\langle K^{\geq 0} \cup Q\rangle_0)$ für jedes $i \in \{1,\dots,d\}$. Dies gilt wegen $s_i \pm \overline{X_i} \in Q$ für alle $i \in \{1,\dots,d\}$. □

Beispiel 3.22. Betrachte den Semiring $P := \langle \mathbb{R}^{\geq 0} \cup E\rangle_0$ mit $E := \{a + X \mid 0 < a \in \mathbb{R}\} \cup \{a - X \mid 0 < a \in \mathbb{R}\}$ in $\mathbb{R}[X]$. Es liegt also die Situation des obigen Lemmas vor mit $K = \mathbb{R}$, $d = 1$ und dem Nullideal I. Der Semiring P ist archimedisch, da $\mathbb{R}^{\geq 0} \subseteq P$ und $1 \pm X \in P$. Jeder über $\mathbb{R}^{\geq 0}$ endlich erzeugte Untersemiring P' von P ist sogar von der Form $P' = \langle \mathbb{R}^{\geq 0} \cup E'\rangle_0$ mit einer endlichen Teilmenge E' von E (dies ist ein Standardargument, denn $\bigcup\{\langle \mathbb{R}^{\geq 0} \cup E_0\rangle_0 \mid E_0 \subseteq E \text{ endlich}\}$ ist offenbar ein Semiring und damit gleich P). Daher gilt für diese P', daß $S(P')$ ein nichtleeres Inneres hat im Gegensatz zu $S(P) = \{0\}$. Hier ist auch ohne das obige Lemma ersichtlich, daß es zu jedem $f \in \mathbb{R}[X]$ mit $f > 0$ auf $S(P)$ (also $f(0) > 0$) eine endliche Teilmenge E' von E gibt, sodaß $P' := \langle \mathbb{R}^{\geq 0} \cup E'\rangle_0$ archimedisch ist und $f > 0$ auf $S(E')$ ist. Es muß E' natürlich von f abhängen dürfen.

Nach dem letzten Lemma können wir uns beim Beweis des angestrebten archimedischen Positivstellensatzes 3.24 für die nichttriviale Implikation beschränken auf über $K^{\geq 0}$ endlich erzeugte archimedische Semiringe P. Diese werden im nächsten Lemma charakterisiert. Erstmals werden jetzt die Parallelen zum letzten Abschnitt deutlich.

Lemma 3.23. *Sei K ein Unterkörper von $\mathbb{R}$, I ein Ideal von $K[X_1,\dots,X_d]$ und Q eine endliche Teilmenge von $K[X_1,\dots,X_d]/I$. Genau dann ist $\langle K^{\geq 0}\cup Q\rangle_0$ archimedisch, wenn es $p_1,\dots,p_n\in K[X_1,\dots,X_d]/I$ gibt mit folgenden Eigenschaften:*

(i) $p_1+\cdots+p_n=1$

(ii) $\langle K^{\geq 0}\cup Q\rangle_0=\langle K^{\geq 0},p_1,\dots,p_n\rangle_0$

(iii) $K[X_1,\dots,X_d]/I=K[p_1,\dots,p_n]$

Beweis: Sei $\langle K^{\geq 0}\cup Q\rangle_0$ archimedisch. Dann gibt es ein $s\in\mathbb{N}$ mit $s-\sum_{q\in Q}q\in\langle K^{\geq 0}\cup Q\rangle_0$. Wir bezeichnen die verschiedenen $\frac{q}{s}$ mit $q\in Q$ durch $p_1,\dots,p_{n-1}$ und setzen

$$p_n=1-(p_1+\cdots+p_{n-1})\in\langle K^{\geq 0}\cup Q\rangle_0.$$

Offenbar gelten (i) und (ii). Eigenschaft (iii) ergibt sich aus

$$K[X_1,\dots,X_d]/I=A(\langle K^{\geq 0}\cup Q\rangle_0)\subseteq K[p_1,\dots,p_n]\subseteq K[X_1,\dots,X_d]/I.$$

Es gebe umgekehrt $p_1,\dots,p_n$ mit (i), (ii) und (iii). Für jedes $i\in\{1,\dots,n\}$ ist dann $p_i\in\langle K^{\geq 0}\cup Q\rangle_0$ und $1-p_i=\sum_{j\neq i}p_j\in\langle K^{\geq 0}\cup Q\rangle_0$, also $p_i\in A(\langle K^{\geq 0}\cup Q\rangle_0)$. Nach Lemma 3.20 ist $A(\langle K^{\geq 0}\cup Q\rangle_0)$ eine K-Unteralgebra von $K[X_1,\dots,X_d]/I$. Es folgt

$$K[p_1,\dots,p_n]\subseteq A(\langle K^{\geq 0}\cup Q\rangle_0).$$

Gemäß (iii) ist $K[p_1,\dots,p_n]=K[X_1,\dots,X_d]/I$, woraus $K[X_1,\dots,X_d]/I=A(\langle K^{\geq 0}\cup Q\rangle_0)$, also die Archimedizität von $\langle K^{\geq 0}\cup Q\rangle_0$ folgt. □

Durch Zusammenklauben der bisherigen Resultate folgt nun das Hauptresultat dieses Kapitels, der archimedische Positivstellensatz. Wie in der Einleitung zu diesem Kapitel bereits geschehen, kann man seinen Gehalt (für über $K^{\geq 0}$ endlich erzeugtes P) zusammen mit Lemma 3.20 unter Vermeidung algebraischer Terminologie so formulieren:

Wenn es Zahlen $s_1,\dots,s_d\in K^{\geq 0}$ gibt, sodaß es für die $2d$ Polynome $s_1+X_1, s_1-X_1,\dots,s_d+X_d, s_d-X_d$ eine Darstellung $(\mathcal{R}_{\geq})$ gibt (insbesondere ist dann $S\subseteq\prod_{i=1}^d[-s_i,s_i]$ kompakt), so gibt es für jedes $f\in K[X_1,\dots,X_d]$ mit $f>0$ auf S eine Darstellung $(\mathcal{R}_{>})$.

Man beachte, daß wir hier nicht $s_1,\dots,s_d\in\mathbb{N}$ fordern müßen. Die Existenz entsprechender $s_1,\dots,s_d\in K^{\geq 0}$ impliziert schon die Existenz entsprechender $s_1,\dots,s_d\in\mathbb{N}$. Obige populäre Formulierung ist etwas unflexibel angesichts der Tatsache, daß man $X_1,\dots,X_d$ durch irgendwelche andere Polynome $f_1,\dots,f_l\in K[X_1,\dots,X_d]$ mit

$$K[\overline{f_1},\dots,\overline{f_l}]=K[X_1,\dots,X_d]/(q_1,\dots,q_m)$$

ersetzen könnte.

Der archimedische Positivstellensatz, den wir gleich etwas eleganter formulieren werden, dürfte im Prinzip bekannt sein, seitdem der Darstellungssatz von Kadison-Dubois bekannt ist (siehe Kapitel 4), denn man erkennt ihn leicht als Spezialfall von diesem (siehe Abschnitt 4.1). Explizit als Positivstellensatz formuliert findet man ihn in [Wör]. In [Ha2] beobachtet Handelman unabhängig davon für $K=\mathbb{R}$, $I=(0)$ und über $K^{\geq 0}$ endlich erzeugtes P ebenfalls dieses Resultat. Der hier vorliegende Beweis ist neu und hat ganz im Gegensatz zu den bisherigen erstaunliche algorithmische Konsequenzen, denen wir uns dann in den nächsten Abschnitten zuwenden werden.

Satz 3.24 (archimedischer Positivstellensatz). *Sei K ein Unterkörper von $\mathbb{R}$, I ein Ideal in $K[X_1,\dots,X_d]$ und P ein archimedischer Semiring in $K[X_1,\dots,X_d]/I$ mit $K^{\geq 0}\subseteq P$. Dann gilt für jedes $f\in K[X_1,\dots,X_d]/I$:*

$$f>0 \text{ auf } S(P) \iff \text{es gibt } a\in K^{>0} \text{ mit } f-a\in P$$

Beweis: Die Implikation „$\Longleftarrow$“ ist trivial. Um „ $\Longrightarrow$ “ zu zeigen, sei $f \in K[X_1, \ldots, X_d]/I$ mit $f > 0$ auf $S(P)$. Nach Lemma 3.21 und Lemma 3.23 können wir o.B.d.A. davon ausgehen, daß P von der Form $P = \langle K^{\geq 0}, p_1, \ldots, p_n \rangle_0$ ist mit $p_1, \ldots, p_n \in K[X_1, \ldots, X_d]/I$, für die $p_1 + \cdots + p_n = 1$ und $K[X_1, \ldots, X_d]/I = K[p_1, \ldots, p_n]$ gilt. Dann ist der K-Algebrenhomomorphismus

$$\psi : K[Y_1, \ldots, Y_n] \to K[X_1, \ldots, X_d]/I : Y_1 \mapsto p_1, \ldots, Y_n \mapsto p_n$$

surjektiv und in seinem Kern J ist das Polynom $Y_1 + \cdots + Y_n - 1$ enthalten. Durch ψ wird ein K-Algebrenisomorphismus

$$\Psi : K[Y_1, \ldots, Y_n]/J \to K[X_1, \ldots, X_d]/I : \overline{Y_1} \mapsto p_1, \ldots, \overline{Y_n} \mapsto p_n$$

induziert. Nach Lemma 3.18 gilt $\Psi^{-1}(f) > 0$ auf $S(\Psi^{-1}(P))$. Nun gilt

$$\Psi^{-1}(P) = \langle K^{\geq 0}, \overline{Y_1}, \ldots, \overline{Y_n} \rangle_0, \quad \text{also} \quad S(\Psi^{-1}(P)) = V_{\mathbb{R}}(J) \cap (\mathbb{R}^{\geq 0})^n.$$

Es ist also $\Psi^{-1}(f) > 0$ auf $V_{\mathbb{R}}(J) \cap (\mathbb{R}^{\geq 0})^n$. Wegen $Y_1 + \cdots + Y_n - 1 \in J$ enthält nach Lemma 3.13 die Restklasse $\Psi^{-1}(f)$ eine Positivform $F \in K[Y_1, \ldots, Y_n]$, d.h. es gilt $\Psi^{-1}(f) = \overline{F} = F(\overline{Y_1}, \ldots, \overline{Y_n})$. Wähle ein $a \in K^{>0}$ so klein, daß die Form $G := F - a(Y_1 + \cdots + Y_n)^{\deg F} \in K[Y_1, \ldots, Y_n]$ keine negativen Koeffizienten besitzt. Dann gilt

$$\Psi^{-1}(f) = G(\overline{Y_1}, \ldots, \overline{Y_n}) + a(\underbrace{\overline{Y_1 + \cdots + Y_n}}_{=1})^{\deg F}, \text{ also}$$

$$\Psi^{-1}(f) - a = G(\overline{Y_1}, \ldots, \overline{Y_n}) \in \langle K^{\geq 0}, \overline{Y_1}, \ldots, \overline{Y_n} \rangle_0 = \Psi^{-1}(P),$$

d.h. $f - a = \Psi(\Psi^{-1}(f) - a) \in P$. □

Die Argumentation am Schluß des gerade beendeten Beweises hätten wir verkürzen können. Es hätte nämlich gereicht, die scheinbar schwächere Implikation

$$(\star) \qquad f > 0 \text{ auf } S(P) \implies f \in P$$

zu beweisen. Da $S(P)$ wegen der Archimedizität von P kompakt ist, gibt es nämlich zu jedem f mit $f > 0$ auf $S(P)$ ein $a \in K^{>0}$, sodaß auch noch $f - a > 0$ auf $S(P)$ ist. Die Implikation $(\star)$ angewandt auf $f - a$ statt f liefert dann $f - a \in P$. Im Hinblick darauf, wie man ein geeignetes $a \in K^{>0}$ tatsächlich algorithmisch bestimmt, haben wir allerdings die vorliegende Beweisvariante vorgezogen. Leider gilt nicht die Implikation

$$f \geq 0 \text{ auf } S(P) \implies f \in P,$$

denn z.B. ist nach Lemma 3.23 der Semiring $P := \langle \mathbb{R}^{\geq 0}, X, 1 - X \rangle_0$ archimedisch in $\mathbb{R}[X]$, aber offensichtlich ist $f := (X - \frac{1}{2})^2 \notin P$, obwohl $f \geq 0$ auf $S(P) = [0, 1]$ ist.

Die Aussage des archimedischen Positivstellensatzes gilt auch dann manchmal, wenn P die Voraussetzung verletzt, archimedisch zu sein. Ein Beispiel hierfür erhält man, wenn P gerade aus den Summen von Quadraten in $\mathbb{R}[X]$ besteht (also $K = \mathbb{R}$, $d = 1$, I das Nullideal und $p = \langle \emptyset \rangle_1$). Es ist dann $S(P) = \mathbb{R}$, also $S(P)$ nicht kompakt und daher P nicht archimedisch. Trotzdem gilt die Aussage des Satzes, denn wenn $f > 0$ auf $\mathbb{R}$ ist, so gibt es ein $a \in \mathbb{R}^{>0}$ mit $f - a \geq 0$ auf $\mathbb{R}$ und zu Beginn des Kapitels 2 wurde gezeigt, daß dann $f - a$ eine Summe von zwei Quadraten in $\mathbb{R}[X]$ ist.

Wenn allerdings $S(P)$ kompakt ist, $K^{\geq 0} \subseteq P$ und die Aussage des Satzes gelten soll, so muß P archimedisch sein. Für jedes $f \in K[X_1, \ldots, X_d]/I$ gibt es dann nämlich wegen der Beschränktheit von f auf $S(P)$ ein $s \in \mathbb{N}$, sodaß $s \pm f > 0$ auf $S(P)$, also $s \pm f \in P$ nach der Aussage des Satzes und wegen $K^{\geq 0} \subseteq P$.

Eine schwierigere Frage scheint zu sein, ob es über $K^{\geq 0}$ endlich erzeugte Semiringe P gibt, die nicht archimedisch sind und für die der Satz trotzdem gilt.

Die nächste Bemerkung zeigt, daß die durch den archimedischen Positivstellensatz garantierte Darstellung ($\mathcal{R}_>$) (sogar ($\mathcal{R}_\geq$)) unter gewissen, leicht zu erfüllenden Umständen beliebig groß werden muß, wenn f auf S genügend kleine positive Werte annimmt. Dies ist ganz analog zum entsprechenden Resultat beim Pólyaschen Satz (siehe Bemerkung 3.8) und zeigt, daß der Satz von Pólya vom komplexitätstheoretischen Standpunkt her kein zu großes Geschütz war, um den archimedischen Positivstellensatz zu beweisen.

Bemerkung 3.25. Seien $p_1, \ldots, p_n \in \mathbb{R}[X_1, \ldots, X_d]$. Für jedes $f \in \mathbb{R}[X_1, \ldots, X_d]$, für das es ein $N \in \mathbb{N}$ gibt, sodaß f eine Darstellung der Form

$$f = \sum_{l_1+\cdots+l_n \leq N} \lambda_l p_1^{l_1} \cdots p_n^{l_n}$$

mit $\lambda_l \in \mathbb{R}^{\geq 0}$ hat, bezeichnen wir das kleinste solche N als Komplexität von f (bzgl. $p_1, \ldots, p_n$). Gibt es kein solches N, so legen wir die Komplexität von f als ∞ fest. Diese Komplexität verhält sich im folgenden Sinn bösartig:

Sei E eine endliche Teilmenge von $\mathbb{N}^d$. Sei $f = \sum_{e\in E} c_e X_1^{e_1} \cdots X_d^{e_d} \neq 0$ mit $c_e \in \mathbb{R}$ für alle $e \in E$. Es gebe ein $\xi \in \mathbb{R}^d$ mit $p_1(\xi) > 0, \ldots, p_n(\xi) > 0$ und $f(\xi) = 0$. Strebt dann $(a_e)_{e\in E}$ in $\mathbb{R}^E$ gegen $(c_e)_{e\in E}$, so strebt die Komplexität von $\sum_{e\in E} a_e X_1^{e_1} \cdots X_d^{e_d}$ gegen ∞.

Dafür geben wir zwei Beweise: Einen analytischen (vgl. VI, Problem 48 in [PS]) und einen modelltheoretischen (ähnlich zur Bemerkung 3.8). Zunächst der analytische Beweis: Sei $N \in \mathbb{N}$. Wir setzen

$$L := \{(l_1, \ldots, l_n) \in \mathbb{N}^n \mid l_1 + \cdots + l_n \leq N\}.$$

Angenommen für jedes $k \in \mathbb{N}$ gibt es ein $(a_{ke})_{e\in E}$ mit

$$\|(a_{ke})_{e\in E} - (c_e)_{e\in E}\| < \frac{1}{k+1} \text{ in } \mathbb{R}^E,$$

sodaß $\sum_{e\in E} a_{ke} X_1^{e_1} \cdots X_d^{e_d}$ Komplexität $\leq N$ hat. Wähle dann zu jedem $k \in \mathbb{N}$ eine Familie $(\lambda_{kl})_{l\in L}$ von nichtnegativen reellen Zahlen mit

$$(\star) \qquad \sum_{e\in E} a_{ke} X_1^{e_1} \cdots X_d^{e_d} = \sum_{l\in L} \lambda_{kl} p_1^{l_1} \cdots p_n^{l_n} \text{ für alle } k \in \mathbb{N}.$$

Wertet man diese Gleichung für jedes $k \in \mathbb{N}$ an der Stelle ξ aus, so sieht man, daß wegen $p_1(\xi) > 0, \ldots, p_n(\xi) > 0$ die Folge $((\lambda_{kl})_{l\in L})_{k\in\mathbb{N}}$ in R^L beschränkt ist und daher nach dem Satz von Bolzano-Weierstraß eine konvergente Teilfolge besitzt. Indem wir zu einer solchen übergehen und auch bei $((a_{ke})_{e\in E})_{k\in\mathbb{N}}$ zur entsprechenden Teilfolge übergehen, können wir o.B.d.A. annehmen, daß $((\lambda_{kl})_{l\in L})_{k\in\mathbb{N}}$ in $\mathbb{R}^L$ gegen einen Punkt $(\lambda_l)_{l\in L}$ konvergiert. Da außerdem $((a_{ke})_{e\in E})_{k\in\mathbb{N}}$ in R^E gegen $(c_e)_{e\in E}$ konvergiert, erhält man durch koeffizientenweisen Grenzübergang $k \to \infty$ in $(\star)$, daß gilt

$$f = \sum_{e\in E} c_e X_1^{e_1} \cdots X_d^{e_d} = \sum_{l\in L} \lambda_l p_1^{l_1} \cdots p_n^{l_n}.$$

Hierbei gilt $\lambda_l = \lim_{k\to\infty} \lambda_{kl} \geq 0$. Durch Auswerten dieser Gleichung an der Stelle ξ erhält man wegen $p_1(\xi) > 0, \ldots, p_n(\xi) > 0$ und $f(\xi) = 0$ also $\lambda_l = 0$ für alle $l \in L$. Dies impliziert $f = 0$ im Widerspruch zur Voraussetzung.

Nun bringen wir den modelltheoretischen Beweis: Wie in Bemerkung 3.8 definieren wir vorweg: Ein Element a eines angeordneten Körpers heißt *endlich* (bzw. *infinitesimal*), wenn der Betrag von a kleiner als eine (bzw. jede) positive rationale Zahl ist. K heißt *archimedisch*, wenn K nur endliche Elemente enthält. Die Menge K_{fin} der endlichen Elemente von K bildet in jedem Fall einen Unterring von K, und die Menge K_{inf} der infinitesimalen Elemente von K bildet ein Primideal von K_{fin}.

Sei nun $N \in \mathbb{N}$. Wir setzen $L := \{(l_1, \ldots, l_n) \in \mathbb{N}^n \mid l_1 + \cdots + l_n \leq N\}$. Zu zeigen ist, daß es eine Umgebung von $(c_e)_{e\in E}$ in $\mathbb{R}^E$ gibt, sodaß für alle $(a_e)_{e\in E}$ aus dieser Umgebung die Komplexität von $\sum_{e\in E} a_e X_1^{e_1} \cdots X_d^{e_d}$ größer als N ist. Es reicht also zu zeigen, daß folgende Bedingung $(\star_K)$ für $K = \mathbb{R}$ gilt:

$(\star_K)$ Es gibt eine Umgebung U von $(c_e)_{e\in E}$ in K^E (bezüglich der Produkttopologie, die von der Ordnungstopologie auf K stammt), sodaß es für kein $(a_e)_{e\in E} \in U$ ein $(\lambda_l)_{l\in L} \in (K^{\geq 0})^L$ gibt mit $\sum_{e\in E} a_e X_1^{e_1} \cdots X_d^{e_d} = \sum_{l\in L} \lambda_l p_1^{l_1} \cdots p_n^{l_n}$.

Nun können wir $(\star_K)$ zumindest für alle nicht archimedischen Erweiterungen K geordneter Körper von $\mathbb{R}$ zeigen, indem wir U einfach „infinitesimal klein“ wählen, also etwa $U = \prod_{e\in E}]c_e - \varepsilon, c_e + \varepsilon[$ setzen mit einem $0 < \varepsilon \in K_{\text{inf}}$ (ein solches existiert). Sei dann nämlich $(a_e)_{e\in E} \in U$. Mit c_e ist auch a_e für jedes $e \in E$ endlich. Angenommen $(\lambda_l)_{l\in L} \in (K^{\geq 0})^L$ mit $\sum_{e\in E} a_e X_1^{e_1} \cdots X_d^{e_d} = \sum_{l\in L} \lambda_l p_1^{l_1} \cdots p_n^{l_n}$. Weil a_e für jedes $e \in E$ kongruent zu c_e modulo dem Ideal K_{inf} von K_{fin} ist, gilt dann in K_{fin}, daß $\sum_{l\in L} \lambda_l (p_1(\xi))^{l_1} \cdots (p_n(\xi))^{l_n} = \sum_{e\in E} a_e \xi_1^{e_1} \cdots \xi_d^{e_d} \equiv f(\xi) = 0$ modulo K_{inf}. Also ist $\sum_{l\in L} \lambda_l (p_1(\xi))^{l_1} \cdots (p_n(\xi))^{l_n} \in K_{\text{inf}}$. Da in dieser Summe jeder Summand ≥ 0 ist, ist jeder Summand aus K_{inf}. Da K_{inf} ein Primideal von K_{fin} ist und $p_i(\xi) \in K_{\text{fin}} \setminus K_{\text{inf}}$ für jedes $i \in \{1, \ldots, n\}$, folgt $\lambda_l \in K_{\text{inf}}$ für alle $l \in L$. Dies impliziert aber, daß $\sum_{e\in E} a_e x_1^{e_1} \cdots x_d^{e_d} = \sum_{l\in L} \lambda_l (p_1(x))^{l_1} \cdots (p_n(x))^{l_n} \in K_{\text{inf}}$ für alle $x \in K_{\text{fin}}^d$. Für alle diese x gilt auch $f(x) \equiv \sum_{e\in E} a_e x_1^{e_1} \cdots x_d^{e_d}$ (modulo K_{inf}) und damit $f(x) \in K_{\text{inf}}$. Insbesondere folgt für alle $x \in \mathbb{R}^d$, daß $f(x) \in K_{\text{inf}} \cap \mathbb{R} = \{0\}$. Dies impliziert bekanntlich $f = 0$ im Widerspruch zur Voraussetzung.

Die gerade bewiesene Tatsache, daß $(\star_K)$ für alle nicht archimedischen Erweiterungen K geordneter Körper von $\mathbb{R}$ gilt, drücken wir nun mit Hilfe des semantischen Folgerungsbegriffs $\models$ der Logik erster Stufe aus. Wir betrachten dazu eine Signatur mit Konstantensymbolen für jede reelle Zahl, Funktionssymbolen für $+$, $-$ und $\cdot$, sowie einem Relationssymbol für $<$. Offenbar läßt sich in dieser Signatur ein Satz φ erster Stufe formulieren, der in jeder Erweiterung K angeordneter Körper von $\mathbb{R}$ (bei natürlicher Interpretation aller Symbole) nichts anderes als $(\star_K)$ besagt. Dann gilt $\{n < c \mid n \in \mathbb{N}\} \cup D \cup \Sigma \models \varphi$, wobei c ein neues Konstantensymbol, D das Basisdiagramm (oft auch nur Diagramm genannt) von $\mathbb{R}$ und Σ ein Axiomensystem für angeordnete Körper sei. Nach dem Endlichkeitssatz der Logik erster Stufe gibt es eine endliche Teilmenge Φ von $\{n < c \mid n \in \mathbb{N}\} \cup D \cup \Sigma$ mit $\Phi \models \varphi$. Durch genügend große Interpretation von c kann man $\mathbb{R}$ zu einem Modell von Φ expandieren. Also gilt φ in $\mathbb{R}$. Das heißt, es gilt $(\star_K)$ für $K = \mathbb{R}$, und die Aussage ist bewiesen.

Wir illustrieren die Aussage mit einem Beispiel: Man betrachte $p_1 := 1 + X$ und $p_2 := 1 - X$. Das von $a \in \mathbb{R}$ abhängige Polynom $X^2 + a$ hat für jedes $a > 0$ nach dem archimedischen Positivstellensatz 3.24 endliche Komplexität bezüglich p_1, p_2. Strebt a gegen 0, so strebt diese allerdings gegen ∞.

Wir beenden diese Betrachtungen mit der Bemerkung, daß wir in der hergeleiteten Aussage für kein einziges $i \in \{1, \ldots, n\}$ die Voraussetzung $p_i(\xi) > 0$ zu $p_i(\xi) \geq 0$ abschwächen können. Gilt nämlich $p_i(\xi) = 0$, so hat das von $a \in \mathbb{R}$ abhängige Polynom $p_i + a$ für jedes $a > 0$ eine Komplexität ≤ 1, obwohl es für $a = 0$ eine Nullstelle in ξ hat.

Aus obiger Bemerkung folgt auch, daß der archimedische Positivstellensatz 3.24 nicht für beliebige reell abgeschlossene Körper (siehe [Lor], [Jac], [We2]) anstelle von $\mathbb{R}$ gelten kann. Sonst bekäme man mit dem Endlichkeitssatz der Logik erster Stufe nämlich obere Schranken für die Komplexität von Polynomen $f \in \mathbb{R}[X_1, \ldots, X_d]$ mit $f > 0$ auf $S(P)$, welche nur vom Grad von f und von d abhängten (siehe [Pr3], vgl. auch Satz 5.5.4 in [We2]). Dies widerspricht offenbar obiger Bemerkung.

3.3 Umsetzung in einen Algorithmus

Es sei K ein Unterkörper von $\mathbb{R}$ und es seien $q_1, \ldots, q_m, p_1, \ldots, p_n, f \in K[X_1, \ldots, X_d]$ mit $f > 0$ auf der Menge

$$S := \{x \in \mathbb{R}^d \mid q_1(x) = 0, \ldots, q_m(x) = 0, p_1(x) \geq 0, \ldots, p_n(x) \geq 0\}.$$

Wir wollen nach Möglichkeit eine Darstellung ($\mathcal{R}_>$) (siehe Seite 33) von f berechnen. Hierfür wollen wir ein Verfahren aus dem Beweis des archimedischen Positivstellensatzes 3.24 extrahieren. Der archimedische Positivstellensatz besagt ja angewandt auf $I := (q_1, \ldots, q_m)$ und $P := \langle K^{\geq 0}, \overline{p_1}, \ldots, \overline{p_n}\rangle_0 \subseteq K[X_1, \ldots, X_d]/I$, daß eine Darstellung ($\mathcal{R}_>$) von f wenigstens *existiert* unter der Voraussetzung, daß P archimedisch ist. Natürlich sollten wir hier die Archimedizität von P auch voraussetzen. Leider brauchen wir sogar eine effektive Version dieser Voraussetzung: Unser Algorithmus wird als Eingabe einen Nachweis der Archimedizität von P in gewisser Form brauchen. Hierzu folgende Definitionen:

Definition 3.26. *Sei K ein Unterkörper von $\mathbb{R}$, I ein Ideal von $K[X_1, \ldots, X_d]$ und seien $p_1, \ldots, p_n, f \in K[X_1, \ldots, X_d]$. Eine* $\{p_1, \ldots, p_n\}$-Darstellung modulo I von f *ist eine Darstellung (wir verzichten hier auf die offensichtliche Formalisierung dieses Begriffs) eines zu f modulo I kongruenten Polynoms in der Form*

$$\sum_e a_e p_1^{e_1} \cdots p_n^{e_n} \qquad (a_e \in K^{\geq 0}).$$

Genau dann besitzt f eine $\{p_1, \ldots, p_n\}$-Darstellung modulo I, wenn $\overline{f} \in P$ gilt. Hat man $\{p_1, \ldots, p_n\}$-Darstellungen ϱ_f bzw. ϱ_g modulo I von f bzw. von g, so lassen sich in offensichtlicher Weise ebensolche $\varrho_f + \varrho_g$ bzw. $\varrho_f \varrho_g$ von $f + g$ bzw. von fg berechnen. Hat man ein $a \in K^{\geq 0}$ und eine $\{p_1, \ldots, p_n\}$-Darstellung ϱ_f modulo I von f, so läßt sich in offensichtlicher Weise eine ebensolche $a\varrho_f$ von af berechnen.

Definition 3.27. *Sei K ein Unterkörper von $\mathbb{R}$, I ein Ideal von $K[X_1, \ldots, X_d]$ und seien $p_1, \ldots, p_n, f \in K[X_1, \ldots, X_d]$. Ein* $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis modulo I für f *ist ein Tripel $(s, \varrho_+, \varrho_-)$ bestehend aus einer Zahl $s \in K^{\geq 0}$ und $\{p_1, \ldots, p_n\}$-Darstellungen ϱ_+ bzw. ϱ_- modulo I von $s + f$ bzw. $s - f$.*

Genau dann besitzt f einen $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis modulo I, wenn $f \in A(P) \subseteq K[X_1, \ldots, X_d]/I$ (dabei nutzen wir $K^{\geq 0} \subseteq P$ aus). Wir haben in obiger Definition nicht $s \in \mathbb{N}$ gefordert, um für praktische Zwecke die Definition flexibel zu halten. Nun können wir das Format definieren, in dem unser Algorithmus einen Nachweis der Archimedizität von P als Eingabe verlangen wird:

Definition 3.28. *Sei K ein Unterkörper von $\mathbb{R}$, I ein Ideal von $K[X_1, \ldots, X_d]$ und seien $p_1, \ldots, p_n \in K[X_1, \ldots, X_d]$. Ein* $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis modulo I *ist ein Tupel $(f_1, \alpha_1, \ldots, f_l, \alpha_l)$ mit $f_1, \ldots, f_l \in K[X_1, \ldots, X_d]$, sodaß $K[X_1, \ldots, X_d]/I = K[\overline{f_1}, \ldots, \overline{f_l}]$ und α_i ein $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis modulo I für f_i ist für jedes $i \in \{1, \ldots, l\}$.*

Wenn P archimedisch ist, so existiert natürlich ein $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis $(f_1, \alpha_1, \ldots, f_l, \alpha_l)$ modulo I. Dabei kann man $f_1, \ldots, f_l$ sogar beliebig wählen mit der Eigenschaft $K[X_1, \ldots, X_d]/I = K[\overline{f_1}, \ldots, \overline{f_l}]$. Wenn umgekehrt ein $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis modulo I existiert, so folgt aus Lemma 3.20, daß P tatsächlich archimedisch ist.

Als Vorbereitung brauchen wir noch einen Algorithmus, der für einen Algebrenhomomorphismus von einer Polynomalgebra in eine Faktoralgebra einer Polynomalgebra den Kern und Urbilder berechnet. Man muß mit Gröbnerbasen (siehe [BW],[Mis],[We1]) vertraut sein,

um zu verstehen, wie dieser Algorithmus arbeitet. Der Teil des Algorithmus, der den Kern berechnet, ist für den Fall, daß der Wertebereich des Homomorphismus keine Faktoralgebra, sondern wieder eine Polynomalgebra ist, als Algorithmus von Spear bekannt. Für den selben Fall ist der Teil, der die Urbilder berechnet, in [BW] (als Algorithmus SUBRING-MEMTEST) zu finden. Es handelt sich um eine geradlinige Verallgemeinerung dieser beiden Algorithmen.

Satz 3.29. *Sei K ein Körper. Der folgende Algorithmus (modulo Rechnen im Körper K) berechnet nach Eingabe von $q_1, \ldots, q_m \in K[X_1, \ldots, X_d]$ und $p_1, \ldots, p_n \in K[Y_1, \ldots, Y_n]$ ein endliches Erzeugendensystem des Kerns des K-Algebrenhomomorphismus*

$$\psi : K[Y_1, \ldots, Y_n] \to K[X_1, \ldots, X_d]/(q_1, \ldots, q_m) : Y_1 \mapsto \overline{p_1}, \ldots, Y_n \mapsto \overline{p_n}.$$

Gibt man zusätzlich noch ein $f \in K[X_1, \ldots, X_d]$ ein, so gibt er auf die Frage, ob $\overline{f}$ ein Urbild g unter ψ besitzt, die richtige Antwort „ja“ oder „nein“ aus und berechnet bei positiver Antwort ein solches g, also ein $g \in K[Y_1, \ldots, Y_n]$ mit $g(\overline{p_1}, \ldots, \overline{p_n}) = \overline{f}$.

(1) Berechne eine Gröbnerbasis G des in $K[X_1, \ldots, X_d, Y_1, \ldots, Y_n]$ von den Polynomen $p_1 - Y_1, \ldots, p_n - Y_n, q_1, \ldots, q_m$ erzeugten Ideals bezüglich einer Termordnung $\leq$ mit der folgenden Eigenschaft: Jeder Term, in dem ein X_i vorkommt ist, ist größer als jeder Term, in dem nur Y_i vorkommen. (Man kann also zum Beispiel die lexikographische Termordnung $\leq$ mit $X_1 > \cdots > X_d > Y_1 > \cdots > Y_n$ verwenden.)

(2) Es ist $G \cap K[Y_1, \ldots, Y_n]$ ein endliches Erzeugendensystem (sogar eine Gröbnerbasis) bezüglich der Einschränkung der Termordnung $\leq$ auf die Terme in $Y_1, \ldots, Y_n$) des Kerns von ψ. Gebe dieses aus.

(3) Falls zusätzlich ein $f \in K[X_1, \ldots, X_d]$ eingegeben wurde: Reduziere f modulo G bezüglich $\leq$ auf eine Normalform $g \in K[X_1, \ldots, X_d, Y_1, \ldots, Y_n]$. Falls $g \in K[Y_1, \ldots, Y_n]$ antworte „ja“ und gebe g aus. Sonst antworte „nein“.

Beweis: Es bezeichne I das von $q_1, \ldots, q_m$ erzeugte Ideal in $K[X_1, \ldots, X_d]$ und L das von $p_1 - Y_1, \ldots, p_n - Y_n, q_1, \ldots, q_m$ erzeugte Ideal in $K[X_1, \ldots, X_d, Y_1, \ldots, Y_n]$. Wir zeigen zunächst folgende Hilfsbehauptung:

$$(\star) \qquad L \cap K[X_1, \ldots, X_d] = I$$

Dazu wählen wir eine weitere Termordnung $\leq'$ in umgekehrter Manier wie $\leq$: Diesmal soll jeder Term, in dem ein Y_i vorkommt, größer sein als jeder Term, in dem nur X_i vorkommen. Sodann wählen wir eine Gröbnerbasis G' von I bezüglich $\leq'$. Nun ist

$$G'' := G' \cup \{Y_1 - p_1, \ldots, Y_n - p_n\}$$

eine Gröbnerbasis von L bezüglich $\leq'$.

Dies begründet man wie folgt: Sicher erzeugt G'' das Ideal L. Außerdem läßt sich jedes aus zwei verschiedenen Polynomen von G'' gebildete S-Polynom modulo G'' (bezüglich $\leq'$) zu 0 reduzieren. Sind nämlich beide Polynome aus G', so läßt sich das S-Polynom ja sogar modulo G' zu 0 reduzieren. Andernfalls haben die beiden Polynome disjunkte höchste Terme (bezüglich $\leq'$) und das S-Polynom läßt sich nach dem bekannten „ersten Kriterium“ von Buchberger (siehe etwa [BW]) modulo G'' zu 0 reduzieren. Folglich ist G'' nach dem S-Polynom-Kriterium eine Gröbnerbasis von L bezüglich $\leq'$.

Wir zeigen nun die Gleichung $(\star)$. Nur die Inklusion „$\subseteq$“ ist fraglich. Sei also $h \in L \cap K[X_1, \ldots, X_d]$. Wegen $h \in L$ reduziert sich h modulo G'' zu 0 (bezüglich $\leq'$). Da in h kein Y_i auftaucht, fließt nach Wahl von $\leq'$ bei diesem Reduktionsprozeß auch kein Y_i ein. Dann können aber nur diejenigen Polynome aus G'' zur Reduktion beigetragen haben, in denen kein Y_i auftaucht, also die Polynome aus G'. Also reduziert sich h sogar modulo G' zu 0, woraus $h \in I$ folgt.

Damit ist $(\star)$ gezeigt. Wir verifizieren nun (2), indem wir erstens zeigen, daß $G \cap K[Y_1, \ldots, Y_n]$ im Kern von ψ enthalten ist, und zweitens, daß sich jedes h aus dem Kern von ψ modulo $G \cap K[Y_1, \ldots, Y_n]$ bezüglich $\leq$ auf 0 reduzieren läßt. Für die erste Behauptung sei $h \in G \cap K[Y_1, \ldots, Y_n]$. Dann ist $h(p_1, \ldots, p_n) \equiv h \equiv 0$ modulo L, also $h(p_1, \ldots, p_n) \in L \cap K[X_1, \ldots, X_d] = I$ nach $(\star)$. Also ist $h(\overline{p_1}, \ldots, \overline{p_n}) = 0$, das heißt h liegt im Kern von ψ. Für die zweite Behauptung sei h aus dem Kern von ψ, also $h \in K[Y_1, \ldots, Y_n]$ mit $h(p_1, \ldots, p_n) \equiv 0$ modulo I. Dann gilt $h \equiv h(p_1, \ldots, p_n) \equiv 0$ modulo L. Bezüglich $\leq$ ist G eine Gröbnerbasis von L, und es reduziert sich h zu 0 modulo G. Da in h kein X_i auftaucht, fließt nach Wahl von $\leq$ bei diesem Reduktionsprozeß auch kein X_i ein. Dann können aber nur diejenigen Polynome aus G zur Reduktion beigetragen haben, in denen kein X_i auftaucht, also die Polynome aus $G \cap K[Y_1, \ldots, Y_n]$.

Schließlich verifizieren wir (3): Zunächst gebe es ein Urbild g' von $\overline{f}$ unter ψ, also ein $g' \in K[Y_1, \ldots, Y_n]$ mit $g'(p_1, \ldots, p_n) \equiv f$ modulo I. Dann gilt $g' \equiv g'(p_1, \ldots, p_n) \equiv f \equiv g$ modulo L. Es ist also g die eindeutig bestimmte Normalform von g' bei Reduktion modulo der Gröbnerbasis G von L bezüglich $\leq$. Bei einem entsprechenden Reduktionsprozeß von g' auf g beginnt man mit dem Polynom g', in dem kein X_i vorkommt, und es wird nach Wahl von $\leq$ auch kein X_i eingeschleppt. Also tritt auch in g kein X_i auf. Der Algorithmus gibt also in diesem Fall die korrekte Antwort „ja“. Umgekehrt setzen wir nun voraus, daß der Algorithmus die Antwort „ja“ ausgibt. Dann gilt $g \in K[Y_1, \ldots, Y_n]$ und es gilt $g(p_1, \ldots, p_n) \equiv g \equiv f$ modulo L. Daraus folgt $g(p_1, \ldots, p_n) - f \in L \cap K[X_1, \ldots, X_d] = I$ nach $(\star)$, also $g(\overline{p_1}, \ldots, \overline{p_n}) = \overline{f}$, d.h. $\psi(g) = \overline{f}$. □

Das folgende Verfahren greift auf den gerade vorgestellten Algorithmus zurück und arbeitet analog zu Lemma 3.20:

Lemma 3.30. *Sei K ein Unterkörper von $\mathbb{R}$. Seien Polynome $q_1, \ldots, q_m, p_1, \ldots, p_n, f \in K[X_1, \ldots, X_d]$ gegeben. Bezeichne I das von $q_1, \ldots, q_m$ erzeugte Ideal in $K[X_1, \ldots, X_d]$. Sei $(f_1, \alpha_1, \ldots, f_l, \alpha_l)$ ein $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis modulo I. Der folgende Algorithmus (modulo Rechnen im angeordneten Körper K) berechnet nach Eingabe von $q_1, \ldots, q_m$, $(f_1, \alpha_1, \ldots, f_l, \alpha_l)$ und f einen $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis α modulo I für f.*

(1) Berechne mit dem Algorithmus aus Satz 3.29 ein Polynom $g \in K[Z_1, \ldots, Z_l]$ mit $g(\overline{f_1}, \ldots, \overline{f_l}) = \overline{f} \in K[X_1, \ldots, X_d]/I$.

(2) Berechne anhand der folgenden Regeln (a), (b), (c) aus den vorliegenden $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweisen a_i für f_i modulo I einen $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis α modulo I für $g(f_1, \ldots, f_l)$:

   (a) Ist $a \in K$, so ist $(|a|, |a| + a, |a| - a)$ ein $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis modulo I für a.

   (b) Sind $(s, \varrho_+, \varrho_-)$ bzw. (t, σ_+, σ_-) $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweise modulo I für h bzw. h', so ist $(s + t, \varrho_+ + \sigma_+, \varrho_- + \sigma_-)$ ein solcher für $h + h'$.

   (c) Sind $(s, \varrho_+, \varrho_-)$ bzw. (t, σ_+, σ_-) $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweise modulo I für h bzw. h', so ist $(st, \frac{1}{2}(\varrho_+\sigma_+ + \varrho_-\sigma_-), \frac{1}{2}(\varrho_-\sigma_+ + \varrho_+\sigma_-))$ ein solcher für hh'.

(3) Gebe α aus.

Beweis: In (1) kann das Polynom g tatsächlich berechnet werden, da $K[X_1, \ldots, X_d]/I = K[\overline{f_1}, \ldots, \overline{f_l}]$ nach Definition 3.28. Die Gültigkeit der Regeln (a) bis (c) in (2) rechnet man sofort wie im Beweis von Lemma 3.20 nach. Das in (3) ausgegebene α ist gemäß (2) ein $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis modulo I für $g(f_1, \ldots, f_l)$ und damit auch für f, denn f ist modulo I kongruent zu $g(f_1, \ldots, f_l)$. □

Als nächstes bringen wir eine algorithmische Version derjenigen Richtung von Lemma 3.23, die für den Beweis des archimedischen Positivstellensatzes 3.24 relevant war.

Lemma 3.31. *Sei K ein Unterkörper von $\mathbb{R}$. Seien $q_1, \ldots, q_m, p_1, \ldots, p_n \in K[X_1, \ldots, X_d]$. Bezeichne I das von $q_1, \ldots, q_m$ erzeugte Ideal in $K[X_1, \ldots, X_d]$. Sei $(f_1, \alpha_1, \ldots, f_l, \alpha_l)$ ein $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis modulo I. Nach Eingabe von $q_1, \ldots, q_m$ sowie von $(f_1, \alpha_1, \ldots, f_l, \alpha_l)$ berechnet der folgende Algorithmus (modulo Rechnen im angeordneten Körper K) $\{p_1, \ldots, p_n\}$-Darstellungen $\varrho_1, \ldots, \varrho_{n+1}$ modulo I von Polynomen $p'_1, \ldots, p'_{n+1} \in K[X_1, \ldots, X_d]$, sodaß gilt:*

(i) $p'_1 + \cdots + p'_{n+1} \equiv 1 \quad$ *modulo* I

(ii) $\langle K^{\geq 0}, \overline{p_1}, \ldots, \overline{p_n} \rangle_0 = \langle K^{\geq 0}, \overline{p'_1}, \ldots, \overline{p'_{n+1}} \rangle_0 \quad$ *in* $K[X_1, \ldots, X_d]/I$

(iii) $K[X_1, \ldots, X_d]/I = K[\overline{p'_1}, \ldots, \overline{p'_{n+1}}]$

(1) Berechne mit dem Algorithmus aus Lemma 3.30 einen $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis $(s, \varrho_+, \varrho_-)$ modulo I für $p_1 + \cdots + p_n$.

(2) O.B.d.A. ist $s > 0$, ersetze sonst $(s, \varrho_+, \varrho_-)$ durch $(s+1, \varrho_+ + 1, \varrho_- + 1)$.

(3) Gebe $\varrho_1 := \frac{1}{s} p_1, \ldots, \varrho_n := \frac{1}{s} p_n$ und $\varrho_{n+1} := \frac{1}{s} \varrho_-$ aus.

Beweis: Es ist ϱ_- eine $\{p_1, \ldots, p_n\}$-Darstellung modulo I von $s - (p_1 + \cdots + p_n)$. Für die durch $\varrho_1, \ldots, \varrho_{n+1}$ dargestellten Polynome $p'_1, \ldots, p'_{n+1}$ gilt daher $p'_1 \equiv \frac{1}{s} p_1, \ldots, p'_n \equiv \frac{1}{s} p_n$ und $p'_{n+1} \equiv 1 - (p'_1 + \cdots + p'_n)$ modulo I. Daraus folgen sofort (i) und (ii). Da ein $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis modulo I eingegeben wurde und daher insbesondere existiert, ist der Semiring $\langle K^{\geq 0}, \overline{p_1}, \ldots, \overline{p_n} \rangle_0 \subseteq K[X_1, \ldots, X_d]/I$ archimedisch. Es folgt

$$K[X_1, \ldots, X_d]/I = A(\langle K^{\geq 0}, \overline{p_1}, \ldots, \overline{p_n} \rangle_0) \subseteq K[\overline{p'_1}, \ldots, \overline{p'_{n+1}}] \subseteq K[X_1, \ldots, X_d]/I,$$

also (iii). □

Nun haben wir alles beieinander, um den Beweis des archimedischen Positivstellensatzes 3.24 in einen Algorithmus umzusetzen. Es handelt sich um das Hauptergebnis dieses Abschnitts.

Satz 3.32. *Sei K ein Unterkörper von $\mathbb{R}$. Durch die Polynome $q_1, \ldots, q_m, p_1, \ldots, p_n \in K[X_1, \ldots, X_d]$ sei die Menge*

$$S := \{x \in \mathbb{R}^d \mid q_1(x) = 0, \ldots, q_m(x) = 0, p_1(x) \geq 0, \ldots, p_n(x) \geq 0\}$$

definiert. Es sei $f \in K[X_1, \ldots, X_d]$ mit $f > 0$ auf S. Es bezeichne I das von $q_1, \ldots, q_m$ in $K[X_1, \ldots, X_d]$ erzeugte Ideal. Nach Eingabe von $q_1, \ldots, q_m, p_1, \ldots, p_n, f$ und einem $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis modulo I berechnet der folgende Algorithmus (modulo Rechnen im angeordneten Körper K) eine Darstellung von f in der Form

$$(\mathcal{R}_>) \qquad f = a + \sum_e a_e p_1^{e_1} \cdots p_n^{e_n} + \sum_{i=1}^m h_i q_i$$

mit $a \in K^{>0}$, $a_e \in K^{\geq 0}$ und $h_i \in K[X_1, \ldots, X_d]$.

(1) Wie in Bemerkung 3.1 auf Seite 34 beschrieben, reicht es aus, ein $a \in K^{>0}$ und eine $\{p_1, \ldots, p_n\}$-Darstellung ϱ modulo I von $f - a$ zu berechnen. Dies geschieht im Folgenden.

(2) Berechne mit dem Algorithmus aus dem letzten Lemma $\{p_1, \ldots, p_n\}$-Darstellungen $\varrho_1, \ldots, \varrho_{n+1}$ modulo I von Polynomen $p'_1, \ldots, p'_{n+1}$ mit den Eigenschaften (i), (ii) und (iii) aus dem letzten Lemma. Es reicht aus, ein $a \in K^{>0}$ und eine $\{p'_1, \ldots, p'_{n+1}\}$-Darstellung ϱ' modulo I von $f - a$ zu berechnen. Durch Komposition von ϱ' mit $\varrho_1, \ldots, \varrho_{n+1}$ erhält man dann nämlich ein ϱ wie in (1). Indem wir von $p_1, \ldots, p_n$ zu

$p'_1, \ldots, p'_{n+1}$ übergehen (dabei ändert sich nach (ii) aus dem letzten Lemma die Menge S nicht), können wir also o.B.d.A. annehmen, daß

$$K[X_1, \ldots, X_d]/I = K[\overline{p_1}, \ldots, \overline{p_n}] \qquad \text{und} \qquad \overline{p_1} + \cdots + \overline{p_n} = 1.$$

(3) Berechne mit dem Algorithmus aus Satz 3.29 ein Erzeugendensystem $r_1, \ldots, r_t$ des Kerns J des K-Algebrenhomomorphismus

$$\psi : K[Y_1, \ldots, Y_n] \to K[X_1, \ldots, X_d]/I : Y_1 \mapsto \overline{p_1}, \ldots, Y_n \mapsto \overline{p_n}$$

und ein $g \in K[Y_1, \ldots, Y_n]$ mit $\psi(g) = \overline{f}$.

(4) Multipliziere in dem Polynom

$$g' := g + C(r_1^2 + \cdots + r_t^2) \in K[C, Y_1, \ldots, Y_n]$$

diejenigen Monome, die in $Y_1, \ldots, Y_n$ einen Grad e haben, der kleiner als der Grad e' von g' in $Y_1, \ldots, Y_n$ ist, mit $(Y_1 + \cdots + Y_n)^{e'-e}$. Wir erhalten ein Polynom $G \in K[C, Y_1, \ldots, Y_n]$ von der Form

$$(\star) \qquad \sum_{e_1+\cdots+e_n=k} (\lambda_e + \mu_e C) Y_1^{e_1} \cdots Y_n^{e_n} \qquad (k \in \mathbb{N}, \lambda_e, \mu_e \in K).$$

(5) Für jedes Polynom $H \in K[C, Y_1, \ldots, Y_n]$ von der Form $(\star)$ läßt sich offenbar leicht entscheiden, ob es ein $c \in K$ gibt, sodaß $H(c, Y_1, \ldots, Y_n) \in K[Y_1, \ldots, Y_n]$ eine Positivform ist, und gegebenfalls läßt sich ebenso leicht ein solches c berechnen. Überprüfe nacheinander für $N = 0, 1, 2, \ldots$ die Existenz eines solchen $c \in K$ für das Polynom

$$H_N := G(Y_1 + \cdots + Y_n)^N \in K[C, Y_1, \ldots, Y_n],$$

welches ebenfalls von der Form $(\star)$ ist. Für das kleinste N, für das ein geeignetes c existiert, bestimme ein solches und definiere eine Positivform F durch

$$F := H_N(c, Y_1, \ldots, Y_n) \in K[Y_1, \ldots, Y_n].$$

(6) Berechne ein $a \in K^{>0}$, sodaß die Form

$$F - a(Y_1 + \cdots + Y_n)^{\deg F} \in K[Y_1, \ldots, Y_n]$$

keine negativen Koeffizienten besitzt und damit via Ersetzung von Y_i durch p_i eine $\{p_1, \ldots, p_n\}$-Darstellung ϱ modulo I definiert. Mit a und ϱ haben wir die in (1) geforderten Daten gefunden.

Beweis: Der Beweis besteht eigentlich nur aus einer Durchsicht des Beweises des archimedischen Positivstellensatzes 3.24, des Lemmas 3.13 und des Lemmas 3.10. Wir wollen ihn hier trotzdem ausführen.

(1) und (2) sind klar. Nach den beiden am Ende von (2) o.B.d.A. vorausgesetzten Beziehungen induziert der K-Algebrenhomomorphismus ψ aus (3) einen K-Algebrenisomorphismus

$$\Psi : K[Y_1, \ldots, Y_n]/J \to K[X_1, \ldots, X_d]/I : Y_1 \mapsto p_1, \ldots, Y_n \mapsto p_n$$

mit $Y_1 + \cdots + Y_n \in J$. Nach Lemma 3.18 gilt $\Psi^{-1}(\overline{f}) > 0$ auf $S(\Psi^{-1}(\{\overline{p_1}, \ldots, \overline{p_n}\}))$. Wegen $\Psi^{-1}(\{\overline{p_1}, \ldots, \overline{p_n}\}) = \{\overline{Y_1}, \ldots, \overline{Y_n}\}$ gilt $S(\Psi^{-1}(\{\overline{p_1}, \ldots, \overline{p_n}\})) = V_{\mathbb{R}}(J) \cap (\mathbb{R}^{\geq 0})^n$. Da außerdem für das in (3) berechnete g gilt $\Psi^{-1}(\overline{f}) = \overline{g}$, erhalten wir

$$g > 0 \quad \text{auf} \quad V_{\mathbb{R}}(J) \cap (\mathbb{R}^{\geq 0})^n.$$

Lemma 3.11 angewendet auf $U := V_{\mathbb{R}}(J) \cap (\mathbb{R}^{\geq 0})^n \subseteq V_{\mathbb{R}}(\{Y_1 + \cdots + Y_n - 1\}) \cap (\mathbb{R}^{\geq 0})^n =: V$, $g|_V$ statt f und $r := (r_1^2 + \cdots + r_t^2)|_V$. garantiert nun die Existenz eines $c' \in K$, sodaß die Spezialisierung $g'(c', Y_1, \ldots, Y_n)$ des in (4) definierten Polynoms g' auf $V_{\mathbb{R}}(\{Y_1 + \cdots + Y_n - 1\}) \cap (\mathbb{R}^{\geq 0})^n$ nur positive Werte annimmt. Das Polynom G entsteht in (4) auf eine solche Weise aus g', daß $G(d, y) = g'(d, y)$ gilt für alle $d \in K$ und alle $y \in V_{\mathbb{R}}(\{Y_1 + \cdots + Y_n - 1\})$. Insbesondere gilt $G(c', y) = g'(c', y) > 0$ für alle $y \in V_{\mathbb{R}}(\{Y_1 + \cdots + Y_n - 1\}) \cap (\mathbb{R}^{\geq 0})^n$. Da $G(c', Y_1, \ldots, Y_n)$ eine Form ist, gilt nach Bemerkung 3.5 auf Seite 35 sogar

$$G(c', y) > 0 \quad \text{für alle} \quad y \in (\mathbb{R}^{\geq 0})^n \setminus \{0\}.$$

Nach dem Satz von Pólya 3.6 gibt es also ein $N' \in \mathbb{N}$, sodaß $G(c', Y_1, \ldots, Y_n)(Y_1 + \cdots + Y_n)^{N'}$ eine Positivform ist. Spätestens, wenn das in (5) schrittweise erhöhte N so groß wie N' geworden ist, stellt der Algorithmus die Existenz eines $c \in K$ fest, für das

$$H_N(c, Y_1, \ldots, Y_n) = G(c, Y_1, \ldots, Y_n)(Y_1 + \cdots + Y_n)^N$$

eine Positivform ist, und berechnet ein solches c. Schritt (5) unseres Algorithmus terminiert also. Daß schließlich die in (6) berechneten a und ϱ wie dort behauptet, die in (1) geforderte Eigenschaft erfüllen, daß ϱ eine $\{p_1, \ldots, p_n\}$-Darstellung modulo I von $f - a$ ist, zeigen wir durch folgende Rechnung:

$$\begin{aligned}
\psi(F - a(Y_1 + \cdots + Y_n)^{\deg F}) &= \psi(H_N(c, Y_1, \ldots, Y_n) - a(Y_1 + \cdots + Y_n)^{\deg F}) \\
&= \psi(G(c, Y_1, \ldots, Y_n)(Y_1 + \cdots + Y_n)^N) - a \\
&= \psi(G(c, Y_1, \ldots, Y_n)) - a \\
&= \psi(g'(c, Y_1, \ldots, Y_n)) - a \\
&= \psi(g + c(r_1^2 + \cdots + r_t^2)) - a \\
&= \psi(g) - a = \overline{f} - a
\end{aligned}$$

Wir haben dabei in der zweiten, dritten und vierten Gleichheit $Y_1 + \cdots + Y_n - 1 \in J$ ausgenutzt. In der vorletzten Gleichheit haben wir $r_1^2 + \cdots + r_t^2 \in J$ ausgenutzt. □

Eine zentrale Idee des obigen Algorithmus war es, das c aus Lemma 3.11 auf Seite 39 einfach als neue Unbestimmte C anzusetzen. So wie der Algorithmus jetzt formuliert ist, kommt diese neue Unbestimmte im Polynom g' aus (4) tatsächlich immer vor, denn $r_1^2 + \cdots + r_t^2$ ist nicht das Nullpolynom (sonst würde jedes r_i auf $\mathbb{R}^n$ verschwinden, damit bekanntlich selber das Nullpolynom sein, was $J = (r_1, \ldots, r_t) = (0)$ im Widerspruch zu $Y_1 + \cdots + Y_n \in J$ nach sich zöge). Man beachte aber, daß es in (3) nicht nötig gewesen wäre, $r_1, \ldots, r_t$ als ein Erzeugendensystem von J zu wählen. Es hätte schon gereicht, die $r_1, \ldots, r_t$ so zu wählen, daß sie *zusammen mit* $Y_1 + \cdots + Y_n - 1$ den Kern J von ψ erzeugen (wie im Beweis von Lemma 3.13). Insbesondere könnte man im Fall $J = (Y_1 + \cdots + Y_n - 1)$ die Zahl t als 0 wählen und damit auf den Einsatz der Unbestimmten C verzichten. In Abschnitt 3.6 werden wir noch mehr Fälle kennenlernen, in denen man durch Einsatz einer Verallgemeinerung des Satzes von Pólya die Unbestimmte C vermeiden kann.

Durch eine Vermeidung von C erreichen wir nicht etwa eine nennenswerte Beschleunigung unseres Algorithmus. Der Aufwand zu prüfen, ob ein Polynom der in (4) angebenen Form ($\star$) durch Spezialisierung von C zu einer Positivform gemacht werden kann, dürfte ja kaum größer sein, als der Aufwand zu prüfen, ob eine Form vom entsprechenden Grad eine Positivform ist.

Aus einem anderen Grund könnte es jedoch erstrebenswert sein, ohne C auszukommen: Es komme in dem in (4) berechneten Polynom G kein C vor (man habe also $t = 0$ gewählt, es ist also $J = (Y_1 + \cdots + Y_n - 1)$). Mit der Schranke (i) aus Bemerkung 3.9 kann man dann

die Anzahl der Schleifendurchläufe in (5) abschätzen durch n, den Grad von G, die Größe der Koeffizienten von G und dem Minimum von G auf $V_{\mathbb{R}}(\{Y_1 + \cdots + Y_n - 1\}) \cap (\mathbb{R}^{\geq 0})^n$. Das genannte Minimum ist nach Lemma 3.18 dann gleich dem Minimum von f auf S (beachte $J = (Y_1 + \cdots + Y_n - 1)$, also $V_{\mathbb{R}}(\{Y_1 + \cdots + Y_n - 1\}) = V_{\mathbb{R}}(J)$). Nun kann man hoffen, daß sich für spezielle $q_1, \ldots, q_m, p_1, \ldots, p_n$ der Grad von G und die Größe der Koeffizienten von G auch in vernünftiger Weise durch die entsprechenden Daten von f abschätzen lassen. Auf diese Weise könnte man für Spezialfälle also eine Komplexitätsabschätzung des Algorithmus erhalten, die lauter vernünftige Maße der Eingabe verwendet, bis auf ein unvernünftiges, nämlich das Minimum von f auf S. In noch spezielleren Fällen könnte man vielleicht sicherstellen, daß G immer ganzzahlige Koeffizienten hat, und dann mit der Schranke (ii) aus Bemerkung 3.9 die Abhängigkeit der Abschätzung vom Minimum von f auf S (welches ja gleich dem Minimum von G auf $V_{\mathbb{R}}(\{Y_1 + \cdots + Y_n - 1\}) \cap (\mathbb{R}^{\geq 0})^n$ ist) auch noch auflösen.

Im Prinzip ist diese Abhängigkeit der Laufzeit unseres Algorithmus vom Minimum von f auf S aber tatsächlich da. Dies zeigt die Bemerkung 3.8 über die Komplexität des Pólyaschen Satzes. Trotzdem ist unser Algorithmus nicht schlecht. Es gibt keinen anderen Algorithmus für dasselbe Problem, der nicht dieselben Eigenheiten aufweisen würde. Die genannte Abhängigkeit ist dem Problem innewohnend. Dies zeigt die Bemerkung 3.25 über die Komplexität des archimedischen Positivstellensatzes.

Wenn man dem Algorithmus entgegen seiner Spezifikation ein f als Eingabe gibt, für welches nicht $f > 0$ auf S gilt, so terminiert er nicht. Die Rechnung am Schluß des Beweises von Satz 3.32 zeigt nämlich, daß der Algorithmus auch dann noch eine Darstellung $(\mathcal{R}_>)$ von f berechnen müßte, wenn er terminieren würde. Diese Darstellung kann es aber nicht geben, wenn nicht $f > 0$ auf S gilt.

Unser Algorithmus ist also kein Entscheidungsverfahren dafür, ob $f > 0$ auf S gilt. Es bringt auch nichts zu versuchen, mit den Schranken aus Lemma 3.9 ein $N_0 \in \mathbb{N}$ vorauszuberechnen, sodaß man sicher sein kann, daß nicht $f > 0$ auf S gilt, wenn die Schleife in (5) bei $N = N_0$ angekommen ist: Im kritischen Fall nämlich, daß nicht $f > 0$ auf S gilt, nimmt für jedes $c \in K$ die Form $G(c, Y_1, \ldots, Y_n)$ auf $V_{\mathbb{R}}(J) \cap (\mathbb{R}^{\geq 0})^n$ ebenfalls nicht nur positive Werte an, und man kann die Schranke (i) aus diesem Lemma auf $G(c, Y_1, \ldots, Y_n)$ gar nicht anwenden. In der Schranke (ii) dieses Lemmas kommt eine feste Zahl vor, die vermutlich niemandem bekannt ist. Würde man sie doch kennen, so wäre sie vermutlich viel zu groß um ein praktikables Entscheidungsverfahren (für den Fall $K = \mathbb{Q}$) zu bekommen. Prinzipiell ist es aber möglich, zu entscheiden, ob $f > 0$ auf S gilt, nämlich durch effektive reelle Quantorenelimination (siehe [rqe],[Mis]).

Das größte Problem an unserem Algorithmus ist aber, daß er einen $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis modulo $\{q_1, \ldots, q_m\}$ als Eingabe braucht. Die Frage, wann ein solcher existiert, haben wir schon im Anschluß an den archimedischen Positivstellensatz 3.24 diskutiert. Man muß hier unterscheiden, ob S kompakt ist oder nicht: Wenn S kompakt ist, existiert ein solcher Archimedizitätsnachweis genau dann, wenn es zu jedem f mit $f > 0$ auf S die Darstellung $(\mathcal{R}_>)$ gibt. Für kompaktes S haben wir also durch die Forderung der Existenz des Archimedizitätsnachweises noch nichts verloren. Wenn S nicht kompakt ist, ist dem Autor weder ein Fall bekannt, in dem die Darstellung $(\mathcal{R}_>)$ für jedes f mit $f > 0$ auf S existiert, noch ein Beweis dafür, daß es einen solchen Fall nicht gäbe. Es ist dem Autor also unbekannt, ob durch die Forderung der Existenz hier bereits etwas verloren wurde.

Als nächstes stellt sich die Frage, wo man den benötigten Archimedizitätsnachweis herbekommt, falls er existiert. Dazu werden wir im nächsten Abschnitt zeigen, daß man bereits für sehr brauchbare Teilklassen unserer Problemstellung einen solchen Archimedizitätsnachweis automatisch berechnen kann. Im Allgemeinen ist es zumindest so, daß man für feste $q_1, \ldots, q_m, p_1, \ldots, p_n$ versuchen kann, per Hand einen solchen zu finden. Dies hat einen gewissen Wert, denn der Archimedizitätsnachweis ist ja nicht von f abhängig. Hat man also

einen gefunden, so hat man nachher für die festen $q_1, \ldots, q_m, p_1, \ldots, p_n$ einen Algorithmus, der für jedes f mit $f > 0$ auf S eine Darstellung $(\mathcal{R}_>)$ berechnet.

3.4 Archimedizitätsnachweise

Sei K ein Unterkörper von $\mathbb{R}$. In diesem Abschnitt befassen wir uns in verschiedenen Beispielen für die Polynome $q_1, \ldots, q_m, p_1, \ldots, p_n \in K[X_1, \ldots, X_d]$ mit der Frage nach der Existenz und der Berechnung von $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweisen modulo $(q_1, \ldots, q_m)$, die durch Satz 3.32 aufgeworfen wird. Es liegt nahe, zunächst einen möglichst einfachen, nichttrivialen Fall zu betrachten, nämlich daß $q_1, \ldots, q_m, p_1, \ldots, p_n$ linear sind, d.h. einen Grad ≤ 1 haben. Die Menge

$$S := \{x \in \mathbb{R}^d \mid q_1(x) = 0, \ldots, q_m(x) = 0, p_1(x) \geq 0, \ldots, p_n(x) \geq 0\}$$

ist dann ein konvexer Polyeder (und umgekehrt läßt sich jeder konvexer Polyeder auf diese Weise definieren). Wir werden für diesen Fall das bestmögliche Resultat erhalten, nämlich ein Verfahren, welches bei nichtleerem kompaktem S nach Eingabe von $q_1, \ldots, q_m, p_1, \ldots, p_n$ einen $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis modulo $(q_1, \ldots, q_m)$ ausgibt. Ein besseres Resultat konnte man nicht erwarten: Für nicht kompaktes S kann ein solcher Archimedizitätsnachweis von vorneherein nicht existieren. Für leeres S haben wir auch schnell ein Beispiel, wo ein solcher Archimedizitätsnachweis nicht existieren kann: Wähle einfach $m = 0$, $n = 1$ und $p_1 = -1$. Der fragliche Archimedizitätsnachweis existiert genau dann, wenn der Semiring $K = \langle K^{\geq 0}, -1\rangle_0 = \langle K^{\geq 0}, p_1\rangle_0 \subseteq K[X_1, \ldots, X_d]$ archimedisch ist, wenn also $A(K) = K[X_1, \ldots, X_d]$ ist. Offensichtlich gilt aber $A(K) = K$. Wenn $d \geq 1$ ist, haben wir also ein Gegenbeispiel.

Eine bisher nie erwähnte, aber völlig offensichtliche, gute Eigenschaft von Archimedizitätsnachweisen ist die folgende:

Bemerkung 3.33. Sei K ein Unterkörper von $\mathbb{R}$. Seien $m, m', n, n' \in \mathbb{N}$ mit $m \leq m'$ und $n \leq n'$. Seien $q_1, \ldots, q_{m'}, p_1, \ldots, p_{n'} \in K[X_1, \ldots, X_d]$. Dann ist jeder $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis modulo $(q_1, \ldots, q_m)$ auch ein $\{p_1, \ldots, p_{n'}\}$-Archimedizitätsnachweis modulo $(q_1, \ldots, q_{m'})$.

Mit dieser Bemerkung liegt dann das oben erwähnte Resultat bei weitem nicht nur dann vor, wenn S ein nichtleerer kompakter konvexer Polyeder ist. Im Prinzip geht es aber um konvexe Polyeder. Wir begeben uns also in die Theorie der linearen Ungleichungen. Um hervorzuheben, wo für die reellen Zahlen typische Eigenschaften eingehen, arbeiten wir solange wie möglich über einem beliebigen angeordneten Körper. Zunächst allerdings noch ein trivialer Hilfsalgorithmus, der mit einer Anordnung des Körpers noch gar nichts zu tun hat:

Lemma 3.34. *Sei K ein Körper. Seien $u_1, \ldots, u_n, v \in K^d$ und $0 \neq a \in K$. Der folgende Algorithmus (modulo Rechnen im Körper K) gibt nach Eingabe von $u_1, \ldots, u_n, v, a$ auf die Frage, ob v in dem von $u_1, \ldots, u_n$ erzeugten Unterraum des K^d liegt, die richtige Antwort „ja“ oder „nein“ aus. Falls er „nein“ ausgibt, berechnet er zusätzlich ein $c \in K^d$ mit $c^T u_1 = \cdots = c^T u_n = 0$ und $c^T v = a$.*

(1) Berechne eine Zeilenstufenform M der Matrix

$$\begin{pmatrix} u_1^T & 0 \\ \vdots & \vdots \\ u_n^T & 0 \\ v^T & a \end{pmatrix} \in K^{(n+1)\times(d+1)}.$$

Schreibe $M = (M'\ b)$ mit $M' \in K^{(n+1)\times d}$ und $b \in K^{(n+1)\times 1}$.

(2) Falls M' weniger Stufen hat als M: Gebe „ja“ aus.

(3) Falls M' genausoviel Stufen hat wie M: Gebe „nein“ aus. Berechne ein $c \in K^d$ mit $M'c = b$.

Beweis: Die Anzahl der Stufen von M ist gleich dem Rang der Matrix

$$\begin{pmatrix} u_1^T & 0 \\ \vdots & \vdots \\ u_n^T & 0 \\ v^T & a \end{pmatrix},$$

also wegen $a \neq 0$ um 1 größer als die Dimension des von $u_1, \ldots, u_n$ erzeugten Unterraums des K^d. Die Anzahl der Stufen von M' ist gleich dem Rang der Matrix

$$\begin{pmatrix} u_1^T \\ \vdots \\ u_n^T \\ v^T \end{pmatrix},$$

also gleich der Dimension des von $u_1, \ldots, u_n, v$ erzeugten Unterraums des K^d. Daraus folgt, daß der Algorithmus stets die richtige Antwort auf die Frage gibt, ob v in dem von $u_1, \ldots, u_n$ erzeugten Unterraum liegt. In (3) existiert ein c, wie es dort berechnet werden soll, da dort der Rang von M gleich dem Rang von M' ist, die letzte Spalte von M also eine Linearkombination der übrigen Spalten von M ist. Schließlich leistet das berechnete c das Gewünschte, denn es gilt

$$\begin{aligned} M'c = b &\iff \begin{pmatrix} u_1^T \\ \vdots \\ u_n^T \\ v^T \end{pmatrix} c = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ a \end{pmatrix} \\ &\iff u_1^T c = \cdots = u_n^T c = 0 \text{ und } v^T c = a \\ &\iff c^T u_1 = \cdots = c^T u_n = 0 \text{ und } c^T v = a. \end{aligned}$$

□

Ganz im Gegensatz zu dem gerade vorgestellten, handelt es sich bei dem nächsten Algorithmus, der sich in [Scr] findet, um einen sehr gehaltvollen. Allein schon die Existenz seiner Ausgabe ist interessant: Sei K ein angeordneter Körper. Wenn ein gegebener Vektor $v \in K^d$ in jedem Halbraum $\{x \in K^d \mid c^T x \geq 0\}$ mit $c \in K^d \setminus \{0\}$ liegt, in dem endlich viele andere gegebene Vektoren $u_1, \ldots, u_n \in K^d$ liegen, dann gibt es eine Darstellung von v, die dies offensichtlich macht, nämlich eine Darstellung der Form $v = \lambda_1 u_1 + \cdots + \lambda_n u_n$ mit $\lambda_1, \ldots, \lambda_n \in K^{\geq 0}$, also als nichtnegative Linearkombination von $u_1, \ldots, u_n$. Für die Idee, wie man den Algorithmus aufzieht ganz wesentlich, für unsere spätere Anwendung allerdings unwesentlich ist, daß man sogar eine Darstellung von der Form $v = \lambda_{i_1} u_{i_1} + \cdots + \lambda_{i_t} u_{i_t}$ mit $\lambda_{i_1}, \ldots, \lambda_{i_t} \in K^{\geq 0}$ und *linear unabhängigen* $u_{i_1}, \ldots, u_{i_t}$ finden kann.

Satz 3.35 (Fundamentalsatz über lineare Ungleichungen). *Sei K ein angeordneter Körper. Seien $u_1, \ldots, u_n, v \in K^d$ und t die Dimension des von diesen Vektoren erzeugten Unterraums von K^d. Der folgende Algorithmus (modulo Rechnen im angeordneten Körper K) berechnet nach Eingabe von $u_1, \ldots, u_n, v$*

- entweder *eine Darstellung von v als nichtnegative Linearkombination linear unabhängiger Vektoren aus $u_1, \ldots, u_n$*

- oder *ein* $c \in K^d \setminus \{0\}$*, sodaß die Hyperebene* $\{x \in K^d \mid c^T x = 0\}$ $t-1$ *linear unabhängige Vektoren aus* $u_1, \ldots, u_m$ *enthält und* $c^T u_1 \geq 0, \ldots, c^T u_n \geq 0$*, aber* $c^T v < 0$ *gilt:*

(1) Überprüfe mit dem Algorithmus aus Lemma 3.34, ob v in dem von $u_1, \ldots, u_m$ erzeugten Unterraum liegt. Berechne, falls das nicht der Fall ist, mit demselben Algorithmus ein $c \in K^d$ mit $c^T u_1 = \cdots = c^T u_n = 0$ und $c^T v = -1$, gebe dieses aus und breche ab.

(2) Es liege also nun v in dem von $u_1, \ldots, u_m$ erzeugten Unterraum. Wähle eine t-elementige Teilmenge $I = \{i_1, \ldots, i_t\}$ von $\{1, \ldots, n\}$, sodaß $u_{i_1}, \ldots, u_{i_t}$ linear unabhängig sind. Wiederhole folgende Schleife:

   (a) Berechne die eindeutig bestimmte Darstellung $v = \lambda_{i_1} u_{i_1} + \cdots + \lambda_{i_t} u_{i_t}$ mit $\lambda_{i_1}, \ldots, \lambda_{i_t} \in K$. Falls $\lambda_{i_1}, \ldots, \lambda_{i_t} \geq 0$, gebe diese Darstellung aus und breche ab.

   (b) Wähle sonst das kleinste h unter $i_1, \ldots, i_t$ mit $\lambda_h < 0$. Berechne mit dem Algorithmus aus Lemma 3.34 ein $c \in K^d$ mit $c^T u_h = 1$ und $c^T u_i = 0$ für alle $i \in I \setminus \{h\}$. Es gilt also $c^T v = \lambda_h c^T u_h = \lambda_h < 0$.

   (c) Falls $c^T u_1, \ldots, c^T u_n \geq 0$, gebe dieses c aus und breche ab.

   (d) Wähle sonst das kleinste $s \in I$ mit $c^T u_s < 0$. Ersetze I durch $(I \setminus \{h\}) \cup \{s\}$.

Beweis: Falls der Algorithmus irgendwann abbricht, liefert er offenbar eine korrekte Antwort. Bei der Ersetzung in (d) von I durch $(I \setminus \{h\}) \cup \{s\}$ ist zu beachten, daß u_s nicht in dem Unterraum liegt, der von den u_i mit $i \in I \setminus \{h\}$ erzeugt wird, denn für diese u_i gilt $c^T u_i = 0$, während $c^T u_s < 0$ gilt. Damit ist zu Beginn des nächsten Schleifendurchlaufs I wieder eine t-elementige Teilmenge von $\{1, \ldots, n\}$ derart, daß $u_{i_1}, \ldots, u_{i_t}$ linear unabhängig sind.

Es bleibt zu zeigen, daß der Algorithmus terminiert. Dazu bezeichne I_k die Menge I zu Beginn des k-ten Schleifendurchlaufs. Angenommen der Algorithmus terminiert nicht. Dann gibt es k und l mit $I_k = I_l$ und $k < l$. Jede Zahl aus $\{1, \ldots, n\}$, die am Ende eines der Schleifendurchläufe $k, k+1, \ldots, l-1$ aus I entfernt wird, muß auch am Ende eines dieser Schleifendurchläufe zu I hinzugenommen werden, und umgekehrt. Bezeichne $r \in \{1, \ldots, n\}$ die höchste Zahl, die überhaupt in einem dieser Schleifendurchläufe zu I hinzu- oder weggenommen wird. Wähle Schleifendurchläufe $p, q \in \{k, k+1, \ldots, l-1\}$, sodaß r am Ende des p-ten Durchlaufs aus I rausgenommen und am Ende des q-ten Durchlaufs in I reingenommen wird. Offenbar gilt $p < q$ oder $q < p$. Nach Wahl von r gilt außerdem $I_p \cap \{r+1, \ldots, n\} = I_q \cap \{r+1, \ldots, n\}$.

Schreibe $I_p = \{i_1, \ldots, i_t\}$ und $v = \lambda_{i_1} u_{i_1} + \cdots + \lambda_{i_t} u_{i_t}$ mit $\lambda_{i_1}, \ldots, \lambda_{i_t} \in K$. Dies ist die zu Beginn des p-ten Schleifendurchlaufs berechnete Darstellung von v. Außerdem betrachten wir den Vektor c, der beim q-ten Schleifendurchlauf in (b) berechnet wird. Dann haben wir den Widerspruch

$$0 > c^T v = c^T(\lambda_{i_1} u_{i_1} + \cdots + \lambda_{i_t} u_{i_t}) = \lambda_{i_1}(c^T u_{i_1}) + \cdots + \lambda_{i_t}(c^T u_{i_t}) > 0$$

Die erste Ungleichung haben wir in (b) schon bemerkt. Die zweite folgt so:

- Für die $i \in I_p = \{i_1, \ldots, i_t\}$ mit $i < r$ gilt $\lambda_i \geq 0$ (wegen der kleinstmöglichen Wahl von $h = r$ in (b) beim p-ten Schleifendurchlauf) und $c^T u_i \geq 0$ (wegen der kleinstmöglichen Wahl von $s = r$ in (d) beim q-ten Schleifendurchlauf).

- Es ist $r \in I_p$ und es gilt $\lambda_r < 0$ (nach Wahl von $h = r$ in (b) beim p-ten Schleifendurchlauf) sowie $c^T u_r < 0$ (nach Wahl von $s = r$ in (d) beim q-ten Schleifendurchlauf).

- Für die $i \in I_p$ mit $i > r$ gilt $c^T u_i = 0$. Wegen $I_p \cap \{r+1, \ldots, n\} = I_q \cap \{r+1, \ldots, n\}$ sind diese i nämlich auch in I_q enthalten. Beim q-ten Schleifendurchlauf in (b) sind diese daher Elemente von $I \setminus \{h\}$ und es gilt $c^T u_i = 0$.

□

Es stellt sich dem Leser vielleicht die Frage, warum der gerade bewiesene Satz als „Fundamentalsatz über lineare Ungleichungen“ bezeichnet wurde. Dies klärt sich bereits etwas auf, wenn man die Punkte $u_1, \ldots, u_n, v$ von K^d als Linearformen $p_1, \ldots, p_n, f$ mit Koeffizienten aus K in den Unbestimmten $X_1, \ldots, X_d$ interpretiert (also als 1-Formen aus $K[X_1, \ldots, X_d]$). Dies geschieht mittels des K-Vektorraumisomorphismus $a \mapsto a_1X^1 + \cdots + a_nX^n$ von K^d in den K-Vektorraum aller Linearformen von $K[X_1, \ldots, X_d]$. Die in der Aussage des Satzes vorkommenden Produkte $c^Tu_1, \ldots, c^Tu_n$ und c^Tv gehen dann bei der neuen Interpretation über in $p_1(c), \ldots, p_n(c)$ und $f(c)$. Wenn wir nun noch x statt c schreiben, so gibt obiger Algorithmus nach Eingabe von Linearformen $p_1, \ldots, p_n, f \in K[X_1, \ldots, X_d]$ entweder eine Darstellung von f als nichtnegative Linearkombination von $p_1, \ldots, p_n$ oder ein $x \in K^d$ mit $p_1(x) \geq 0, \ldots, p_n(x) \geq 0$ und $f(x) < 0$ aus. Man bekommt also eine in jeder Hinsicht algorithmische Version des folgenden Nichtnegativstellensatzes für Linearformen:

Sei K ein angeordneter Körper. Durch die Linearformen $p_1, \ldots, p_n \in K[X_1, \ldots, X_d]$ sei (der polyedrische konvexe Kegel)

$$S := \{x \in \mathbb{R}^d \mid p_1(x) \geq 0, \ldots, p_n(x) \geq 0\}$$

definiert. Sei $f \in K[X_1, \ldots, X_d]$ eine weitere Linearform. Genau dann gilt $f \geq 0$ auf S, wenn f eine nichtnegative Linearkombination (über K) von $p_1, \ldots, p_n$ ist.

Wenn man statt Linearformen sogar lineare Polynome zuläßt (also zusätzlich nichtverschwindende konstante Koeffizienten erlaubt), stimmt dieser Satz leider nicht mehr. Ein Gegenbeispiel ist $n = 2$, $d = 1$, $p_1 = X$, $p_2 = -X$, $f = 1$. Wenn man allerdings zusätzlich voraussetzt, daß die Menge S ein nichtleeres Inneres besitzt, so stimmt der Satz wieder. Dies ist wiederum ein nichttriviales Resultat (siehe [Ha2], beachte: dort werden lineare Polynome als „linear form“ bezeichnet). Für unsere Zwecke angemessener wird jedoch ein Satz sein, in dem nicht das Innere, sondern nur S selber als nichtleer vorausgesetzt ist, in dem es dafür zusätzlich erlaubt ist, daß auch das konstante Polynom 1 zur nichtnegativen Linearkombination beiträgt. Da in [Ha2] ganz ähnliche Zwecke verfolgt wurden, wäre es auch dort angemessener gewesen, diesen Satz zu benutzen (siehe die Notizen im Anschluß zu Satz 3.39). Wir haben diesen Satz nirgends gefunden. Er dürfte aber wohl bekannt sein.

Satz 3.36 (linearer Nichtnegativstellensatz). *Sei K ein angeordneter Körper. Seien $p_1, \ldots, p_n, f \in K[X_1, \ldots, X_d]$ lineare Polynome, sodaß es ein $\xi \in K^d$ gibt mit $p_1(\xi) \geq 0, \ldots, p_n(\xi) \geq 0$. Der folgende Algorithmus (modulo Rechnen im angeordneten Körper K) gibt nach Eingabe von $p_1, \ldots, p_n, f$ auf die Frage, ob für alle $x \in K^d$ gilt*

$$p_1(x) \geq 0, \ldots, p_n(x) \geq 0 \implies f(x) \geq 0,$$

die richtige Antwort „ja“ oder „nein“ aus. Gibt er „ja“ aus, so berechnet er zusätzlich eine Darstellung von f als nichtnegative Linearkombination linear unabhängiger Polynome aus $p_1, \ldots, p_n, 1$.

(1) Wende den Algorithmus aus dem letzten Satz auf die Bilder $u_1, \ldots, u_n, u_{n+1}, v$ von $p_1, \ldots, p_n, 1, f$ unter dem Isomorphismus

$$a_1X_1 + \cdots + a_dX_d + a_0 \mapsto (a_0, a_1, \ldots, a_d)$$

vom Vektorraum der linearen Polynome aus $K[X_1, \ldots, X_d]$ auf den Vektorraum K^{d+1} an.

(2) Falls dieser Algorithmus eine Darstellung von v als nichtnegative Linearkombination linear unabhängiger Vektoren aus $u_1, \ldots, u_n, u_{n+1}$ ausgibt: Antworte „ja“ und gebe die entsprechende Darstellung von f als nichtnegative Linearkombination linear unabhängiger Polynome aus $p_1, \ldots, p_n, 1$ aus. Breche ab.

(3) Antworte sonst „nein“.

Beweis: Zu zeigen ist nur, daß der Algorithmus die richtige Antwort gibt, falls er (3) erreicht. Er erreiche also (3). Nach Satz 3.35 gibt es dann ein $c = (c_0, c_1, \ldots, c_d) \in K^{d+1}$ mit $c^T u_1 \geq 0, \ldots, c^T u_n \geq 0, c^T u_{n+1} \geq 0$ und $c^T v < 0$. Zurückübersetzt heißt dies nichts anderes als

$$
\begin{aligned}
&(\diamond) && p_i(c_1, \ldots, c_d) - p_i(0) + c_0 p_i(0) \geq 0 \quad \text{für } i \in \{1, \ldots, n\},\\
&(\star) && c_0 \geq 0 \quad \text{und}\\
&(\square) && f(c_1, \ldots, c_d) - f(0) + c_0 f(0) < 0.
\end{aligned}
$$

Wir unterscheiden gemäß $(\star)$ die Fälle $c_0 > 0$ und $c_0 = 0$.

Sei zunächst $c_0 > 0$. Da $p_i - p(0)$ für jedes $i \in \{1, \ldots, n\}$ genauso wie $f - f(0)$ eine Linearform ist, folgt dann durch Teilen der Gleichungen $(\diamond)$ und $(\square)$ durch c_0 mit der Bezeichnung $x := \frac{1}{c_0}(c_1, \ldots, c_d)$, daß $p_i(x) \geq 0$ für $i \in \{1, \ldots, n\}$ und $f(x) < 0$. In diesem Fall gibt also der Algorithmus die korrekte Antwort „nein“.

Sei nun $c_0 = 0$. Dann besagt $(\square)$, daß f vom Nullpunkt ausgehend in Richtung $(c_1, \ldots, c_d)$ fällt. Da f eine affin-lineare Abbildung ist, fällt f von *jedem* Punkt ausgehend in Richtung $(c_1, \ldots, c_d)$. Analog fällt nach $(\diamond)$ in derselben Richtung aber keines der p_i. Nun gibt es nach Voraussetzung einen Punkt $\xi \in K^d$ mit $p_1(\xi) \geq 0, \ldots, p_n(\xi) \geq 0$. Geht man von ξ aus in Richtung $(c_1, \ldots, c_d)$, so bleiben also $p_1, \ldots, p_n$ positiv, während f irgendwann schließlich negativ wird. Um dieses Argument präzise zu machen, setzen wir

$$x = \xi + \frac{1 + \max\{f(\xi), 0\}}{f(0) - f(c_1, \ldots, c_d)}(c_1, \ldots, c_d).$$

Dann gilt für jedes $i \in \{1, \ldots, n\}$

$$
\begin{aligned}
p_i(x) &= p_i(\xi) + \left(p_i\left(\frac{1 + \max\{f(\xi), 0\}}{f(0) - f(c_1, \ldots, c_d)}(c_1, \ldots, c_d) \right) - p_i(0) \right)\\
&= \underbrace{p_i(\xi)}_{\geq 0} + \underbrace{\frac{1 + \max\{f(\xi), 0\}}{f(0) - f(c_1, \ldots, c_d)}}_{> 0 \text{ nach } (\square)} \underbrace{(p_i(c_1, \ldots, c_d) - p_i(0))}_{\geq 0 \text{ nach } (\diamond)} \geq 0
\end{aligned}
$$

und gleichzeitig

$$
\begin{aligned}
f(x) &= f(\xi) + \left(f\left(\frac{1 + \max\{f(\xi), 0\}}{f(0) - f(c_1, \ldots, c_d)}(c_1, \ldots, c_d) \right) - f(0) \right)\\
&= f(\xi) + \frac{1 + \max\{f(\xi), 0\}}{f(0) - f(c_1, \ldots, c_d)}(f(c_1, \ldots, c_d) - f(0))\\
&= f(\xi) - \max\{f(\xi), 0\} - 1 \leq -1 < 0.
\end{aligned}
$$

Auch in diesem Fall ist also die Antwort „nein“ korrekt. □

Auf die in den Voraussetzungen des Satzes geforderte Existenz von ξ kann man nicht verzichten. Dies zeigt das Beispiel $n = 1$, $p_1 = -1$, $f \notin K$.

Sei nun K nicht mehr nur ein angeordneter Körper, sondern sogar wieder ein Unterkörper von $\mathbb{R}$. Der durch die linearen Polynome $q_1, \ldots, q_m, p_1, \ldots, p_n \in K[X_1, \ldots, X_d]$ definierte konvexe Polyeder

$$S := \{x \in \mathbb{R}^d \mid q_1(x) = 0, \ldots, q_m(x) = 0, p_1(x) \geq 0, \ldots, p_n(x) \geq 0\}$$

sei kompakt und nichtleer. Nach dem letzten Satz ist klar, wie man einen $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis $(f_1, \alpha_1, \ldots, f_l, \alpha_l)$ modulo $(q_1, \ldots, q_m)$ berechnen kann: Man wähle

lineare Polynome $f_1, \ldots, f_l$ mit der Eigenschaft $K[X_1, \ldots, X_d]/(q_1, \ldots, q_m) = K[\overline{f_1}, \ldots, \overline{f_l}]$ (etwa $l = d, f_1 = \overline{X_1}, \ldots, f_l = \overline{X_d}$). Für jedes f_i läßt sich ein $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis α_i modulo $(q_1, \ldots, q_m)$ leicht berechnen, da f_i linear ist. Dann ist $(f_1, \alpha_1, \ldots, f_l, \alpha_l)$ ein $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis modulo $(q_1, \ldots, q_m)$.

Wir müssen nur noch nachtragen, wie man für jedes lineare Polynom f einen $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis modulo $(q_1, \ldots, q_m)$ berechnet. Dies ist ganz einfach: Für $s = 0, 1, 2, \ldots$ überprüfe man, ob $s + f \geq 0$ auf S und $s - f \geq 0$ auf S gilt. Dies bewerkstelligt man mit dem Algorithmus aus dem letzten Satz, indem man die Menge S schreibt als

$$S = \{x \in \mathbb{R}^d \mid p_1(x) \geq 0, \ldots, p_n(x) \geq 0, q_1(x) \geq 0, -q_1(x) \geq 0, \ldots, q_m(x) \geq 0, -q_m(x) \geq 0\}.$$

Da S kompakt ist, erhält man für genügend großes $s \in \mathbb{N}$ schließlich $s + f \geq 0$ auf S und $s - f \geq 0$ auf S. Außerdem berechnet der Algorithmus aus dem obigen Satz Darstellungen von $s + f$ und $s - f$ als nichtnegative Linearkombinationen der n+2m+1 Polynome $p_1, \ldots, p_n, q_1, -q_1, \ldots, q_m, -q_m, 1$. Indem man in diesen Darstellungen den Teil der Linearkombination, der zu den 2m Polynomen $q_1, -q_1, \ldots, q_m, -q_m$ gehört, einfach wegläßt, erhält man offenbar $\{p_1, \ldots, p_n\}$-Darstellungen ϱ_+ und ϱ_- modulo $(q_1, \ldots, q_m)$ von $s+f$ bzw. $s-f$. Dann ist $\alpha := (s, \varrho_+, \varrho_-)$ ein $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis modulo $(q_1, \ldots, q_m)$ für das lineare Polynom f.

Im nächsten Satz schreiben wir eine etwas ausgefeiltere Version dieses Algorithmus auf, die noch um einiges schneller sein dürfte. Außerdem wird noch berücksichtigt, daß gemäß Bemerkung 3.33 keineswegs S immer ein konvexer Polyeder sein muß. Zu diesem Zweck ist noch eine Beobachtung über das Verhalten des Algorithmus aus dem linearen Nichtnegativstellensatzes 3.36 nützlich:

Lemma 3.37. *Sei K ein angeordneter Körper. Seien $p_1, \ldots, p_n, f \in K[X_1, \ldots, X_d]$ lineare Polynome. Es gebe $i_1, \ldots, i_k \in \{1, \ldots, n\}$, sodaß für alle $x \in K^d$ gilt*

$$p_{i_1}(x) \geq 0, \ldots, p_{i_k}(x) \geq 0 \implies f(x) \geq 0$$

und sodaß es ein $\xi \in K^d$ gibt mit $p_{i_1}(\xi) \geq 0, \ldots, p_{i_k}(\xi) \geq 0$. Dann gibt der Algorithmus (modulo Rechnen im angeordneten Körper K) von Satz 3.36 nach Eingabe von $p_1, \ldots, p_n, f$ eine Darstellung von f als nichtnegative Linearkombination linear unabhängiger Polynome aus $p_1, \ldots, p_n, 1$ aus.

Beweis: Nach Satz 3.36 angewandt auf $p_{i_1}, \ldots, p_{i_k}$ statt $p_1, \ldots, p_n$ gibt es eine Darstellung von f als nichtnegative Linearkombination der Polynome $p_{i_1}, \ldots, p_{i_k}, 1$. Insbesondere gibt es eine Darstellung von f als nichtnegative Linearkombination von $p_1, \ldots, p_n, 1$. Das heißt, es gibt mit den Bezeichnungen aus (1) des diskutierten Algorithmus eine Darstellung von v als nichtnegative Linearkombination von $u_1, \ldots, u_{n+1}$. Der in (1) aufgerufene Algorithmus aus dem Fundamentalsatz über lineare Ungleichungen 3.35 kann daher kein $c \in K^d$ finden mit $c^T u_1 \geq 0, \ldots, c^T u_{n+1} \geq 0$ und $c^T v < 0$. Der Rest ist klar. □

Wir haben nun alles aus der Theorie der linearen Ungleichungen beieinander, um das gewünschte Resultat zu beweisen:

Satz 3.38. *Sei K ein Unterkörper von $\mathbb{R}$. Unter den Polynomen $q_1, \ldots, q_m, p_1, \ldots, p_n \in K[X_1, \ldots, X_d]$ seien lineare Polynome $q_{i_1}, \ldots, q_{i_k}, p_{j_1}, \ldots, p_{j_l}$, sodaß die Menge*

$$\{x \in \mathbb{R}^d \mid q_{i_1}(x) = 0, \ldots, q_{i_k}(x) = 0, p_{j_1}(x) \geq 0, \ldots, p_{j_l} \geq 0\}$$

kompakt und nichtleer ist. Dann berechnet der folgende Algorithmus (modulo Rechnen in K) nach Eingabe von $q_1, \ldots, q_m, p_1, \ldots, p_n$ einen $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis modulo $(q_1, \ldots, q_m)$.

(1) Man streiche alle nichtlinearen Polynome aus $q_1, \ldots, q_m, p_1, \ldots, p_n$. Dies kann nach Bemerkung 3.33 o.B.d.A. gemacht werden. Es seien also nun $q_1, \ldots, q_m, p_1, \ldots, p_n$ lineare Polynome.

(2) Wende für $s = 0, 1, 2, 4, 8, 16, \ldots$ den Algorithmus aus Satz 3.36 auf die Polynome $p_1, \ldots, p_n, q_1, -q_1, \ldots, q_m, -q_m$ statt $p_1, \ldots, p_n$ und auf $s - (p_1 + \cdots + p_n)$ statt f an, solange bis er eine Darstellung von $s-(p_1+\cdots+p_n)$ als nichtnegative Linearkombination der Polynome $p_1, \ldots, p_n, q_1, -q_1, \ldots, q_m, -q_m, 1$ ausgibt.

(3) Es bezeichne ϱ diejenige $\{p_1, \ldots, p_n\}$-Darstellung modulo $(q_1, \ldots, q_m)$ von $s-(p_1+\cdots+p_n)$, die man erhält, wenn man in dieser Linearkombination die zu $q_1, -q_1, \ldots, q_m, -q_m$ gehörigen Summanden vergißt.

(4) Für jedes $i \in \{1, \ldots, n\}$ ist $\alpha_i := \left(s, s + p_i, \varrho + \sum_{j \neq i} p_j\right)$ eine $\{p_1, \ldots, p_n\}$-Darstellung modulo $(q_1, \ldots, q_m)$ von p_i.

(5) Gebe $(p_1, \alpha_1, \ldots, p_n, \alpha_n)$ als den gewünschten $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis modulo $(q_1, \ldots, q_m)$ aus.

Beweis: Schritt (1) ist klar. Als nächstes zeigen wir, daß die Schleife in Schritt (2) terminiert: Nach Voraussetzung ist die Menge

$$\{x \in \mathbb{R}^d \mid p_{j_1}(x) \geq 0, \ldots, p_{j_l}(x) \geq 0, q_{i_1}(x) \geq 0, -q_{i_1}(x) \geq 0, \ldots, q_{i_k}(x) \geq 0, -q_{i_k}(x) \geq 0\}$$

kompakt und nichtleer. Für genügend großes s ist also $s-(p_1+\cdots+p_n) \geq 0$ auf dieser Menge, und nach Lemma 3.37 ist für solche s das Abbruchkriterium der Schleife aus (2) erfüllt. Die Schritte (3) und (4) sind wieder klar. Um Schritt (5) zu verifizieren, müßen wir noch nachweisen, daß $K[X_1, \ldots, X_d]/(q_1, \ldots, q_m) = K[\overline{p_1}, \ldots, \overline{p_n}]$ gilt. Hierzu sei $i \in \{1, \ldots, d\}$. Wir zeigen $\overline{X_i} \in K[\overline{p_1}, \ldots, \overline{p_n}]$. Für genügend großes $s \in \mathbb{N}$ ist $s + X_i \geq 0$ auf der kompakten nichtleeren Menge

$$\{x \in \mathbb{R}^d \mid p_{j_1}(x) \geq 0, \ldots, p_{j_l}(x) \geq 0, q_{i_1}(x) \geq 0, -q_{i_1}(x) \geq 0, \ldots, q_{i_k}(x) \geq 0, -q_{i_k}(x) \geq 0\}.$$

Nach dem linearen Nichtnegativstellensatz 3.36 ist dann $s + X_i$ eine nichtnegative Linearkombination der Polynome $p_1, \ldots, p_n, q_1, -q_1, \ldots, q_m, -q_m, 1$. Insbesondere gilt $\overline{X_i} \in K[\overline{p_1}, \ldots, \overline{p_n}]$. □

Durch Kombination des obigen Satzes mit Satz 3.32 erhalten wir das Hauptergebnis dieses Abschnitts:

Satz 3.39. *Sei K ein Unterkörper von $\mathbb{R}$. Unter den Polynomen $q_1, \ldots, q_m, p_1, \ldots, p_n \in K[X_1, \ldots, X_d]$, durch welche die Menge*

$$S := \{x \in \mathbb{R}^d \mid q_1(x) = 0, \ldots, q_n(x) = 0, p_1(x) \geq 0, \ldots, p_n(x) \geq 0\}$$

definiert sei, gebe es lineare Polynome $q_{i_1}, \ldots, q_{i_k}, p_{j_1}, \ldots, p_{j_l}$, sodaß die Obermenge

$$\{x \in \mathbb{R}^d \mid q_{i_1}(x) = 0, \ldots, q_{i_k}(x) = 0, p_{j_1}(x) \geq 0, \ldots, p_{j_l} \geq 0\}$$

von S kompakt und nichtleer ist. Es sei $f \in K[X_1, \ldots, X_d]$ mit $f > 0$ auf S. Nach Eingabe von $q_1, \ldots, q_m, p_1, \ldots, p_n, f$ berechnet der folgende Algorithmus (modulo Rechnen im angeordneten Körper K) eine Darstellung von f in der Form

$$(\mathcal{R}_>) \qquad f = a + \sum_e a_e p_1^{e_1} \cdots p_n^{e_n} + \sum_{i=1}^m h_i q_i$$

mit $a \in K^{>0}$, $a_e \in K^{\geq 0}$ und $h_i \in K[X_1, \ldots, X_d]$.

(1) Berechne mit dem Algorithmus aus Satz 3.38 einen $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis modulo $(q_1, \ldots, q_m)$.

(2) Verwende diesen als Eingabe für den Algorithmus aus Satz 3.32, um eine Darstellung von f in der Form $(\mathcal{R}_>)$ zu berechnen.

Soweit uns die bisherige Entwicklungsgeschichte des obigen Satzes bekannt ist, wollen wir sie hier wiedergeben:

- Im Jahr 1921 bemerkt Hausdorff die Existenz der der Darstellung für den Fall $K = \mathbb{R}, d = 1, m = 0, n = 2, p_1 = X, p_2 = 1 - X$ (siehe die Seiten 98 und 99 in [Hau] oder VI, Problem 49, 1. Lösung in [PS]). Er benutzt dabei, daß sich in $\mathbb{R}[X]$ jedes Polynom in lineare und quadratische Faktoren zerlegen läßt.

- In dem 1924 zum ersten Mal aufgelegten Buch [PS] (VI, Problem 49, 2. Lösung) wird ein zweiter Beweis vorgestellt für den Spezialfall $d = 1, m = 0, n = 2, p_1 = X, p_2 = 1 - X$, nun jedoch für einen beliebigen Unterkörper K von $\mathbb{R}$. Dieser zweite Beweis benutzt etliche andere Resultate. Man kann aus ihm einen Algorithmus extrahieren, der bei genauer Betrachtung nichts anderes macht als unserer, obwohl der Satz von Pólya erst 1927 bewiesen wird.

- In dem Buch [HLP], dessen erste Auflage von 1934 stammt, wird am Ende von Kapitel II als Problem 57 die Übungsaufgabe gestellt, mit Hilfe des inzwischen zur Verfügung stehenden Satzes von Pólya, einen Beweis zu geben für den allgemeineren Fall $m = 0, n = d + 1, p_1 = X_1, \ldots, p_{n-1} = X_d, p_n = 1 - (X_1 + \cdots + X_d)$. Danach gerät der Zusammenhang mit dem Satz von Pólya offenbar wieder in Vergessenheit.

- 1985 veröffentlicht Handelman die Abhandlung [Ha1], in der nichtkonstruktiv die reine Existenzaussage für den Fall $m = 0, n = 2d, p_1 = 1 + X_1, p_2 = 1 - X_1, \ldots, p_{2d-1} = 1 + X_d, p_{2d} = 1 - X_d$ bewiesen wird.

- 1988 wird der Artikel [Ha2] desselben Autors veröffentlicht, in dem die reine Existenzaussage zwar nicht so allgemein wie bei uns formuliert wird, im Grunde genommen aber (wiederum nichtkonstruktiv) bewiesen wird unter der zusätzlichen Voraussetzung, daß die Menge

  $$\{x \in \mathbb{R}^d \mid q_{i_1}(x) = 0, \ldots, q_{i_k}(x) = 0, p_{j_1}(x) \geq 0, \ldots, p_{j_l} \geq 0\}$$

  sogar nichtleeres Inneres hat.

- Unter derselben zusätzlichen Voraussetzung wird in der 1998 veröffentlichten Dissertation [Wör] von Wörmann der Satz auf den archimedischen Positivstellensatz 3.24 zurückgeführt, der dort aber wieder nichtkonstruktiv bewiesen wird.

Im nächsten Beispiel wollen wir exemplarisch Polynome $q_1, \ldots, q_m, p_1, \ldots, p_n$ betrachten, für die die Voraussetzungen von Satz 3.39 nicht erfüllt sind. Das Beispiel stammt aus [Ha2], wo allerdings sehr umständlich argumentiert wird.

Beispiel 3.40. Sei K ein Unterkörper von $\mathbb{R}$. Betrachte die Situation

$$d = 2, m = 0, n = 3, p_1 = X_1, p_2 = X_2, p_3 = 1 - (X_1)^2 - (X_2)^2.$$

Die Menge $S := \{x \in \mathbb{R}^2 \mid p_1(x) \geq 0, p_2(x) \geq 0, p_3(x) \geq 0\}$ ist also das rechte obere Viertel der abgeschlossenen Einheitskreisscheibe. Dann ist $\langle K^{\geq 0}, p_1, p_2, p_3 \rangle_0$ nicht archimedisch, es

existiert also kein $\{p_1, p_2, p_3\}$-Archimedizitätsnachweis (modulo dem Nullideal). Angenommen, dies wäre doch so. Dann gäbe es eine Zahl $s \in K^{\geq 0}$ und eine $\{p_1, p_2, p_3\}$-Darstellung von $s - X_1$, also eine Darstellung von $s - X_1$ in der Form

$$s - X_1 = \sum_{i,j,k} a_{(i,j,k)} X_1^i X_2^j (1 - X_1^2 - X_2^2)^k$$

mit $a_{(i,j,k)} \in K^{\geq 0}$ für alle $(i, j, k) \in \mathbb{N}^3$, über die summiert wird. Zum Koeffizienten von X_1 auf der rechten Seite dieser Gleichung tragen offensichtlich nur die zu $i = 1, j = 0$ gehörigen Summanden etwas bei. Genauer gesagt ist der Koeffizient von X_1 auf der rechten Seite der Gleichung gleich $\sum_k a_{(1,0,k)}$. Dieser Koeffizient ist natürlich nicht negativ, da jedes $a_{(i,j,k)}$ nichtnegativ ist. Auf der linken Seite der Gleichung hat aber X_1 einen negativen Koeffizienten. Widerspruch!

Seien nun $a, b \in K^{>0}$. Wir betrachten nun die (für kleines a und b wenig abgeänderte) Situation

$$d = 2, m = 0, n = 3, p_1 = X_1, p_2 = X_2, p_3 = 1 - (X_1 + a)^2 - (X_2 + b)^2.$$

Die Menge $S := \{x \in \mathbb{R}^2 \mid p_1(x) \geq 0, p_2(x) \geq 0, p_3(x) \geq 0\}$ ist nun der Schnitt des rechten oberen Quadranten der Anschauungsebene mit der abgeschlossenen Kreisscheibe um den Mittelpunkt $(-a, -b) \in \mathbb{R}^2$ mit Radius 1. Dann ist $\langle K^{\geq 0}, p_1, p_2, p_3 \rangle_0$ archimedisch. Es ist nämlich

$$\left(\frac{1}{2a}, \frac{1}{2a} + X_1, \frac{1}{2a} \left(\left(1 - (X_1 + a)^2 - (X_2 + b)^2\right) + X_1^2 + a^2 + X_2^2 + 2bX_2 + b^2 \right) \right)$$

eine $\{p_1, p_2, p_3\}$-Archimedizitätsnachweis α_1 für X_1. Analog erhält man einen $\{p_1, p_2, p_3\}$-Archimedizitätsnachweis α_2 für X_2. Dann ist $(X_1, \alpha_1, X_2, \alpha_2)$ ein $\{p_1, p_2, p_3\}$-Archimedizitätsnachweis.

3.5 Der Positivstellensatz von Schmüdgen

Im Jahr 1990 bewies Schmüdgen auf funktionalanalytische Weise den nach ihm benannten Positivstellensatz: Sei durch die endlich vielen Polynome $p_1, \ldots, p_n \in \mathbb{R}[X_1, \ldots, X_d]$ eine *kompakte* Menge $S := \{x \in \mathbb{R}^d \mid p_1(x) \geq 0, \ldots, p_n(x) \geq 0\}$ definiert. Sei $f \in \mathbb{R}[X_1, \ldots, X_d]$ ein weiteres Polynom. Genau dann gilt $f > 0$ auf S, wenn f von der Form

$$f = a + \sum_e \left(\sum_j g_{ej}^2 \right) p_1^{e_1} \cdots p_n^{e_n}$$

ist mit $g_{ej} \in \mathbb{R}[X_1, \ldots, X_d]$ und $a \in \mathbb{R}^{>0}$ (alle Summen sind endlich, o.B.d.A. reicht es über alle $e \in \{0, 1\}^n$ zu summieren). Diese Darstellung entspricht der auf Seite 34 beschriebenen Darstellung $(\mathcal{R}_>^k)$ für $k = 1$, $m = 0$ und $K = \mathbb{R}$, denn in $\mathbb{R}$ ist jede nichtnegative Zahl ein Quadrat.

Man kann dieses Resultat etwas eleganter ausdrücken: Sei Q eine endliche Teilmenge von $\mathbb{R}[X_1, \ldots, X_d]$. Sei $S(Q)$ kompakt. Dann gilt für jedes $f \in \mathbb{R}[X_1, \ldots, X_d]$:

$$f > 0 \text{ auf } S(Q) \iff \text{es gibt } a \in K^{>0} \text{ mit } f - a \in \langle Q \rangle_1$$

Wenn man diese Aussage schon kennt, so wird klar, daß man den archimedischen Positivstellensatz 3.24 von Seite 44 ausnutzen kann, um einen neuen Beweis dafür zu geben. Denn die Aussage sagt offenbar insbesondere aus, daß $\langle Q \rangle_1$ die Voraussetzung des archimedischen Positivstellensatzes erfüllt, archimedisch zu sein. Für jedes $f \in \mathbb{R}[X_1, \ldots, X_d]$ gibt es wegen der Kompaktheit von $S(Q)$ eine Zahl $s \in \mathbb{N}$ mit $s + f > 0$ auf $S(Q)$ und $s - f > 0$ auf $S(Q)$.

Nach der Aussage gibt es dann positive reelle Zahlen a_1 und a_2 mit $s + f - a_1 \in \langle Q \rangle_1$ und $s - f - a_2 \in \langle Q \rangle_1$. Da a_1 und a_2 Quadrate in $\mathbb{R}$ sind, sind sie in $\langle Q \rangle_1$ enthalten, und es folgt $s + f \in \langle Q \rangle_1$ sowie $s - f \in \langle Q \rangle_1$, also $f \in A(\langle Q \rangle_1)$.

In seiner 1998 veröffentlichten Dissertation [Wör] gibt Wörmann tatsächlich auf diese Weise einen neuen Beweis für den Schmüdgenschen Positivstellensatz, indem er nur noch zeigt, daß für jede endliche Teilmenge Q von $\mathbb{R}[X_1, \ldots, X_d]$ mit kompaktem $S(Q)$ der Semiring erster Stufe $\langle Q \rangle_1$ archimedisch ist. Daraus folgt dann mit dem archimedischen Positivstellensatz die Aussage des Satzes von Schmüdgen. Mit Hilfe einer von Becker entwickelten Theorie für höhere Potenzen (siehe [Be2]) zeigt Wörmann sogar mehr:

Satz 3.41. *Sei K ein Unterkörper von $\mathbb{R}$, I ein Ideal von $K[X_1, \ldots, X_d]$ und Q eine endliche Teilmenge von $K[X_1, \ldots, X_d]/I$. Sei $k \in \mathbb{N}$ eine ungerade Zahl. Dann gilt:*

$$S(Q) \text{ ist kompakt} \iff \langle K^{\geq 0} \cup Q \rangle_k \text{ ist archimedisch}$$

Beweis: Siehe [Wör]. (Die Implikation „$\Longleftarrow$“ ist trivial.) □

Wendet man auf diesen Satz den archimedischen Positivstellensatz an, so erhält man die folgende Verallgemeinerung des Satzes von Schmüdgen auf Faktoralgebren von Polynomalgebren über einem beliebigen Unterkörper von $\mathbb{R}$ und $2k$-te Potenzen mit einer beliebigen *ungeraden* Zahl k:

Satz 3.42. *Sei K ein Unterkörper von $\mathbb{R}$, I ein Ideal von $K[X_1, \ldots, X_d]$ und Q eine endliche Teilmenge von $K[X_1, \ldots, X_d]/I$, sodaß $S(Q)$ kompakt ist. Sei $k \in \mathbb{N}$ eine ungerade Zahl. Dann gilt für jedes $f \in K[X_1, \ldots, X_d]/I$:*

$$f > 0 \text{ auf } S(Q) \iff \text{es gibt } a \in K^{>0} \text{ mit } f - a \in \langle K^{\geq 0} \cup Q \rangle_k$$

Bemerkung 3.43. Für kein einziges gerades $k \in \mathbb{N}$ gilt der Satz 3.41 oder der Satz 3.42. Für $k = 0$ haben wir das bereits in Bemerkung 3.40 gesehen. Für gerades $k \geq 2$ erhalten wir ein Gegenbeispiel, indem wir wie in [Wör] die Menge $Q := \{1 - X^k\} \subseteq K[X]$ betrachten (es ist also $d = 1$ und I das Nullideal). Da k gerade ist, ist dann nämlich die Menge $S(Q)$ kompakt. Dennoch liegt für kein $s \in \mathbb{N}$ das Polynom $s + X$ im Semiring $\langle K^{\geq 0} \cup Q \rangle_k$. Denn es besteht sogar für jedes $k \in \mathbb{N}$ der Semiring $\langle K^{\geq 0}, 1 - X^k \rangle_k$ nur aus dem Nullpolynom und aus Polynomen, deren Grad von k geteilt wird. Genauer gilt sogar $\langle K^{\geq 0}, 1 - X^k \rangle_k \subseteq P$, wobei P die Menge aller Polynome $p \in K[X]$ ist, für die gilt: Wenn p einen positiven höchsten Koeffizienten hat, so gilt $\deg p \equiv 0$ modulo $2k$. Wenn p einen negativen höchsten Koeffizienten hat, so gilt $\deg p \equiv k$ modulo $2k$. Offensichtlich gilt $0, 1 \in P$, $P + P \subseteq P$ (beim Addieren zweier Elemente von P können sich niemals die höchsten Koeffizienten auslöschen) und $P \cdot P \subseteq P$. Also ist P ein Semiring, sogar einer von k-ter Stufe, denn jede $2k$-te Potenz eines Elements von $K[X_1, \ldots, X_d]$ hat einen positiven höchsten Koeffizienten und einen Grad, der kongruent zu 0 modulo $2k$ ist, sofern es nicht das Nullpolynom ist. Schließlich gilt offenbar $K^{\geq 0} \cup \{1 - X^k\} \subseteq P$.

Bemerkung 3.44. Auf die Voraussetzung der Endlichkeit von Q kann man in den Sätzen 3.41 und 3.42 nicht verzichten. Um dies einzusehen, betrachten wir die unendliche Menge $Q := \{X - n \mid n \in \mathbb{N}\} \subseteq K[X]$ (es ist also $d = 1$ und I das Nullideal). Dann ist $S(Q) = \emptyset$ kompakt. Für kein $s \in \mathbb{N}$ und kein $k \in \mathbb{N}$ gilt aber $s - X \in \langle Q \rangle_k$. Sonst gäbe es nämlich eine endliche Teilmenge E von Q mit $s - X \in \langle E \rangle_k$. Man könnte offenbar ein $n \in \mathbb{N}$ wählen mit $[n, \infty[\subseteq S(E)$. Es folgte $s - X \geq 0$ auf $[n, \infty[$, was nicht möglich ist.

Bemerkung 3.45. In der Situation des Positivstellensatzes 3.42 gilt für kein einziges $k \in \mathbb{N}$ die Aussage

$$f \geq 0 \text{ auf } S(Q) \iff f \in \langle K^{\geq 0} \cup Q \rangle_k.$$

Wie in [St2] betrachten wir hierzu $Q := \{(1-X^2)^3\} \subseteq K[X]$ und $f := 1-X^2$ (es ist also $d = 1$ und I das Nullideal). Dann ist $S(Q)$ kompakt und es gilt $f \geq 0$ auf $S(Q)$. Dennoch ist f kein Element des Semirings k-ter Stufe $\langle K^{\geq 0}, (1-X^2)^3\rangle_k$. Es ist nämlich $\langle K^{\geq 0}, (1-X^2)^3\rangle_k$ enthalten im Semiring k-ter Stufe P aller Polynome $p \in K[X]$, die an der Stelle 1 einen positiven Wert annehmen oder von $(X-1)^2$ geteilt werden, und f ist nicht einmal ein Element von P.

Wir wollen nun auf den von Wörmann bewiesenen Satz 3.41 nicht nur den archimedischen Positivstellensatz 3.24, sondern sogar dessen algorithmische Version 3.32 anwenden. Dazu sind kleine Modifikationen nötig, da der Algorithmus aus Satz 3.32 ja als Eingabe einen $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis modulo I gebraucht hat, also einen Nachweis der Archimedizität des Semirings (0-ter Stufe) $\langle K^{\geq 0}, \overline{p_1}, \ldots, \overline{p_n}\rangle_0$ in gewisser Form. Satz 3.41 sagt aber für $Q = \{\overline{p_1}, \ldots, \overline{p_n}\}$ nur aus, daß der Semiring *k-ter Stufe* $\langle K^{\geq 0}, \overline{p_1}, \ldots, \overline{p_n}\rangle_k$ archimedisch ist. Deswegen brauchen wir jetzt eine Verallgemeinerung des Begriffs Archimedizitätsnachweis. Dies wird dann der Begriff eines Archimedizitätsnachweis k-ter Stufe sein. In Verallgemeinerung der Definitionen 3.26, 3.27 und 3.28 legen wir also fest:

Definition 3.46. *Sei K ein Unterkörper von $\mathbb{R}$, I ein Ideal von $K[X_1, \ldots, X_d]$ und seien $p_1, \ldots, p_n, f \in K[X_1, \ldots, X_d]$. Sei $k \in \mathbb{N}$. Eine* $\{p_1, \ldots, p_n\}$-Darstellung k-ter Stufe modulo I von f *ist eine Darstellung (wir verzichten hier auf die offensichtliche Formalisierung dieses Begriffs) eines zu f modulo I kongruenten Polynoms in der Form*

$$\sum_e \left(\sum_j a_{ej} g_{ej}^{2k} \right) p_1^{e_1} \cdots p_n^{e_n} \qquad (a_{ej} \in K^{\geq 0}, g_{ej} \in K[X_1, \ldots, X_d]).$$

Genau dann besitzt f eine $\{p_1, \ldots, p_n\}$-Darstellung k-ter Stufe modulo I, wenn

$$\overline{f} \in \langle K^{\geq 0}, \overline{p_1}, \ldots, \overline{p_n}\rangle_k \subseteq K[X_1, \ldots, X_d]/I.$$

Definition 3.47. *Sei K ein Unterkörper von $\mathbb{R}$, I ein Ideal von $K[X_1, \ldots, X_d]$ und seien $p_1, \ldots, p_n, f \in K[X_1, \ldots, X_d]$. Sei $k \in \mathbb{N}$. Ein* $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis k-ter Stufe modulo I für f *ist ein Tripel $(s, \varrho_+, \varrho_-)$ bestehend aus einer Zahl $s \in K^{\geq 0}$ und $\{p_1, \ldots, p_n\}$-Darstellungen k-ter Stufe ϱ_+ bzw. ϱ_- modulo I von $s+f$ bzw. $s-f$.*

Genau dann besitzt f einen $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis k-ter Stufe modulo I, wenn $\overline{f} \in A(\langle K^{\geq 0}, \overline{p_1}, \ldots, \overline{p_n}\rangle_k) \subseteq K[X_1, \ldots, X_d]/I$.

Definition 3.48. *Sei K ein Unterkörper von $\mathbb{R}$, I ein Ideal von $K[X_1, \ldots, X_d]$ und seien $p_1, \ldots, p_n \in K[X_1, \ldots, X_d]$. Sei $k \in \mathbb{N}$. Ein* $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis k-ter Stufe modulo I *ist ein Tupel $(f_1, \alpha_1, \ldots, f_l, \alpha_l)$ mit $f_1, \ldots, f_l \in K[X_1, \ldots, X_d]$, sodaß $K[X_1, \ldots, X_d]/I = K[\overline{f_1}, \ldots, \overline{f_l}]$ und α_i ein $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis k-ter Stufe modulo I für f_i ist für jedes $i \in \{1, \ldots, l\}$.*

Wenn $\langle K^{\geq 0}, \overline{p_1}, \ldots, \overline{p_n}\rangle_k$ archimedisch ist, so existiert natürlich ein $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis k-ter Stufe $(f_1, \alpha_1, \ldots, f_l, \alpha_l)$ modulo I. Dabei kann man $f_1, \ldots, f_l$ sogar beliebig wählen mit der Eigenschaft $K[X_1, \ldots, X_d]/I = K[\overline{f_1}, \ldots, \overline{f_l}]$. Wenn umgekehrt ein $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis k-ter Stufe modulo I existiert, so folgt aus Lemma 3.20, daß $\langle K^{\geq 0}, \overline{p_1}, \ldots, \overline{p_n}\rangle_k$ tatsächlich archimedisch ist. Nach Satz 3.41 gibt es also für ungerades $k \in \mathbb{N}$ genau dann einen $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis k-ter Stufe modulo I, wenn $S(\{p_1, \ldots, p_n\})$ kompakt ist.

Lemma 3.49. *Sei K ein Unterkörper von $\mathbb{R}$. Seien Polynome $q_1, \ldots, q_m, p_1, \ldots, p_n, f \in K[X_1, \ldots, X_d]$ gegeben. Bezeichne I das von $q_1, \ldots, q_m$ erzeugte Ideal in $K[X_1, \ldots, X_d]$. Sei $(f_1, \alpha_1, \ldots, f_l, \alpha_l)$ ein $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis k-ter Stufe modulo I. Ersetzt man im Algorithmus aus Lemma 3.30 überall „Archimedizitätsnachweis " durch „Archimedizitätsnachweis k-ter Stufe", so erhält man einen Algorithmus (modulo Rechnen im angeordneten Körper K), der nach Eingabe von $q_1, \ldots, q_m, (f_1, \alpha_1, \ldots, f_l, \alpha_l)$ und f einen $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis k-ter Stufe modulo I für f berechnet.*

An einer einzigen Stelle unseres Algorithmus aus Satz 3.32 müßen wir eine Änderung vornehmen, die über das bloße Austauschen eines Archimedizitätsnachweises durch einen Archimedizitätsnachweis k-ter Stufe hinausgeht, und zwar bei der Unterprozedur aus Lemma 3.31. Wir müßen uns nun ein bißchen mehr anstrengen, um die Bedingung (iii) sicherzustellen:

Lemma 3.50. *Sei K ein Unterkörper von $\mathbb{R}$. Seien $q_1, \ldots, q_m, p_1, \ldots, p_n \in K[X_1, \ldots, X_d]$. Bezeichne I das von $q_1, \ldots, q_m$ erzeugte Ideal in $K[X_1, \ldots, X_d]$. Es sei $k \in \mathbb{N}$ und es sei $(f_1, \alpha_1, \ldots, f_l, \alpha_l)$ ein $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis k-ter Stufe modulo I. Nach Eingabe von $q_1, \ldots, q_m$ und $(f_1, \alpha_1, \ldots, f_l, \alpha_l)$ berechnet der folgende Algorithmus (modulo Rechnen im angeordneten Körper K) $\{p_1, \ldots, p_n\}$-Darstellungen k-ter Stufe $\varrho_1, \ldots, \varrho_{n+l+1}$ modulo I von Polynomen $p'_1, \ldots, p'_{n+l+1} \in K[X_1, \ldots, X_d]$, sodaß gilt:*

(i) $p'_1 + \cdots + p'_{n+l+1} \equiv 1 \quad$ *modulo* I

(ii) $\langle K^{\geq 0}, \overline{p_1}, \ldots, \overline{p_n} \rangle_0 = \langle K^{\geq 0}, \overline{p'_1}, \ldots, \overline{p'_{n+l+1}} \rangle_0 \quad$ *in* $K[X_1, \ldots, X_d]/I$

(iii) $K[X_1, \ldots, X_d]/I = K[\overline{p'_1}, \ldots, \overline{p'_{n+l+1}}]$

(1) Schreibe $\alpha_i = (s_i, \sigma_i, \tau_i)$ für $i \in \{1, \ldots, l\}$.

(2) Berechne mit dem Algorithmus aus Lemma 3.49 einen $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis k-ter Stufe $(s, \varrho_+, \varrho_-)$ modulo I für $p_1 + \cdots + p_n + (s_1 + f_1) + \cdots + (s_l + f_l)$.

(3) O.B.d.A. ist $s > 0$, ersetze sonst $(s, \varrho_+, \varrho_-)$ durch $(s + 1, \varrho_+ + 1, \varrho_- + 1)$.

(4) Gebe $\varrho_1 := \frac{1}{s} p_1, \ldots, \varrho_n := \frac{1}{s} p_n, \varrho_{n+1} := \frac{1}{s} \sigma_1, \ldots, \varrho_{n+l} := \frac{1}{s} \sigma_l$ und $\varrho_{n+l+1} := \frac{1}{s} \varrho_-$ aus.

Beweis: Es ist ϱ_- eine $\{p_1, \ldots, p_n\}$-Darstellung k-ter Stufe modulo I von

$$s - (p_1 + \cdots + p_n + (s_1 + f_1) + \cdots + (s_l + f_l)).$$

Für die durch $\varrho_1, \ldots, \varrho_{n+l+1}$ dargestellten Polynome $p'_1, \ldots, p'_{n+l+1}$ gilt daher

$$p'_1 \equiv \frac{1}{s} p_1, \ldots, p'_n \equiv \frac{1}{s} p_n, p'_{n+1} \equiv \frac{1}{s}(s_1 + f_1), \ldots, p'_{n+l} \equiv \frac{1}{s}(s_l + f_l)$$

und $p'_{n+l+1} \equiv 1 - (p'_1 + \cdots + p'_{n+l})$ modulo I. Daraus folgen sofort (i) und (ii). Nach Definition 3.48 haben wir $K[X_1, \ldots, X_d]/I = K[\overline{f_1}, \ldots, \overline{f_l}]$. Wegen $K[\overline{p'_{n+1}}, \ldots, \overline{p'_{n+l}}] = K[\overline{f_1}, \ldots, \overline{f_l}]$ gilt also $K[X_1, \ldots, X_d]/I = K[\overline{p'_{n+1}}, \ldots, \overline{p'_{n+l}}]$. Erst recht gilt dann (iii). □

Damit erhalten wir nun das Hauptergebnis dieses Abschnitts:

Satz 3.51. *Sei K ein Unterkörper von $\mathbb{R}$. Durch die Polynome $q_1, \ldots, q_m, p_1, \ldots, p_n \in K[X_1, \ldots, X_d]$ sei die Menge*

$$S := \{x \in \mathbb{R}^d \mid q_1(x) = 0, \ldots, q_m(x) = 0, p_1(x) \geq 0, \ldots, p_n(x) \geq 0\}$$

definiert. Es sei $f \in K[X_1, \ldots, X_d]$ mit $f > 0$ auf S. Es sei $k \in \mathbb{N}$. Es bezeichne I das von $q_1, \ldots, q_m$ in $K[X_1, \ldots, X_d]$ erzeugte Ideal. Nach Eingabe von $q_1, \ldots, q_m, p_1, \ldots, p_n, f$ und

einem $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis k-ter Stufe modulo I (wenn k ungerade und S kompakt ist, existiert ein solcher!) berechnet der folgende Algorithmus (modulo Rechnen im angeordneten Körper K) eine Darstellung von f in der Form

$$(\mathcal{R}^k_>) \qquad f = a + \sum_e \left(\sum_j a_{ej} g_{ej}^{2k} \right) p_1^{e_1} \cdots p_n^{e_n} + \sum_{i=1}^m h_i q_i$$

mit $g_{ej}, h_i \in K[X_1, \ldots, X_d]$, $a \in K^{>0}$ und $a_{ej} \in K^{\geq 0}$.

(1) Wie in Bemerkung 3.1 auf Seite 34 beschrieben, reicht es aus, ein $a \in K^{>0}$ und eine $\{p_1, \ldots, p_n\}$-Darstellung k-ter Stufe ϱ modulo I von $f - a$ zu berechnen. Dies geschieht im Folgenden.

(2) Berechne mit dem Algorithmus aus dem letzten Lemma $\{p_1, \ldots, p_n\}$-Darstellungen k-ter Stufe $\varrho_1, \ldots, \varrho_{n+l+1}$ modulo I von Polynomen $p'_1, \ldots, p'_{n+l+1}$ mit den Eigenschaften (i), (ii) und (iii) aus dem letzten Lemma. Es reicht aus, ein $a \in K^{>0}$ und eine $\{p'_1, \ldots, p'_{n+l+1}\}$-Darstellung ϱ' modulo I von $f - a$ zu berechnen. Durch Komposition von ϱ' mit $\varrho_1, \ldots, \varrho_{n+l+1}$ erhält man dann nämlich ein ϱ wie in (1). Indem wir von $p_1, \ldots, p_n$ zu $p'_1, \ldots, p'_{n+l+1}$ übergehen (dabei ändert sich nach (ii) aus dem letzten Lemma die Menge S nicht), können wir also o.B.d.A. annehmen, daß

$$K[X_1, \ldots, X_d]/I = K[\overline{p_1}, \ldots, \overline{p_n}] \qquad \text{und} \qquad \overline{p_1} + \cdots + \overline{p_n} = 1.$$

(3)-(6) Wörtlich wie im Algorithmus aus Satz 3.32.

Beweis: Nur die Schritte (3)-(6) sind zu verifizieren. Dies wurde aber im Beweis zu Satz 3.32 schon gemacht. □

Wir wollen es uns hier nicht nehmen lassen, eine abgespeckte Version des obigen Satzes zu bringen, die wir gleichzeitig mit einem kleinen Trick aus [Wör] verbinden, sodaß sie neben Satz 3.39 vielleicht am einprägsamsten verdeutlicht, worum es in dieser Arbeit geht.

Durch Polynome $p_1, \ldots, p_n \in \mathbb{Q}[X_1, \ldots, X_d]$ sei die Menge

$$S := \{x \in \mathbb{R}^d \mid p_1(x) \geq 0, \ldots, p_n(x) \geq 0\}$$

definiert. Es sei S beschränkt, d.h. es gebe ein $s \in \mathbb{Q}$ mit

$$(\diamond) \qquad \|x\|^2 < s \quad \textit{für alle} \quad x \in S.$$

Für jedes s mit $(\diamond)$ gibt es dann einen Nachweis der schwächeren Bedingung

$$(\square) \qquad \|x\|^2 \leq s \quad \textit{für alle} \quad x \in S$$

in der besonders schönen Form

$$(\star) \qquad s - (X_1^2 + \cdots + X_d^2) = \sum_e \left(\sum_j g_{ej}^2 \right) p_1^{e_1} \cdots p_n^{e_n} \qquad (g_{ej} \in \mathbb{Q}[X_1, \ldots, X_d]).$$

Der folgende Algorithmus braucht als Eingabe $p_1, \ldots, p_n$, ein $f \in K[X_1, \ldots, X_d]$ mit $f > 0$ auf S und leider einen Nachweis der Beschränktheit von S in der Form $(\star)$. Er liefert dann eine Darstellung von f, die offensichtlich macht, daß $f > 0$ auf S gilt, in der besonders schönen Form

$$f = a + \sum_e \left(\sum_j g_{ej}^2 \right) p_1^{e_1} \cdots p_n^{e_n} \qquad (g_{ej} \in \mathbb{Q}[X_1, \ldots, X_d], a \in \mathbb{Q}^{>0}).$$

(1) Berechne aus dem eingegebenen Nachweis der Beschränktheit von S in der Form $(\star)$ mit Hilfe der $2d$ Gleichungen

$$\frac{s+1}{2} \pm X_i = \frac{1}{2}\left((X_i \pm 1)^2 + (s - (X_1^2 + \cdots + X_d^2)) + \sum_{j \neq i} X_j^2\right) \qquad (i \in \{1, \ldots, d\})$$

einen $\{p_1, \ldots, p_n\}$-Archimedizitätsnachweis erster Stufe (modulo dem Nullideal).

(2) Der Algorithmus aus Satz 3.51 angewandt mit $K = \mathbb{Q}$ und $m = 0$ liefert nach Eingabe dieses Archimedizitätsnachweises erster Stufe eine Darstellung von f in der Form

$$f = a + \sum_e \left(\sum_j a_{ej} g_{ej}^2\right) p_1^{e_1} \cdots p_n^{e_n}$$

mit $g_{ej} \in \mathbb{Q}[X_1, \ldots, X_d]$, $a \in \mathbb{Q}^{>0}$ und $a_{ej} \in \mathbb{Q}^{\geq 0}$.

(3) Schreibe ein jedes a_{ej} in der Form $a_{ej} = \frac{p}{q}$ mit $p \in \mathbb{N}$ und $q \in \mathbb{N}^{>0}$. Wandle die in (2) berechnete Darstellung von f um in eine Darstellung der Form $(\star)$ vermöge der Expansion $a_{ej} = \frac{pq}{q^2} = \left(\frac{1}{q}\right)^2 + \cdots + \left(\frac{1}{q}\right)^2$ von a_{ej} in pq Quadrate.

Für den Beweis ist nur zu bemerken, warum es für jedes s mit $(\diamond)$ die Darstellung $(\star)$ von $s - (X_1^2 + \cdots + X_d^2)$ gibt. Dies folgt aber sofort aus dem Schmüdgen-Positivstellensatz zusammen mit der in (3) beschriebenen Expansion von nichtnegativen rationalen Zahlen in eine Summe von Quadraten.

3.6 Variation über direkte Produkte von Simplizes

Wir betrachten die 4-Form $F \in \mathbb{R}[X_{11}, X_{12}, X_{21}, X_{22}]$ definiert durch

$$\begin{aligned} F :&= X_{11}^2 X_{21}^2 + X_{11}^2 X_{22}^2 + X_{12}^2 X_{21}^2 + X_{12}^2 X_{22}^2 - 3X_{11}X_{12}X_{21}X_{22} \\ &= (X_{11}X_{21} - X_{12}X_{22})^2 + (X_{11}X_{22} - X_{12}X_{21})^2 + X_{11}X_{12}X_{21}X_{22}. \end{aligned}$$

Es gilt nicht $F > 0$ auf $(\mathbb{R}^{\geq 0})^4 \setminus \{0\}$, denn es ist zum Beispiel $F(1, 1, 0, 0) = 0$. Nach der trivialen Richtung des Satzes von Pólya 3.6 gibt es also kein $N \in \mathbb{N}$, sodaß

$$F(X_{11} + X_{12} + X_{21} + X_{22})^N$$

eine Positivform ist. Wie man leicht sieht, gilt jedoch $F > 0$ auf der kleineren Menge

$$((\mathbb{R}^{\geq 0})^2 \setminus \{0\}) \times ((\mathbb{R}^{\geq 0})^2 \setminus \{0\}).$$

Für feste $(x_{11}, x_{12}) \in (\mathbb{R}^{\geq 0})^2 \setminus \{0\}$ nimmt also die 2-Form $F(x_{11}, x_{12}, X_{21}, X_{22}) \in \mathbb{R}[X_{21}, X_{22}]$ nur positive Werte auf $(\mathbb{R}^{\geq 0})^2 \setminus \{0\}$ an. Nach dem Satz von Pólya gibt es daher für alle diese (x_{11}, x_{12}) ein $N_2 \in \mathbb{N}$, sodaß $F(x_{11}, x_{12}, X_{21}, X_{22})(X_{21} + X_{22})^{N_2} \in \mathbb{R}[X_{21}, X_{22}]$ eine Positivform ist. Optimistisch wollen wir einmal unter der Annahme arbeiten, daß wir sogar ein $N_2 \in \mathbb{N}$ finden können, sodaß für alle $(x_{11}, x_{12}) \in (\mathbb{R}^{\geq 0})^2 \setminus \{0\}$ das Polynom $F(x_{11}, x_{12}, X_{21}, X_{22})(X_{21} + X_{22})^{N_2} \in \mathbb{R}[X_{21}, X_{22}]$ eine Positivform ist, mit anderen Worten: Faßt man $F(X_{21} + X_{22})^{N_2}$ als Polynom in den beiden Unbestimmten X_{21} und X_{22} mit Koeffizienten aus $\mathbb{R}[X_{11}, X_{12}]$ auf, so nehmen die Koeffizienten nur positive Werte auf $(\mathbb{R}^{\geq 0})^2 \setminus \{0\}$ an. Da jedes Monom von F in den Unbestimmten X_{11} und X_{12} denselben Grad 2 hat, sind diese Koeffizienten 2-Formen in $\mathbb{R}[X_{11}, X_{12}]$, lassen sich also nach dem Satz von Pólya durch Multiplizieren mit einer genügend großen Potenz von $X_{11} + X_{12}$ zu

einer Positivform in $\mathbb{R}[X_{11}, X_{12}]$ machen. Da es nur endlich viele solche Koeffizienten gibt, finden wir insgesamt ein $N_1 \in \mathbb{N}$, sodaß $F(X_{11}+X_{12})^{N_1}(X_{21}+X_{22})^{N_2}$ von der Form

$$(\star) \qquad \sum_{e_{11}+e_{12}=2+N_1} \sum_{e_{21}+e_{22}=2+N_2} a_e X_{11}^{e_{11}} X_{12}^{e_{12}} X_{21}^{e_{21}} X_{22}^{e_{22}} \qquad (a_e \in \mathbb{R}^{>0})$$

ist. Unter unserer Annahme über die Existenz eines uniformen N_2 haben wir also eine algebraische Beziehung gefunden, die offensichtlich macht, daß $F > 0$ auf $((\mathbb{R}^{\geq 0})^2 \setminus \{0\}) \times ((\mathbb{R}^{\geq 0})^2 \setminus \{0\})$ gilt. Wenn umgekehrt N_1 und N_2 existieren, sodaß $F(X_{11}+X_{12})^{N_1}(X_{21}+X_{22})^{N_2}$ von der Form $(\star)$ ist, so muß offenbar auch unsere Annahme über die Existenz eines uniformen N_2 richtig gewesen sein. Neben dieser Annahme ist entscheidend miteingegangen, daß F nicht nur eine 4-Form, sondern sogar eine $(2,2)$-Form im Sinne der nächsten Definition war.

Zur Notation in diesem Abschnitt: In diesem Abschnitt seien stets u und $d_1, \ldots, d_u$ natürliche Zahlen ≥ 1. Stets sei $d = d_1 + \cdots + d_u$. Es seien d paarweise verschiedene Unbestimmte $X_{11}, \ldots, X_{1d_1}, \ldots, X_{u1}, \ldots, X_{ud_u}$ gegeben. Für jedes $i \in \{1, \ldots, u\}$ schreiben wir $\underline{X_i}$ als Abkürzung für $X_{i1}, \ldots, X_{id_i}$. Wir schreiben $\underline{X}$ als Abkürzung für $\underline{X_1}, \ldots, \underline{X_u}$ (also als Abkürzung für $X_{11}, \ldots, X_{1d_1}, \ldots, X_{u1}, \ldots, X_{ud_u}$). Entsprechend liege für jedes d-Tupel $x \in \mathbb{R}^d$ automatisch die Indizierung $x = (x_{11}, \ldots, x_{1d_1}, \ldots, x_{u1}, \ldots, x_{ud_u})$ vor. Für jedes $i \in \{1, \ldots, u\}$ bezeichnen wir das d_i-Tupel $(x_{i1}, \ldots, x_{id_i})$ durch x_i.

Definition 3.52. *Sei K ein angeordneter Körper. Sei $F \in K[\underline{X}]$. Seien $k_1, \ldots, k_u \in \mathbb{N}$. F heißt $(k_1, \ldots, k_u)$-Form (bezüglich $(\{\underline{X_1}\}, \ldots, \{\underline{X_u}\})$, wenn für alle $i \in \{1, \ldots, n\}$ ein jedes in F vorkommende Monom denselben Grad k_i in den Unbestimmten $\underline{X_i}$ hat. Eine $(k_1, \ldots, k_u)$-Form $F \in K[\underline{X}]$ nennen wir Positivform bezüglich $(\underline{X_1}, \ldots, \underline{X_u})$, wenn sie die Gestalt*

$$F = \sum_{e_{11}+\cdots+e_{1d_1}=k_1} \cdots \sum_{e_{u1}+\cdots+e_{ud_u}=k_u} a_e X_{11}^{e_{11}} \ldots X_{1d_1}^{e_{1d_1}} \ldots X_{u1}^{e_{u1}} \ldots X_{ud_u}^{e_{ud_u}} \qquad (a_e \in K^{>0})$$

hat.

Bemerkung 3.53. Sei $F \in \mathbb{R}[\underline{X}]$ eine $(k_1, \ldots, k_u)$-Form. Dann gilt

$$F(\lambda_1 x_1, \ldots, \lambda_u x_u) = \lambda_1^{k_1} \cdots \lambda_u^{k_u} F(x)$$

für alle $x \in \mathbb{R}^d$ und $\lambda_1, \ldots, \lambda_u \in \mathbb{R}$. Für $\lambda_1 > 0, \ldots, \lambda_u > 0$ ergibt sich: $F(\lambda_1 x_1, \ldots, \lambda_u x_u)$ hat dasselbe Vorzeichen wie $F(x)$. Es folgt, daß für jede $(k_1, \ldots, k_u)$-Form F folgende Äquivalenz gilt:

$$F > 0 \text{ auf } ((\mathbb{R}^{\geq 0})^{d_1} \setminus \{0\}) \times \cdots \times ((\mathbb{R}^{\geq 0})^{d_u} \setminus \{0\}) \iff F > 0 \text{ auf } \Delta_{d_1} \times \cdots \times \Delta_{d_u}.$$

Wir können nun eine unseren Überlegungen entsprechende Vermutung formulieren und beweisen. Wir geben drei Beweise, als erstes einen, der in unserer Situation am bequemsten ist, weil wir schon den archimedischen Positivstellensatz 3.24 zur Verfügung haben. Es handelt sich um die Verallgemeinerung eines in [Wör] von Wörmann gegebenen Beweises des Satzes von Pólya aus dem archimedischen Positivstellensatz (Wörmann ist also die umgekehrte Richtung gegangen wie wir).

Satz 3.54. *Sei $F \in \mathbb{R}[\underline{X}]$ eine $(k_1, \ldots, k_u)$-Form. Genau dann ist*

$$F > 0 \text{ auf } ((\mathbb{R}^{\geq 0})^{d_1} \setminus \{0\}) \times \cdots \times ((\mathbb{R}^{\geq 0})^{d_u} \setminus \{0\}),$$

wenn es ein $N \in \mathbb{N}$ gibt, sodaß

$$F((X_{11} + \cdots + X_{1d_1}) \cdots (X_{u1} + \cdots + X_{ud_u}))^N$$

eine Positivform bezüglich $(\{\underline{X_1}\}, \ldots, \{\underline{X_u}\})$ ist.

Beweis: Die eine Richtung ist trivial. Für die andere Richtung sei vorausgesetzt, daß $F > 0$ auf $((\mathbb{R}^{\geq 0})^{d_1} \setminus \{0\}) \times \cdots \times ((\mathbb{R}^{\geq 0})^{d_u} \setminus \{0\})$ gilt. Betrachte das Ideal

$$I := (X_{11} + \cdots + X_{1d_1} - 1, \ldots, X_{u1} + \cdots + X_{ud_u} - 1) \subseteq \mathbb{R}[\underline{X}]$$

und den Semiring

$$P := \langle \mathbb{R}^{\geq 0}, \overline{X_{11}}, \ldots, \overline{X_{1d_1}}, \ldots, \overline{X_{u1}}, \ldots, \overline{X_{ud_u}} \rangle_0 \subseteq \mathbb{R}[\underline{X}]/I.$$

Für jedes $i \in \{1, \ldots, u\}$ und $j \in \{1, \ldots, d_i\}$ gilt $X_{ij} \in A(P)$, denn es ist $1 + \overline{X_{ij}} \in P$ und $1 - \overline{X_{ij}} = \sum_{l \neq j} \overline{X_{il}} \in P$. Nach Lemma 3.20 ist also P archimedisch. Da nach Voraussetzung $\overline{F} > 0$ gilt auf $\Delta_{d_1} \times \cdots \times \Delta_{d_u} = S(P)$, gibt es nach dem archimedischen Positivstellensatz 3.24 ein $a \in \mathbb{R}^{>0}$ mit $\overline{F} - a \in P$. Es ist also $\overline{F}$ von der Form

$$(\star) \qquad \overline{F} = a + \sum_e a_e \overline{X_{11}}^{e_{11}} \cdots \overline{X_{1d_1}}^{e_{1d_1}} \cdots \overline{X_{u1}}^{e_{u1}} \cdots \overline{X_{ud_u}}^{e_{ud_u}} \qquad (a \in \mathbb{R}^{>0}, a_e \in \mathbb{R}^{\geq 0}).$$

Es ist das Ideal I enthalten im Kern des $\mathbb{R}$-Algebrenhomomorphismus

$$\varphi : \mathbb{R}[\underline{X}] \to \mathbb{R}(\underline{X}) : X_{ij} \mapsto \frac{X_{ij}}{X_{i1} + \cdots + X_{id_i}} \qquad (i \in \{1, \ldots, u\}, j \in \{1, \ldots, d_i\}).$$

Nach dem Homomorphiesatz induziert daher φ einen $\mathbb{R}$-Algebrenhomomorphismus

$$\Phi : \mathbb{R}[\underline{X}]/I \to \mathbb{R}(\underline{X}) : \overline{X_{ij}} \mapsto \frac{X_{ij}}{X_{i1} + \cdots + X_{id_i}} \qquad (i \in \{1, \ldots, u\}, j \in \{1, \ldots, d_i\}).$$

Wir wenden Φ auf beiden Seiten von $(\star)$ an. Da F eine $(k_1, \ldots, k_u)$-Form ist, erhalten wir

$$\frac{F}{(X_{11} + \cdots + X_{1d_1})^{k_1} \cdots (X_{u1} + \cdots + X_{ud_u})^{k_u}} = a + \sum_e a_e \frac{X_{11}^{e_{11}} \cdots X_{1d_1}^{e_{1d_1}} \cdots X_{u1}^{e_{u1}} \cdots X_{ud_u}^{e_{ud_u}}}{(X_{11} + \cdots + X_{1d_1})^{e_{11} + \cdots + e_{1d_1}} \cdots (X_{u1} + \cdots + X_{1d_u})^{e_{u1} + \cdots + e_{ud_u}}}.$$

Wegen $a \in \mathbb{R}^{>0}, a_e \in \mathbb{R}^{\geq 0}$ erhalten wir durch Multiplikation dieser Gleichung mit

$$(X_{11} + \cdots + X_{1d_1})^{k_1 + N} \cdots (X_{u1} + \cdots + X_{ud_u})^{k_u + N}$$

das Gewünschte, wenn wir N so groß wählen, daß alle Nenner weggehen. □

Bemerkung 3.55. Einen zweiten Beweis erhält man durch Verallgemeinerung des Originalbeweises des Satzes von Pólya: Sei wieder $F > 0$ auf $((\mathbb{R}^{\geq 0})^{d_1} \setminus \{0\}) \times \cdots \times ((\mathbb{R}^{\geq 0})^{d_u} \setminus \{0\})$ vorausgesetzt. Wir schreiben

$$F = \sum_{e_{11} + \cdots + e_{1d_1} = k_1} \cdots \sum_{e_{u1} + \cdots + e_{ud_u} = k_u} a_e \prod_{i=1}^{u} \prod_{j=1}^{d_i} X_{ij}^{e_{ij}} \qquad (a_e \in \mathbb{R})$$

und definieren eine neue $(k_1, \ldots, k_u)$-Form $G \in \mathbb{R}[\underline{X}, T_1, \ldots, T_u]$ bezüglich $(\{\underline{X_1}, T_1\}, \ldots, \{\underline{X_u}, T_u\})$ durch

$$G = \sum_{e_{11} + \cdots + e_{1d_1} = k_1} \cdots \sum_{e_{u1} + \cdots + e_{ud_u} = k_u} a_e \prod_{i=1}^{u} \prod_{j=1}^{d_i} \underbrace{X_{ij}(X_{ij} - T_i) \cdots (X_{ij} - (e_{ij} - 1)T_i)}_{e_{ij} \text{ Faktoren}}.$$

Man beachte $G(\underline{X},0)=F$. Wir rechnen für beliebiges $N\in\mathbb{N}$:

$$
\begin{aligned}
&F((X_{11}+\cdots+X_{1d_1})\cdots(X_{u1}+\cdots+X_{ud_u}))^N\\
&=\left(\prod_{i=1}^{u}(X_{i1}+\cdots+X_{id_i})^N\right)\sum_{e_{11}+\cdots+e_{1d_1}=k_1}\cdots\sum_{e_{u1}+\cdots+e_{ud_u}=k_u}a_e\prod_{i=1}^{u}\prod_{j=1}^{d_i}X_{ij}^{e_{ij}}\\
&=\left(\prod_{i=1}^{u}\sum_{l_{i1}+\cdots+l_{id_i}=N}\binom{N}{l_{i1}\dots l_{id_i}}X_{i1}^{l_{i1}}\cdots X_{id_i}^{l_{id_i}}\right)\\
&\qquad\cdot\sum_{e_{11}+\cdots+e_{1d_1}=k_1}\cdots\sum_{e_{u1}+\cdots+e_{ud_u}=k_u}a_e\prod_{i=1}^{u}\prod_{j=1}^{d_i}X_{ij}^{e_{ij}}\\
&=\sum_{e_{11}+\cdots+e_{1d_1}=k_1}\cdots\sum_{e_{u1}+\cdots+e_{ud_u}=k_u}a_e\prod_{i=1}^{u}\sum_{l_{i1}+\cdots+l_{id_i}=N}\binom{N}{l_{i1}\dots l_{id_i}}\prod_{j=1}^{d_i}X_{ij}^{l_{ij}+e_{ij}}\\
&=\sum_{e_{11}+\cdots+e_{1d_1}=k_1}\cdots\sum_{e_{u1}+\cdots+e_{ud_u}=k_u}\\
&\qquad a_e\sum_{l_{11}+\cdots+l_{1d_1}=N}\cdots\sum_{l_{u1}+\cdots+l_{ud_u}=N}\left(\prod_{i=1}^{u}\binom{N}{l_{i1}\dots l_{id_i}}\right)\prod_{i=1}^{u}\prod_{j=1}^{d_i}X_{ij}^{l_{ij}+e_{ij}}\\
&=\sum_{e_{11}+\cdots+e_{1d_1}=k_1}\cdots\sum_{e_{u1}+\cdots+e_{ud_u}=k_u}\ \sum_{\substack{r_{11}+\cdots+r_{1d_1}=k_1+N\\ r_{11}\ge e_{11},\dots,r_{1d_1}\ge e_{1d_1}}}\cdots\sum_{\substack{r_{u1}+\cdots+r_{ud_u}=k_u+N\\ r_{u1}\ge e_{u1},\dots,r_{ud_u}\ge e_{ud_u}}}\\
&\qquad a_e\left(\prod_{i=1}^{u}\binom{N}{(r_{i1}-e_{i1})\dots(r_{id_i}-e_{id_i})}\right)\prod_{i=1}^{u}\prod_{j=1}^{d_i}X_{ij}^{r_{ij}}\\
&=\sum_{r_{11}+\cdots+r_{1d_1}=k_1+N}\cdots\sum_{r_{u1}+\cdots+r_{ud_u}=k_u+N}\\
&\quad\left(\sum_{\substack{e_{11}+\cdots+e_{1d_1}=k_1\\ e_{11}\le r_{11},\dots,e_{1d_1}\le r_{1d_1}}}\cdots\sum_{\substack{e_{u1}+\cdots+e_{ud_u}=k_u\\ e_{u1}\le r_{u1},\dots,e_{ud_u}\le r_{ud_u}}}a_e\left(\prod_{i=1}^{u}\frac{N!}{(r_{i1}-e_{i1})!\cdots(r_{id_i}-e_{id_i})!}\right)\right)\\
&\qquad\cdot\prod_{i=1}^{u}\prod_{j=1}^{d_i}X_{ij}^{r_{ij}}\\
&=\sum_{r_{11}+\cdots+r_{1d_1}=k_1+N}\cdots\sum_{r_{u1}+\cdots+r_{ud_u}=k_u+N}\left(\prod_{i=1}^{u}\frac{N!}{r_{i1}!\cdots r_{id_i}!}\right)\\
&\quad\left(\sum_{\substack{e_{11}+\cdots+e_{1d_1}=k_1\\ e_{11}\le r_{11},\dots,e_{1d_1}\le r_{1d_1}}}\cdots\sum_{\substack{e_{u1}+\cdots+e_{ud_u}=k_u\\ e_{u1}\le r_{u1},\dots,e_{ud_u}\le r_{ud_u}}}a_e\prod_{i=1}^{u}\frac{r_{i1}!\cdots r_{id_i}!}{(r_{i1}-e_{i1})!\cdots(r_{id_i}-e_{id_i})!}\right)\\
&\qquad\cdot\prod_{i=1}^{u}\prod_{j=1}^{d_i}X_{ij}^{r_{ij}}\\
&=\sum_{r_{11}+\cdots+r_{1d_1}=k_1+N}\cdots\sum_{r_{u1}+\cdots+r_{ud_u}=k_u+N}\left(\prod_{i=1}^{u}\frac{N!}{r_{i1}!\cdots r_{id_i}!}\right)
\end{aligned}
$$

$$\left(\sum_{\substack{e_{11}+\cdots+e_{1d_1}=k_1\\ e_{11}\le r_{11},\ldots,e_{1d_1}\le r_{1d_1}}}\cdots\sum_{\substack{e_{u1}+\cdots+e_{ud_u}=k_u\\ e_{u1}\le r_{u1},\ldots,e_{ud_u}\le r_{ud_u}}} a_e\prod_{i=1}^{u}\prod_{j=1}^{d_i}(r_{ij}(r_{ij}-1)\cdots(r_{ij}-(e_{ij}-1)))\right)$$
$$\cdot\prod_{i=1}^{u}\prod_{j=1}^{d_i}X_{ij}^{r_{ij}}$$
$$=\sum_{r_{11}+\cdots+r_{1d_1}=k_1+N}\cdots\sum_{r_{u1}+\cdots+r_{ud_u}=k_u+N}\left(\prod_{i=1}^{u}\frac{N!}{r_{i1}!\cdots r_{id_i}!}\right)$$
$$\left(\sum_{e_{11}+\cdots+e_{1d_1}=k_1}\cdots\sum_{e_{u1}+\cdots+e_{ud_u}=k_u} a_e\prod_{i=1}^{u}\prod_{j=1}^{d_i}(r_{ij}(r_{ij}-1)\cdots(r_{ij}-(e_{ij}-1)))\right)$$
$$\cdot\prod_{i=1}^{u}\prod_{j=1}^{d_i}X_{ij}^{r_{ij}}$$
$$=\sum_{r_{11}+\cdots+r_{1d_1}=k_1+N}\cdots\sum_{r_{u1}+\cdots+r_{ud_u}=k_u+N}\left(\prod_{i=1}^{u}\frac{N!}{r_{i1}!\cdots r_{id_i}!}\right)G(r,\underbrace{1,\ldots,1}_{u\text{-mal}})\prod_{i=1}^{u}\prod_{j=1}^{d_i}X_{ij}^{r_{ij}}$$
$$=\sum_{r_{11}+\cdots+r_{1d_1}=k_1+N}\cdots\sum_{r_{u1}+\cdots+r_{ud_u}=k_u+N}\left(\prod_{i=1}^{u}\frac{N!(k_i+N)^{k_i}}{r_{i1}!\cdots r_{id_i}!}\right)$$
$$\cdot G(\frac{r_1}{k_1+N},\ldots,\frac{r_u}{k_u+N},\frac{1}{k_1+N},\ldots,\frac{1}{k_u+N})\prod_{i=1}^{u}\prod_{j=1}^{d_i}X_{ij}^{r_{ij}}$$

Weil für alle $r\in\mathbb{N}^d$, über die summiert wird, gilt $(\frac{r_1}{k_1+N},\ldots,\frac{r_u}{k_u+N})\in\Delta_{d_1}\times\cdots\times\Delta_{d_u}$, und weil $(\frac{1}{k_1+N},\ldots,\frac{1}{k_u+N})$ für $N\to\infty$ in $\mathbb{R}^u$ gegen 0 konvergiert, genügt es folgendes zu zeigen: Es gibt eine Umgebung V von 0 in $\mathbb{R}^u$, sodaß für alle $x\in\Delta_{d_1}\times\cdots\times\Delta_{d_u}$ und $t\in V$ gilt $G(x,t)>0$.

Diese Umgebung V erhält man wie folgt: Für jedes $x\in\Delta_{d_1}\times\cdots\times\Delta_{d_u}$ ist $G(x,0)=F(x)>0$, also gibt es eine Umgebung von $(x,0)$ in $\mathbb{R}^d\times\mathbb{R}^u$, auf der $G>0$ gilt. Wir wählen zu jedem $x\in\Delta_{d_1}\times\cdots\times\Delta_{d_u}$ eine solche Umgebung von $(x,0)$, o.B.d.A. eine von der Form $U_x\times V_x$, wobei U_x eine offene Umgebung von x in $\mathbb{R}^d$ und V_x eine Umgebung von 0 in $\mathbb{R}^u$ ist. Die Menge $\{U_x\mid x\in\Delta_{d_1}\times\cdots\times\Delta_{d_u}\}$ bildet eine Überdeckung der kompakten Menge $\Delta_{d_1}\times\cdots\times\Delta_{d_u}$ durch offene Mengen. Daher gibt es eine Teilüberdeckung $\{U_x\mid x\in X\}$ mit einer endlichen Teilmenge X von $\Delta_{d_1}\times\cdots\times\Delta_{d_u}$. Wir setzen $V=\bigcap\{V_x\mid x\in X\}$. Da X endlich ist, ist V wieder eine Umgebung von 0 in $\mathbb{R}^u$.

Außerdem leistet V das Gewünschte: Sei $x\in\Delta_{d_1}\times\cdots\times\Delta_{d_u}$ und $t\in V$. Wir zeigen $G(x,t)>0$. Wähle $y\in X$ mit $x\in U_y$. Dann gilt $(x,t)\in U_y\times V\subseteq U_y\times V_y$. Wegen $G>0$ auf $U_y\times V_y$ folgt $G(x,t)>0$.

Bemerkung 3.56. Einen dritten Beweis für Satz 3.54 erhält man, indem man die Existenz des im einleitenden Beispiel „uniformen N_2“ direkt beweist. Dazu brauchen wir allerdings die Schranke (i) aus der Bemerkung 3.9, die wir in dieser Arbeit nicht bewiesen haben.

Wir zeigen durch Induktion nach u, daß es zu jeder $(k_1,\ldots,k_u)$-Form $F\in\mathbb{R}[\underline{X_1},\ldots,\underline{X_u}]$ mit $F>0$ auf $((\mathbb{R}^{\ge0})^{d_1}\setminus\{0\})\times\cdots\times((\mathbb{R}^{\ge0})^{d_u}\setminus\{0\})$ Exponenten $N_1,\ldots,N_u\in\mathbb{N}$ gibt, sodaß $F(X_{11}+\cdots+X_{1d_1})^{N_1}\cdots(X_{u1}+\cdots+X_{ud_u})^{N_u}$ eine Positivform bezüglich $(\{\underline{X_1}\},\ldots,\{\underline{X_u}\})$ ist.

Für den Induktionsanfang $u=1$ ist das gerade der Satz von Pólya. Sei nun im Induktionsschritt $u\ge2$ und die Behauptung schon für $u-1$ statt u gezeigt. Sei $F\in\mathbb{R}[\underline{X_1},\ldots,\underline{X_u}]$ mit $F>0$ auf $((\mathbb{R}^{\ge0})^{d_1}\setminus\{0\})\times\cdots\times((\mathbb{R}^{\ge0})^{d_u}\setminus\{0\})$. Für jedes $x\in\Delta_{d_1}\times\cdots\times\Delta_{d_{u-1}}$ ist $F(x,\underline{X_u})\in\mathbb{R}[\underline{X_u}]$ eine k_u-Form, die auf $(\mathbb{R}^{\ge0})^{d_u}\setminus\{0\}$ nur positive Werte annimmt.

Nach der Schranke (i) aus Bemerkung 3.9 gilt dann für diese x, daß für jedes $N \in \mathbb{N}$ mit $N \geq \frac{d_u k_u (b_x k_u + 1)}{\mu_x}$ das Polynom $F(x, \underline{X_u})(X_{u1} + \cdots + X_{ud_u})^N \in \mathbb{R}[\underline{X_u}]$ eine Positivform ist. Hierbei bezeichne $\mu_x > 0$ das Minimum von $F(x, \underline{X_u})$ auf Δ_{d_u} und b_x das Maximum der Beträge der Koeffizienten von $F(x, \underline{X_u})$. Auf der kompakten Menge $\Delta_{d_1} \times \cdots \times \Delta_{d_u}$ nimmt F ein Minimum $\mu > 0$ an, nämlich $\mu = \min\{\mu_x \mid x \in \Delta_{d_1} \times \cdots \times \Delta_{d_{u-1}}\}$. Da die auf der kompakten Menge $\Delta_{d_1} \times \cdots \times \Delta_{d_{u-1}}$ definierte Abbildung $x \mapsto b_x$ stetig ist, können wir außerdem $b = \max\{b_x \mid x \in \Delta_{d_1} \times \cdots \times \Delta_{d_{u-1}}\}$ setzen. Wir wählen nun $N_u \in \mathbb{N}$ so, daß $N_u \geq \frac{d_u k_u (b k_u + 1)}{\mu}$. Für alle $x \in \Delta_{d_1} \times \cdots \times \Delta_{d_{u-1}}$ ist dann $N_u \geq \frac{d_u k_u (b_x k_u + 1)}{\mu_x}$ und damit $F(x, \underline{X_u})(X_{u1} + \cdots + X_{ud_u})^{N_u} \in \mathbb{R}[\underline{X_u}]$ eine Positivform. Mit anderen Worten: Faßt man $F(X_{u1} + \cdots + X_{ud_u})^{N_u}$ als Polynom in den Unbestimmten $\underline{X_u}$ mit $(k_1, \ldots, k_{u-1})$-Formen aus $\mathbb{R}[\underline{X_1}, \ldots, \underline{X_{u-1}}]$ als Koeffizienten auf, so nehmen alle diese Koeffizienten auf $\Delta_{d_1} \times \cdots \times \Delta_{d_{u-1}}$ nur positive Werte an. Nach Induktionsvoraussetzung findet man also zu jedem dieser Koeffizienten Exponenten $N_1, \ldots, N_{u-1} \in \mathbb{N}$, sodaß der Koeffizient nach Multiplizieren mit $(X_{11} + \cdots + X_{1d_1})^{N_1} \cdots (X_{(u-1)1} + \cdots + X_{(u-1)d_{u-1}})^{N_{u-1}}$ eine Positivform bezüglich $(\{\underline{X_1}\}, \ldots, \{\underline{X_{u-1}}\})$ wird. Da es nur endlich viele solche Koeffizienten gibt, können wir $N_1, \ldots, N_{u-1} \in \mathbb{N}$ sogar so wählen, daß jeder der Koeffizienten von $F(X_{11} + \cdots + X_{1d_1})^{N_1} \cdots (X_{u1} + \cdots + X_{ud_u})^{N_u}$ (aufgefaßt als Polynom in $\underline{X_u}$) eine Positivform bezüglich $(\{\underline{X_1}\}, \ldots, \{\underline{X_{u-1}}\})$ ist. Dann ist $F(X_{11} + \cdots + X_{1d_1})^{N_1} \cdots (X_{u1} + \cdots + X_{ud_u})^{N_u}$ eine Positivform bezüglich $(\{\underline{X_1}\}, \ldots, \{\underline{X_u}\})$.

Die dreifach bewiesene Verallgemeinerung des Satzes von Pólya kann man nun verwenden, um für Spezialfälle Varianten unseres Algorithmus aus Satz 3.32 anzugeben, deren einziger Vorteil gegenüber dem schon bekannten Algorithmus es ist, daß sie vergleichsweise sehr übersichtliche Transformationen durchführen, insbesondere auf den Einsatz einer neuen Unbestimmten C verzichten. Würde man für die Verallgemeinerung des Satzes von Pólya dann Komplexitätsabschätzungen analog zu Bemerkung 3.9 machen, so könnte man wie nach Satz 3.32 diskutiert wohl die Komplexität dieser Varianten unseres Algorithmus abschätzen. Ursprünglich entstanden sind diese Varianten noch vor dem Algorithmus selber, als dem Autor noch nicht aufgefallen ist, daß man Lemma 3.10 durch Lemma 3.11, aus dem ja die Unbestimmte C stammt, in erstaunlicherweise zu Lemma 3.13 verallgemeinern kann.

Wir präsentieren nun als Anwendungsbeispiel der Verallgemeinerung des Satzes von Pólya den folgenden Satz zur Darstellung von Polynomen, die auf einem direkten Produkt von Simplizes (in kanonischer Position) positiv sind. Es sei ausdrücklich noch einmal vermerkt, daß der einzige neue Gehalt dieses Satzes ist, daß wir einen *übersichtlichen* Algorithmus zur Berechnung der Darstellung haben. Schon in Satz 3.39 haben wir ja ein viel allgemeineres Problem mit einem höchstens unmerklich langsameren Algorithmus gelößt.

Satz 3.57. *Sei K ein Unterkörper von $\mathbb{R}$. Durch $1 \leq d_1, \ldots, d_u \in \mathbb{N}$ sei die Menge*

$$\begin{aligned} S := \{x \in \mathbb{R}^d \mid \; & x_{11} \geq 0, \ldots, x_{1d_1} \geq 0, x_{11} + \cdots + x_{1d_1} \leq 1, \\ & \vdots \\ & x_{u1} \geq 0, \ldots, x_{ud_u} \geq 0, x_{u1} + \cdots + x_{ud_u} \leq 1\} \end{aligned}$$

definiert. Es sei $f \in K[\underline{X}]$ mit $f > 0$ auf S. Nach Eingabe von $d_1, \ldots, d_u$ und f berechnet der folgende Algorithmus (modulo Rechnen im angeordneten Körper K) eine Darstellung von f in der Form

$$f = a + \sum_e a_e \prod_{i=1}^{u} (1 - (X_{i1} + \cdots + X_{id_i}))^{e_{i0}} X_{i1}^{e_{i1}} \cdots X_{id_i}^{e_{id_i}} \qquad (a \in K^{>0}, a_e \in K^{\geq 0}).$$

(1) Für jedes $i \in \{1, \ldots, u\}$ bezeichne k_i den Grad von f in den Unbestimmten $\underline{X_i}$. Schreibe

$$f = \sum_e b_e X_{11}^{e_{11}} \cdots X_{1d_1}^{e_{1d_1}} \cdots X_{u1}^{e_{u1}} \cdots X_{ud_u}^{e_{ud_u}} \qquad (b_e \in K)$$

und definiere eine $(k_1,\dots,k_u)$-Form $G \in K[\underline{X},Y_1,\dots,Y_u]$ bezüglich $(\{\underline{X_1},Y_1\},\dots,\{\underline{X_u},Y_u\})$ durch

$$G := \sum_e b_e \prod_{i=1}^{u} X_{i1}^{e_{i1}} \cdots X_{id_i}^{e_{id_i}} (X_{i1}+\cdots+X_{id_i}+Y_i)^{k_i-(e_{i1}+\cdots+e_{id_i})}.$$

(2) Bilde für $N = 0,1,2,3,\dots$ die $(k_1+N,\dots,k_u+N)$-Form

$$H_N := G((X_{11}+\cdots+X_{1d_1}+Y_1)\cdots(X_{u1}+\cdots+X_{ud_u}+Y_u))^N \in K[\underline{X},Y_1,\dots,Y_u]$$

bezüglich $(\{\underline{X_1},Y_1\},\dots,\{\underline{X_u},Y_u\})$ und überprüfe, ob sie eine Positivform bezüglich $(\{\underline{X_1},Y_1\},\dots,\{\underline{X_u},Y_u\})$ ist. Für das kleinste N, für das dies der Fall ist, setze $F := H_N$.

(3) Berechne ein $a \in K^{>0}$, sodaß die $(k_1+N,\dots,k_u+N)$-Form

$$F - a(X_{11}+\cdots+X_{1d_1}+Y_1)^{k_1+N}\cdots(X_{u1}+\cdots+X_{ud_u}+Y_u)^{k_u+N}$$

keine negativen Koeffizienten besitzt, sich also schreiben läßt in der Form

$$\sum_e a_e {Y_1}^{e_{10}} X_{11}^{e_{11}} \cdots X_{1d_1}^{e_{1d_1}} \cdots {Y_u}^{e_{u0}} X_{u1}^{e_{u1}} \cdots X_{ud_u}^{e_{ud_u}} \qquad (a_e \in K^{\geq 0}).$$

Durch a und a_e ist die gewünschte Darstellung gegeben.

Beweis: Betrachte den K-Algebrenhomomorphismus

$$\psi : \begin{cases} K[\underline{X},Y_1,\dots,Y_u] \to K[\underline{X}] & \\ X_{ij} \mapsto X_{ij} & \text{für } i \in \{1,\dots,u\} \text{ und } j \in \{1,\dots,d_i\} \\ Y_i \mapsto 1-(X_{i1}+\cdots+X_{id_i}) & \text{für } i \in \{1,\dots,u\}. \end{cases}$$

Offensichtlich ist ψ surjektiv. Wir behaupten, daß der Kern von ψ gleich dem Ideal

$$J := (X_{11}+\cdots+X_{1d_1}+Y_1-1,\dots,X_{u1}+\cdots+X_{ud_u}+Y_u-1) \subseteq K[\underline{X},Y_1,\dots,Y_u]$$

ist. Sicherlich ist J in diesem Kern enthalten. Sei umgekehrt r ein Element des Kerns von ψ, also $r \in K[\underline{X},Y_1,\dots,Y_u]$ mit

$$\psi(r) = r(\underline{X}, 1-(X_{11}+\cdots+X_{1d_1}),\dots,1-(X_{u1}+\cdots+X_{ud_u})) = 0.$$

Dann ist $\psi(r) = 0$ modulo J offensichtlich kongruent zu r in $K[\underline{X},Y_1,\dots,Y_u]$. Also ist r ein Element von J. Insgesamt ist also J der Kern von ψ. Durch ψ wird nun ein K-Algebrenisomorphismus

$$\Psi : \begin{cases} K[\underline{X},Y_1,\dots,Y_u]/J \to K[\underline{X}] & \\ \overline{X_{ij}} \mapsto X_{ij} & \text{für } i \in \{1,\dots,u\} \text{ und } j \in \{1,\dots,d_i\} \\ \overline{Y_i} \mapsto 1-(X_{i1}+\cdots+X_{id_i}) & \text{für } i \in \{1,\dots,u\}. \end{cases}$$

induziert. Nach Voraussetzung gilt $f > 0$ auf

$$S(\{\underline{X}, 1-(X_{11}+\cdots+X_{1d_1}),\dots,1-(X_{u1}+\cdots+X_{ud_u})\}).$$

Gemäß Lemma 3.18 ist dies gleichbedeutend mit $\Psi^{-1}(f) > 0$ auf

$$S(\Psi^{-1}(\{\underline{X}, 1-(X_{11}+\cdots+X_{1d_1}),\dots,1-(X_{u1}+\cdots+X_{ud_u})\})).$$

Nun gilt $\Psi^{-1}(f) = \overline{f} = \overline{G}$ und

$$S(\Psi^{-1}(\{\underline{X}, 1-(X_{11}+\cdots+X_{1d_1}),\dots,1-(X_{u1}+\cdots+X_{ud_u})\})) =$$
$$\{(x,y_1,\dots,y_u) \in (\mathbb{R}^{\geq 0})^{d+u} \mid x_{11}+\cdots+x_{1d_1}+y_1 = 1,\dots,x_{u1}+\cdots+x_{ud_u}+y_u = 1\}.$$

Nach Bemerkung 3.53 gilt $G > 0$ auf

$$\{(x, y_1, \ldots, y_u) \in (\mathbb{R}^{\geq 0})^{d+u} \mid x_{11} + \cdots + x_{1d_1} + y_1 \neq 0, \ldots, x_{u1} + \cdots + x_{ud_u} + y_u \neq 0\}.$$

Nach der Verallgemeinerung 3.54 des Satzes von Pólya 3.54 terminiert daher die Schleife in (2). Daß durch a und a_e wie in (3) berechnet die gewünschte Darstellung von F gegeben ist, überprüft man sofort, indem man

$$\psi(F - a(X_{11} + \cdots + X_{1d_1} + Y_1)^{k_1+N} \cdots (X_{u1} + \cdots + X_{ud_u} + Y_u)^{k_u+N}) = f - a$$

nachrechnet. □

Kapitel 4

Der Darstellungssatz von Kadison-Dubois

Ein angeordneter Körper wird bekanntlich genau dann als archimedisch bezeichnet, wenn im Sinne der Definition 3.19 der Semiring $K^{\geq 0}$ der nichtnegativen Elemente von K archimedisch in K ist. Bekannt ist auch der Satz, daß es für jeden archimedisch angeordneten Körper K einen (und nur einen) Ringhomomorphismus $\varphi : K \to \mathbb{R}$ mit $\varphi(K^{\geq 0}) \subseteq \mathbb{R}^{\geq 0}$ gibt (wir erinnern hier daran, daß wir unter einem Ring stets einen kommutativen Ring mit 1 und dementsprechend unter einem Ringhomomorphismus stets einen Homomorphismus von Ringen mit 1 verstehen). Da K ein Körper ist, ist φ dann eine Einbettung des angeordneten Körpers K in den angeordneten Körper $\mathbb{R}$.

Hier werden wir einen viel allgemeineren Satz beweisen. Statt des archimedischen Semirings $K^{\geq 0}$ in einem angeordneten Körper K betrachten wir jetzt einen archimedischen Semiring P in einem Ring R. Genauso wie wir vorher nach den Ringhomomorphismen $\varphi : K \to \mathbb{R}$ mit $\varphi(K^{\geq 0}) \subseteq \mathbb{R}^{\geq 0}$ gefragt haben, betrachten wir jetzt die Menge

$$X(P) := \{\varphi : R \to \mathbb{R} \mid \varphi \text{ Ringhomomorphismus}, \varphi(P) \subseteq \mathbb{R}^{\geq 0}\}.$$

Man beachte, daß diese Definition von $X(P)$ sich nicht ganz mit der bisherigen Definition 3.16 von $X(P)$ verträgt. Wir verwenden ab jetzt die neue Definition. Wenn $-1 \in P$ gilt (wegen der Archimedizität von P gilt dann sogar $P = R$), so ist $X(P)$ offensichtlich leer. Der Satz wird unter anderem sagen, daß in allen anderen Fällen $X(P)$ nicht leer ist.

Beispiele für die betrachtete Situation, daß ein Ring mit einer archimedischen Präprimstelle vorliegt, werden geliefert durch kompakte topologische Räume X (ein kompakter Raum muß bei uns nicht notwendig Hausdorffsch sein): Man betrachte den Ring $\mathcal{C}(X, \mathbb{R})$ der stetigen Funktionen von X nach $\mathbb{R}$. Der Semiring $\mathcal{C}^+(X, \mathbb{R})$ aller stetigen Funktionen von X nach $\mathbb{R}$, die keine negativen Werte annehmen, ist dann archimedisch in $\mathcal{C}(X, \mathbb{R})$.

Nun sind dies nicht die einzigen Beispiele (bis auf Isomorphie) für die betrachtete Situation. Wenn man etwa einen nicht archimedisch angeordneten Körper K nimmt und darin den Unterring $R := A(K^{\geq 0})$ aller endlichen Elemente betrachtet, so ist $P := R^{\geq 0}$ ein archimedischer Semiring in R. Es kann aber (R, P) als Ring mit ausgezeichnetem Semiring nicht isomorph sein zu $(\mathcal{C}(X, \mathbb{R}), \mathcal{C}^+(X, \mathbb{R}))$ mit einem kompakten Raum X: Es gibt nämlich in R ein negatives infinitesimales Element a, also ein $a \in R$ mit $a \notin P$, aber $1 + na \in P$ für alle $n \in \mathbb{N}$. Dagegen muß jede Funktion $A \in \mathcal{C}(X, \mathbb{R})$, für welche alle Funktionen $1 + nA$ ($n \in \mathbb{N}$) in $\mathcal{C}^+(X, \mathbb{R})$ liegen, schon selber ein Element von $\mathcal{C}^+(X, \mathbb{R})$ sein.

Der Darstellungssatz von Kadison-Dubois sagt aber in sehr expliziter Weise, wie weit man im Allgemeinen von diesen Beispielen entfernt ist. Er gibt konkret einen kompakten Raum X

sowie einen Ringhomomorphismus $\Phi : R \to \mathcal{C}(X, \mathbb{R})$ mit $\Phi(P) \subseteq \mathcal{C}^+(X, \mathbb{R})$ an und sagt sehr konkret etwas darüber aus, wie weit Φ davon entfernt ist, ein Isomorphismus von Ringen mit ausgezeichnetem Semiring von (R, P) nach $(\mathcal{C}(X, \mathbb{R}), \mathcal{C}^+(X, \mathbb{R}))$ zu sein: Wieviel zur Injektivität fehlt, wird geklärt, indem der Kern von Φ bestimmt wird. Es wird gezeigt, daß Φ „annähernd surjektiv“ ist. Und es wird $\Phi^{-1}(\mathcal{C}^+(X, \mathbb{R}))$ explizit bestimmt, also diejenige Menge, zu der man P vergrößern müßte, damit die bestehende Inklusion $\Phi(P) \subseteq \mathcal{C}^+(X, \mathbb{R})$ wenigstens zu einer Gleichheit $\Phi(P) = \mathcal{C}^+(X, \mathbb{R}) \cap \Phi(R)$, wenn schon nicht zu der Gleichheit $\Phi(P) = \mathcal{C}^+(X, \mathbb{R})$ würde.

Wie sollten wir nun den topologischen Raum X wählen, damit die Chancen dafür, daß der gewünschte Ringhomomorphismus $\Phi : R \to \mathcal{C}(X, \mathbb{R})$ mit $\Phi(P) \subseteq \mathcal{C}^+(X, \mathbb{R})$ ein Isomorphismus mit $\Phi(P) = \mathcal{C}^+(X, \mathbb{R})$ wird, möglichst gut stehen? Um diese Frage zu klären, wollen wir vorerst die Topologie außer acht lassen. Sei also X eine bloße Menge und $\Phi : R \to \mathbb{R}^X$ ein Ringhomomorphismus mit $\Phi(P) \subseteq (\mathbb{R}^{\geq 0})^X$. Wir müßen zum Beispiel versuchen, den Kern von Φ klein zu halten. Der Kern von Φ ist der Schnitt über die Kerne der Ringhomomorphismen $\varphi : R \to \mathbb{R} : a \mapsto \Phi(a)(x)$ $(x \in X)$. Für alle diese φ gilt $\varphi(P) \subseteq \mathbb{R}^{\geq 0}$, also $\varphi \in X(P)$. Möglichst viele $\varphi \in X(P)$ sollten ihren Kern zu diesem Schnitt beitragen. Wir werden X so wählen, daß *alle* dies tun. Dann muß es zu jedem $\varphi \in X(P)$ ein $x \in X$ geben mit $\Phi(a)(x) = \varphi(a)$ für alle $a \in R$. Die Lösung liegt nahe: Das zu φ gehörige x wird φ selber sein und $\Phi(a)$ wird die Auswertung im Punkt a sein. Wir wählen also $X := X(P)$ und $\Phi : R \to \mathbb{R}^X : a \mapsto \hat{a}$ mit $\hat{a} : X \to \mathbb{R} : \varphi \mapsto \varphi(a)$ für alle $a \in R$. Daß Φ ein Ringhomomorphismus ist, rechnet man sofort nach.

Nun müßen wir uns um die Topologie kümmern. Einerseits soll die Abbildung $\Phi : R \to \mathbb{R}^X : a \mapsto \hat{a}$ auch noch wohldefiniert sein, wenn wir $\Phi : R \to \mathcal{C}(X, \mathbb{R}) : a \mapsto \hat{a}$ schreiben. Die Topologie auf $X = X(P)$ soll also fein genug sein, um alle Funktionen $\hat{a} : X \to \mathbb{R}$ $(a \in R)$ stetig zu machen. Andererseits soll sie grob genug sein, um $X(P)$ zu einem kompakten Raum zu machen. Auch hier liegt die Lösung nahe: Man statte $X(P)$ mit der gröbsten Topologie aus, die alle Funktionen $\hat{a} : X \to \mathbb{R}$ $(a \in R)$ stetig macht, also mit der Initialtopologie (auch schwache Topologie genannt) bezüglich all dieser Funktionen. Nun formulieren wir den Gegenstand dieses Kapitels, den Darstellungssatz von Kadison-Dubois für Ringe mit einem ausgezeichneten Semiring.

Satz 4.1 (Kadison-Dubois). *Sei R ein Ring und P ein archimedischer Semiring in R mit $-1 \notin P$. Wir statten die Menge*

$$X(P) := \{\varphi : R \to \mathbb{R} \mid \varphi \text{ Ringhomomorphismus}, \varphi(P) \subseteq \mathbb{R}^{\geq 0}\}.$$

mit der Initialtopologie bezüglich aller Funktionen $\hat{a} : X(P) \to \mathbb{R} : \varphi \mapsto \varphi(a)$ $(a \in R)$ aus und betrachten den Ringhomomorphismus

$$\Phi : R \to \mathcal{C}(X(P), \mathbb{R}) : a \mapsto \hat{a}.$$

Dann gilt:

(i) $X(P)$ ist ein nichtleerer kompakter Hausdorffraum.

(ii) $\mathcal{C}^+(X(P), \mathbb{R}) := \{A \in \mathcal{C}(X(P), \mathbb{R}) \mid A(\varphi) \geq 0$ für alle $\varphi \in X(P)\}$ ist ein archimedischer Semiring in $\mathcal{C}(X(P), \mathbb{R})$, für den gilt:

$$\Phi^{-1}(\mathcal{C}^+(X(P), \mathbb{R})) = \{a \in R \mid \text{für alle } n \in \mathbb{N} \text{ gibt es ein } t \in \mathbb{N}^{>0} \text{ mit } t(1 + na) \in P\}.$$

(iii) Φ hat den Kern

$$\{a \in R \mid \text{für alle } n \in \mathbb{N} \text{ gibt es ein } t \in \mathbb{N}^{>0} \text{ mit } t(1 + na) \in P \text{ und } t(1 - na) \in P\}.$$

(iv) $\mathbb{Q} \cdot \Phi(R)$ liegt dicht in $\mathcal{C}(X(P), \mathbb{R})$ bezüglich der Supremumsnorm auf $\mathcal{C}(X(P), \mathbb{R})$.

Den Beweis des Satzes verschieben wir noch. Wir können jedoch sofort einsehen:

Bemerkung 4.2. Gilt in obigem Satz entgegen der Voraussetzung doch $-1 \in P$, so gelten alle Aussagen des Satzes weiterhin mit folgenden beiden Ausnahmen: In (i) ist $X(P) = \emptyset$. Deswegen macht es in (iv) keinen Sinn mehr von der Supremumsnorm zu sprechen.

Man sieht dies wie folgt: Daß $X(P)$ leer ist, ist klar. Damit ist $X(P)$ ein kompakter Hausdorffraum. Die Menge $\mathcal{C}(X(P), \mathbb{R})$ hat dann genau ein einziges Element, nämlich die leere Funktion. Diese ist auch in $\mathcal{C}^+(X(P), \mathbb{R})$ enthalten. Also gilt $\mathcal{C}(X(P), \mathbb{R}) = \mathcal{C}^+(X(P), \mathbb{R})$. Insbesondere ist dann $\mathcal{C}^+(X(P), \mathbb{R})$ ein archimedischer Semiring in $\mathcal{C}(X(P), \mathbb{R})$ und es gilt $\Phi^{-1}(\mathcal{C}^+(X(P), \mathbb{R})) = R$. Wenn man noch $P = R$ beachtet ($-1 \in P$ impliziert dies zusammen mit der Archimedizität von P), so werden die in (ii) und (iii) behaupteten Gleichheiten trivial.

Für den Darstellungssatz von Kadison-Dubois in obiger Formulierung waren bis jetzt zwei Beweise bekannt, zu denen wir hier einen dritten hinzufügen werden. Der erste und längste Beweis benutzt funktionalanalytische Methoden und zeigt im Gegensatz zu den beiden anderen Beweisen, daß der Satz sogar gelten würde, wenn wir von der Multiplikation eines Ringes weder Kommutativität noch Assoziativität fordern würden. Dieser Beweis liegt in drei Arbeiten verstreut. In [Kad] findet man den ersten Spezialfall, der von Stone 1941 behandelt wurde, sowie die Erweiterung dieses Resultats durch Kadison im Jahre 1951. Dubois verallgemeinerte dies 1967 in [Dub]. Schließlich brachte 1979 Becker in [Be1] den Satz in die obige Form. Seit 1983 gibt es einen sehr viel kürzeren und weitgehend algebraischen Beweis von Becker und Schwartz (siehe [BS]). Dort wird zwar auch nur der Fall behandelt, daß ein Ring im Sinne dieser Arbeit vorliegt, jedoch wird statt eines archimedischen Semirings sogar ein „Modul“ über einem archimedischen Semiring betrachtet.

In diesem Kapitel geben wir einen neuen Beweis, der eine durch technische Details angereicherte Version unseres Beweises des archimedischen Positivstellensatzes 3.24 ist, also wesentlich den Satz von Pólya verwendet. Der Beweis kann an Kürze gut konkurrieren mit dem Beweis von Becker und Schwartz und ist im Gegensatz zu den bisherigen Beweisen in gewissem Sinne konstruktiv. Wir gehen allerdings weder über Ringe im Sinne dieser Arbeit hinaus, noch haben wir die Verallgemeinerung auf Moduln über einem archimedischen Semring (der Autor hat aber grobe Ideen, wie man letzteres noch erreichen könnte). Für die meisten Anwendungen reicht die hier bewiesene Version des Darstellungssatzes von Kadison-Dubois aus. Zu diesen Anwendungen zählt zum Beispiel die Untersuchung $2k$-ter Potenzen in Körpern, allgemeiner ein Analogon zur Artin-Schreier-Theorie für höhere Potenzen (siehe [Be2]).

4.1 Der archimedische Positivstellensatz als Spezialfall

Wie kommen wir auf die Idee, den Satz von Kadison-Dubois 4.1 ähnlich wie den archimedischen Positivstellensatz zu beweisen? Dies liegt daran, daß der Satz von Kadison-Dubois angewandt auf einen Ring R von der Form $R = K[X_1, \ldots, X_d]/I$ (K ein Unterkörper von $\mathbb{R}$ und I ein Ideal von $K[X_1, \ldots, X_d]$) und einen Semiring P in R mit $K^{\geq 0} \subseteq P$ plötzlich geometrische Gestalt annimmt und im Wesentlichen zu unserem archimedischen Positivstellensatz wird.

In dieser Situation ist nämlich jeder Ringhomomorphismus $\varphi : K[X_1, \ldots, X_d]/I \to \mathbb{R}$ mit $\varphi(P) \subseteq \mathbb{R}^{\geq 0}$ sogar ein K-Algebrenhomomorphismus. Jedes solche φ ist nämlich eingeschränkt auf $\mathbb{Q}$ (genauer auf $\mathbb{Q} \cdot 1$) die identische Abbildung und außerdem wegen $\varphi(K^{\geq 0}) \subseteq \mathbb{R}^{\geq 0}$ (hier geht $K^{\geq 0} \subseteq P$ ein!) eingeschränkt auf K monoton steigend. Daraus folgert man leicht, daß φ sogar auf K die identische Abbildung ist. Dies heißt nichts anderes, als daß φ ein K-Algebrenhomomorphismus ist.

Es stimmt also in dieser Situation die jetzige Definition von $X(P)$ überein mit der früheren Definition aus 3.16. Folglich gilt Lemma 3.17:

$$h : S(P) \to X(P) : x \mapsto \varepsilon_x \quad \text{ist eine Bijektion.}$$

Diese Abbildung ist sogar ein Homöomorphismus. Auf $X(P)$ haben wir nämlich die gröbste Topologie, die alle Funktionen $\Phi(a) = \hat{a}$ $(a \in K[X_1, \ldots, X_d]/I)$ stetig macht. Jede dieser Funktionen ist aber von der Form $\Phi(f(\overline{X_1}, \ldots, \overline{X_d})) = f(\Phi(\overline{X_1}), \ldots, \Phi(\overline{X_d}))$ mit einem Polynom $f \in K[X_1, \ldots, X_d]$, also schon stetig, wenn nur die d Funktionen

$$\Phi(\overline{X_i}) : X(P) \to \mathbb{R} : \varphi \mapsto \varphi(\overline{X_i}) \qquad (i \in \{1, \ldots, d\})$$

stetig sind. Weil wir schon wissen, daß h eine Bijektion ist, können wir die i-te dieser d Abbildungen auch schreiben als $\varepsilon_x \mapsto \varepsilon_x(\overline{X_i}) = x_i$ $(x \in S(P))$. Bei der Topologie auf $X(P)$ handelt es sich also um die Initialtopologie bezüglich der d Funktionen $\varepsilon_x \mapsto x_i$ $(i \in \{1, ..., d\})$. Auf $S(P)$ haben wir die natürliche Topologie. Dies ist die Initialtopologie bezüglich der Koordinatenprojektionen $x \mapsto x_i$ $(i \in \{1, ..., d\})$. Unter der Bijektion h entspricht die Funktion $\varepsilon_x \mapsto x_i$ gerade der Projektion $x \mapsto x_i$. Also erhalten wir sogar:

$$h : S(P) \to X(P) : x \mapsto \varepsilon_x \quad \text{ist ein Homöomorphismus.}$$

Um den Satz von Kadison-Dubois besser zu verstehen, übersetzen wir ihn nun wörtlich mit Hilfe dieses Homöomorphismus h. Dabei verkommen einige Teile der ursprünglichen Formulierung zu überflüssigen Phrasen.

Sei K ein Unterkörper von $\mathbb{R}$, I ein Ideal von $K[X_1, \ldots, X_d]$ und P ein archimedischer Semiring $K[X_1, \ldots, X_d]/I$ mit $-1 \notin P$. Wir statten die Menge die Menge

$$S(P) = \{x \in V_{\mathbb{R}}(I) \mid \textit{für alle } p \in P \textit{ gilt } p(x) \geq 0\}$$

mit der natürlichen Topologie aus und betrachten den Ringhomomorphismus

$$\Phi : K[X_1, \ldots, X_d]/I \to \mathcal{C}(S(P), \mathbb{R}) : a \mapsto a|_{S(P)}.$$

Dann gilt:

(i) $S(P)$ ist ein nichtleerer kompakter Hausdorffraum.

(ii) $\mathcal{C}^+(S(P), \mathbb{R}) := \{A \in \mathcal{C}(S(P), \mathbb{R}) \mid A(x) \geq 0$ für alle $x \in S(P)\}$ ist ein archimedischer Semiring in $\mathcal{C}(S(P), \mathbb{R})$, für den gilt:

$$\Phi^{-1}(\mathcal{C}^+(S(P), \mathbb{R})) = \{a \in K[X_1, \ldots, X_d]/I \mid \textit{für alle } n \in \mathbb{N} \textit{ gibt es ein } t \in \mathbb{N}^{>0} \textit{ mit } t(1 + na) \in P\}.$$

(iii) Φ hat den Kern

$$\{a \in K[X_1, \ldots, X_d]/I \mid \textit{für alle } n \in \mathbb{N} \textit{ gibt es ein } t \in \mathbb{N}^{>0} \textit{ mit } t(1 \pm na) \in P\}.$$

(iv) $\mathbb{Q} \cdot \Phi(R)$ liegt dicht in $\mathcal{C}(S(P), \mathbb{R})$ bezüglich der Supremumsnorm auf $\mathcal{C}(S(P), \mathbb{R})$.

Hier kann man wegen $\mathbb{Q}^{\geq 0} \subseteq P$ (ii) und (iii) viel einfacher schreiben. Wegen $\mathbb{Q} \subseteq K$ gilt in (iv) $\mathbb{Q} \cdot \Phi(R) = \Phi(R)$. Indem wir nun noch einige Selbstverständlichkeiten weglassen und die Definition von Φ explizit ausnutzen, erhalten wir die folgende äquivalente Aussage:

Sei K ein Unterkörper von $\mathbb{R}$, I ein Ideal von $K[X_1,\ldots,X_d]$ und P ein archimedischer Semiring in $K[X_1,\ldots,X_d]/I$ mit $-1 \notin P$. Dann gilt:

(i) $S(P)$ *ist nichtleer und kompakt.*

(ii) Für alle $a \in K[X_1,\ldots,X_d]/I$ *gilt:*

$$a \geq 0 \text{ auf } S(P) \iff \frac{1}{n} + a \in P \text{ für alle } n \in \mathbb{N}$$

(iii) Für alle $a \in K[X_1,\ldots,X_d]/I$ *gilt:*

$$a = 0 \text{ auf } S(P) \iff \frac{1}{n} \pm a \in P \text{ für alle } n \in \mathbb{N}$$

(iv) Die Menge der reellen Polynomfunktionen auf $S(P)$ liegt dicht in der Menge der stetigen reellen Funktionen auf $S(P)$.

Wir zeigen, wie man diese Aussage aus dem archimedischen Positivstellensatz gewinnt und umgekehrt: Sei zunächst der archimedische Positivstellensatz als bekannt vorausgesetzt. Wir zeigen als erstes (ii): Gilt die rechte Seite der Äquivalenz, so haben wir insbesondere $\frac{1}{n} + a \geq 0$ auf $S(P)$ für alle $n \in \mathbb{N}$, woraus $a \geq 0$ auf $S(P)$ folgt. Sei nun $a \geq 0$ auf $S(P)$ und $n \in \mathbb{N}$. Dann ist $\frac{1}{n} + a > 0$ auf $S(P)$, woraus mit dem archimedischen Positivstellensatz $\frac{1}{n} + a \in P$ folgt. (iii) ist eine direkte Konsequenz von (ii). Wäre $S(P)$ anders als in (i) behauptet leer, so wäre $-1 \geq 0$ auf $S(P)$ und nach dem schon bewiesenen (ii) $-1 \in P$ im Widerspruch zur Voraussetzung (denn $-1 = 2(\frac{1}{n} - 1) \in P \cdot P \subseteq P$ für $n = 2$). Daß $S(P)$ kompakt ist, folgt sofort aus der Archimedizität von P. Damit erhält man dann Teil (iv) der Aussage aus dem Satz von Weierstraß , der besagt, daß man jede stetige Funktion auf einer kompakten Teilmenge des $\mathbb{R}^d$ beliebig genau durch eine Polynomfunktion gleichmäßig approximieren kann (sieh etwa [Scu]).

Umgekehrt sei nun obige Aussage als bekannt vorausgesetzt. Um die nichttriviale Richtung des archimedischen Positivstellensatzes zu zeigen, sei $f \in K[X_1,\ldots,X_d]/I$ mit $f > 0$ auf $S(P)$. Wegen der Archimedizität von P ist $S(P)$ kompakt. Daher gibt es ein $n \in \mathbb{N}$, sodaß sogar $f - \frac{2}{n} \geq 0$ auf $S(P)$ gilt. Nach (ii) aus obiger Aussage gilt dann $\frac{1}{n} + (f - \frac{2}{n}) \in P$. Setzen wir $a := \frac{1}{n} \in K^{>0}$, so folgt also $f - a \in P$.

4.2 Ein neuer Beweis des Satzes von Kadison-Dubois

Wie versprochen geben wir jetzt einen neuen Beweis des Satzes von Kadison-Dubois. Der Leser, der die bisherigen Kapitel nicht gelesen hat, muß dazu nur den Satz von Pólya 3.6, das Lemma 3.11 und das Lemma 3.17 auf den Seiten 35, 39 und 41 nachlesen. Als erstes geben wir eine kleine Variante von Lemma 3.10 auf Seite 39:

Lemma 4.3. *Sei $s \in \mathbb{N}$ und I das von $X_1+\cdots+X_d-s$ erzeugte Hauptideal in $\mathbb{Z}[X_1,\ldots,X_d]$. Sei $f \in \mathbb{Z}[X_1,\ldots,X_d]$ mit $f > 0$ auf $V_{\mathbb{R}}(I) \cap (\mathbb{R}^{\geq 0})^d$. Dann gibt es ein $N \in \mathbb{N}$, sodaß fs^N modulo I in $\mathbb{Z}[X_1,\ldots,X_d]$ äquivalent ist zu einer Positivform.*

Beweis: Für $s = 0$ ist die Aussage trivial. Sei also $s > 0$. Bezeichne M den maximalen Grad der in f vorkommenden Monome (also den Grad von f) und m den minimalen Grad der in f vorkommenden Monome. Das Polynom fs^{M-m} ist dann modulo I äquivalent zu der M-Form

$$F := \sum_{a \text{ Monom in } f} a(X_1 + \cdots + X_d)^{M - \deg a} s^{(\deg a) - m}.$$

Es gilt $F > 0$ auf $V_{\mathbb{R}}(I) \cap (\mathbb{R}^{\geq 0})^d$ und, weil F eine Form ist, sogar auf $(\mathbb{R}^{\geq 0})^d \setminus \{0\}$. Nach dem Satz von Pólya 3.6 gibt es ein $n \in \mathbb{N}$, sodaß $F(X_1 + \cdots + X_d)^n \in \mathbb{Z}[X_1, \ldots, X_d]$ eine Positivform ist. Mit $N := (M - m) + n$ gilt dann modulo I

$$fs^N = (fs^{M-m})s^n \equiv Fs^n \equiv F(X_1 + \cdots + X_d)^n.$$

□

Die folgende Variante von Lemma 3.13 auf Seite 39 zeigt, daß das eben bewiesene Lemma auch dann noch gilt, wenn wir das Ideal I vergrößern:

Lemma 4.4. *Sei $s \in \mathbb{N}$ und I ein Ideal in $\mathbb{Z}[X_1, \ldots, X_d]$ mit $X_1 + \cdots + X_d - s \in I$. Sei $f \in \mathbb{Z}[X_1, \ldots, X_d]$ mit $f > 0$ auf $V_{\mathbb{R}}(I) \cap (\mathbb{R}^{\geq 0})^d$. Dann gibt es ein $N \in \mathbb{N}$, sodaß fs^N modulo I in $\mathbb{Z}[X_1, \ldots, X_d]$ äquivalent ist zu einer Positivform.*

Beweis: Da der Ring $\mathbb{Z}$ noethersch ist, ist nach dem Hilbertschen Basissatz auch der Ring $\mathbb{Z}[X_1, \ldots, X_d]$ noethersch. Also ist das Ideal I endlich erzeugt, etwa

$$I = (X_1 + \cdots + X_d - s, r_1, \ldots, r_t)$$

mit $r_1, \ldots, r_t \in \mathbb{Z}[X_1, \ldots, X_d]$. Wir wenden Lemma 3.11 auf Seite 39 an mit

$$U := V_{\mathbb{R}}(I) \cap (\mathbb{R}^{\geq 0})^d \subseteq V_{\mathbb{R}}(\{X_1 + \cdots + X_d - s\}) =: V, \quad f|_V \quad \text{und} \quad r := (r_1^2 + \cdots + r_t^2)|_V.$$

Es gibt also ein $c \in \mathbb{Z}$, sodaß $f + c(r_1^2 + \cdots + r_t^2) > 0$ auf V. Nach Lemma 4.3 gibt es ein $N \in \mathbb{N}$, sodaß $(f + c(r_1^2 + \cdots + r_t^2))s^N$ modulo $X_1 + \cdots + X_d - s$, insbesondere modulo I äquivalent ist zu einer Positivform in $\mathbb{Z}[X_1, \ldots, X_d]$. Da fs^N modulo I äquivalent ist zu $(f + c(r_1^2 + \cdots + r_t^2))s^N$, folgt die Behauptung. □

Wir schließen die Arbeit ab mit:

Beweis des Satzes von Kadison-Dubois 4.1: Da P archimedisch ist, können wir zu jedem $a \in R$ ein $s_a \in \mathbb{N}$ mit $s_a \pm a \in P$ wählen. Für jedes $\varphi \in X(P)$ gilt dann $s_a \pm \varphi(a) \geq 0$, also $\varphi(a) \in [-s_a, s_a]$. Daher ist $X(P) \subseteq \prod_{a \in R}[-s_a, s_a]$ als Menge, aber auch als topologischer Raum, wie man sofort sieht. Da $\prod_{a \in R}[-s_a, s_a]$ nach dem Satz von Tychonoff kompakt ist, ist auch der Unterraum $X(P)$ kompakt, denn

$$\begin{aligned} X(P) = &\left\{ \varphi \in \prod_{a \in R}[-s_a, s_a] \;\middle|\; \varphi(0) = 0 \right\} \cap \left\{ \varphi \in \prod_{a \in R}[-s_a, s_a] \;\middle|\; \varphi(1) = 1 \right\} \cap \\ &\bigcap \left\{ \left\{ \varphi \in \prod_{a \in R}[-s_a, s_a] \;\middle|\; \varphi(b + c) - \varphi(b) - \varphi(c) = 0 \right\} \;\middle|\; b, c \in R \right\} \cap \\ &\bigcap \left\{ \left\{ \varphi \in \prod_{a \in R}[-s_a, s_a] \;\middle|\; \varphi(bc) - \varphi(b) - \varphi(c) = 0 \right\} \;\middle|\; b, c \in R \right\} \cap \\ &\bigcap \left\{ \left\{ \varphi \in \prod_{a \in R}[-s_a, s_a] \;\middle|\; \varphi(b) \geq 0 \right\} \;\middle|\; b \in P \right\} \end{aligned}$$

ist als Schnitt von abgeschlossenen Mengen (nämlich Urbildern von abgeschlossenen Teilmengen von $\mathbb{R}$ unter stetigen Funktionen) abgeschlossen in $\prod_{a \in R}[-s_a, s_a]$. Es ist klar, daß $X(P)$ Hausdorffsch ist. Damit ist (i) bis auf $X(P) \neq \emptyset$ gezeigt.

Nun zeigen wir (ii). Daß $\mathcal{C}^+(X(P), \mathbb{R})$ archimedisch ist, folgt aus der Kompaktheit von $X(P)$. Von der behaupteten Gleichheit zeigen wir zunächst die einfache Inklusion „$\supseteq$“: Sei $a \in R$, sodaß es für alle $n \in \mathbb{N}$ ein $t \in \mathbb{N}^{>0}$ gibt mit der Beziehung $t(1+na) \in P$. Wendet man ein $\varphi \in X(P)$ auf diese Beziehung an, so folgt $1 + n\varphi(a) \geq 0$ für alle $n \in \mathbb{N}$, also $\varphi(a) \geq 0$. Das bedeutet nichts anderes als $\hat{a}(\varphi) \geq 0$ für alle $\varphi \in X(P)$, also $\Phi(a) = \hat{a} \in \mathcal{C}^+(X(P), \mathbb{R})$.

Nun zum wesentlichen Gehalt des Satzes, zur Inklusion „$\subseteq$“ in (ii): Sei $a \in R$ mit $\Phi(a) \in \mathcal{C}^+(X(P), \mathbb{R})$, d.h. $\varphi(a) = \hat{a}(\varphi) \geq 0$ für alle $\varphi \in X(P)$. Sei $n \in \mathbb{N}$. Zu zeigen ist die Existenz eines $t \in \mathbb{N}^>$ mit $t(1 + na) \in P$. Für eine Familie paarweise verschiedener Unbestimmter $(Y_p)_{p \in P}$ betrachten wir nun den Polynomring $\mathbb{Z}[(Y_p)_{p \in P}]$ und den $\mathbb{Z}$-Algebrenhomomorphismus (= Ringhomomorphismus) $\psi : \mathbb{Z}[(Y_p)_{p \in P}] \to R : Y_p \mapsto p$. Wegen der Archimedizität von P gilt $R = \mathbb{Z}[P]$, d.h. ψ ist surjektiv, und wir können ein $f \in \mathbb{Z}[(Y_p)_{p \in P}]$ mit $\psi(f) = a$ wählen. Bezeichnen wir mit J den Kern von ψ, so wird durch ψ ein Isomorphismus $\mathbb{Z}[(Y_p)_{p \in P}]/J \to R$ induziert. Das Urbild des Semirings P unter diesem Isomorphismus ist die Menge $\{\overline{Y_p} \mid p \in P\}$. Die an a gemachte Voraussetzung besagt übertragen auf das Urbild $\overline{f}$ von a unter diesem Isomorphismus, daß für alle Ringhomomorphismen (= $\mathbb{Z}$-Algebrenhomomorphismen) $\varphi : \mathbb{Z}[(Y_p)_{p \in P}]/J \to R$ mit $\varphi(\{\overline{Y_p} \mid p \in P\}) \subseteq \mathbb{R}^{\geq 0}$ gilt $\varphi(\overline{f}) \geq 0$. Völlig analog zu Lemma 3.17 beweist man, daß dies äquivalent ist zu $f \geq 0$ auf $V_{\mathbb{R}}(J) \cap (\mathbb{R}^{\geq 0})^P$. Daraus folgt

$$(\diamond) \qquad 1 + nf > 0 \quad \text{auf} \quad V_{\mathbb{R}}(J) \cap (\mathbb{R}^{\geq 0})^P.$$

Im Folgenden zeigen wir, daß wir in ($\diamond$) die Menge J durch eine endliche Teilmenge H von J ersetzen können (vgl. Lemma 3.21). Es gilt ($V_{\mathbb{R}}(g)$ sei die Menge der Nullstellen von g in $\mathbb{R}^P$)

$$\left(\prod_{p \in P} [0, s_p]\right) \cap \left(\bigcap \{V_{\mathbb{R}}(g) \mid g \in J\}\right) \cap \{y \in \mathbb{R}^P \mid 1 + nf(y) \leq 0\} = \emptyset,$$

denn sogar, wenn wir jeweils $[0, \infty[$ statt $[0, s_p]$ geschrieben hätten, bliebe dies richtig (weil es nichts anderes als ($\diamond$) besagte). Wir wollten jedoch erreichen, daß eine der Mengen, die zu dem Schnitt beitragen, kompakt ist. Dies haben wir erreicht: Nach dem Satz von Tychonoff ist $\prod_{p \in P}[0, s_p]$ kompakt. Außerdem sind alle anderen Mengen, die zu dem Schnitt beitragen abgeschlossen in $\mathbb{R}^P$. Daher können wir eine endliche Teilmenge G von J wählen, sodaß sogar

$$\left(\prod_{p \in P} [0, s_p]\right) \cap \left(\bigcap \{V_{\mathbb{R}}(g) \mid g \in G\}\right) \cap \{y \in \mathbb{R}^P \mid 1 + nf(y) \leq 0\} = \emptyset$$

gilt. Nun würden wir statt $[0, s_p]$ gerne wieder $[0, \infty[$ schreiben. Bezeichne Q diejenige Menge, für die $\{Y_p \mid p \in Q\}$ gerade die Menge der Unbestimmten ist, die in f und in den Elementen von G tatsächlich vorkommen. Q ist endlich, da G endlich ist und in jedem Polynom nur endlich viele Unbestimmte vorkommen. Für alle $p \in P \setminus Q$ können wir ohne Bedenken wieder $[0, \infty[$ statt $[0, s_p]$ schreiben. Wir erhalten dann

$$(\mathbb{R}^{\geq 0})^P \cap \{(y_p)_{p \in P} \in \mathbb{R}^P \mid y_p \leq s_p \text{ für alle } p \in Q\} \cap V_{\mathbb{R}}(G) \cap \{y \in \mathbb{R}^P \mid 1 + nf(y) \leq 0\} = \emptyset.$$

Um in diesem Schnitt die Menge $\{(y_p)_{p \in P} \mid y_p \leq s_p \text{ für alle } p \in Q\}$ weglassen zu können, vergrößern wir G wieder zur Menge $H := G \cup \{Y_{s_p - p} + Y_p - s_p \mid p \in Q\}$ (beachte $s_a - a \in P$ für alle $a \in R$), welche immer noch endlich und eine Teilmenge von J ist. Weil aus $(y_p)_{p \in P} \in (\mathbb{R}^{\geq 0})^P \cap V_{\mathbb{R}}(\{Y_{s_p - p} + Y_p - s_p \mid p \in Q\})$ nämlich $s_p - y_p = y_{s_p - p} \geq 0$ und damit $y_p \leq s_p$ für alle $p \in Q$ folgt, haben wir erst recht

$$(\mathbb{R}^{\geq 0})^P \cap V_{\mathbb{R}}(H) \cap \{y \in \mathbb{R}^P \mid 1 + nf(y) \leq 0\} = \emptyset.$$

Damit haben wir in ($\diamond$) nun J durch die endliche Teilmenge H von J ersetzt und erhalten

$$(\square) \qquad 1 + nf > 0 \quad \text{auf} \quad V_{\mathbb{R}}(H) \cap (\mathbb{R}^{\geq 0})^P.$$

Seien nun $X_1, \ldots, X_{d-1} \in \{Y_p \mid p \in P\}$ die endlich vielen paarweise verschiedenen Unbestimmten, die in f und in den Polynomen aus H tatsächlich vorkommen. Wegen der

Archimedizität des Semirings $\{\overline{Y_p} \mid p \in P\}$ in $\mathbb{Z}[(Y_p)_{p\in P}]/J$ (dieser ist ja das Urbild des archimedischen Semirings P unter dem durch ψ induzierten Isomorphismus) gibt es ein $s \in \mathbb{N}$, sodaß

$$s - (\overline{X_1} + \cdots + \overline{X_{d-1}}) \in \{\overline{Y_p} \mid p \in P\} \subseteq \mathbb{Z}[(Y_p)_{p\in P}]/J.$$

Damit sind sogar $s-(\overline{X_1}+\cdots+\overline{X_{d-1}})$, $s+1-(\overline{X_1}+\cdots+\overline{X_{d-1}})$, $s+2-(\overline{X_1}+\cdots+\overline{X_{d-1}})$, ... Elemente von $\{\overline{Y_p} \mid p \in P\}$, und zwar paarweise verschiedene (denn sonst gäbe es ein $n \in \mathbb{N}^{>0}$ mit $n \in J$, also $n = 0$ in R, und es wäre $-1 = n - 1 \in P$. Wir können s daher o.B.d.A. so wählen, daß

$$s \geq 1 \qquad \text{und} \qquad s - (\overline{X_1} + \cdots + \overline{X_{d-1}}) \in \{\overline{Y_p} \mid p \in P\} \setminus \{\overline{X_1}, \ldots, \overline{X_{d-1}}\}.$$

Wir wählen nun $X_d \in \{Y_p \mid p \in P\}$ mit $s-(\overline{X_1}+\cdots+\overline{X_{d-1}}) = \overline{X_d}$. Aus $\overline{X_d} \notin \{\overline{X_1}, \ldots, \overline{X_{d-1}}\}$ folgt insbesondere $X_d \notin \{X_1, \ldots, X_{d-1}\}$. Also sind nun $X_1, \ldots, X_d$ paarweise verschiedene Unbestimmte, sodaß

$$\begin{aligned}
&H \cup \{f\} \subseteq \mathbb{Z}[X_1, \ldots, X_d],\\
&I := (H \cup \{X_1 + \cdots + X_d - s\}) \subseteq J \cap \mathbb{Z}[X_1, \ldots, X_d] \quad \text{und}\\
&1 + nf > 0 \text{ auf } V_{\mathbb{R}}(I) \cap (\mathbb{R}^{\geq 0})^d.
\end{aligned}$$

Nach Lemma 4.4 gibt es ein $N \in \mathbb{N}$, sodaß $(1+nf)s^N$ modulo $I \subseteq J$ äquivalent ist zu einer Positivform in $\mathbb{Z}[X_1, \ldots, X_d]$, also insbesondere zu einem Element aus $\langle X_1, \ldots, X_d\rangle_0$. Mit $t := s^N > 0$ folgt (beachte J ist der Kern von ψ)

$$t(1 + na) = \psi(t(1 + nf)) \in \langle \psi(X_1, \ldots, \psi(X_d)\rangle_0 \subseteq P.$$

Damit ist nun (ii) bewiesen. Für (i) müssen wir noch $X(P) \neq \emptyset$ nachtragen. Das folgt jetzt aus (ii): Angenommen $X(P) = \emptyset$. Dann ist die leere Abbildung das einzige Element von $\mathcal{C}(X(P), \mathbb{R})$. Diese ist auch in $\mathcal{C}^+(X(P), \mathbb{R})$ enthalten. Also gilt $\mathcal{C}(X(P), \mathbb{R}) = \mathcal{C}^+(X(P), \mathbb{R})$ und folglich $\Phi^{-1}(\mathcal{C}^+(X(P), \mathbb{R})) = R$, insbesondere $-1 \in \Phi^{-1}(\mathcal{C}^+(X(P), \mathbb{R}))$. Laut (ii) gibt es dann zu $n = 2$ ein $t \in \mathbb{N}^{>0}$ mit $t(1 + n(-1)) = -t \in P$. Es folgt $-1 = -t + (t - 1) \in P$ im Widerspruch zu $-1 \notin P$.

Auch (iii) erhält man leicht aus (ii): Sei $a \in R$ aus der Menge, von der in (iii) behauptet wird, daß sie der Kern von Φ sei. Dann sind sowohl a als auch $-a$ in der Menge, die auf der rechten Seite der in (ii) bewiesenen Gleichheit steht. Es folgt $a, -a \in \Phi^{-1}(\mathcal{C}^+(X(P), \mathbb{R})$, d.h. $\Phi(a), -\Phi(a) \in \mathcal{C}^+(X(P), \mathbb{R})$. Es folgt $\Phi(a) = 0$.

Sei umgekehrt $a \in R$ mit $\Phi(a) = 0$. Sei $n \in \mathbb{N}$. Wegen $a, -a \in \Phi^{-1}(\mathcal{C}^+(X(P), \mathbb{R})$ gibt es nach (ii) ein $t_+ \in \mathbb{N}^{>0}$ und ein $t_- \in \mathbb{N}^{>0}$ mit $t_+(1 + na) \in P$ und $t_-(1 - na) \in P$. Setzen wir dann etwa $t := t_+t_- \in \mathbb{N}^{>0}$, so gilt $t(1 + na) \in P$ und $t(1 - na) \in P$.

Schließlich zeigen wir noch (iv): Zu je zwei verschiedenen Punkten φ_1 und φ_2 des topologischen Raumes $X(P)$ gibt es ein $a \in R$ mit $\varphi_1(a) \neq \varphi_2(a)$, d.h. $\Phi(a)(\varphi_1) \neq \Phi(a)(\varphi_2)$. Also ist $\Phi(R)$ eine punktetrennende Menge stetiger reellwertiger Funktionen auf dem kompakten Raum $X(P)$. Gemäß der üblichen Formulierung des Satzes von Weierstraß-Stone (siehe etwa [Scu]) liegt nun die von $\Phi(R)$ erzeugte $\mathbb{R}$-Algebra dicht in $\mathcal{C}^+(X(P), \mathbb{R})$. Man überlegt sich aber sofort, daß es wegen der Dichte von $\mathbb{Q}$ in $\mathbb{R}$ schon reicht, statt der erzeugten $\mathbb{R}$-Algebra die erzeugte $\mathbb{Q}$-Algebra zu nehmen, und das ist in unserem Fall $\mathbb{Q} \cdot \Phi(R)$. □

In den bisherigen Beweisen des Satzes von Kadison-Dubois war die Hauptarbeit zu zeigen, daß $X(P)$ nicht leer ist. Es ist erstaunlich, daß diese Tatsache sich hier erst offenbart, nachdem man die Gleichheit in (ii) gezeigt hat.

Erstaunlich ist auch folgende Beobachtung von Matthias Aschenbrenner [Asc]: Man kann beim Beweis der Inklusion „$\subseteq$“ von (ii) ein Polynom $F \in \mathbb{Z}[(Y_p)_{p\in P}]$ mit $\psi(f) = a$ konkret abgeben, und zwar sogar ein lineares Polynom, nämlich $f = Y_{s_a+a} - s_a$. Deswegen würde es ausreichen, Lemma 4.4 nur für den Spezialfall zu beweisen, daß f von der Form $X_1 - t$ mit einem $t \in \mathbb{N}$ ist! Bisher gelang es dem Autor jedoch nicht, dies auszunutzen.

Anhang: Quellcode `sos.red`

```
module sos;

% This file is part of Markus Schweighofer's Diplomarbeit under the
% supervision of Volker Weispfenning at the University of Passau
% March 1999
%
% This is a program for the computer algebra system REDUCE (version 3.6)
%
% Computation of a representation of a univariate polynomial with
% rational coefficients having no negative values on the real axis
% as a sum of squares of polynomials with rational coefficients.
%
% Filename: sos.red
% Author: Markus Schweighofer

load roots; % we will use this package for computing Sturm sequences

exports sos, sosprint, sos1;

imports sturm; % import procedure to compute Sturm sequences from package
               % 'roots'

symbolic operator sos, sosprint;



procedure sos(f);
% Sum of squares - entry point for REDUCE's algebraic mode.
% f is anything.
% Terminates the program or returns a list whose first entry is a
% polynomial in several new indeterminates and whose other entries are
% equations. Prints out a message and terminates the program if f does not
% define a univariate polynomial with rational coefficients in an
% indeterminate v. Prints out a message and terminates the program if f
% considered as a univariate polynomial takes on negative values on the
% real axis. Otherwise returns a list: The first entry of this list is a
% polynomial in several new indeterminates with non-negative rational
% coefficients in lisp-prefix-form. This polynomial has at most (deg f)
% many terms. In each term of this polynomial each indeterminate occurs
% with an even power. Each of the remaining entries of the list is an
% equation. The left hand sides of these equations correspond one to one
% with the new indeterminates. The right hand sides are univariate
% polynomials in v with rational coefficients in lisp-prefix-form. If you
% substitute the new indeterminates in the first entry according to these
% equations you get the original polynomial f.

begin scalar sf, il, help, sosf, ssosf, result, old_!*rational;
  old_!*rational := !*rational;
  on rational;
  if not polynomialp f then <<
     !*rational := old_!*rational;
     typerr(f, "polynomial") >>
```

```
  else <<
     sf := numr simp f;
     il := idlist sf;
     if il and cdr il then <<
        !*rational := old_!*rational;
        typerr(f, "univariate polynomial") >>
     else <<
        help := sos1 sf;
        if null help then <<
           !*rational := old_!*rational;
           typerr(f, "polynomial having no negative values") >>
        else <<
           sosf := car help; % a representation of f as sum of squares
                             % "in new indeterminates"
           ssosf := cdr help; % ssosf is the alist giving the substitutions
                              % for these new indeterminates
           result := 'list . prepf sosf .
                     (for each x in ssosf collect
                         {'equal, car x, prepf cdr x}) >> >> >>;
   !*rational := old_!*rational;
   return result
end;



procedure sosprint(f);
% Print sum of squares - entry point for REDUCE's algebraic mode.
% f is anything.
% Returns nil or terminates the program. Prints out something.
% Prints out a message and terminates the program if f does not define a
% univariate polynomial with rational coefficients in an indeterminate v.
% Prints out a message and terminates the program if f considered as a
% univariate polynomial takes on negative values on the real axis.
% Otherwise returns nil and prints out a representation of f as a weighted
% sum (with non-negative rational weights) of at most (deg f) many products
% of even powers of univariate polynomials with rational coefficients.

begin scalar help;
   help := cdr sos f;
   mathprint substitute(car help, cdr help);
   return nil
end;



procedure substitute(f, sublist);
% Substitute.
% f is a lisp-prefix-form. sublist is a list of equations whose left hand
% sides are pairwise distinct identifiers and whose right hand sides are
% lisp-prefix-forms. Returns a lisp-prefix-form.
% Substitutes the left hand sides of the equations in sublist by their
% right hand sides in the lisp-prefix-form f. Does not simplify the
% resulting lisp-prefix-form!
```

```
begin scalar scsublist;
   if idp f then <<
      scsublist := sublist;
      while scsublist and not(cadar scsublist eq f) do
         scsublist := cdr scsublist;
      if scsublist then % i.e. cadar scsublist eq f
         return caddar scsublist
      else
         return f >>
   else if not listp f then
      return f
   else
      return(car f . for each x in cdr f collect substitute(x, sublist))
end;



procedure sos1(f);
% Sum of squares - entry point for REDUCE's symbolic mode.
% f is a univariate polynomial in an indeterminate v in standard form
% with rational coefficients.
% Returns nil if f takes on negative values on the real axis. If f doesn't
% take on negative values on the real axis it returns a pair:
% The car of this pair is a polynomial with non-negative rational
% coefficients in several new indeterminates in standard form. In each
% term of this polynomial each indeterminate occurs with an even power. The
% cdr of this pair is an alist.
% The cars of the entries in this alist correspond one to one with the new
% indeterminates. The cdrs are univariate polynomials in v with rational
% coefficients in standard form. If you substitute the new indeterminates
% in the car of the returned pair according to these entries in the alist
% you get the original polynomial f.

begin scalar help, af, cf, df, of, bof;
   if domainp f and minusf f then
      return nil
   else if domainp f and not minusf f then
      return f . nil
   else if (not evenp ldeg f) or (minusf lc f) then
      return nil
   else << % Now we know that the derivative of f has only finitely many
           % roots but at least one because f has a global minimum point
           % on the real axis.
      help := even_part f;
      af := car help; % af is the even part of f
      cf := cadr help; % cf is the even part of f "in new indeterminates"
      df := cddr help; % df is the alist giving the substitutions for
                       % these
      of := quotf(f, af); % of is square-free - the "odd part" of f
      if count_zeros(sturm_sequence of, 'mininf, 'inf) > 0 then
         return nil
      else <<
         bof := sos2 of;
         return(multf(cf, car bof) .
```

```
                append(df, cdr bof)) >> >>
end;



procedure sos2(f);
% Sum of squares - no entry point for the user.
% f is a pointwise positive univariate polynomial in an indeterminate v in
% standard form with rational coefficients.
% Returns a pair: The car of this pair is a polynomial with non-negative
% rational coefficients in several new indeterminates in standard form.
% In each term of this polynomial each indeterminate occurs with an even
% power. The cdr of this pair is an alist.
% The cars of the entries in this alist correspond one to one with the new
% indeterminates. The cdrs are univariate polynomials in v with rational
% coefficients in standard form. If you substitute the new indeterminates
% in the car of the returned pair according to the entries in the alist
% you get the original polynomial f.

begin scalar sturm, intv, scintv, fder, v, success, tang, parabola,
             bparabola, help, d, ad, cd, dd, od, bod;
   if domainp f then
      return(f . nil)
   else if ldeg f eq 2 then
      return sos_parabola f
   else <<
      v := mvar f;
      fder := diffuni f;
      sturm := sturm_sequence square!-free_part fder;
      intv := isolate_roots(sturm); % The roots of the derivative of f are
      % the only candidates for the global minimum points of f on the real
      % axis. Conversely, the derivative of f has at least one such root
      % because f has at least one global minimum point.
      success := nil; % success is boolean, success is true iff an
                      % appropriate tangent point tang has been found
      repeat <<
         scintv := intv; % the purpose of scintv is to scan intv
         while scintv and not success do <<
            tang := caar scintv; % the left margin of the first interval in
                                 % scintv
            parabola := tangent_parabola(f, tang);
            d := addf(f, negf parabola); % It cannot happen that d is the
            % zero polynomial because that would imply that f equals
            % parabola. But that is impossible because f has no root and
            % parabola has one.
            help := even_part d;
            ad := car help; % ad is the even part of d
            cd := cadr help; % cd is the even part of d
                             % "in new indeterminates"
            dd := cddr help; % dd is the alist giving the substitutions
                             % for these
            od := quotf(d, ad); % od is square-free - the "odd part" of d
            if count_zeros(sturm_sequence od, 'mininf, 'inf) eq 0 then
               success := t;
```

```
            scintv := cdr scintv >>;
         if not success then
            intv := refine(sturm, intv) >>
      until success;
      bod := sos2 od;
      bparabola := sos2 parabola;
      return(addf(car bparabola, multf(cd, car bod)) .
             append(dd, append(cdr bparabola, cdr bod))) >>
end;



procedure sos_parabola(f);
% Sum of squares for parabolas.
% f is a pointwise non-negative parabola (i.e. a univariate polynomial of
% degree 2) in an indeterminate v in standard form with rational
% coefficients.
% Returns a pair: The car of this pair is a univariate polynomial of the
% form a * x^2 + b with non-negative rational numbers a and b and a new
% indeterminate x. The cdr is an alist whose only entry is a pair
% x . (v - r) where r is a rational number and v - r is represented in
% standard form. a * (v - r)^2 + b is equal to the original polynomial f.

begin scalar apex, apex_value, v, fder, x;
   v := mvar f;
   x := intern gensym();
   fder := diffuni f;
   apex := quotf(negf red fder, lc fder);
   apex_value := numr subf(f, {v . prepf apex});
   return(addf(multf(lc f, x .** 2 .* 1 .+ nil), apex_value) .
          {(x . addf(v .** 1 .* 1 .+ nil, negf apex))}
         )
end;



procedure tangent_parabola(f, tang);
% Tangent parabola.
% f is a pointwise non-negative univariate polynomial in standard form with
% rational coefficients. tang is a rational number.
% Returns a standard form.
% Computes the unique polynomial p of degree at most 2 with the following
% properties:
% - p(tang) = f(tang)
% - p'(tang) = f'(tang)
% - p either has exactly one real zero or is constant

begin scalar v, fder, help;
   if domainp f then
      return f
   else <<
      v := mvar f;
      fder := diffuni f;
      help := multf(numr subf(fder, {v . prepf tang}),
```

```
                    addf(v .** 1 .* 1 .+ nil, negf tang)
                   );
      return addf(addf(numr subf(f, {v . prepf tang}),
                       multf(addf(v .** 1 .* 1 .+ nil, negf tang),
                             numr subf(fder, {v . prepf tang})
                            )
                      ),
                  quotf(multf(help, help),
                        multf(4, numr subf(f, {v . prepf tang}))
                       )
                 ) >>
end;



procedure polynomialp(f);
% Check on polynomial.
% f is anything.
% Returns a boolean.
% Checks if f is a lisp-prefix-form defining a multivariate polynomial with
% rational coefficients as far as it is worth the trouble. Inspects
% everything that a user of the algebraic mode might get wrong without
% bad intention but does not notice whether f is a cyclic structure for
% example (which is illegal).

begin scalar ok;
  return fixp f or idp f or
         (pairp f and (((car f memq '(plus times)) and
                        << ok := t;
                           for each g in cdr f do
                             ok := ok and polynomialp g;
                           ok >>) or
                       ((car f eq 'minus) and cdr f and not cddr f and
                        polynomialp cadr f) or
                       ((car f eq 'quotient) and cdr f and cddr f and
                        not cdddr f and polynomialp cadr f and
                        fixp caddr f) or
                       ((car f eq 'expt) and cdr f and cddr f and
                        not cdddr f and polynomialp cadr f and
                        fixp caddr f and (caddr f >= 0))
                      )
         )
end;



procedure idlist(f);
% List of indeterminates.
% f is a polynomial in standard form.
% Returns a list of pairwise distinct identifiers which are exactly the
% identifiers occuring in f.

idlist1(f, nil);
```

```
procedure idlist1(f, exc);
% List of indeterminates.
% f is a polynomial in standard form. exc is a list of identifiers.
% Returns a list of pairwise distinct identifiers which are exactly the
% identifiers occuring in f but not in exc.

begin scalar redlist, lclist;
   if domainp f then
      return nil
   else <<
      redlist := idlist1(red f, (mvar f).exc);
      lclist := idlist1(lc f, (mvar f).append(redlist, exc));
      if mvar f memq exc then
         return append(redlist, lclist)
      else
         return (mvar f).append(redlist, lclist) >>
end;



procedure square!-free_decomposition(f);
% Square-free decomposition.
% f is a univariate non-zero polynomial with rational coefficients in
% standard form.
% Returns a list (g_1 ... g_n) of pairwise relatively prime monic
% square-free polynomials g_i with rational coefficients in standard
% form such that f = a * g_1^1 * ... * g_n^n.

begin scalar leading_coefficient, list1, list2, list3;
   if domainp f then
      return nil
   else <<
      leading_coefficient := lc f;
      list1 := {monic f};
      while not domainp car list1 do
         list1 := (monic gcdf!*(car list1, diffuni car list1)) . list1;
      while cdr list1 do <<
         list2 := quotf(cadr list1, car list1) . list2;
         list1 := cdr list1 >>;
      while cdr list2 do <<
         list3 := monic quotf(car list2, cadr list2) . list3;
         list2 := cdr list2 >>;
      list3 := monic car list2 . list3;
      return reversip list3 >>
end;



procedure even_part(f);
% Even part.
% f is a univariate non-zero polynomial with rational coefficients in
% standard form. Let f = a * g_1^1 * ... * g_n^n with pairwise relatively
```

```
% prime, monic and square-free g_i (square-free decomposition).
% This function returns a pair ( a . b ) whose cdr is again a pair
% b = (c . d). Here a is the "even part" of f, namely
% g_2^2 * g_3^2 * g_4^4 * g_5^4 * ..., in standard form.
% b is a monomial in new identifiers each of which occurs with an even
% power in standard form. d is an alist. The cars of the entries of this
% alist correspond one to one with the new identifiers in c.
% The cdrs are polynomials in standard form. If you substitute the new
% indeterminates in c according to the entries of d you get a, i.e. the
% "even part" of f

begin scalar sqfreedecomp, a, c, d, x; integer exponent;
   exponent := 1;
   a := 1;
   c := 1;
   d := nil;
   sqfreedecomp := square!-free_decomposition f;
   while sqfreedecomp do <<
      if (car sqfreedecomp neq 1) and (exponent neq 1) then <<
         x := intern gensym();
         a := multf(a, if evenp exponent then
                           numr simp {'expt, prepf car sqfreedecomp,
                                      exponent}
                       else
                           numr simp {'expt, prepf car sqfreedecomp,
                                      exponent - 1}
                   );
         c := multf(c, if evenp exponent then
                           numr simp {'expt, x, exponent}
                       else
                           numr simp {'expt, x, exponent - 1}
                   );
         d := (x . car sqfreedecomp) . d >>;
      exponent := exponent + 1;
      sqfreedecomp := cdr sqfreedecomp >>;
   return(a . (c . d))
end;




procedure square!-free_part(f);
% Square-free part.
% f is a univariate non-zero polynomial with rational coefficients in
% standard form.
% Returns the monic square-free part of f in standard form.

monic quotf(f, gcdf!*(f, diffuni f));



procedure monic(f);
% Monic.
% f is a univariate non-zero polynomial with rational coefficients in
```

```
% standard form.
% Returns f divided by its leading coefficient.

if domainp f then
   1
else
   quotf(f, lc f);



procedure diffuni(f);
% Differentiate univariate polynomial.
% f is a univariate polynomial with rational coefficients in standard
% form.
% Returns the derivative of f in standard form.

if domainp f then
   nil
else
   addf(multf(lc f,
              numr simp {'times, ldeg f, {'expt, mvar f, ldeg f - 1}}),
        diffuni red f);



procedure sturm_sequence(f);
% Sturm sequence.
% f is a non-zero square-free univariate polynomial with rational
% coefficients in standard form.
% Returns a list of monic standard forms.
% Computes a Sturm sequence of f.

begin scalar p, v; integer d;
   if domainp f then
      if minusf f then
         return {-1}
      else
         return {1}
   else <<
      v := mvar f;
      return for each x in cdr sturm0 {prepf f} join
                if fixp x then
                   if x eq 0 then
                      nil
                   else
                      {x}
                else <<
                   d := cadr x;
                   p := nil;
                   for each y in cddr x do <<
                      p := addf(p,
                                multf(y, mksp!*(v .** 1 .* 1 .+ nil, d)));
                      d := d - 1 >>;
                   {p} >> >>
```

```
end;



procedure variations_in_sign(sturm, a);
% Variations in sign.
% sturm is a list of univariate polynomials with rational coefficients in
% standard form which represents a Sturm sequence. a is a rational number
% or 'minusinf or 'inf.
% Returns a natural number.
% Computes the variations in sign of the Sturm sequence sturm evaluated
% at the point a (where 'minusinf means -infinity and 'inf means infinity).

begin integer oldsign, newsign, result; scalar value;
   while sturm do <<
      if domainp car sturm then
         if minusf car sturm then
            newsign := -1
         else % note that f is not zero, because sturm is a sturm sequence
            newsign := 1
      else if a eq 'inf then
         if minusf lc car sturm then
            newsign := -1
         else
            newsign := 1
      else if a eq 'mininf then
         if (evenp ldeg car sturm and not minusf lc car sturm) or
            (not evenp ldeg car sturm and minusf lc car sturm) then
            newsign := 1
         else
            newsign := -1
      else << % if a is a rational number
         value := numr subf(car sturm, {(mvar car sturm) . prepf a});
         if minusf value then
            newsign := -1
         else if null value then
            newsign := 0
         else
            newsign := 1 >>;
      if newsign eq 0 then % zeros are ignored
         newsign := oldsign;
      if (oldsign neq 0) and (oldsign neq newsign) then
         result := result + 1;
      oldsign := newsign;
      sturm := cdr sturm >>;
   return result
end;



procedure isolate_roots(sturm);
% Isolate roots.
% sturm is a Sturm sequence of the square-free part of a non-zero
% univariate polynomial f with rational coefficients.
```

```
% Returns a list of pairs of rational numbers.
% The pairs of this list correspond one-to-one with the different zeros
% of f: For each pair (a . b) f has exactly one root in the interval
% ]a, b[.

begin scalar bound;
   bound := addf(1, bound_on_roots car sturm);
   return isolate_roots1(sturm, negf bound, bound)
end;



procedure isolate_roots1(sturm, a, b);
% Isolate roots.
% sturm is a Sturm sequence of the square-free part of a non-zero
% univariate polynomial f with rational coefficients. a and b are
% rational numbers such that f(a) and f(b) are both not zero and a < b.
% Returns a list of pairs of rational numbers.
% The pairs of this list correspond one-to-one with the different zeros
% of f in the interval [a, b]: For each pair (c . d) it holds that a <= c
% and d <= b and f has exactly one root in the interval ]c, d[.

begin scalar ratio, cut, v;
      integer zeros_left_cut, zeros_right_cut, total_zeros, i;
   if domainp car sturm then
      return nil
   else <<
      v := mvar car sturm;
      ratio := quotf(1, 2); % ratio can take the values 1/2, 1/2 + 1/16,
      % 1/2 + 1/32, 1/2 + 1/64, 1/2 + 1/128 and so on. We fix ratio so
      % that a + ratio * (b - a) (the point where we want to cut the
      % interval ]a, b[) is no root of car sturm. As car sturm
      % has only finitely many roots this can always be done. Note that
      % in every case 1/2 <= ratio <= 1/2 + 1/16 < 1. When n tends to
      % infinity, the n-th power of (1/2 + 1/16) tends to zero.
      % Therefore if we successively cut an interval according to
      % ratio, the length of the interval tends to zero. This fact
      % is important to observe that our isolating algorithm
      % eventually terminates.
      cut := addf(a, multf(ratio, addf(b, negf a)));
      if null numr subf(car sturm, {v . prepf cut}) then <<
         % ratio = 1/2 not possible, try other ratio
         i := 4; % 16 = 2^4
         repeat <<
            ratio := addf(quotf(1, 2), quotf(1, 2^i));
            cut := addf(a, multf(ratio, addf(b, negf a)));
            i := i + 1 >>
         until numr subf(car sturm, {v . prepf cut}) >>;
      zeros_left_cut := count_zeros(sturm, a, cut);
      total_zeros := count_zeros(sturm, a, b);
      zeros_right_cut := total_zeros - zeros_left_cut;
      if total_zeros eq 0 then
         return nil
      else if total_zeros eq 1 then
```

```
         return {a . b}
      else
         return append(isolate_roots1(sturm, a, cut),
                       isolate_roots1(sturm, cut, b)) >>
end;



procedure refine(sturm, intv);
% Refine roots.
% sturm is a Sturm sequence of the square-free part of a non-zero
% univariate polynomial f with rational coefficients.
% intv is a list of pairs of rational numbers whose entries correspond
% with the different zeros of f (see comment on isolate_roots).
% Returns again such a list with the property that the length of the
% interval represented by the n-th entry is at most half the length of the
% interval represented by the n-th entry of intv.

begin scalar mean, v;
   if domainp car sturm then
      return nil
   else <<
      v := mvar car sturm;
      return for each x in intv collect <<
                mean := quotf(addf(car x, cdr x), 2);
                if null numr subf(car sturm, {v . prepf mean}) then
                   quotf(addf(car x, mean), 2) . quotf(addf(mean, cdr x),
                         2)
                else if count_zeros(sturm, car x, mean) > 0 then
                   car x . mean
                else
                   mean . cdr x >> >>
end;



procedure bound_on_roots(f);
% Bound on roots.
% f is a non-zero univariate polynomial with rational coefficients in
% standard form.
% Returns a non-negative rational number C such that all real zeros of f
% lie in the interval [-C, C].

begin scalar ff, c;
   if domainp f then
      return 0
   else <<
      ff := red monic f;
      while not domainp ff do <<
         c := addf(c, absolute_value lc ff);
         ff := red ff >>;
      c := addf(c, absolute_value ff);
      if minusf addf(c, negf 1) then
         return 1
```

```
      else
         return c >>
end;



procedure absolute_value(r);
% Absolute value.
% r is a rational number.
% Returns a rational number.
% Computes the absolute value of f.

if null r then
   nil
else if minusf r then
   negf r
else
   r;



procedure count_zeros(sturm, a, b);
% Count zeros.
% sturm is a Sturm sequence of the square-free part of a non-zero
% univariate polynomial f with rational coefficients. a and b are
% rational numbers or 'inf or 'mininf such that a < b and f(a) and f(b)
% are both different from zero.
% Returns a natural number.
% Computes the number of zeros (multiplicities not counted) of p in the
% interval ]a, b[.

variations_in_sign(sturm, a) - variations_in_sign(sturm, b);



endmodule;

end;
```

Literaturverzeichnis

[Asc] M. Aschenbrenner, mündliche Mitteilung, maschenb@math.uiuc.edu, Urbana-Champaign: University of Illinois

[Be1] E. Becker: Partial orders on a field and valuation rings, Commun. Algebra **7**, 1933-1976 (1979)

[Be2] E. Becker: Extended Artin-Schreier theory of fields, Rocky Mt. J. Math. **14**, 881-897 (1984)

[BCR] J. Bochnak, M. Coste, M. Roy: Real algebraic geometry, Ergebnisse der Mathematik und ihrer Grenzgebiete, 3. Folge **36**, Berlin: Springer (1998)

[Brö] T. Bröcker: Analysis II, Mannheim: B. I. Wissenschaftsverlag (1992)

[BS] E. Becker, N. Schwartz: Zum Darstellungssatz von Kadison-Dubois, Arch. Math. **40**, 421-428 (1983)

[BW] T. Becker, V. Weispfenning: Gröbner bases, Graduate Texts in Mathematics **141**, New York: Springer-Verlag (1993)

[Cas] J.W.S. Cassels: On the representation of rational functions as sums of squares, Acta Arith. **9**, 79-82 (1964)

[Dub] D.W. Dubois: A note on David Harrison's theory of preprimes, Pacific J. Math. **21**, 15-19 (1967)

[Ha1] D. Handelman: Positive polynomials and product type actions of compact groups, Mem. Am. Math. Soc. **320** (1985)

[Ha2] D. Handelman: Representing polynomials by positive linear functions on compact convex polyhedra, Pac. J. Math. **132**, No.1, 35-62 (1988)

[Hab] W. Habicht: Über die Zerlegung strikter definiter Formen in Quadrate, Comment. Math. Helv. **12**, 317-322 (1940)

[Hau] F. Hausdorff: Summationsmethoden und Momentfolgen I, Mathematische Zeitschrift **9**, 74-109 (1921)

[Hea] A.C. Hearn: REDUCE User's Manual, Version 3.6, Santa Monica, CA: RAND Publication (1995), also available from: Konrad-Zuse-Zentrum Berlin, Germany

[HLP] G.H. Hardy, J.E. Littlewood, G. Pólya: Inequalities, second edition, Cambridge: Cambridge University Press (1967)

[Jac] N. Jacobson: Basic algebra I, 2nd ed., New York: W.H. Freeman and Company (1995)

[Kad] R.V. Kadison: A representation theory for commutative topological algebra, Mem. Am. Math. Soc. **7** (1951)

[Kam] S.L. Kameny: The REDUCE Root Finding Package, Bath: Codemist Ltd. (1993)

[Lor] F. Lorenz: Einführung in die Algebra, Teil II, Mannheim: B.I.-Wissenschaftsverlag (1990)

[LS] J.A. de Loera, F. Santos: An effective version of Polya's theorem on positive definite forms, J. Pure Appl. Algebra **108**, No.3, 231-240 (1996)

[Mel] H. Melenk: REDUCE Symbolic Mode Primer, Berlin: Konrad-Zuse-Zentrum für Informationstechnik (1993)

[Mis] B. Mishra: Algorithmic algebra, Texts and Monographs in Computer Science, Berlin: Springer-Verlag (1993)

[Neu] W. Neun, persönliche Mitteilung, neun@zib.de, Berlin: Konrad-Zuse-Zentrum für Informationstechnik

[Pfi] A. Pfister: Hilbert's seventeenth problem and related problems on definite forms, Math. Dev. Hilbert Probl., Proc. Symp. Pure Math. 28, De Kalb 1974, 483-489 (1976)

[Pól] G. Pólya: Über positive Darstellung von Polynomen, Vierteljahresschrift der Naturforschenden Gesellschaft in Zürich **73** (1928), 141-145, siehe auch: Collected Papers, Volume 2, 309-313, Cambridge: MIT press (1974)

[Pou] Y. Pourchet: Sur la représentation en somme de carrés des polynômes à une indéterminée sur un corps de nombres algébriques, Acta Arith. **19**, 89-104 (1971)

[Pr1] A. Prestel: Einführung in die mathematische Logik und Modelltheorie, Vieweg Studium, Aufbaukurs Mathematik **60**, Braunschweig/Wiesbaden: Friedr. Vieweg & Sohn (1986)

[Pr2] A. Prestel: Lectures on formally real fields, Lecture Notes in Mathematics **1093**, Berlin etc.: Springer-Verlag (1984)

[Pr3] A. Prestel: Model theory for the real algebraic geometer, Dipartimento di Matematica dell'università di Pisa (1997)

[PS] G. Pólya, G. Szegö: Problems and theorems in analysis, Vol. II, Theory of functions - zeros - polynomials - determinants - number theory - geometry, Rev. and enl. translation of the 4th ed, Springer Study Editon, New York - Heidelberg - Berlin: Springer-Verlag (1976)

[Rob] A. Robinson: Algorithms in Algebra, D.H. Saracino, V. Weispfenning (ed.), Model Theory and Algebra, Lecture Notes in Mathematics **498**, Berlin etc.: Springer-Verlag, 28-33 (1975)

[Rot] P. Rothmaler: Einführung in die Modelltheorie, Vorlesungen, ausgearbeitet von Frank Reitmaier, Spektrum Lehrbuch. Heidelberg: Spektrum Akademischer Verlag (1995)

[rqe] Quantifier elimination and cylindrical algebraic decomposition, Proceedings of a symposium, Linz, Austria, October 6–8, 1993, Wien: Springer, Texts and Monographs in Symbolic Computation (1998)

[Sch] K. Schmüdgen: The K-moment problem for compact semi-algebraic sets, Math. Ann. **289**, No.2, 203-206 (1991)

[Scr] A. Schrijver: Theory of linear and integer programming, Wiley-Interscience Series in Discrete Mathematics, A Wiley-Interscience Publication, Chichester: Wiley & Sons Ltd (1986)

[Scu] H. Schubert: Topologie, 4. Aufl., Stuttgart: B. G. Teubner (1975)

[St1] G. Stengle: A Nullstellensatz and a Positivstellensatz in semialgebraic geometry, Math. Ann. **207**, 87-97 (1974)

[St2] G. Stengle: Complexity estimates for the Schmuedgen Positivstellensatz, J. Complexity **12**, No.2, 167-174 (1996)

[We1] V. Weispfenning: Computeralgebra, eine Einführung, Skriptum einer Vorlesung gehalten an der Universität Passau

[We2] V. Weispfenning: Algebraische Modelltheorie, Skriptum einer Vorlesung gehalten im Wintersemester 1995/96 an der Universität Passau, ausgearbeitet und ergänzt von Matthias Aschenbrenner

[Wör] T. Wörmann: Strikt positive Polynome in der semialgebraischen Geometrie, Dissertation, Universität Dortmund (1998)

Zusammenfassung

Im ersten Teil dieser Arbeit haben wir einen bekannten Beweis dafür, daß jedes Polynom $f \in \mathbb{R}[X]$ vom Grad $n \geq 1$ mit $f \geq 0$ auf $\mathbb{R}$ eine Summe von n Quadraten in $\mathbb{R}[X]$ ist, modifiziert. Wir konnten zeigen: Wenn K ein angeordneter Körper ist, der dicht in seinem reellen Abschluß R liegt, dann ist jedes Polynom $f \in K[X]$ vom Grad $n \geq 1$ mit $f \geq 0$ auf R eine gewichtete Summe von n Quadraten in $K[X]$. Für den Fall, daß K nicht dicht in $\mathbb{R}$ liegt, konnten wir mit Hilfe einer leichten Verallgemeinerung eines Satzes von Cassels zeigen, daß diese Aussage zumindest noch gilt, wenn man sie dahingehend abschwächt, daß man nichts mehr über die Anzahl der Summanden aussagt. Für den Fall, daß K ein Unterkörper von $\mathbb{R}$ ist, konnten wir aus dem Beweis einen Algorithmus extrahieren, der die garantierte Darstellung berechnet. Für den Fall $K = \mathbb{Q}$ haben wir diesen Algorithmus implementiert.

Der zweite Teil dieser Arbeit kann gesehen werden als ein neuer Beweis des Darstellungssatzes von Kadison-Dubois für Ringe mit einem ausgezeichneten archimedischen Semiring durch Rückführung auf ein Theorem von Pólya. Dieser Beweis eröffnet neue Möglichkeiten zur Berechnung von Darstellungen, wie sie durch verschiedene Positivstellensätze garantiert wird, in deren Beweis der Satz von Kadison-Dubois einfließt. Wir haben dabei deutlich gemacht, daß theoretische Resultate über die Komplexität dieser Darstellungen einer Berechnung, anders als es bisher schien, keineswegs im Wege stehen.

Eidesstattliche Erklärung

Ich versichere, daß ich diese Arbeit selbstständig und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt habe. Alle Ausführungen, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet. Diese Arbeit habe ich in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegt.

Passau, 25. März 1999

Markus Schweighofer