JOURNAL OF APPROXIMATION THEORY 15, 124–131 (1975)

Nichtnegative Polynome in Zwei Veraenderlichen

CLAUS GÜNTHER

Kernforschungszentrum Karlsruhe, 75 Karlsruhe 1, West Germany

Communicated by Oved Shisha

Ein Darstellungssatz fuer reelle Polynome in zwei Veraenderlichen, die auf einem Ellipsenbereich nichtnegativ sind, wird angegeben. Es wird auch untersucht, wo solche Polynome verschwinden koennen.

1. EINLEITUNG

Fuer nichtnegative Polynome in einer reellen Veraenderlichen gibt es fuer verschiedene Definitionsbereiche Darstellungssaetze, die in Abhaengigkeit vom Grad N dieser Polynome die Anzahl und die moegliche Lage der Nullstellen dieser Polynome charakterisieren, z.B. Polya u. Szego [8] and Karlin u. Studden [4]. Diese Saetze finden in verschiedenen Bereichen der Mathematik Anwendung. In dem zuletzt genannten Werk werden diese Ergebnisse, einer Methode von Krein [6] folgend, zur Darstellung nichtnegativer linearer Funktionale auf Raeumen, die ein Tschebyscheffsystem von Funktionen als Basis haben, benuetzt. Diese Methode gestattet es u. a., die Existenz und die Form der sogenannten "Gausschen Quadraturformeln" zu beweisen, ohne von der Zerlegung von Polynomen in einer Veraenderlichen in Linearfaktoren Gebrauch zu machen. Das Beduerfnis, mit derselben Methode entsprechende Ergebnisse in mehreren Dimensionen zu erhalten, hat darauf gefuehrt, nach Darstellungssaetzen fuer nichtnegative Polynome in zwei Veraenderlichen zu suchen, z.B. Günther [2]. Offensichtlich sind bisher keine solchen Saetze bekannt, wenn man vom Fall $N = 2$ absieht, zu dessen Behandlung man die Theorie der quadratischen Formen heranziehen kann. Neben dem genannten Beispiel bestehen auch Zusammenhaenge von nichtnegativen Polynomen mit Approximationsproblemen und Optimalitaetsfragen [6]. Verwandte Fragestellungen hat auch Motzkin [7] untersucht. Es hat sich als vorteilhaft erwiesen, dass man in zwei Dimensionen entweder die ganze euklidsche Ebene oder eine Ellipse D als Definitionsbereich verwendet. Entsprechende Aussagen sind, wie spaeter noch genauer erlaeutert wird, auch fuer andere kompakte Bereiche moeglich, deren Raender algebraischen Gleichungen genuegen.

Das Hauptergebnis dieser Arbeit laesst sich folgendermassen zusammenfassen: Die reellen Teiler eines ueber D nichtnegativen Polynoms p haben entweder nur endlich viele Nullstellen in D oder sind, falls sie unendlich viele Nullstellen im Innern von D haben, in gerader Vielfachheit in p enthalten. Im erstgenannten Fall wird die Anzahl der (isolierten) Nullstellen von p in D in Abhaengigkeit von Grad p angegeben.

2. Die Gemeinsamen Nullstellen von Zwei Polynomen in Zwei Veraenderlichen

Dieser Abschnitt enthaelt Definitionen und Saetze ueber Polynome, die vor allem aus der Algebraischen Geometrie entnommen sind. Sie spielen beim Beweis der Saetze ueber nichtnegative Polynome eine wesentliche Rolle. Diese Saetze sind dem Buch von Walker [9] entnommen. Auf dieses Werk sei auch verwiesen bezueglich der Definitionen einer algebraischen Kurve, von singulaeren Punkten einer algebraischen Kurve, von singulaeren Punkte einer solchen Kurve, speziell der Vielfachheit eines singulaeren Punktes.

Satz 2.1. *Eine algebraische Kurve der Ordnung N ohne mehrfache Komponenten habe in den Punkten X_1, X_2,..., X_l die Vielfachheiten r_1, r_2,..., r_l; dann gilt*

$$N(N-1) \geqslant \sum_{j=1}^{l} r_j (r_j - 1).$$

Ueber die Anzahl gemeinsamer Nullstellen zweier Polynome gibt der folgende Satz Auskunft.

Satz 2.2 (Satz von Bézout). *$p(x, y)$ und $q(x, y)$ aus $\mathbb{C}[x, y]$, dem Ring der Polynome in x und y mit komplexen Koeffizienten, seien vom Grad $N1$ und $N2$. Die beiden Polynome seien teilerfremd. Dann haben $p(x, y)$ und $q(x, y)$ genau $N1 \cdot N2$ gemeinsame Nullstellen.*

Definiert man nach Walker [9, p. 54], die Tangente, bzw. die Tangenten eines Polynoms p in einem Punkt X, so kann man zeigen:

Satz 2.3. *Ist X ein r-facher Punkt von p und ein s-facher Punkt von q, so ist X eine mindestens $r \cdot s$-fache gemeinsame Nullstelle von p und q. Die Vielfachheit betraegt genau $r \cdot s$, wenn keine Tangente an p in X Tangente an q in X ist,* Walker [9, p. 114].

Eine unmittelbare Folgerung ist

KOROLLAR 2.1. *Ist X ein einfacher Punkt von p und ein einfacher Punkt von q und faellt in diesem Punkt eine Tangente von q mit einer Tangente von p zusammen, so ist X eine mindestens zweifache gemeinsame Nullstelle von p und q.*

Wir weisen noch auf folgende Dinge hin:

1. Es sei $\mathbb{R}[x, y]$ der Ring der Polynome in x und y mit Koeffizienten aus dem Koerper $\mathbb{R}$ der reellen Zahlen, $\mathbb{C}[x, y]$ sei der entsprechende Polynomring mit komplexen Koeffizienten. In beiden Ringen gilt der Satz von der eindeutigen Faktorzerlegung.

2. Es gibt $d = (1/2)(N + 1)(N + 2)$ linear unabhaengige Monome $x^i y^j$, $i \geqslant 0, j \geqslant 0, i + j \leqslant N$, vom Grad $\leqslant N$. Daher gibt es zu $d - 1$ Punkten aus $\mathbb{R}^2$ bzw. $\mathbb{C}^2$ jeweils ein Polynom (vom Grad $\leqslant N$) aus dem zugehoerigen Polynomring, das in diesen Punkten verschwindet.

3. NICHTNEGATIVE POLYNOME 1

Es sei jetzt D ein kompakter Bereich des $\mathbb{R}^2$, der von einer nichtdegenerierten Ellipse berandet wird. Wir bezeichnen mit $P_N^2(D)$ den reellen linearen topologischen Raum der Funktionen, deren Zuordnungsvorschrift durch ein Polynom vom Grade $\leqslant N$ gegeben ist und die von D nach $\mathbb{R}$ abbilden. Den in D liegenden Teil des Nullstellengebildes eines Polynoms p nennen wir D_p, also $D_p = \{(x, y) \in D / p(x, y) = 0\}$.

Wir wollen im folgenden die nichtnegativen Elemente auf $P_N^2(D)$ untersuchen. Die Teiler solcher Funktionen klassifizieren wir bei den folgenden Betrachtungen nach der Anzahl ihrer Nullstellen in Int D und auf ∂D (= Rand von D).

Es sei f_∂ das Polynom zweiten Grades, das die Punkte von ∂D als reelle Nullstellen hat. Dann gilt

SATZ 3.1. *Es sei $p(x, y) \in \mathbb{R}[x, y]$ vom* Grade $N' \geqslant 1$. *Hat D_p mit ∂D mehr als $2 \cdot N'$ Punkte gemeinsam, so ist f_∂ Teiler von p.*

Beweis. Die Aussage dieses Satzes ist eine unmittelbare Folgerung des Satzes von Bézout.

Wir kommen nun zu einem weiteren wichtigen Teilergebnis.

SATZ 3.2. *Es sei* $p(x, y) \in \mathbb{R}[x, y]$, Grad $p \geqslant 2$ *und* $p(x, y) \geqslant 0$ *fuer* $(x, y) \in D$. f_∂ *sei nicht in* p *enthalten.* D_p *bestehe aus unendlich vielen verschiedenen Punkten.*

Dann enthaelt p *bei der Faktorzerlegung ueber* $\mathbb{R}$ *mindestens einen Teiler* q *mit unendlich vielen Nullstellen in Int* D. *Jeder solche Teiler* q *ist mehrfacher Teiler von* p *und zwar von gerader Vielfachheit.*

Beweis. Bei der Faktorzerlegung ueber $\mathbb{R}$ enthaelt p trivialerweise mindestens einen (ueber $\mathbb{R}$) irreduziblen Teiler mit unendlich vielen Nullstellen in D, den wir q nennen und der in irgendeiner Vielfachheit v in p enthalten sei. Der Einfachheit halber nehmen wir noch an, dass q der einzige solche Teiler sei.

Waere q auch ueber $\mathbb{C}$ irreduzibel, so koennte man schliessen, dass q in Punkten verschwindet, die eine genuegend klein waehlbare, offene Umgebung besitzen, in der q weder einen singulaeren Punkt noch q mit anderen Teilern von p gemeinsame Nullstellen hat. Da das Vorzeichenverhalten von p in dieser Umgebung nur von der Vielfachheit v abhaengt, folgt, dass v geradzahlig sein muss.

Wir zeigen nun noch, dass q tatsaechlich auch ueber $\mathbb{C}$ irreduzibel sein muss. Es wird dabei nur benuetzt, das D_p aus unendlich vielen Punkten besteht. Wir nehmen an, es gelte das Gegenteil. Dann sei $q = q_1^{v_1} q_2^{v_2} \cdots q_r^{v_r}$, $r \geqslant 2$, $q_j \in \mathbb{C}[x, y]$, $q_j \notin \mathbb{R}[x, y]$, q_j irreduzibel ueber $\mathbb{C}$, $N_j = \text{Grad}\, q_j$. Wir koennen das Polynom $\bar{q}_j$ bilden, dessen Koeffizienten die konjugiert komplexen Koeffizienten von q_j sind. Es gilt auch $q_j \in \mathbb{C}[x, y]$ und $q_j \in \mathbb{R}[x, y]$. q_j und $\bar{q}_j$ sind teilerfremd, da beide irreduzibel ueber $\mathbb{C}$ sind und sich nur um einen trivialen Teiler unterscheiden koennten; im letztgenannten Falle waere jedoch q_j bis auf einen trivialen komplexen Teiler auch aus $\mathbb{R}[x, y]$ im Widerspruch zur Voraussetzung. q_j und $\bar{q}_j$ haben daher N_j^2 gemeinsame Nullstellen. Da jede reelle Nullstelle von q_j auch reelle Nullstelle von $\bar{q}_j$ ist, sind die reellen Nullstellen von q_j in der Menge der gemeinsamen Nullstellen von q_j und $\bar{q}_j$ enthalten. Es gibt daher nur endlich viele reelle Punkte, fuer die q_j verschwindet. Dies gilt fuer alle j. Ware nun q ueber $\mathbb{C}$ reduzibel, so enthielte D_q nur endlich viele Punkte im Widerspruch zur Voraussetzung.

Wir haben damit gezeigt, dass die Teiler aus $\mathbb{R}[x, y]$ eines in D nichtnegativen Polynoms, die unendlich viele Nullstellen in D haben, entweder Potenzen von f_∂ oder mehrfache Teiler sind, welche in gerader Vielfachheit in diesem Polynom enthalten sind. Dieses Polynom kann daneben nur noch Teiler aus $\mathbb{R}[x, y]$ enthalten, die in D endlich viele oder ueberhaupt keine Nullstellen haben.

Die bisherigen Ergebnisse von 2. ueber nichtnegative Polynome lassen sich zusammenfassen in

SATZ 3.3. *Es sei* $p \in P_N^2(D)$, $N \geqslant 0$ *und* $p(x, y) \geqslant 0$ *fuer alle* $(x, y) \in D$. *Dann gilt folgende Zerlegung von p in nichtnegative Faktoren*

$$p = \alpha \cdot f_\partial^\lambda \cdot p_1^{2\mu_1} \cdots p_k^{2\mu_k} q_1^{v_1} \cdots q_r^{v_r}.$$

1. f_∂, *die* p_j *und die* q_j *sind irreduzibel ueber* $\mathbb{R}$.
2. λ, *die* u_j *und* v_j *sind positive ganze Zahlen.*
3. α *ist ein in D positives Polynom,* α *kann eine Konstante sein.*
4. *Falls* λ *ungerade ist, muss* $f_\partial(x, y) \geqslant 0$ *sein fuer* $(x, y) \in D$.
5. *Die* p_j *sind untereinander und von* f_∂ *verschieden, sie haben unendlich viele Nullstellen in D.*
6. *Die* q_j *sind untereinander verschieden, sind o.B.d.A. nichtnegativ in D. Sie haben endlich viele Nullstellen in D.*
7. *Alle Bestandteile der Zerlegung ausser* α *koennen fehlen.*
8. *Bis auf konstante Faktoren und die Reihenfolge ist diese Darstellung eindeutig.*

4. NICHTNEGATIVE POLYNOME II

In diesem Abschnitt untersuchen wir auf D nichtnegative Funktionen $q(x, y)$ aus $P_N^2(D)$, die in D nur endlich viele Nullstellen haben. Liegt eine solche Nullstelle X in Int D, so ist X eine isolierte Nullstelle von q und ist daher ein mindestens zweifacher Punkt von $q = 0$. Ist $X \in \partial D$ ein einfacher Punkt von $q = 0$, so haben f_∂ und q in X eine gemeinsame Tangente und nach Korollar 2.1 ist X eine mindestens zweifache gemeinsame Nullstelle von f_∂ und q. Wenn $X \in \partial D$ ein mindestens zweifacher Punkt von $q = 0$ ist, ist X, wie aus Satz 2.3 folgt ebenfalls eine mindestens zweifache gemeinsame Nullstelle von f_∂ und q. Fuer ein $q \in P_N^2(D)$ bestehe D_q aus endlich vielen Punkten. Wir nennen dann $a(q)$ die Anzahl der Punkte von $D_q \cap$ Int D, die Anzahl der Punkte von $D_q \cap \partial D$ nennen wir $b(q)$. Nach dem Satz von Bézout ist $b(q') \leqslant N'$, wenn $q' \in P_N^2(D)$ vom Grade N'.

Wir beweisen zuerst einen Satz ueber Polynome aus $\mathbb{R}[x, y]$, die in $\mathbb{R}^2$ nichtnegative sind. Wir wollen solche Polynome auch "totalnichtnegativ" nennen.

SATZ 4.1. *Es sei* $p(x, y) \in \mathbb{R}[x, y]$ *vom* Grade $N \leqslant 4$. *p sei totalnichtnegativ und habe nur endlich viele Nullstellen in* $\mathbb{R}^2$. *Dann ist N gerade und die Anzahl dieser Nullstellen ist* $\leqslant N^2/4$.

Beweis. Wie bei Polynomen in einer Veraenderlichen ueber der reellen Achse zeigt man, dass p nicht von ungeradem Grad N sein kann. Es kommen daher nur die Werte $N = 2$ und $N = 4$ in Frage.

Im Falle $N = 2$ kann man annehmen, dass p irreduzibel ueber $\mathbb{R}$ ist. Ein ueber $\mathbb{R}$ reduzibles p koennte keine isolierte Nullstelle in $\mathbb{R}^2$ haben. Haette man zwei isolierte Nullstellen p in $\mathbb{R}^2$, so koennte man ein lineares reelles Polynom q' finden, das in den beiden Nullstellen von p verschwindet und zu p teilerfremd ist. Dann haetten p und q' vier Nullstellen im Widerspruch zu satz 2.2.

Beispiele zeigen, dass die in Satz 2.1 genannten Schranken angenommen werden koennen. Seien $P1$ und $P2$ zwei Polynome in x und y mit reellen Koeffizienten vom Grade N', die $N' \cdot N'$ verschiedene reelle gemeinsame Nullstellen haben. Dann ist $P3 = \alpha \cdot P1 + \beta \cdot P2 + 2_\gamma \cdot P1 \cdot P2$ vom Grade $2N'$, totalnichtnegativ und hat $N' \cdot N'$ isolierte Nullstellen, wenn $\alpha > 0$ und $\alpha \cdot \beta - \gamma^2 > 0$.

Fuer $t \in \mathbb{R}$ sei $[t]$ die groesste ganze Zahl $\leqslant t$. Wir schaetzen nun $a(q)$ in Abhaengigkeit vom Grad q ab.

SATZ 4.2. *Es sei $q(x, y) \in \mathbb{R}[x, y]$ vom Grade $N \leqslant 4$, nichtnegativ in D. q habe nur isolierte Nullstellen in Int D. Dann ist $a(q) \leqslant [N/2]^2$.*

Beweis. $N = 2$ und $N = 4$ sind beim Beweis von Satz 4.1 erledigt worden, da dort beim Abschaetzen der Anzahl der isolierten Nullstellen nicht benuetzt wurde, dass p totalnichtnegativ war. Der Fall $N = 1$ ist trivial, der Fall $N = 3$ wird wie der Fall $N = 2$ in Satz 4.1 bewiesen.

Eine Konsequenz unsrer Betrachtungen ist noch, dass fuer ein in D nichtnegatives Polynom p, das in D nur endlich viele Nullstellen hat, $a(p)$ und $b(p)$ nicht voneinander unabhaengig zu sein brauchen. Man kann sofort einsehen, dass im Falle $N = 2$ $b(p) = 0$ sein muss, wenn $a(p) = 1$ ist. Entsprechend zeigt man, dass im Falle $N = 4$ auch $b(p) = 0$ sein muss, wenn $a(p) = 4$ ist.

Wir schreiben als Satz 4.3 eine Tabelle der moeglichen maximalen Kombinationen von $a(p)$ und $b(p)$ in Abhaengigkeit von N fuer $N \leqslant 5$.

SATZ 4.3. *Es sei $p(x, y) \in \mathbb{R}[x, y]$ nichtnegativ ueber D, vom* Grade $N \leqslant 5$. *p habe nur endlich viele Nullstellen in D. Dann koennen die folgenden gemeinsamen Schranken fuer $a(p)$ und $b(p)$ nicht ueberschritten werden.*

$N = 1:$	(a) $a(p) = 0$	$b(p) = 1$
$N = 2:$	(a) $a(p) = 1$	$b(p) = 0$
	(b) $a(p) = 0$	$b(p) = 2$
$N = 3:$	(a) $a(p) = 1$	$b(p) = 3$
$N = 4:$	(a) $a(p) = 4$	$b(p) = 0$
	(b) $a(p) = 3$	$b(p) = 4$
$N = 5:$	(a) $a(p) = 6$	$b(p) = 5$

Wir muessen nur noch beweisen, dass fuer $N = 5$ $a(p)$ nicht groesser als sechs sein kann. Dazu unterscheiden wir drei Faelle.

(a) Ist p ueber $\mathbb{R}$ reduzibel, so kann man $a(p) \leqslant 6$ aus den Abschaetzungen fuer $N \leqslant 4$ herleiten.

(b) Ist p irreduzibel ueber $\mathbb{R}$, jedoch reduzibel ueber $\mathbb{C}$, so schliesst man aehnlich wie beim Beweis von Satz 3.2, dass p mit einem Teiler q, $q \in \mathbb{C}[x, y]$, $q \notin \mathbb{R}[x, y]$, auch den Teiler $\bar{q}$ mit den konjugiert komplexen Koeffizienten enthaelt. Dies fuehrt auf das Ergebnis. dass $a(p)$ in diesem Fall hoechstens den Wert 4 annehmen kann.

(c) Es sei jetzt p irreduzibel ueber $\mathbb{C}$. Wir wollen dann annehmen, dass p mindestens sieben isolierte Nullstellen X_1, X_2,..., X_7 hat. Wir konstruieren ein Polynom q vom Grad drei, das sechs dieser X_i als zumindest einfache Nullestellen und einen der x_i als mindestens doppelte Nullstelle hat. Man kann dies durch 9 Bedingungen an die 10 Koeffizienten von q schaffen. Die Polynome p und q sind teilerfremd, da p irreduzibel ueber $\mathbb{C}$ ist. Waere unsre Annahme ueber die Anzahl isolierter Nullstellen korrekt, so haetten p und q mindestens 16 gemeinsame Nullstellen gemaess Satz 2.3 im Widerspruch zu Satz 2.2.

Der allgemeine Satz, den wir hier nicht beweisen, lautet:

SATZ 4.4. *Es sei $p(x, y) \in \mathbb{R}[x, y]$ nichtnegativ ueber D, vom* Grade $N \geqslant 1$. *p habe nur endlich viele Nullstellen in D. Dann koennen die folgenden Schranken fuer $a(p)$ und $b(p)$ in Abhaengigkeit von N nicht ueberschritten werden.*

$$N \textit{ gerade } (\mathrm{a})\ a(p) = \frac{(N-1)(N-2)}{2} + 1 \qquad b(p) = 0$$

$$(\mathrm{b})\ a(p) = \frac{(N-1)(N-2)}{2} \qquad b(p) = N$$

$$N \textit{ ungerade } (\mathrm{a})\ a(p) = \frac{(N-1)(N-2)}{2} \qquad b(p) = N$$

5. AUSBLICK

Zweierlei Verallgemeinerungen der hier erarbeiteten Ergebnisse scheinen von besonderem Interesse zu sein. 1. Verallgemeinerungen auf mehr als zwei Dimensionen, wenn D ein Ellipsoid in n Dimensionen ist, 2. Die Verallgemeinerung auf den Fall, dass D aus einer Ellipsen-, bzw. aus einer Kugeloberflaeche besteht. Die Schwierigkeit bei der Behandlung dieser Faelle liegt darin, dass als Nullstellenkurven nichtnegativer Polynome verschiedendimensionale Punktmengen auftreten koennen.

Ein einfaches, nahezu triviales Beispiel fuer Polynomgrad zwei und in [3] angegeben.

Verallgemeinerungen auf andere zweidimensionale Bereiche D, z.B. Dreiecke, sind offensichtlich. Es ist unmittelbar einzusehen, dass die Werte von $a(p)$ unveraendert gelten, waehrend $b(p)$ i.a. groesser wird.

Literaturverzeichnis

1. P. J. Davis, A construction of nonnegative approximate quadratures, *Math. Comp.* **21** (1967), 578–582.
2. C. Günther, Third degree integration formulas with four real points and positive weights in two dimensions, *SIAM J. Num. Anal.*, **11** (1974), 480–493.
3. C. Günther, Ueber die Anzahl von Stuetzstellen in mehrdimensionalen Integrationsformeln mit positiven Gewichten, Interner Bericht 1973/5, Institut für Praktische Mathematik, Universität Karlsruhe.
4. S. J. Karlin und W. Studden, "Tchebycheff Systems." Interscience, New York, 1966.
5. M. G. Krein, The ideas of P. L. Cebysev and A. A. Markov in the theory of limiting values of integrals and their further developments, *Amer. Math. Soc. Transl.*, *Ser. 2*, **12** (1951), 1–122.
6. F. Locher, Optimale definite Polynome und Quadraturformeln, Proc. Symp. Numerische Methoden bei Optimierungsaufgaben, Oberwolfach, November 1971 Birkhaeuser Verlag, Basel, *ISNM* **17** (1973), 111–121.
7. T. S. Motzkin, Algebraic Inequalities, *in* "Inequalities" (O. Shisha, Ed.), Vol. 1, pp. 199–202, Academic Press, New York, London, 1967.
8. G. Pólya und G. Szegö, "Aufgaben und Lehrsaetze aus der Analysis," Band II. Springer-Verlag, Berlin, 1964.
9. R. J. Walker, "Algebraic Curves," Dover, New York, 1962.