ISSN 0138 - 3248

Rostocker Mathematisches Kolloquium

Heft 26

WILHELM-PIECK-UNIVERSITÄT
ROSTOCK

ROSTOCKER MATHEMATISCHES KOLLOQUIUM

Heft 26

1984

Wilhelm-Pieck-Universität Rostock
Sektion Mathematik

Herausgeber: Der Rektor der Wilhelm-Pieck-Universität Rostock

Schriftleitung: Prof. Dr. Wolfgang Engel, Direktor der Sektion Mathematik
Prof. Dr. Gerhard Maeß, Schriftleiter
Dr. Werner Plischke, Lektor
Dorothea Meyer, Herstellung der
Christiane Remer, Druckvorlage

Sektion Mathematik der
Wilhelm-Pieck-Universität Rostock,
DDR-2500 Rostock, Universitätsplatz 1

Das Rostocker Mathematische Kolloquium erscheint in der Regel dreimal im Jahr und ist im Rahmen des Schriftentausches über die Universitätsbibliothek, Tauschstelle, DDR-2500 Rostock, Universitätsplatz 5, zu beziehen.

Veröffentlicht durch die Abt. Wissenschaftspublizistik der Wilhelm-Pieck-Universität Rostock, DDR-2500 Rostock
Vogelsang 13/14, Fernruf 369 577

Genehmigungs-Nr.: C 70/84
Druck: Ostsee-Druck Rostock, Werk II Ribnitz

https://doi.org/10.18453/rosdok_id00003837

Inhalt

Rostock. Math. Kolloq. 26, 5 - 16 (1984)

35Q20
73D35

Peter Möbius

Methode der Spektraltransformation[1]

In den letzten beiden Jahrzehnten wurden bei der Entwicklung allgemeiner Verfahren zur Lösung von nichtlinearen partiellen Differentialgleichungen beachtliche Fortschritte erzielt. Die Methode der Spektraltransformation zählt gegenwärtig zu den bedeutendsten Erfolgen in dieser Richtung, zu dem "großen Durchbruch" /1/. Die Ursachen für die verstärkten Anstrengungen hierzu dürften auf der Tatsache beruhen, daß es sich infolge der raschen Zunahme der allgemeinen Experimentiertechnik, die sich z.B. im Bau moderner Großgeräte und Hochenergiebeschleuniger, aber auch in der Verfeinerung der Meßverfahren widerspiegelt, als notwendig erwies, nichtlineare Differentialgleichungen zur Modellierung physikalischer Erscheinungen zu verwenden. Das bewirkte einen Übergang von der linearen zur nichtlinearen Feldtheorie, der in kurzer Zeit von anerkannten Erfolgen begleitet war. Besonders augenfällig wurden sie im Rahmen der nichtabelschen Eichfeldtheorie, einer hochgradig nichtlinearen Feldtheorie, die zu einer grundsätzlich neuen Sicht in der Vereinheitlichung der fundamentalen Wechselwirkungen der elementaren Bausteine führte. Im Falle der Eichfeldtheorie der elektroschwachen Wechselwirkungen /2/ wurden diese Bemühungen sogar mit dem Nobelpreis für Physik für Glashow, Salam und Weinberg belohnt. Sehr interessante Ergebnisse sind ferner bei der Beschreibung kollektiver Bewegungen mit großer Amplitude, die zum Beispiel in der Schwerionenphysik auftreten, bei den Untersuchungen über die Ausbreitung verschiedenartiger nichtlinearer Wellen und vor allem bei der Behandlung der technischen Selbstwechselwirkungseffekte zu

1 Vortrag, gehalten am 20.1.1984 im Mathematischen Kolloquium anläßlich des 75. Geburtstages von Prof. em. Dr. Adam Schmidt

verzeichnen, wie etwa der Selbstmodulation von Schwingungen, der Selbststabilisierung von Schwingungsamplituden und der Selbstfokussierung von Strahlung in dispersiven Medien /3/. Eine große Rolle spielen hierbei nichtlineare Evolutionsgleichungen, die nichtlineare partielle Differentialgleichungen in Raum und Zeit sind und die zeitliche Entwicklung physikalischer Größen modellieren. Für eine Anzahl spezieller derartiger Gleichungen in einer Raum- und Zeitdimension gelang es nun, ein systematisches Verfahren zu entwickeln, die "Methode der Spektraltransformation", die es gestattet, das zugehörige Cauchyproblem zu behandeln /1/, also die zeitliche Entwicklung der Lösung aus einer vorgegebenen Anfangsverteilung zu ermitteln. Es ist durchaus berechtigt, dieses Verfahren als eine Erweiterung der Methode der Fouriertransformation anzusehen, die ja zur Lösung des Cauchyproblems für lineare partielle Evolutionsgleichungen dient und ursprünglich von Fourier zur Behandlung der Wärmeleitung verwendet wurde.
Die Grundlagen dieser Methode sind in der inversen Streutheorie zu finden. Die ersten systematischen Betrachtungen hierzu stammen aus dem Jahre 1967, als Gardner, Greene, Kruskal und Miura /4/ begannen, allgemeine Lösungen der Korteweg - de Vries - Gleichung (KdV-GL)

$$\frac{\partial v}{\partial t} + \alpha v \frac{\partial v}{\partial x} + \beta \frac{\partial^3 v}{\partial x^3} = 0$$

zu ermitteln, wobei $v(x,t)$ die gesuchte Funktion ist und α und β Konstanten sind. Bekannt ist die Anwendung dieser Gleichung zur Beschreibung von Wellen an der Oberfläche seichten Wassers, wobei $v(x,t)$ das Geschwindigkeitsfeld bedeutet.
Einige Jahre später, etwa 1972, wurde die nichtlineare parabolische Gleichung der Gestalt

$$i \frac{\partial \psi}{\partial t} + a \frac{\partial^2 \psi}{\partial x^2} + k^2 (\psi^* \psi)\psi = 0,$$

oft auch als nichtlineare Schrödingergleichung bezeichnet, zur Behandlung der zweidimensionalen Selbstfokussierung und der eindimensionalen Selbstmodulation von Wellen in nichtlinearen Medien exakt gelöst, wobei zur KdV-Gleichung analoge Verfahren

Anwendung fanden.
Im letzten Jahrzehnt entstand hieraus nun ein umfassendes Verfahren, derartige spezielle nichtlineare Evolutionsgleichungen bei vorgegebener Anfangsverteilung zu lösen, das jetzt als die Methode der Spektraltransformation bezeichnet wird. Sie beruht im wesentlichen darauf, daß die Lösung des Cauchyproblemes in drei Schritten erreicht wird. Zunächst wird der nichtlinearen Evolutionsgleichung auf funktionentheoretischem Wege ein lineares Eigenwertproblem zugeordnet, welches die vorgegebene Anfangsverteilung enthält. Hieraus folgt die "Spektraltransformierte", die die Eigenschaften des zugehörigen diskreten und kontinuierlichen Teiles des Spektrums widerspiegelt. Dann wird die zeitliche Entwicklung derjenigen Größen, die in der Spektraltransformierten auftreten, ermittelt und letztlich durch die Anwendung der inversen Spektraltransformation die gesuchte Lösungsfunktion gewonnen. Dieses Verfahren hieß anfänglich (1974) "AKNS-Methode" um anzudeuten, daß die vier Autoren Ablowitz, Kaup, Newell und Segur /5/ wesentliche Beiträge dazu lieferten.
Gegenwärtig sind folgende vier Grundtypen von nichtlinearen Evolutionsgleichungen

$$\frac{\partial \Phi}{\partial t} + \alpha\Phi\frac{\partial \Phi}{\partial x} + \beta\frac{\partial^3 \Phi}{\partial x^3} = 0$$

Korteweg-de Vries-Gleichung, (1)

$$\frac{\partial \Phi}{\partial t} + \alpha_n\Phi^n\frac{\partial \Phi}{\partial x} + \beta\frac{\partial^3 \Phi}{\partial x^3} = 0 \quad (n = 2,3,\ldots)$$

modifizierte Korteweg-de Vries-Gleichung, (2)

$$\frac{1}{c^2}\frac{\partial^2 \Phi}{\partial t^2} - \frac{\partial^2 \Phi}{\partial x^2} + \varkappa^2 \sin\Phi = 0$$

Sinus-Gordon-Gleichung, (3)

$$i\frac{\partial \psi}{\partial t} + a\frac{\partial^2 \psi}{\partial x^2} + k(\psi^*\psi)\psi = 0$$

nichtlineare parabolische Gleichung (nichtlineare Schrödingergleichung) (4)

mit ihren zugehörigen "Hierarchien" umfassend untersucht, da das entsprechende lineare Eigenwertproblem zu (1) durch die eindimensionale gewöhnliche Schrödingergleichung der

Quantentheorie gegeben ist, während es in den drei anderen Fällen durch die eindimensionale mehrkomponentige Schrödingergleichung, oft auch Matrix-Schrödingergleichung genannt, gegeben ist.

1. Grundlagen der Methode

Zunächst ist es zweckmäßig, die Grundlagen der Methode der Fouriertransformation nochmals zu skizzieren und sowohl die Analogien als auch die notwendigen Erweiterungen klar herauszustellen. Gegeben sei eine lineare Evolutionsgleichung der Gestalt

$$\frac{\partial q}{\partial t} = -i\omega(-i\frac{\partial}{\partial x},t)\, q(x,t), \qquad x \in R\ ,\ t \in R\ ,\ q(x,t) \in C\ , \tag{5}$$

wobei $\omega(z,t)$ eine ganze Funktion ist und $q(x,0)$ die vorgegebene Anfangsverteilung. Zunächst gehen wir vom x-Raum in den k-Raum mittels der Fouriertransformation

$$\hat{q}(k,t) = \int_{-\infty}^{+\infty} dx\ \ q(x,t)e^{-ikx} \tag{6}$$

über, die jeder Funktion $q(x,t)$ eine andere Funktion $\hat{q}(k,t)$, die sogenannte Fouriertransformierte, zuordnet.
Die Zeitentwicklung von $\hat{q}(k,t)$ folgt durch Integration der Gleichung

$$\frac{\partial \hat{q}}{\partial t} = -i\omega(k,t)\ \hat{q}(k,t)\ , \tag{7}$$

die mittels der Bezeichnung

$$\hat{q}(k,0) = \int_{-\infty}^{+\infty} q(x,0)e^{-ikx}dx \tag{8}$$

zu dem Ergebnis führt

$$\hat{q}(k,t) = \hat{q}(k,0)\ \exp\left[-i\int_{0}^{t} dt'\omega(k,t')\right]. \tag{9}$$

Die Lösung für beliebige Zeiten ergibt sich nun durch die inverse Fouriertransformation aus der Anfangsverteilung auf folgende Weise

$$q(x,t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} dk \quad \hat{q}(k,t)e^{ikx}$$

$$= \frac{1}{2\pi}\int_{-\infty}^{+\infty} dk \quad \hat{q}(k,0)e^{ikx}\exp\left[-i\int_0^t dt'\omega(k,t')\right]. \qquad (10)$$

Das läßt sich durch die folgenden Schritte übersichtlich charakterisieren

$$q(x,0)\rightarrow\hat{q}(k,0)\rightarrow\hat{q}(k,t)\rightarrow q(x,t), \qquad (11)$$

bei denen aus einer vorgegebenen Anfangsverteilung eine andere Funktion, die Fouriertransformierte, folgt, deren Zeitentwicklung berechnet wird und die dann durch Anwendung der inversen Fouriertransformation die gesuchte Lösungsfunktion liefert. Dieses Schema wird nun im Prinzip in der Methode der Spektraltransformation übernommen. Aus der Anfangsverteilung $q(x,0) \equiv q(x)$ konstruieren wir im Falle von Gleichung (1) den Operator

$$\mathcal{H}(x) = -\frac{d^2}{dx^2} + q(x)\ , \qquad (12)$$

stellen dann das Eigenwertproblem auf,

$$\mathcal{H}(x)\varphi(x) = \lambda\varphi(x), \qquad (13)$$

und suchen quadratisch integrable Eigenfunktionen unter der Voraussetzung, daß die Anfangsverteilung absolut integrierbar ist:

$$\int_{-\infty}^{+\infty} |q(x)|\,dx < \infty.$$

Nun ist (13) der Schrödingergleichung für die Bewegung eines Teilchens im Potential $q(x)$ ähnlich. Folglich gibt es für $\lambda < 0$ diskrete Eigenwerte λ_n, und die zugehörigen Eigenfunktionen weisen mit der Abkürzung $\sqrt{-\lambda_n} = p_n$ folgende asymptotische Gestalt auf /1/

$$\varphi_n(x) = C_n e^{-p_n x}(1 + O(1)) \quad \text{für } x\rightarrow\infty, \qquad (14)$$

während für $\lambda>0$ ein kontinuierliches Spektrum auftritt, dessen Lösungsfunktionen mittels $\lambda = k^2$ die Form annehmen

$$\varphi_k(x) = (e^{-ikx} + R(k)e^{ikx})(1 + O(1)) \quad \text{für } x \to \infty, \tag{15}$$

wobei R(k) den Reflexionskoeffizienten darstellt.
Die Größen

$$\left\{\sqrt{-\lambda_n} \equiv p_n, \quad c_n^2, \quad R(k)\right\} = S_q(0)$$

bilden die Spektraltransformierte $S_q(0)$ von q(x) zur Zeit t = 0, und die Spektraltransformation ist die Zuordnung $q(x) \to S_q(0)$. In Analogie zur Fouriertransformation erhebt sich die Frage, ob es auch eine inverse Spektraltransformation gibt, die jedem $S_q(0)$ ein q(x) zuordnet. Unter gewissen Bedingungen ist die Aufgabe eindeutig lösbar. Hierzu wird eine Hilfsgröße M(x) benötigt /1/

$$M(x) = \sum_{n=1}^{N} c_n^2 e^{-p_n x} + \frac{1}{2\pi} \int_{-\infty}^{+\infty} R(k)e^{-ikx}dx, \tag{16}$$

wobei die Summe über alle Eigenwerte des diskreten Spektrums läuft, die in diesem Falle einfach sind. Mittels M(x) wird folgende Integralgleichung aufgestellt

$$K(x,y) + M(x + y) + \int_x^{\infty} K(x,z)M(z + y)dz = 0, \tag{17}$$

die häufig als Gelfand-Levitan-Marchenko-Gleichung bezeichnet wird. Deren Lösung hängt nun mit der gesuchten Funktion folgendermaßen zusammen

$$q(x) = -2\frac{d}{dx}K(x,x). \tag{18}$$

Somit ist es prinzipiell möglich, aus der Spektraltransformierten die ursprüngliche Funktion zurückzugewinnen. Ein einfaches Beispiel soll den Sachverhalt erläutern. Die Spektraltransformierte bestehe nur aus einem diskreten Eigenwert $\sqrt{-\lambda} = p$, und der Reflexionskoeffizient sei Null, $S(0) = \{p, c^2, 0\}$. Das ergibt für (17)

$$K(x,y) + c^2 e^{-p(x+y)} + \int_x^{\infty} c^2 K(x,z)\, e^{-p(z+y)}dz = 0 \tag{19}$$

und liefert für q(x) die Funktion

$$q(x) = \frac{-2p^2}{\cosh^2 p(x-x_o)} \qquad \text{mit} \quad e^{2px_o} = C^2/2p \ , \tag{20}$$

welche eine "solitäre" Lösung der KdV-Gleichung verkörpert, die weiter unten noch erläutert wird. Besitzt die Spektraltransformierte mehrere diskrete Eigenwerte λ_n und ebenfalls wieder einen verschwindenden Reflexionskoeffizienten $R(k) \equiv 0$, so ergibt sich als Lösung von (17)

$$q(x) = -2 \frac{d^2}{dx^2} \ln \det \left| I + A(x) \right| , \tag{21}$$

bei der I die Einheitsmatrix und $A(x) = (A_{mn}(x))$ eine N×N-Matrix ist mit den Elementen

$$A_{mn} = C_m C_n \frac{e^{-(p_m + p_n)x}}{p_m + p_n} \ . \tag{22}$$

Nachdem die beiden Schritte, die Bestimmung der Spektraltransformierten $S_q(0)$ zur Zeit $t = 0$ aus der Anfangsverteilung $q(x)$ und die Umkehrung, die Berechnung der Funktion $q(x)$ aus $S_q(0)$, behandelt wurden, wenden wir uns nun dem Problem zu, die Zeitabhängigkeit der Größen, die in die Spektraltransformierte eingehen, aufzusuchen, d.h., die zugehörigen gewöhnlichen Differentialgleichungen aufzustellen, die sich aus der nichtlinearen Evolutionsgleichung ergeben. Wir wollen das jedoch nicht nur für die KdV-Gleichung (1) allein ausführen, sondern gleich für alle Evolutionsgleichungen, die zur "KdV-Hierarchie" gehören. Hierzu führen wir den Operator L ein

$$LF(x,t) = F_{xx}(x,t) - 4q(x,t)F(x,t) + 2q_x(x,t) \int_x^\infty F(x',t)dx' \ , \tag{23}$$

wobei $F(x,t)$ eine reelle Funktion von x und t ist und der Index "x" die partielle Ableitung nach x bedeutet. Nun seien $\alpha(z,t)$ und $\beta(z,t)$ ganze Funktionen in z und $q(x,t)$ eine beliebige reelle Funktion. Mit ihrer Hilfe läßt sich jetzt die Gesamtheit aller zur Hierarchie der Gleichung (1) gehörenden nichtlinearen partiellen Differentialgleichungen in der Gestalt schreiben

$$\beta(L,t) \frac{\partial q}{\partial t} = \alpha(L,t) \frac{\partial q}{\partial x} \ , \tag{24}$$

die bereits eine Vielzahl von Gleichungen umfaßt, wenn $\alpha=\alpha(z)$ alle Polynome beliebigen Grades durchläuft. Für alle zur Hierarchie (24) gehörenden Gleichungen wird die Zeitabhängigkeit der zur Spektraltransformierten gehörenden Größen nun durch folgendes lineare gewöhnliche Differentialgleichungssystem erfaßt

$$\frac{dp_n}{dt} = 0,$$

$$\beta(4p_n^2,t)\,\frac{dC_n}{dt} = -\,p_n\,\alpha(4p_n^2,t)C_n(t), \tag{25}$$

$$\beta(-4k^2,t)\,\frac{dR}{dt}(k,t) = 2ik\,\alpha(-4k^2,t)R(k,t),$$

so daß auf Grund der ersten Gleichung die Eigenwerte zeitlich konstant sind. Diese "Isospektralität" gehört zur Grundlage der Methode. Die allgemeine Integration der beiden anderen Gleichungen erfordert jedoch die Kenntnis der Funktionen $\alpha(z,t)$, $\beta(z,t)$. Da es sich aber um lineare Gleichungen handelt, ist es also prinzipiell möglich, die Spektraltransformierte zur Zeit t zu berechnen

$$S_q(t) = \{p_n,\ C_n(t),\ R(k,t)\}.$$

Zwei instruktive Beispiele wollen wir diskutieren, um die Bedeutung und den Umfang dieser Methode zu veranschaulichen.

1. Beispiel: $\alpha(z,t) = z$, $\beta = 1$ ergibt $\frac{\partial q}{\partial t} = L\,\frac{\partial q}{\partial x}$,

was ausführlich geschrieben die KdV-Gleichung (1) mit speziellen Koeffizienten liefert

$$\frac{\partial q}{\partial t} + 6q\,\frac{\partial q}{\partial t} - \frac{\partial^3 q}{\partial x^3} = 0. \tag{26}$$

Die Zeitabhängigkeit der Spektraltransformierten läßt sich hier explizit angeben

$$p_n(t) = p_n,\quad C_n(t) = C_n(t_o)\,e^{-4p_n^3(t-t_o)},$$
$$R(k,t) = R(k,t_o)\,e^{-8ik^3(t-t_o)}, \tag{27}$$

so daß damit alle Schritte für die KdV-Gleichung (1) erörtert sind.

2. Beispiel: $\alpha(z,t) = \alpha(z)$, $\beta = 1$:

$$\frac{\partial q}{\partial t} = \alpha(L)\,\frac{\partial q}{\partial x}\,, \tag{28}$$

woraus sich Gleichungen 5., 7., 9.,.... Ordnung in x ergeben. Hier gelingt es noch, die Zeitabhängigkeit der Spektraltransformierten allgemein anzugeben

$$\begin{aligned} p_n(t) &= p_n, \qquad C_n(t) = C_n(t_o)\, e^{-p_n \alpha(4p_n^2)(t-t_o)}\,, \\ R(k,t) &= R(k,t_o)\, e^{2ik\,\alpha(-4k^2)(t-t_o)}\,. \end{aligned} \tag{29}$$

Es lohnt sich, folgenden Spezialfall der Spektraltransformierten zu diskutieren. Es sei nur ein Eigenwert $\sqrt{-\lambda} = p$ vorhanden und der Reflexionskoeffizient verschwinde wieder:

$$S_q(t) = \{p,\ C(t),\quad R \equiv 0\}. \tag{30}$$

Dann ergibt sich zunächst für die Lösung

$$q(x,t) = -\frac{\partial^2}{\partial x^2}\,\ln\!\left(1 + \frac{C^2(t)}{2p}\, e^{-2px}\right)\,, \tag{31}$$

bei der sich die Abkürzung anbietet

$$\frac{C^2(t)}{2p} = \frac{C^2(t_o)}{2p}\, e^{-2p\,\alpha(4p^2)(t-t_o)} \Rightarrow e^{2p(x_o + v(t-t_o))} \tag{32}$$

und in der $v = -\alpha(4p^2)$ die Ausbreitungsgeschwindigkeit der Störung darstellt. Nun ist es instruktiv, (31) in die bekannte Form einer "Einsolitonenlösung" der KdV-Gleichung umzuschreiben

$$q(x,t) = \frac{-2p^2}{\cosh^2 p(x - x_o - v(t - t_o))}\,. \tag{33}$$

Treten aber N diskrete Eigenwerte in der Spektraltransformierten auf

$$S_q(t) = \{p_1 \ldots p_N,\ C_1(t) \ldots C_n(t),\quad R \equiv 0\}, \tag{34}$$

lassen sich leicht überschaubare Aussagen nur über das asymptotische Verhalten der Lösung angeben. Bezeichnen $v_n=-\alpha(4p_n^2)$ $(n = 1,\ldots,N)$ die unterschiedlichen Geschwindigkeiten der Solitonen, so besitzt die N-Solitonenlösung für $t\to-\infty$ die Gestalt

$$\lim_{t\to-\infty} q(x,t) = \sum_{1}^{N} \frac{-2p_n^2}{\cosh^2 p_n(x - x_n - v_n(t - t_o))}, \qquad (35)$$

während sie für $t\to\infty$ übergeht in

$$\lim_{t\to\infty} q(x,t) = \sum_{1}^{N} \frac{-2p_n^2}{\cosh^2 p_n(x - x_n - \Delta_n - v_n(t - t_o))}, \qquad (36)$$

wobei

$$C_n^2(t_o) = 2p_n\, e^{2p_n x_n}$$

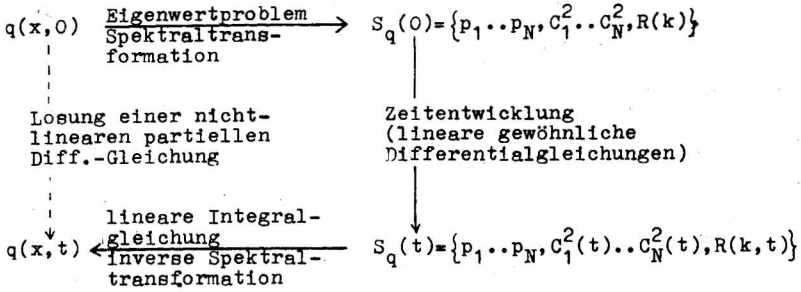
ist und $\Delta_n(p_1\ldots p_N)$ eine Phasenverschiebung bedeutet. Enthält die Spektraltransformierte nun noch einen nichtverschwindenden Reflexionskoeffizienten R(k,t), so treten in der Lösung zusätzliche, rasch veränderliche Wellenzüge auf, die im Laufe der Zeit abklingen. Der hier verwendete Begriff des Solitons kennzeichnet einen bestimmten Lösungstyp von speziellen nichtlinearen Differentialgleichungen in Raum und Zeit, der teilchenartigen Charakter besitzt und sich durch besondere Lokalisierungs- und Stabilitätseigenschaften auszeichnet. Im Rahmen der Methode der Spektraltransformation läßt sich diese physikalische Festlegung des Begriffes Soliton durch die mathematische Forderung "ergänzen", daß in diesem Falle die Spektraltransformierte einen verschwindenden Reflexionskoeffizienten besitzt; d.h., die Spektraltransformierte

$$S_q(t) = \{p_1\ldots p_N,\ C_1(t)\ \ldots\ C_N(t),\ 0\} \qquad (37)$$

führt auf eine N-Solitonenlösung. Leider ist es bislang noch nicht gelungen zu zeigen, daß die anschauliche physikalische "Definition" mit der mathematischen über die Spektraltransformierte übereinstimmt.

Es ist abschließend zweckmäßig, nochmals die wesentlichen

Schritte dieser Methode in Form eines Diagrammes darzustellen /6/.



Dabei wird ersichtlich, daß mit dem Auffinden des Eigenwertproblemes, welches der nichtlinearen Evolutionsgleichung mit Hilfe eines Riemann-Hilbert-Problemes zugeordnet werden kann /5/, sich alle weiteren Schritte, also die Lösung des Eigenwertproblemes, das Aufstellen der Spektraltransformierten, die Ermittlung ihrer Zeitentwicklung, die Berechnung der inversen Spektraltransformation, durch lineare Verfahren bewältigen lassen. Somit liefert diese Methode das überraschende Resultat, daß die Lösungen spezieller nichtlinearer partieller Differentialgleichungen mittels einer geschickten Kombination linearer Verfahren zu gewinnen sind und damit die Zahl der "integrablen Systeme" unerwartet rasch gestiegen ist. Das eröffnet sowohl für die nichtlineare Feldtheorie mit ihren vielseitigen Anwendungen in verschiedenen Gebieten der Physik als auch für die allgemeine mathematische Lösungstheorie spezieller nichtlinearer partieller Differentialgleichungen eine ungeahnte Perspektive, die hiermit angedeutet werden sollte.

Literatur

/1/ Rañada, A. F. (Ed.): Nonlinear Problems in Theoretical Physics. Lecture Notes in Physics, Bd. 98. Berlin 1979

/2/ Möbius, P.: Unifikation der Kräfte. In: Stendel, H., und Möbius, P.: Aspekte nichtlinearer Feldtheorie III. ZIE Preprint 82 - 5, Berlin 1982

/3/ Scott, A. C. , Chu, F. Y. F. , and McLaughlin, W.: The soliton: a new concept in applied science. Proc. IEE-E 61, 1443 - 1483 (1973)

/4/ Gardner, C. S., Greene, J. M., Kruskal, M. D., and Miura, R. M.: Methods for solving the Korteweg-de Vries equation. Phys. Rev. Lett. 19, 1095 - 1097 (1967)

/5/ Ablowitz, M. J., Kaup, D. J., Newell, A. C., and Segur, H.: The inverse scattering transform-Fourier analysis for nonlinear problems. Stud. Appl. Math. 53, 249 - 315 (1974)

/6/ Möbius, P.: Nonlinear field theory. In: Kurke, H., et al. (Eds.): Recent Trends in Mathematics. Conference in Reinhardsbrunn, Oct. 11 - Oct. 13, 1982. Teubner - Texte zur Mathematik, Bd. 50, pp. 193 - 203. Leipzig 1983

eingegangen: 24. 04. 1984

Anschrift des Verfassers:

Prof. Dr. P. Möbius
Technische Universität Dresden
Sektion Physik
Mommsenstraße 13

DDR-8027 Dresden

Rostock. Math. Kolloq. 26, 17 - 26 (1984)

35J40
35A35

Günther Wildenhain

Approximation durch Lösungen elliptischer Randwertprobleme[1]

1. Die Thematik des folgenden Vortrages geht zurück auf eine Arbeit von H. Beckert /2/, in der die folgende Aussage bewiesen wurde. Es sei L ein elliptischer Differentialoperator zweiter Ordnung mit hinreichend glatten Koeffizienten, $\Omega \subset R^n$ ein beschränktes Gebiet mit glattem Rand $\partial\Omega$, $\Gamma \subset \Omega$ eine glatte (n-1)-dimensionale Fläche, welche Ω nicht zerlegt (d. h. $\Omega \setminus \Gamma$ ist zusammenhängend), $V \subset \partial\Omega$ eine offene Teilmenge des Randes und

$$L_V(\Omega) = \{u \in C^2(\Omega) \cap C(\bar{\Omega}): Lu=0 \text{ in } \Omega,\ u|_{\partial\Omega\setminus V} = 0\}.$$

Ferner wird angenommen, daß das homogene Dirichlet-Problem der Gleichung $Lu = 0$ in Ω nur die triviale Lösung besitzt. Dann liegt der Raum $L_V(\Gamma)$ aller Einschränkungen von $L_V(\Omega)$ auf Γ dicht im Raum $L^2(\Gamma)$, d. h., durch Variation der Randwerte auf einem festen (aber beliebig klein vorgebbaren) Teil V des Randes kann eine in $L^2(\Gamma)$ dichte Lösungsmenge erzeugt werden. Von A. Göpfert /6/, /7/, G. Wanka /10/ und G. Anger /1/ wurde unter anderem dieses Resultat auf den Fall einer geschlossenen Fläche sowie auf den Fall gleichmäßiger Approximation verallgemeinert. Das Anliegen dieses Vortrages besteht darin, über Möglichkeiten der Verallgemeinerung dieser Aussagen auf elliptische Differentialgleichungen beliebiger Ordnung zu berichten. Beiläufig sei erwähnt, daß verwandte Approximationssätze (auch für elliptische Gleichungen höherer Ordnung) bei F. E. Browder /5/ sowie bei B.-W. Schulze und G. Wildenhain /9/ bewiesen sind.

2. Wir formulieren zunächst die Voraussetzungen, die von uns gemacht werden.

---

1 Vortrag, gehalten am 20. 1. 1984 im Mathematischen Kolloquium anläßlich des 75. Geburtstags von Prof. em. Dr. Adam Schmidt

1° Es sei

$$L = \sum_{|\alpha| \leq 2m} a_\alpha(x)\, D^\alpha$$

($m > 0$ ganz, $\alpha = (\alpha_1, \ldots, \alpha_n)$, $\alpha_i \geq 0$ ganz, $|\alpha| = \alpha_1 + \ldots + \alpha_n$, $D^\alpha = D_1^{\alpha_1} \ldots D_n^{\alpha_n}$, $D_i = \partial/\partial x_i$, $x = (x_1, \ldots, x_n) \in R^n$) ein eigentlich elliptischer Differentialoperator (vgl. /9/) mit beliebig oft differenzierbaren Koeffizienten.

2° Der adjungierte Operator

$$L^* u = \sum_{|\alpha| \leq 2m} (-1)^{|\alpha|} D^\alpha (a_\alpha(x)\, u)$$

besitze die eindeutige Fortsetzungseigenschaft. Dies bedeutet folgendes: Ist u eine Lösung der Gleichung $L^* u = 0$ in der zusammenhängenden offenen Menge Ω und gilt $u \equiv 0$ in einer nichtleeren offenen Teilmenge $\Omega' \subset \Omega$, so folgt $u \equiv 0$ in Ω.

Diese Voraussetzung ist zum Beispiel erfüllt, falls L analytische Koeffizienten besitzt.

3° Auf dem Rand $\partial\Omega$ des beschränkten Gebietes Ω mit glattem Rand sei ein normales System von Randoperatoren $B_1, \ldots, B_m$ mit beliebig oft differenzierbaren Koeffizienten und $m_j = \operatorname{ord} B_j \leq 2m-1$ ($j = 1, \ldots, m$; $m_i \neq m_j$ für $i \neq j$) vorgegeben. Ferner wird angenommen, daß die sogenannte Wurzelbedingung erfüllt ist und daß der Index des damit definierten Randwertproblems gleich Null ist.

Zur Definition der benutzten Begriffe verweisen wir auf /9/.

4° Der Operator L besitze eine globale, lokal integrierbare Fundamentallösung $\Phi(x,y)$, d. h., es gelte

$$\varphi(x) = \int_{R^n} \Phi(x,y)\, L\varphi(y)\, dy$$

für beliebige $\varphi \in C_0^\infty(R^n)$.

Die Bedingung 4^o ist zum Beispiel für Operatoren mit konstanten Koeffizienten stets erfüllt.

$W^{m-1}(\Gamma)$ bezeichne den Vektorraum der Whitneyschen Taylorfelder der Ordnung m-1 auf Γ, d. h. den Raum aller Vektoren $g = (g_\alpha)_{|\alpha|\leq m-1}$ stetiger Funktionen $g_\alpha \in C(\Gamma)$, für die eine Funktion $\varphi \in C^{m-1}(R^n)$ existiert mit $D^\alpha \varphi|_\Gamma = g_\alpha$ $(|\alpha| \leq m-1)$. In $W^{m-1}(\Gamma)$ kann man die Norm

$$\|g\| = \sum_{|\alpha|\leq m-1} \|g_\alpha\|_{C(\Gamma)} \tag{2.1}$$

einführen. Die Menge Γ heißt (m-1)-regulär, falls $W^{m-1}(\Gamma)$ bezüglich der Norm (2.1) vollständig ist.

5^o Γ sei (m-1)-regulär.

6^o Γ möge das Gebiet Ω in die Gebiete Ω_i und Ω_a zerlegen, d. h. $\Omega_i \cap \Omega_a = \emptyset$, $\Omega = \Omega_i \cup \Omega_a \cup \Gamma$. Das Dirichlet-Problem zur Gleichung $Lu = 0$ sei in Ω_i eindeutig lösbar.

3. In Analogie zu Punkt 1 führen wir die folgenden Bezeichnungen ein. Dazu seien $V \subset \partial\Omega$ bzw. $G \subset \Omega_a$ beliebig fixierte offene Teilmengen des Randes $\partial\Omega$ bzw. der Komponente Ω_a.

$$L_V(\Omega) = \{u: Lu = 0 \text{ in } \Omega,\ B_j u|_{\partial\Omega \setminus V} = 0\}$$
$$L_V(\Gamma) = L_V(\Omega)|_\Gamma$$
$$L_G(\Omega) = \{u: Lu = g \in C_o^\infty(G),\ B_j u|_{\partial\Omega} = 0\}$$
$$L_G(\Gamma) = L_G(\Omega)|_\Gamma$$

<u>Theorem 1:</u> Unter den Annahmen 1^o bis 6^o gilt die Dichtheit $\overline{L_G(\Gamma)} = W^{m-1}(\Gamma)$.

<u>Bemerkungen zum Beweis:</u> Es werden 2 Fälle unterschieden.
(A): Das homogene Randwertproblem

$$\begin{aligned} &Lu = 0 \text{ in } \Omega, \\ &B_j u|_{\partial\Omega} = 0 \ (j=1,\ldots,m) \end{aligned} \tag{3.1}$$

besitzt nur die triviale Lösung $u \equiv 0$.
(B): Das Problem (3.1) besitzt nichttriviale Lösungen.

Zu (A): Man wählt ein Funktional $l \in (W^{m-1}(\Gamma))'$ mit

$$l(u) = 0 \text{ für alle } u \in L_G(\Gamma). \tag{3.2}$$

Daraus hat man $l = 0$ zu folgern, woraus unmittelbar die Behauptung folgt. Der Beweis verläuft in mehreren Schritten. Für eine ausführliche Darstellung verweisen wir auf /13/ bzw. /14/.

(a) Auf Grund der Annahme 5^0 läßt sich l in der Gestalt

$$l(u) = \sum_{|\alpha| \leq m-1} \int_\Gamma D^\alpha u(x)\, d\mu_\alpha(x) \tag{3.3}$$

durch ein System von Maßen μ_α mit supp $\mu_\alpha \subset \Gamma$ darstellen (vgl. /9/). Andererseits besitzen die Funktionen $u \in L_G(\Omega)$ nach bekannten Ergebnissen der Theorie der elliptischen Randwertprobleme (vgl. /12/) die Darstellung

$$D^\alpha u(x) = \int_G g(y)\, D_x^\alpha G(x,y)\, dy, \tag{3.4}$$

wobei $G(x,y)$ die zum Gebiet Ω gehörige Greensche Funktion bezeichnet. Setzt man (3.4) in (3.3) ein, benutzt (3.2), den Satz von Fubini und berücksichtigt, daß $g \in C_0^\infty(G)$ beliebig wählbar ist, so folgt

$$G^*\mu(y) := \sum_{|\alpha| \leq m-1} \int_\Gamma D_x^\alpha G(x,y)\, d\mu_\alpha(x) = 0 \text{ in } G.$$

Da das "Potential" $G^*\mu(y)$ in $\Omega \setminus \Gamma$ der Gleichung $L^*u = 0$ genügt, folgert man mit Hilfe der Voraussetzung 2^0 hieraus

$$G^*\mu(y) \equiv 0 \text{ in } \Omega_a.$$

(b) Durch den Einsatz potentialtheoretischer Hilfsmittel, insbesondere durch Abschätzung der Singularität der Greenschen Funktion und ihrer Ableitungen erhält man

$$D^\beta G^*\mu(y) = 0 \text{ quasi überall auf } \Gamma \ (|\beta| \leq m-1).$$

"Quasi überall" bedeutet hierbei "mit Ausnahme einer Menge der Wiener-Kapazität Null".

(c) Anwendung von Resultaten aus der Potentialtheorie elliptischer Gleichungen höherer Ordnung (Theorie der verallgemeinerten harmonischen Maße, Balayage-Theorie), die in /11/ und /9/ entwickelt worden sind, liefert

$$G^*\mu(y) \equiv 0 \quad \text{in } \Omega_i.$$

Insbesondere verwendet man die Tatsache, daß die Lösung des Dirichlet-Problems

$$L^*w = 0 \quad \text{in } \Omega_i,$$

$$D^\beta w|_\Gamma = g_\beta \quad (|\beta| \leq m-1),$$

wobei $g = (g_\beta)_{|\beta| \leq m-1} \in W^{m-1}(\Gamma)$ vorgegeben ist, für $z \in \Omega_i$ in der Gestalt

$$w(z) = \sum_{|\beta| \leq m-1} \int_\Gamma g_\beta(y)\, d\tau_z^\beta(y)$$

durch verallgemeinerte harmonische Vektormaße $\tau_z = (\tau_z^\beta)_{|\beta| \leq m-1}$ ($\operatorname{supp} \tau_z^\beta \subset \Gamma$) gegeben ist.

(d) Aus $G^*\mu(y) = 0$ quasi überall in Ω folgert man schließlich unter Benutzung der Definition der Fundamentallösung $l = 0$.

Zu (B): Das homogene Problem (3.1) möge k linear unabhängige Lösungen $u_1,\ldots,u_k$ besitzen. Da der Index des Problems Null ist, besitzt dann auch die Randwertaufgabe

$$L^*v = 0 \quad \text{in } \Omega,$$

$$B_j^* v|_{\partial\Omega} = 0 \quad (j=1,\ldots,m)$$

k linear unabhängige Lösungen $v_1,\ldots,v_k$. Die adjungierten Randoperatoren $B_1^*,\ldots,B_m^*$ sind dabei so zu wählen, daß die Greensche Formel

$$\int_\Omega (Lu)v\,dx - \int_\Omega uL^*v\,dx = \sum_{j=1}^m \int_{\partial\Omega} C_j u B_j^* v\,d\sigma - \sum_{j=1}^m \int_{\partial\Omega} B_j u C_j^* v\,d\sigma$$

gilt (vgl. /12/). In diesem Falle existiert eine verallgemeinerte Greensche Funktion G(x,y) mit folgenden Eigenschaften (vgl. /4/):

$$L^*_{(y)}\tilde{G}(x,y) = \sum_{i=1}^{k} u_i(x)\, u_i(y) \quad (x,y \in \Omega, x \neq y),$$

$$B_j\tilde{G}(x,y)\big|_{y \in \partial\Omega} = 0 \quad (j=1,\dots,m),$$

$$\int \tilde{G}(x,y)\, v_i(y)\,dy = 0 \quad (i=1,\dots,k).$$

Die Überlegungen aus (A) lassen sich dann in modifizierter Form durchführen, wobei $G^*\mu(y)$ durch

$$\tilde{G}^*\mu(y) := \sum_{|\alpha| \leq m-1} \int_\Gamma D_x^\alpha \tilde{G}(x,y)\, d\mu_\alpha(y)$$

zu ersetzen ist. Im Schritt (a) ergibt sich

$$\tilde{G}^*\mu(y) = v(y) \text{ in } G.$$

v ist eine Linearkombination von $v_1,\dots,v_k$.

Mit $w(y) := \tilde{G}^*\mu(y) - v(y)$ gilt $L^*w = 0$ in $\Omega\setminus\Gamma$, und wegen 2^0 schließt man daraus $w \equiv 0$ in Ω_e. Durch analoge Schlüsse wie in (A), technisch aber wesentlich aufwendiger, ergibt sich $w = 0$ quasi überall in Ω und daraus $\overline{L_G(\Gamma)} = W^{m-1}(\Gamma)$.

4. Theorem 2: Es seien die Voraussetzungen 1^0 bis 6^0 erfüllt. Dann gilt $\overline{L_V(\Gamma)} = W^{m-1}(\Gamma)$.

Beweis: Das Gebiet Ω werde so zu einem Gebiet $\Omega_1 \supset \Omega$ vergrößert, daß $\partial\Omega \setminus V \subset \partial\Omega_1$ gilt. Wegen der Glattheitsvoraussetzungen an die Koeffizienten der Randoperatoren kann dies offensichtlich in der Weise geschehen, daß auch für die geeignet fortgesetzten Randoperatoren die Bedingung (3) gilt. Wir wählen nun eine offene Menge $G \subset \Omega_1 \setminus \Omega$ und betrachten den Raum $L_G(\Omega_1)$. Nach Theorem 1 gilt $\overline{L_G(\Omega_1)}\big|_\Gamma = W^{m-1}(\Gamma)$, und wegen

$L_G(\Omega_1)\big|_\Omega \subset L_V(\Omega)$ erhalten wir daher $\overline{L_V(\Gamma)} = W^{m-1}(\Gamma)$.

5. Im folgenden wollen wir andeuten, daß sich auch höhere Ableitungen approximieren lassen, wenn man gleichmäßige Approximation durch Approximation im Mittel ersetzt. Wir nehmen aber jetzt an, daß Γ das Gebiet Ω nicht zerlegt.

7^{o} Es sei $\Gamma \in C^{\infty}$ (d. h. eine C^{∞}-Mannigfaltigkeit) und $\Omega \setminus \Gamma$ zusammenhängend.

Der Sobolev-Raum $W_2^{2m-1}(\Gamma)$ sei definiert als Vervollständigung der über Γ beliebig oft differenzierbaren Funktionen bezüglich der Norm

$$\|u\|_{2m-1,2} = \left\{ \sum_{|\beta| \leq 2m-1} \int_{\Gamma} |D^{\beta} u(x)|^2 \, d\sigma_{\Gamma}(x) \right\}^{1/2}.$$

Hierbei bezeichnet D^{β} Differentiation bezüglich lokaler Koordinaten in der Fläche und $d\sigma_{\Gamma}$ das Flächenelement auf Γ. Wie in Punkt 3 wählen wir eine offene Menge $G \subset \Omega \setminus \Gamma$ bzw. $V \subset \partial\Omega$. Das folgende Resultat ist in /15/ bewiesen.

Theorem 3: Unter den Annahmen 1^{o} bis 3^{o} und 7^{o} gilt

$$\overline{L_G(\Gamma)} = \overline{L_V(\Gamma)} = W_2^{2m-1}(\Gamma).$$

Dieses Resultat kann wesentlich vertieft werden, wie wir im folgenden andeuten wollen. Es handelt sich um bisher unveröffentlichte Resultate, die teilweise in Zusammenarbeit mit U. Hamann erzielt wurden (/8/).
Bekanntlich können die Sobolev-Räume $W_p^s(\Omega)$ bzw. $W_p^s(\Gamma)$ für beliebiges reelles s und $1 < p < \infty$ definiert werden. Man spricht in diesem Falle von den Sobolev-Slobodeckij-Räumen. Die Motivierung dazu ergibt sich unter anderem aus Spurbetrachtungen. Für $u \in W_p^s(\Omega)$ ($s > 0$ ganz) gilt beispielsweise $u|_{\Gamma} \in W_p^{s-1/p}(\Gamma)$. Betrachten wir etwa eine klassische Lösung $u \in C^{2m}(\Omega)$ der Differentialgleichung $Lu = 0$, so gilt für eine Umgebung U von Γ offenbar $u \in W_p^{2m}(U)$ und daher $u|_{\Gamma} \in W_p^{2m-1/p}(\Gamma)$. Folglich kann man vermuten, daß die Einschränkungsräume $L_G(\Gamma)$ und $L_V(\Gamma)$ sogar in $W_p^{2m-1/p}(\Gamma)$ dicht liegen.

6. Um die soeben formulierte Vermutung zu beweisen, ist nicht die allgemeine Theorie der Sobolev-Slobodeckij-Räume erforderlich. Es genügt, sich auf die hier benötigten Spurräume zu be-

schränken (vgl. /3/, /12/). Der Raum $W_p^{2m-1/p}(\Gamma)$ ist erklärt als Raum aller Einschränkungen von Funktionen aus $W_p^{2m}(\Omega)$ auf Γ, wobei die Einschränkung im Sinne der Einbettungssätze zu verstehen ist. Man definiert

$$\|\varphi\|_{2m-1/p,p} := \inf_{\substack{u \in W_p^{2m}(\Omega) \\ u|_\Gamma = \varphi}} \|u\|_{2m,p} .$$

$\|\cdot\|_{2m,p}$ bezeichnet die klassische Sobolev-Norm. Für $\psi \in C^\infty(\Gamma)$ definiert man

$$\|\psi\|_{-2m+1/p,q} := \sup_{\varphi \in W_p^{2m-1/p}(\Gamma)} \frac{|(\psi,\varphi)|}{\|\varphi\|_{2m-1/p,p}}$$

(mit dem gewöhnlichen L^2-Skalarprodukt (ψ,φ) über Γ sowie $1/p + 1/q = 1$) und $W_q^{-2m+1/p}(\Gamma)$ als Vervollständigung von $C^\infty(\Gamma)$ bezüglich dieser Norm. Durch stetige Fortsetzung läßt sich dann das "Skalarprodukt" (ψ,φ) für $\psi \in W_q^{-2m+1/p}(\Gamma)$, $\varphi \in W_p^{2m-1/p}(\Gamma)$ erklären. Es gilt die verallgemeinerte Schwarzsche Ungleichung

$$|(\psi,\varphi)| \leq \|\psi\|_{-2m+1/p,q} \; \|\varphi\|_{2m-1/p,p}.$$

Der Raum $W_q^{-2m+1/p}(\Gamma)$ erweist sich als dualer Raum zu $W_p^{2m-1/p}(\Gamma)$ und $(.,.)$ als zugehörige duale Paarung.

Theorem 4: Es gelte $1 < p < \infty$, und die Annahmen 1^0 bis 3^0 sowie 7^0 seien erfüllt. Dann folgt

$$\overline{L_G(\Gamma)} = \overline{L_V(\Gamma)} = W_p^{2m-1/p}(\Gamma).$$

Bemerkungen zum Beweis: Wir wählen ein Element

$$h \in W_q^{-2m+1/p}(\Gamma) = (W_p^{2m-1/p}(\Gamma))'$$

mit $(h,u) = 0$ für alle $u \in L_G(\Gamma)$ und dazu eine Folge $h_j \in C^\infty(\Gamma)$ mit $\|h_j - h\|_{-2m+1/p,q} \to 0$. Das bedeutet

$$(h,u) = \lim_{j\to\infty} (h_j,u) = \lim_{j\to\infty} \int_\Gamma h_j(x)u(x)d\sigma_\Gamma(x) = 0.$$

Daraus folgert man weiter

$$\lim_{j\to\infty} \int_{\Gamma} \tilde{G}(x,y)h_j(x)d\sigma_\Gamma(x) = v(y) \text{ in } G,$$

wobei die Funktionen $\tilde{G}$ und v die gleiche Bedeutung besitzen wie im Teil (B) des Beweises zu Theorem 1. Setzt man

$$w(y) := \lim_{j\to\infty} \int_{\Gamma} \tilde{G}(x,y)h_j(x)d\sigma_\Gamma(x) - v(y),$$

so ergibt sich $L^*w = 0$ in $\Omega\setminus\Gamma$ und wegen $w \equiv 0$ in G nach Anwendung von 2^o $w \equiv 0$ in $\Omega\setminus\Gamma$. Durch eine Reihe weiterer Abschätzungen, insbesondere durch Ausnutzung der bekannten Agmon-Douglis-Nirenberg-Ungleichungen erhalten wir schließlich $h \equiv 0$, damit $\overline{L_G(\Gamma)} = W_p^{2m-1/p}(\Gamma)$ und wie im Beweis zu Theorem 2

$\overline{L_V(\Gamma)} = W_p^{2m-1/p}(\Gamma)$.

Abschließend formulieren wir noch ein Resultat von U. Hamann.

Theorem 5: Es seien die Annahmen 1^o bis 3^o sowie 7^o erfüllt und es gelte $1 \leq \dim\Gamma \leq n-1$.

(a) Aus $1 < p \leq 2$ folgt $\overline{L_G(\Gamma)} = \overline{L_V(\Gamma)} = W_p^s(\Gamma)$ für $0 \leq s \leq 2m$.

(b) Aus $2 < p < \infty$ folgt $\overline{L_G(\Gamma)} = \overline{L_V(\Gamma)} = W_p^s(\Gamma)$ für $0 \leq s < 2m$.

Literatur

/1/ Anger, G.: Eindeutigkeitssätze und Approximationssätze für Potentiale II. Math. Nachr. 50, 229 - 244 (1971)

/2/ Beckert, H.: Eine bemerkenswerte Eigenschaft der Lösungen des Dirichletschen Problems bei linearen elliptischen Differentialgleichungen. Math. Ann. 139, 255 - 264 (1960)

/3/ Березанский, Ю.М.: Разложение по собственным функциям самосопряженных операторов. Киев 1965

/4/ Березанский, Ю.М., и Ройтберг, Я.А.: Теорема о гомеоморфизмах и функция Грина для общих эллиптических граничных задач. Укр. математ. ж. 19, 3 - 32 (1967)

/5/ Browder, F. E.: Functional analysis and partial differential equations II. Math. Ann. 145, 81 - 226 (1962)

/6/ Göpfert, A.: Über L_p-Approximationssätze - eine Eigenschaft der Lösungen elliptischer Differentialgleichungen. Math. Nachr. 31, 1 - 24 (1966)

/7/ Göpfert, A.: Eine Anwendung des Unitätssatzes von Itô-Yamabe. Beiträge Anal. 1, 29 - 41 (1971)

/8/ Hamann, U., und Wildenhain, G.: Approximation by solutions of elliptic equations. (in Vorbereitung)

/9/ Schulze, B. W., und Wildenhain, G.: Methoden der Potentialtheorie für elliptische Differentialgleichungen beliebiger Ordnung. Berlin 1977, Basel und Stuttgart 1977

/10/ Wanka, G.: Gleichmäßige Approximation von Randwertproblemen elliptischer Differentialgleichungen zweiter Ordnung. Beiträge Anal. 17, 19 - 29 (1981)

/11/ Wildenhain, G.: Potentialtheorie linearer elliptischer Differentialgleichungen beliebiger Ordnung. Berlin 1968

/12/ Wildenhain, G.: Darstellung von Lösungen linearer elliptischer Differentialgleichungen. Berlin 1981

/13/ Wildenhain, G.: Uniform approximation by solutions of general boundary value problems for elliptic equations of arbitrary order I. Erscheint in Zeitschrift für Analysis und ihre Anwendungen 2, 6 (1983)

/14/ Wildenhain, G.: Uniform approximation by solutions of general boundary value problems for elliptic equations of arbitrary order II. Math. Nachr. 113, 225 - 235 (1983)

/15/ Wildenhain, G.: Approxiamtion in Sobolev-Räumen durch Lösungen allgemeiner elliptischer Randwertprobleme bei Gleichungen beliebiger Ordnung. Rostock. Math. Kolloq. 22, 43 - 56 (1983)

eingereicht: 27. 02. 1984

Anschrift des Verfassers:

Prof. Dr. G. Wildenhain
Wilhelm-Pieck-Universität Rostock
Sektion Mathematik
Universitätsplatz 1
DDR-2500 Rostock

Rostock. Math. Kolloq. 26, 27 - 34 (1984)

35J40
35A35

Klaus Bayer

Approximation durch Lösungen elliptischer Randwertprobleme

In dieser Note betrachten wir, wie verschiedene Autoren (/1/-/3/, /5/-/8/) zuvor, die Approximation einer auf einer glatten Fläche Γ definierten Funktion durch die Lösungen gewisser Randwertaufgaben zu linearen elliptischen Differentialgleichungen. Im einfachsten Fall handelt es sich dabei um den folgenden Sachverhalt:

(A) Sei $\Omega \subset R^n$ ein beschränktes Gebiet mit glattem Rand $\partial\Omega$ und $\Omega^- \subset \Omega$ ein Teilgebiet, dessen Rand Γ zu $\partial\Omega$ durchschnittefremd ist. Bezeichne U eine weitere, in $\Omega^+ = \Omega \setminus \overline{\Omega^-}$ vorgegebene offene Menge. Wir betrachten die Lösungen der Dirichletschen Randwertprobleme

$$\Delta u = \varphi \text{ in } \Omega, \; u = 0 \text{ längs } \partial\Omega, \tag{1}$$

bei denen die rechten Seiten φ den Raum $C_0^\infty(U)$ durchlaufen. Es gilt nun, daß die Gesamtheit dieser Funktionen, wenn man sie über der Fläche Γ betrachtet, eine in jedem der Randräume $H^{m-1/2}(\Gamma)$ dichte Menge bilden[1]. Es läßt sich also durch Änderung der Quellen über U allein erreichen, daß eine längs Γ vorgeschriebene Funktion der Klasse $H^{m-1/2}$ durch die zugehörigen Lösungen der Dirichletprobleme beliebig genau in der Metrik dieses Raums approximiert wird.

Eine einfache Schlußweise (vgl. /6/) erlaubt es, von (A) auf die weitere Eigenschaft (B) zu schließen:

(B) Ist V eine gegebene offene Teilmenge des Randes $\partial\Omega$, so bilden auch die auf Γ eingeschränkten Lösungen der Dirichletprobleme

$$\Delta v = 0 \text{ in } \Omega, \quad v = 0 \text{ längs } \partial\Omega \setminus V$$

eine in $H^{m-1/2}(\Gamma)$ dichte Funktionenmenge.

---

1 Zu diesen Räumen s. /4/. Im Text sei $m \geq 1$ ganzzahlig.

Für die Metrik des Raums $H^1(\Gamma)$ wurde (B) erstmals in /1/ bewiesen. Die späteren Verallgemeinerungen dieser ursprünglichen Variante beziehen sich in /2/ auf die Gleichungen der klassischen Elastizitätstheorie bzw. in /6/ auf die Lösungen allgemeiner Randwertaufgaben für elliptische Differentialgleichungen 2m-ter Ordnung, wobei die Approximation auf die Normen der Randräume $H^{2m-1}(\Gamma)$ beschränkt blieb. Wie schon in /1/ für Gleichungen zweiter Ordnung bemerkt, lassen sich die benutzten Methoden aber nicht auf einen Beweis für die Möglichkeit der Approximation höherer als (2m-1)-ter Ableitungen ausdehnen. Es ist ein Vorzug des hier eingeschlagenen rein funktionalanalytischen Beweisweges, diese Schranke zu umgehen. Zur Vermeidung technischer Details führen wir die Überlegungen nur für das obige Dirichletproblem zum Laplaceoperator aus.
Um die später etwas verwischte einfache Beweisidee hervorzuheben, skizzieren wir den Beweis von (A) zunächst für die Approximation über $H^{1/2}(\Gamma)$.

Wir bezeichnen mit $H_0^1(\Omega)$ den Sobolewraum der über Ω meßbaren Funktionen mit quadratisch summierbaren ersten Ableitungen und Randwerten Null, mit $\Delta \in L(H_0^1, H^{-1})$ den Laplaceoperator unter homogenen Dirichletbedingungen, aufgefaßt als Abbildung von H_0^1 in sein Dual H^{-1}. Es gilt $G = \Delta^{-1} \in L(H^{-1}, H_0^1)$: G ist der Greensche Operator. Schließlich bezeichne $\gamma : u \to u|_\Gamma$ den auf H_0^1 definierten Spuroperator. Bekanntermaßen bildet γ den Raum H_0^1 stetig auf den Randraum $H^{1/2}(\Gamma)$ ab. Wäre nun die Menge $\gamma G(C_0^\infty(U))$ nicht dicht in $H^{1/2}(\Gamma)$, so existierte wenigstens ein zu seinem Dualraum $H^{-1/2}(\Gamma)$ gehöriges nichtverschwindendes Funktional T mit der Eigenschaft

$$\langle T, \gamma G \varphi \rangle = 0 \text{ für alle } \varphi \in C_0^\infty(U).$$

(Die eckigen Klammern werden im Text zur Kennzeichnung des Wertes eines linearen Funktionals auch für andere Raumpaarungen benutzt.) Für die adjungierten Abbildungen $G' \in L(H^{-1}, H_0^1)$ und $\gamma' \in L(H^{-1/2}(\Gamma), H^{-1}(\Omega))$ bedeutet das $\langle \varphi, G'\gamma' T \rangle = 0$, woraus

$G'\gamma'T = o$ in U (2)

folgt. Offenbar gilt $G' = G$. Die durch $w = G'\gamma'T$ definierte Funktion genügt daher über Ω der Differentialgleichung $\Delta w = \gamma'T$. In üblicher, nicht ganz korrekter Schreibweise heißt das

$$\int_\Omega \nabla w \nabla \varphi dV = -\int_\Gamma T\varphi d\Gamma;\ w \in H_o^1,\ \forall\, \varphi \in H_o^1$$

($dV = dx_1 \ldots dx_n$). Partielle Integration führt zu

$$\begin{aligned} &\Delta w^+ = o \text{ in } \Omega^+,\ \Delta w^- = o \text{ in } \Omega^-, \\ &w^+ = o \text{ längs } \partial\Omega \ ^2, \\ &w^+ - w^- = o \ ^2,\ \frac{\partial w^+}{\partial n} - \frac{\partial w^-}{\partial n} = -T \text{ längs } \Gamma, \end{aligned} \tag{3}$$

wobei $w^\pm$ die Einschränkungen der Funktion w auf die Gebiete $\Omega^\pm$ sowie n den nach Ω^- gerichteten Normalenvektor an die Fläche Γ bezeichnen. Nach (2) verschwindet die harmonische Funktion w^+ in U und deshalb in ganz Ω^+. Folglich wird $w^+ = \frac{\partial w^+}{\partial n} = o$ und damit auch $w^- = o$ längs Γ. Die harmonische Funktion w^- gehört mithin zu $H_o^1(\Omega^-)$, also $w^- = o$ in Ω^-. Die zweite Koppelungsbedingung in (3) liefert nunmehr $T = \frac{\partial w^-}{\partial n} = o$. Daher ist $\gamma G(C_o^\infty(U))$ dicht in $H^{1/2}(\Gamma)$. Behauptung (A) ist für $n = o$ bewiesen. Daß unsere Schlußweise tatsächlich als streng anzusehen ist, beruht auf der wohlbekannten Möglichkeit, die Normalenableitungen der Variationslösung (3) als "Lagrangesche Multiplikatoren" zu deuten. Wir werden das später noch auszuführen haben.

Wir wenden uns jetzt dem Beweis von (A) für $n \geq 2$ zu und notieren zunächst den (für glatte Ränder gültigen) fundamentalen Isomorphismus

$$\Delta \in L(H^n \cap H_o^1,\ H^{n-2}),\quad G := \Delta^{-1} \in L(H^{n-2},\ H^n \cap H_o^1)$$

(s. /4/).

---

2 wegen $w \in H_o^1(\Omega)$

Unter den neuen Voraussetzungen bezeichne $\gamma_0 \in L(H^m \cap H_0^1,\ H^{m-1/2}(\Gamma))$ die Spurabbildung $u \to u|_\Gamma$ und $\gamma_0' \in L(H^{-m+1/2}(\Gamma),\ (H^m \cap H_0^1)')$ ihre Adjungierte. Wird der Hilbertraum H^{m-2} mit seinem Dual identifiziert, so läßt sich die Adjungierte G^* des Greenschen Operators von vornherein als Abbildung zwischen den Räumen $(H^m \cap H_0^1)'$ und H^{m-2} auffassen: $G^* \in L((H^m \cap H_0^1)',\ H^{m-2})$. Bei diesem Standpunkt gilt für $T \in H^{-m+1/2}(\Gamma)$

$$\langle T,\ \gamma_0 G\varphi\rangle = (G^*\gamma_0' T,\ \varphi) \text{ für alle } \varphi \in H^{m-2},$$

wobei $(\ldots)$ das Skalarprodukt in H^{m-2} bedeutet. Wir definieren schließlich $A \in L(H^{m-2}, H^{-m+2})$ durch die Vorschrift

$$(v,\varphi) = \langle Av,\varphi\rangle \text{ für alle } v \in H^{m-2},\ \varphi \in H_0^{m-2}.$$

Wir schließen wie früher. Sei $T \in H^{-m+1/2}(\Gamma)$ ein zum orthogonalen Komplement von $\gamma_0 G(C_0^\infty(U))$ gehöriges Funktional:

$$\langle T,\ \gamma_0 G\varphi\rangle = (G^*\gamma_0' T,\ \varphi) = 0 \text{ für alle } \varphi \in C_0^\infty(U). \tag{4}$$

Setzen wir $w = G^*\gamma_0' T$, so gilt für jedes $u \in H^m \cap H_0^1$ die Beziehung

$$(w, \Delta u) = \langle T,\ \gamma_0 G\Delta u\rangle = \langle T,\ \gamma_0 u\rangle. \tag{5}$$

Insbesondere folgt hieraus $(w, \Delta\varphi) = 0$ für alle $\varphi \in C_0^\infty(\Omega^+)$ bzw. $\varphi \in C_0^\infty(\Omega^-)$, d. h. $\Delta Aw = 0$ in Ω^+ bzw. in Ω^-. Wir erschließen weiter: $Aw = 0$ in Ω^+, da dies wegen (4) über U zutrifft. Dieser Sachverhalt läßt sich auch durch die Gleichungen

$$(w^+, \varphi)^+ = 0 \quad \text{für alle } \varphi \in H_0^{m-2}(\Omega^+) \tag{6}$$

ausdrücken, wenn unter $(\ldots)^+$ der über Ω^+ geführte Anteil des Skalarprodukts $(\ldots)$ verstanden wird. $(\ldots)^+$ kann als Skalarprodukt von $H^{m-2}(\Omega^+)$ angesehen werden. $(\ldots)^-$ ist in ähnlichem Sinne zu deuten. Nach (5) gilt

$$(w^-, \Delta u)^- = 0 \quad \text{für alle } u \in H_0^m(\Omega^-). \tag{7}$$

Wir legen jetzt durch $\gamma_{2j} = \Delta^j$, $\gamma_{2j+1} = \frac{\partial}{\partial n}\,\Delta^j$ ein System von Randoperatoren über den Rändern $\partial\Omega$ und Γ fest. Ordnet man einer Funktion $\varphi \in H^{m-2}(\Omega^+)$ ihre "Spuren" $\gamma_0\varphi$ bis $\gamma_{m-3}\varphi$ längs $\partial\Omega$ und Γ zu, so bildet der dadurch bestimmte Operator Γ_{m-2} den Raum $H^{m-2}(\Omega^+)$ linear und stetig in das Produkt

$$X = \prod_{j=0}^{m-3} H^{m-5/2-j}(\partial\Omega) \times \prod_{j=0}^{m-3} H^{m-5/2-j}(\Gamma)$$

ab (s. /4/). Wie man unschwer erkennt, gilt $N(\Gamma_{m-2}) = H_0^{m-2}(\Omega^+)$, $R(\Gamma_{m-2}) = X$ für den Nullraum N bzw. Wertevorrat R von Γ_{m-2}. Nach (6) gehört nun das durch die Vorschrift $\varphi \to (w^+, \varphi)^+$ über $H^{m-2}(\Omega^+)$ definierte Funktional f zu $H_0^{m-2}(\Omega^+)^\perp = R(\Gamma'_{m-2})$. Infolgedessen existiert ein zu

$$X' = \prod_{j=0}^{m-3} H^{-m+5/2+j}(\partial\Omega) \times \prod_{j=0}^{m-3} H^{-m+5/2+j}(\Gamma)$$

gehörendes lineares Funktional $\lambda = (\lambda_0, \ldots, \lambda_{m-3}; \lambda_0^+, \ldots, \lambda_{m-3}^+)$ mit $\Gamma'_{m-2}\,\lambda = f$. Wegen $N(\Gamma'_{m-2}) = R(\Gamma_{m-2})^\perp = \{0\}$ ist λ darüberhinaus eindeutig bestimmt. Wir sind somit von (6) zu der für alle $\varphi \in H^{m-2}(\Omega^+)$ gültigen Beziehung

$$(w^+, \varphi)^+ = \sum_{j=0}^{m-3} \langle \lambda_j, \gamma_j\varphi\rangle_{\partial\Omega} + \sum_{j=0}^{m-3} \langle \lambda_j^+, \gamma_j\varphi\rangle_\Gamma$$

gelangt. Es folgt

$$(w^+, \Delta u)^+ = \sum_{j=2}^{m-1} \langle \lambda_{j-2}, \gamma_j u\rangle_{\partial\Omega} + \sum_{j=2}^{m-1} \langle \lambda_{j-2}^+, \gamma_j u\rangle_\Gamma \qquad (8)$$

für alle $u \in H^m(\Omega^+)$; man beachte $\gamma_j \circ \Delta = \gamma_{j+2}$. In ähnlicher Weise erkennt man

$$(w^-, \Delta u)^- = \sum_{j=0}^{m-1} \langle \lambda_j^-, \gamma_j u\rangle_\Gamma \quad \text{für alle } u \in H^m(\Omega^-), \qquad (9)$$

$$(\lambda_0^-, \ldots, \lambda_{m-1}^-) \in \prod_{j=0}^{m-1} H^{-m+1/2+j}(\Gamma)$$

bei Wiederholung unserer Schlußweise; wobei diesmal von (7) auszugehen ist.

Fassen wir (5) bzw.

$$(w,\Delta u) = (w^+,\Delta u^+)^+ + (w^-,\Delta u^-)^- = \langle T, \gamma_o u\rangle_\Gamma$$

erneut ins Auge. Ersetzt man ihre linke Seite gemäß (8) und (9), so liefert anschließender Vergleich sich entsprechender Randterme

$$\lambda_o^- = T, \quad \lambda_1^- = o, \quad \lambda_{j-2}^+ + \lambda_j^- = o \quad \text{für } 2 \leq j \leq m-1. \tag{10}$$

Von $\lambda_1^- = o$ läßt sich sogleich auf $\lambda_o^- = o$ und damit, wie gewünscht, auf $T = o$ weiterschließen. Ist nämlich φ eine beliebige in $H_o^{m-2}(\Omega^-)$ liegende Funktion und bestimmt man $u \in H^m(\Omega^-) \cap H_o^1(\Omega^-)$ gemäß $\Delta u = \varphi$, dann gilt nach (9)

$$(w^-,\varphi)^- = (w^-,\Delta u)^- = \sum_{j=2}^{m-1} \langle \lambda_j^-, \gamma_j u\rangle_\Gamma = \sum_{j=2}^{m-1} \langle \lambda_j^-, \gamma_{j-2}\varphi\rangle_\Gamma = o.$$

Wie schon mehrfach bemerkt, impliziert das die Existenz Lagrangescher Multiplikatoren $\Lambda_j^- \in H^{m-5/2-j}(\Gamma)$ mit

$$(w^-,\varphi)^- = \sum_{j=o}^{m-3} \langle \Lambda_j^-, \gamma_j\varphi\rangle_\Gamma \quad \text{für alle } \varphi \in H^{m-2}(\Omega^-).$$

Anstelle von (9) gilt also tatsächlich

$$(w^-,\Delta u)^- = \sum_{j=2}^{m-1} \langle \Lambda_{j-2}^-, \gamma_j u\rangle_\Gamma \quad \text{für alle } u \in H^m(\Omega^-).$$

Wiederholung des obigen Vergleichs liefert nunmehr $T = o$. Unsere Behauptung ist bewiesen.

<u>Bemerkung:</u> Bisher haben wir die Randfunktionen durch die Funktionswerte $\gamma_o G\varphi$ approximiert. Es ist wegen

$$\int_\Gamma \frac{\partial}{\partial n} G\varphi \, d\Gamma = o \tag{11}$$

von vornherein klar, daß die Normalenableitungen $\frac{\partial}{\partial n} G\varphi$ unserer Lösungen längs Γ keine dichte Menge bilden können. Es gilt aber:

Die in (11) zum Ausdruck gelangende Bedingung ist auch hinreichend für die Approximierbarkeit. Der Abschluß der Menge $\gamma_1 G(C_0^\infty(U))$ in $H^{m-3/2}(\Gamma)$ besteht also aus genau denjenigen Funktionen dieses Raums, deren Mittelwert über Γ verschwindet. Zum Beweis sind die früheren Überlegungen wie folgt zu modifizieren: Zunächst ist es klar, daß für jedes $T \in H^{-m+3/2}(\Gamma)$ im orthogonalen Komplement von $\gamma_1 G(C_0^\infty(U))$ anstelle von (10) neu

$$\lambda_0^- = 0, \quad \lambda_1^- = T, \quad \lambda_{j-2}^+ + \lambda_j^- = 0 \quad \text{für} \quad 2 \leq j \leq m-1$$

zu bestehen hat. Zu jedem $\varphi \in H_0^{m-2}(\Omega^-)$, das der Nebenbedingung

$$\int_{\Omega^-} \varphi \, dV = 0 \tag{12}$$

genügt, bestimme man jetzt eine Lösung $u \in H^m(\Omega^-)$ des Neumannschen Problems $\Delta u = \varphi$ in Ω^-, $\frac{\partial u}{\partial n} = 0$ längs Γ. Wenn wir u in die Darstellungsformel (9) eintragen, dann erkennen wir wie früher: $(w^-, \varphi)^- = (w^-, \Delta u)^- = 0$, wobei allerdings die hinzugekommene Bedingung (12) zu berücksichtigen ist. Es folgt

$$(w^-, \varphi)^- = \lambda \int_{\Omega^-} \varphi \, dV + \sum_{j=0}^{m-3} \langle \Lambda_j^-, \gamma_j \varphi \rangle_\Gamma$$

($\lambda \in R$) für alle $\varphi \in H^{m-2}(\Omega^-)$, und weiter

$$(w^-, \Delta u)^- = -\lambda \int_\Gamma \frac{\partial u}{\partial n} \, d\Gamma + \sum_{j=2}^{m-1} \langle \Lambda_{j-2}^-, \gamma_j u \rangle_\Gamma$$

für alle $u \in H^m(\Omega^-)$. Erneuter Vergleich der Randterme liefert $T = -\lambda = \text{const}$. Q.e.d.

Literatur

/1/ Beckert, H.: Eine bemerkenswerte Eigenschaft der Lösungen des Dirichletschen Problems bei linearen elliptischen Differentialgleichungen. Math. Ann. 139, 255 - 264 (1960)

/2/ Göpfert, A.: Über L_2-Approximationssätze - eine Eigenschaft der Lösungen elliptischer Differentialgleichungen. Math. Nachr. 31, 1 - 24 (1966)

/3/ Göpfert, A.: Eine Anwendung des Unitätssatzes von Itô-Yamabe. Beiträge Anal. 1, 29 - 41 (1971)

/4/ Lions, J.-L., et Magenes, E.: Problèmes aux limites non homogènes et applications I. Paris 1968

/5/ Wanka, G.: Gleichmäßige Approximation durch Lösungen von Randwertproblemen elliptischer Differentialgleichungen zweiter Ordnung. Beiträge Anal. 17, 19 - 29 (1981)

/6/ Wildenhain, G.: Approximation in Sobolev-Räumen durch Lösungen allgemeiner elliptischer Randwertprobleme bei Gleichungen beliebiger Ordnung. Rostock. Math. Kolloq. 22, 43 - 56 (1983)

/7/ Wildenhain, G.: Uniform approximation by solutions of general boundary value problems for elliptic equations of arbitrary order I. Zeitschrift für Analysis und ihre Anwendungen (in print)

/8/ Wildenhain, G.: Uniform approximation by solutions of general boundary value problems for elliptic equations of arbitrary order II. Math. Nachr. 113, 225 - 235 (1983)

eingegangen: 15. 02. 1984

Anschrift des Verfassers:

Prof. Dr. K. Beyer
Wilhelm-Pieck-Universität Rostock
Sektion Mathematik
Universitätsplatz 1
DDR-2500 Rostock

Rostock. Math. Kolloq. 26, 35 - 44 (1984)

35J70
35R05

Ehrhard Herbst

Hebbarkeit von Singularitäten für lineare Differentialoperatoren mit gestörter Elliptizität

0. Einleitung

Im engen Zusammenhang mit der Frage nach der Fortsetzbarkeit von Lösungen elliptischer Differentialgleichungen bzw. mit der Frage nach der Lösbarkeit von Dirichlet-Problemen, bei denen Vorgaben auf Randteilen niedrigerer Dimension gestellt werden, steht der Begriff der hebbaren Singularität (vgl. /2/, /5/, /6/, /7/). Eine kompakte Menge K, die in der offenen Menge $\Omega \subset R^n$ enthalten ist, heißt hebbare Singularität für den Differentialoperator A und die Funktionenklasse H, wenn aus $Au = 0$ in $\Omega \setminus K$ für $u \in H$ die Gültigkeit von $Au = 0$ in Ω folgt. Die Größe der Menge K wird durch verschiedene "Maße" charakterisiert, die von dem Differentialoperator und der Funktionenklasse abhängen: Den Minkowski-Inhalt, das Hausdorff-Maß oder die Kapazität (vgl. /2/, /5/, /6/, /7/). In dem vorliegenden Artikel wird die Hebbarkeit von Singularitäten für lineare Differentialoperatoren mit gestörter Elliptizität untersucht. Die zugrunde liegenden Funktionenklassen sind gewichtete Sobolew-Räume. Zur Charakterisierung der Menge K wird ein entsprechender Kapazitätsbegriff benutzt. Die Beweise der Aussagen stellen Modifizierungen der klassischen Beweismethoden dar (vgl. /2/). Zur Illustration werden zwei Beispiele angeführt.

1. Bezeichnungen und Begriffe

Wir fixieren eine offene Kugel mit dem Mittelpunkt im Koordinatenursprung des Euklidischen Raumes R^n ($n \geq 1$) und dem Radius $r > 0$ und bezeichnen sie mit $\sum_r$. Im Raum R^n sei ein Vektorfeld von Gewichtsfunktionen $g = (g_0, g_1, \ldots, g_n)$ gegeben, wobei die g_j ($j = 0,1,\ldots,n$) meßbare Funktionen im R^n mit der Eigenschaft sind, daß $g_j(x) > 0$ fast überall gilt. M_0 bezeichne die Menge

aller Funktionen $u \in C_0^\infty(\sum)$ mit $\|u\|_H < \infty$, wobei

$$\|u\|_H^2 := \int \left\{ g_0(x)\, u^2(x) + \sum_{i=1}^{n} g_i(x) \left[\frac{\partial u}{\partial x_i}\right]^2 \right\} dx. \qquad (1)$$

Für die Vervollständigung von M_0 in der Norm (1) schreiben wir H.

Definition: Der Differentialoperator A auf $\sum$ mit den Eigenschaften:

(i) $A : M_0 \to \mathcal{D}'$;

(ii) für alle $u, v \in M_0$ gilt $\langle Au, v\rangle = \langle u, Av\rangle$ (Hier stellt $\langle Au, v\rangle$ die Anwendung der Distribution Au auf die Testfunktion v dar.);

(iii) A ist linear;

heißt regulärer Operator zweiter Ordnung auf $\sum$, wenn es eine Konstante $\varkappa$ mit $0 < \varkappa < \infty$ derart gibt, daß für alle $u \in M_0$ gilt: $\varkappa^{-1}\|u\|_H^2 \leq \langle Au, u\rangle \leq \varkappa\|u\|_H^2$.

Damit kann man H als einen Hilbert-Raum mit dem Skalarprodukt $(u;v) = \langle Au, v\rangle$ auffassen, welches eine zu $\|\cdot\|_H$ äquivalente Norm erzeugt. Für $\sum \subset R^2$ und $g(x,y) = (|x|^\alpha, |x|^\beta, |x|^\gamma)$ mit $\alpha, \beta, \gamma \in R^1$ ist der Operator

$$Au = -a_1 \frac{\partial}{\partial x}\left(|x|^\beta \frac{\partial u}{\partial x}\right) - a_2 \frac{\partial}{\partial y}\left(|x|^\gamma \frac{\partial u}{\partial y}\right) + a_0 |x|^\alpha u$$

ein regulärer Operator ($a_1, a_2, a_0 = \text{const.}$).

Definition:

(i) Das Element $u \in H$ heißt Lösung von $Au = 0$ in $\Omega \subset\subset \sum$, falls für alle $v \in M_0 \cap C_0^\infty(\Omega)$ gilt: $\langle Au, v\rangle = 0$.

(ii) Die kompakte Menge $K \subset \Omega \subset\subset \sum$ ist eine hebbare Singularität (kurz: K ist hebbar) für den regulären Operator A und den Raum H, falls für jedes $f \in H$ mit $Af = 0$ in $\Omega \setminus K$ gilt: $Af = 0$ in Ω.

Verschiedene Konstanten, die in Abschätzungen von Zeile zu Zeile unterschiedliche Werte annehmen können, werden wir mit $\varkappa$ bezeichnen.

2. Eine hinreichende Bedingung

Wir fragen nun, unter welchen Bedingungen an die kompakte Menge K diese eine hebbare Singularität für den regulären Operator A und den gewichteten Sobolew-Raum H darstellt. Unsere hinreichende Bedingung beruht auf dem folgenden Kapazitätsbegriff.

Definition: Es sei $K \subset \Sigma$ eine kompakte Menge. Dann heißt die Zahl

$$cap_1(K) = \inf \{ \langle Au,u \rangle : u \in M_0 \text{ und } u \equiv 1 \text{ in einer Umgebung von } K \}$$

Kapazität von K (bezüglich A).

Hier haben wir stillschweigend angenommen, daß das Infimum über eine leere Menge als $+\infty$ erklärt wird. Die Kapazität hat dann die folgenden Eigenschaften:

Lemma 1: Für alle kompakten Mengen $K, K_1 \subset \Sigma$ gilt:

(i) $0 \leq cap_1(K) \leq +\infty$.

(ii) $K \subset K_1$ impliziert $cap_1(K) \leq cap_1(K_1)$.

(iii) Es existiert eine von K unabhängige Konstante $0 < \varkappa < +\infty$ derart, daß $\varkappa^{-1} cap_1(K) \leq$
$\leq \inf \{ \|u\|_H^2 : u \in M_0 \text{ und } u \equiv 1 \text{ in einer Umgebung von } K \}$
$\leq \varkappa cap_1(K)$ gilt.

Aus Lemma 1(iii) folgt im besonderen, daß die Mengen der Kapazität Null nur vom Gewicht abhängen und somit für alle regulären Operatoren gleich sind. Für den Beweis unserer Hebbarkeitsaussage benötigen wir noch das folgende

Lemma 2: Die kompakte Menge $K \subset \Sigma$ habe endliche Kapazität. Es gilt genau dann $cap_1(K) = 0$, wenn die Funktionen aus M_0, die in einer Umgebung von K verschwinden, dicht in H liegen.

Beweis: Es seien die Funktionen aus $M_0 \cap C_0^\infty(\Sigma \setminus K)$ dicht in H. $u \in M_0$ sei eine Funktion aus der Konkurrenzmenge von $cap_1(K)$,

die nach Voraussetzung nicht leer ist. Zu jedem natürlichen n gibt es nun eine Funktion $u_n \in M_o \cap C_o^\infty(\Sigma \setminus K)$ derart, daß $\|u - u_n\|_H \leq n^{-1}$ gilt. $u - u_n$ gehört für alle n zur Konkurrenzmenge von $cap_1(K)$, und damit gilt $cap_1(K) = 0$.

Wir setzen nun $cap_1(K) = 0$ voraus. Weiter sei $u \in H$ und (u_n) eine u approximierende Folge von Funktionen aus M_o. Es sei f eine Funktion aus der Konkurrenzmenge von $cap_1(K)$. $w_n = u_n - fu_n$ verschwindet dann in einer Umgebung von K und gehört zu M_o. Es gilt $\|u_n - w_n\|_H = \|fu_n\|_H \leq \alpha(u_n)\,\|f\|_H$. Wegen $cap_1(K) = 0$ kann nun die rechte Seite der Ungleichung beliebig klein gemacht werden. Ein Dreiecksargument liefert die Behauptung.

Wir kommen nun zur ersten Hebbarkeitsaussage.

Satz 1: Die kompakte Menge $K \subset \Omega \subset\subset \Sigma$ ist für alle regulären Operatoren A und den Raum H hebbar, wenn $cap_1(K) = 0$ gilt.

Beweis: Es sei $u \in H$ mit $\langle Au, w\rangle = 0$ für alle $w \in M_o \cap C_o^\infty(\Omega \setminus K)$ sowie $v \in M_o \cap C_o^\infty(\Omega)$. Nach Lemma 2 gibt es eine Folge $(v_n) \subset M_o \cap C_o^\infty(\Omega \setminus K)$ mit $\|v - v_n\|_H \to 0$ für $n \to \infty$. Wir haben dann $|\langle Au, v\rangle| = |\langle Au, v_n\rangle + \langle Au, v - v_n\rangle| = |\langle Au, v - v_n\rangle| \leq \alpha(u)\,\|v - v_n\|_H$, wobei die rechte Seite für $n \to \infty$ gegen Null strebt. Also gilt $\langle Au, v\rangle = 0$, und die Menge K ist hebbar.

3. Eine notwendige Bedingung

In der klassischen Theorie (vgl. /2/) werden zum Beweis einer notwendigen Bedingung wesentlich die Ergebnisse der Potentialtheorie benutzt. Um analoge Aussagen für unseren Fall zur Anwendung zu bringen, ist der bisher benutzte Kapazitätsbegriff ungeeignet. Wir führen einen neuen Kapazitätsbegriff ein, der es uns gestattet, den Apparat der Potentialtheorie und die damit verbundenen Extremaleigenschaften auszunutzen. Wir setzen im folgenden voraus, daß $M_o = C_o^\infty(\Sigma)$ gilt. Das ist z. B. der Fall, wenn $g_j \in L^1_{loc}(\Sigma)$ für $j = 0,1,\ldots,n$ gilt.

Definition: Es sei $K \subset \Sigma$ eine kompakte Menge. Dann heißt die Zahl

$$cap_2(K) = \inf\{\langle Au, v\rangle : u \in M_o \text{ und } u \geq 1 \text{ in einer Umgebung von } K\}$$

Kapazität von K.

Diese Kapazität hat dann die folgenden Eigenschaften.

Lemma 3: (vgl. /3/, /4/, /8/). Für alle kompakten Mengen $K, K_1 \subset \Sigma$ gilt:

(i) $0 \leq cap_2(K) < +\infty$.

(ii) $K \subset K_1$ impliziert $cap_2(K) \leq cap_2(K_1)$.

(iii) Es existiert eine von K unabhängige Konstante $0 < \alpha < +\infty$ derart, daß $\alpha^{-1} cap_2(K) \leq$
$\leq \inf\{\|u\|_H^2 : u \in M_o \text{ und } u \geq 1 \text{ in einer Umgebung von } K\}$
$\leq \alpha cap_2(K)$ gilt.

(iv) Es existiert eine von K unabhängige Konstante $0 < \alpha < +\infty$ derart, daß $\alpha^{-1} cap_1(K) \leq cap_2(K) \leq cap_1(K)$ gilt.

Das Lemma besagt also insbesondere, daß die beiden Kapazitätsbegriffe äquivalent sind. Daher verwenden wir den Begriff Kapazität für beide. Die neue Kapazität hat aber über Lemma 3 hinaus noch die folgenden Eigenschaften.

Lemma 4: (vgl. /3/, /4/, /8/). Es sei $K \subset \Sigma$ eine kompakte Menge. Es existiert genau ein Element $u_K \in H$ mit den Eigenschaften:

(i) $\langle Au_K, u_K\rangle = cap_2(K)$ und $u_K \geq 1$ auf K bis auf eine Menge der Kapazität Null.

(ii) Für alle $v \in M_o$ mit $v \geq 0$ in einer Umgebung von K gilt $\langle Au_K, v\rangle \geq 0$.

(iii) Es existiert ein positives Maß μ_K mit supp $\mu_K \subset K$ derart, daß $\langle Au_K, v\rangle = \int v\, d\mu_K$ für alle $v \in M_o$ sowie $cap_2(K) = \mu_K(K)$ gilt.

Daraus ergibt sich z. B., daß für das kapazitive Potential u_K in einem gewissen Sinne $Au_K = \mu_K$ gilt, also $Au_K = 0$ außerhalb

von K. Daraus erhalten wir nun unsere notwendige Bedingung für die Hebbarkeit.

Satz 2: Wenn $cap_2(K) > 0$ für die kompakte Menge $K \subset \Omega \subset\subset \Sigma$ gilt, dann ist K nicht hebbar für A und H.

Beweis: Für $u_K \in H$ gilt $\langle Au_K, v\rangle = 0$ für alle $v \in C_0^\infty(\Omega\setminus K)$. Aber für $v \in M_0$ mit $v \geq 1$ in einer Umgebung von K gilt

$$0 < cap_2(K) = \mu_K(K) \leq \int v d\mu_K = \langle Au_K, v\rangle.$$

Aus den Sätzen 1 und 2 ergibt sich die

Folgerung: Falls $M_0 = C_0^\infty(\Sigma)$ gilt, dann ist die kompakte Menge $K \subset \Omega \subset\subset \Sigma$ genau dann hebbar für alle regulären Operatoren A und den Raum H, wenn $cap_2(K) = cap_1(K) = 0$ gilt.

4. Beispiel

Wir werden für den Fall $n = 1$ die Kapazität des Nullpunktes für verschiedene Gewichte abschätzen. Dieses Beispiel stellt eine leichte Modifizierung des Beispiels aus /1/ dar. Wegen der Äquivalenz der Kapazitäten werden wir sie im weiteren nicht unterscheiden. Wir benötigen noch die folgende Eigenschaft der Kapazität.

Lemma 5: (vgl. /4/).

(i) Für jede monoton fallende Folge kompakter Mengen

$$K_n \subset \Sigma \text{ gilt } cap\left(\bigcap_{n\geq 1} K_n\right) = \inf_n cap(K_n). \tag{2}$$

(ii) Es existiert eine von K unabhängige Konstante $\varkappa$ derart, daß

$$\varkappa^{-1} cap(K) \leq \inf\left\{\langle Au, u\rangle : u \in M_0,\ 0 \leq u \leq 1,\ u \equiv 1 \text{ in einer Umgebung von } K\right\} \leq \varkappa cap(K) \text{ gilt.}$$

Damit läßt sich zeigen:

Lemma 6:

(i) Für das Gewicht $g_j(x) = |x|^{2\gamma}$ $(j = 0,1)$ mit $\gamma \geq 2^{-1}$ gilt $cap(\{0\}) = 0$.

(ii) Falls $g_1(x) = x^{2\gamma}$ für $x > 0$ oder $g_1(x) = |x|^{2\gamma}$ für $x < 0$ mit $|\gamma| < 2^{-1}$ gilt, so ist $\mathrm{cap}(\{0\}) > 0$.

Beweis: (i) Wir schätzen die Kapazität für die Menge $K = [c;d] \subset (-r;r)$ ab. Dazu wählen wir die Zahlen c', d' so, daß $-1 < c' < c < 0 < d < d' < 1$ und $[c';d'] \subset (-r;r)$ gilt. Wir haben dann für $L = \{u \in C_0^\infty((c';d')) : u \equiv 1 \text{ auf } K \text{ und } 0 \leq u \leq 1\}$

$$\mathrm{cap}(K) \leq \varkappa \inf\left\{\int u'^2(x)|x|^{2\gamma}\,dx + \int_{c'}^{d'} u^2(x)|x|^{2\gamma}\,dx : u \in L\right\}$$

$$\leq \varkappa \inf\left\{\int_{[c';c]\cup[d;d']} u'^2(x)|x|^{2\gamma}\,dx + \int_{c'}^{d'} |x|^{2\gamma}\,dx : u \in L\right\}$$

$$\leq \varkappa \inf\left\{\int_{[c';c]\cup[d;d']} u'^2(x)|x|\,dx : u \in L\right\} + \varkappa \int_{c'}^{d'} |x|^{2\gamma}\,dx.$$

Das Infimum wird durch die Lösung der Euler-Gleichung $(2u'x)' = 0$ angenommen, und zwar durch

$$u(x) = \begin{cases} (\ln cc'^{-1})^{-1} \ln xc^{-1} & \text{für } c' \leq x < c, \\ 1 & \text{für } c \leq x \leq d, \\ (\ln d'd^{-1})^{-1} \ln d'x^{-1} & \text{für } d < x \leq d'. \end{cases}$$

Nach Integration erhalten wir damit

$$\mathrm{cap}(K) \leq \varkappa\left\{(\ln cc'^{-1})^{-1} + (\ln d'd^{-1})^{-1} + \int_{c'}^{d'} |x|^{2\gamma}\,dx\right\}.$$

Wenn wir nun c, d gegen Null streben lassen, so können wir stets c', d' so wählen, daß gilt: $c' \to 0$, $d' \to 0$, $(\ln cc'^{-1})^{-1} \to 0$, $(\ln d'd^{-1})^{-1} \to 0$, $\int_{c'}^{d'} |x|^{2\gamma}\,dx \to 0$. Aus (2) folgt dann $\mathrm{cap}(\{0\}) = 0$.

(ii) Es sei $K = [0;f]$ mit $0 < f < r$. Dann gilt für $g_1(x) = x^{2\gamma}$ und $x > 0$:

$$\mathrm{cap}(K) \geq \varkappa \inf\left\{\int_f^r u'^2(x)\, x^{2\gamma}\,dx : u \in C_0^\infty(\Sigma),\ u \equiv 1 \text{ auf } K\right\}.$$

Das Infimum wird durch die Lösung der Euler-Gleichung $(2u' x^{2\gamma})' = 0$ mit den Randbedingungen $u(f) = 1$, $u(r) = 0$ angenommen. Das ist $u(x) = (r^{1-2\gamma} - x^{1-2\gamma})(r^{1-2\gamma} - f^{1-2\gamma})$. Daraus erhalten wir $cap(K) \geq \varkappa(r^{1-2\gamma} - f^{1-2\gamma})^{-1} \geq \varkappa r^{2\gamma-1} > 0$ unabhängig von f, also $cap(\{0\}) > 0$. Der Beweis des anderen Falles verläuft völlig analog.

Wir erhalten daraus die

Folgerung: Im Fall $n = 1$ ist der Nullpunkt eine hebbare Singularität, wenn das Gewicht die Gestalt $g_j(x) = |x|^{2\gamma}$ $(j = 0,1)$ mit $\gamma \geq 2^{-1}$ hat. Gilt $g_1(x) = x^{2\gamma}$ für $x > 0$ oder $g_1(x) = |x|^{2\gamma}$ für $x < 0$ mit $|\gamma| < 2^{-1}$, dann ist der Nullpunkt nicht hebbar.

5. Beispiel

Durch eine sich aus Punkt 4 ableitende Betrachtungsweise erhalten wir für den Fall $n = 2$ das nachfolgende Resultat (vgl. /1/). Die Übertragung auf den Fall $n = 3$ verläuft völlig analog.

Lemma 7: Es sei $\sum$ der Kreis $\{(x,y) : x^2 + y^2 < r^2\}$, $C(y)$ eine fast überall im R^1 nicht negative Funktion aus $L^1_{loc}(R^1)$ sowie $-r < \alpha < \beta < r$ ein Intervall auf der y-Achse.

(i) Wenn $g_j(x,y) = |x|^{2\gamma} C(y)$ mit $\gamma \geq 2^{-1}$ $(j = 0,1,2)$ ist, dann gilt $cap(\{0\} \times [\alpha;\beta]) = 0$.

(ii) Wenn $|\gamma| < 2^{-1}$ und entweder $g_1(x,y) = x^{2\gamma} C(y)$ für $x > 0$ oder $g_1(x,y) = |x|^{2\gamma} C(y)$ für $x < 0$ gilt, dann ist $cap(\{0\} \times [\alpha;\beta]) > 0$.

Beweis: (i) Wir schätzen $cap(K)$ für die kompakte Menge $K = [c;d] \times [\alpha;\beta]$ ab. Dazu wählen wir die Zahlen c,d,c',d',α',β' so, daß gilt: $-1 < c' < c < 0 < d < d' < 1$; $\alpha' < \alpha < \beta < \beta'$ und $[c';d'] \times [\alpha';\beta'] \subset \sum$. Wir haben dann für die Menge $L = \{u(x) \in C_0^\infty((c';d')) : 0 \leq u \leq 1,\ u \equiv 1 \text{ auf } [c;d]\}$ und die Funktion $v(y) \in C_0^\infty((\alpha';\beta'))$ mit $v \equiv 1$ auf $[\alpha;\beta]$ sowie $0 \leq v \leq 1$ die Ungleichung

$$\text{cap}(K) \leq \varkappa \inf\Big\{\Big(\int_{[c';c]\cup[d;d']} u'^2(x)|x|^{2\gamma}dx\Big)\Big(\int_{\alpha'}^{\beta'} v^2(y)C(y)\,dy\Big)$$

$$+ \Big(\int_{c'}^{d'} u^2(x)|x|^{2\gamma}\,dx\Big)\Big(\int_{\alpha'}^{\beta'} v'^2(y)C(y)dy\Big) : u \in L\Big\}$$

$$+ \varkappa\Big(\int_{c'}^{d'} |x|^{2\gamma}\,dx\Big)\Big(\int_{\alpha'}^{\beta'} C(y)dy\Big)$$

$$\leq \varkappa \inf\Big\{\Big(\int_{\alpha'}^{\beta'} v^2(y)C(y)dy\Big)\Big(\int_{[c';c]\cup[d;d']} u'^2(x)|x|dx\Big) : u \in L\Big\}$$

$$+ \varkappa\Big(\int_{c'}^{d'} |x|^{2\gamma}\,dx\Big)\Big(\int_{\alpha'}^{\beta'} v'^2(y)C(y)\,dy\Big)$$

$$+ \varkappa\Big(\int_{c'}^{d'} |x|^{2\gamma}\,dx\Big)\Big(\int_{\alpha'}^{\beta'} C(y)\,dy\Big).$$

Wie im Beweis von Lemma 6(i) können wir das Infimum berechnen. Durch geeignete Wahl von c', d' können wir erreichen, daß die rechte Seite der Ungleichung beliebig klein wird, wenn c, d gegen Null streben. Also erhalten wir $\text{cap}(\{0\} \times [\alpha;\beta]) = 0$.

(ii) Es sei $K = [0;f] \times [\alpha;\beta] \subset \sum$,

$D_K = \{u(x,y) \in C_0^\infty((-r;r)\times(-r;r)) : u \equiv 1$ in einer Umgebung von $K\}$.

Dann gilt für $g_1(x,y) = x^{2\gamma'}C(y)$ $(x>0)$: $\text{cap}(K) \geq$

$$\geq \varkappa \inf\Big\{\int_\alpha^\beta \Big(\int_f^r \Big|\frac{\partial u}{\partial x}(x,y)\Big|^2 x^{2\gamma}\,dx\Big)C(y)\,dy : u \in D_K\Big\}$$

$$\geq \varkappa\, r^{2\gamma-1}\int_\alpha^\beta C(y)\,dy > 0$$ auf Grund der Überlegungen im Beweis von Lemma 6(ii). Also gilt $\text{cap}(\{0\} \times [\alpha;\beta]) > 0$. Der andere Fall wird analog bewiesen.

Es ergibt sich damit die

<u>Folgerung:</u> Im Fall $n = 2$ ist die Strecke $[\alpha;\beta]$ auf der y-Achse eine hebbare Singularität, wenn das Gewicht die Gestalt

$g_j(x,y) = |x|^{2\gamma} C(y)$ $(j = 0,1,2)$ hat mit $\gamma \geq 2^{-1}$ und $C(y) > 0$ fast überall im R^1 sowie $C(y) \in L^1_{loc}(R^1)$. Gilt $g_1(x,y) = x^{2\gamma}C(y)$ für $x > 0$ oder $g_1(x,y) = |x|^{2\gamma} C(y)$ für $x < 0$ mit $|\gamma| < 2^{-1}$ und gleichem $C(y)$, dann ist die Strecke $\{0\} \times [\alpha;\beta]$ nicht hebbar.

Literatur

/1/ Albeverio, S., Fukushima, M., Karwowski, W., and Streit, L.: Capacity and quantum mechanical tunneling. Comm. Math. Phys. 81, 501 - 513 (1981)

/2/ Carleson, L.: Selected Problems on Exceptional Sets. Princeton 1967

/3/ Fabes, E., Jerison, D., and Kenig, C.: The Wiener-test for degenerate elliptic equations. Ann. Inst. Fourier (Grenoble) 32, 151 - 182 (1982)

/4/ Fukushima, M.: Dirichlet Forms and Markov Processes. Amsterdam 1980

/5/ Hamann, U.: Hebbarkeit von Singularitäten und ein Dirichlet-Problem. Dissertation (A), Wilhelm-Pieck-Universität Rostock 1979

/6/ Harvey, R., and Polking, J. C.: A notion of capacity which characterizes removable singularities. Trans. Amer. Math. Soc. 169, 183 - 195 (1972)

/7/ Schulze, B.-W., und Wildenhain, G.: Methoden der Potentialtheorie für elliptische Differentialgleichungen beliebiger Ordnung. Berlin 1977, Basel 1977

/8/ Herbst, E.: Spektralsynthese, Stabilität und Konvergenz in gewichteten Sobolew-Räumen. Dissertation (A), Wilhelm-Pieck-Universität Rostock 1982

eingegangen: 16. 04. 1984

Anschrift des Verfassers:
Dr. E. Herbst
Wilhelm-Pieck-Universität Rostock, Sektion Mathematik
Universitätsplatz 1
DDR-2500 Rostock

Rostock. Math. Kolloq. 26, 45 - 50 (1984)

39A10
15A15
34A30

Lothar Berg

Über die Greensche Funktion und die reduzierte Wronskische Determinante

In ihrer Arbeit /2/ hat Frau I. Jagnow bei Differenzengleichungen dritter Ordnung mit konstanten Koeffizienten für die um eine Zeile und Spalte reduzierte Wronskische Determinante gezeigt, daß sie einer Gleichung genügt, die mit der adjungierten Differenzengleichung eng zusammenhängt. Im zweiten Teil dieser Arbeit soll dieses Ergebnis auf entsprechende Gleichungen beliebiger Ordnung und im dritten Teil auf lineare Differentialgleichungen ausgedehnt werden. Zuvor wird über die Verallgemeinerung des Hauptergebnisses aus /2/ über die als Inverse einer tetradiagonalen Toeplitzschen Matrix darstellbaren Greenschen Funktion durch D. S. Meek /3/ berichtet.

1. Explizite Darstellung der Greenschen Funktion

Es sei $A_N = (a_{n-m})$ eine reguläre Toeplitzsche Bandmatrix N-ter Ordnung mit $a_q \neq 0$, $a_{-p} \neq 0$ und $a_k = 0$ für $k > q$ sowie für $k < -p$, d. h.

$$A_N = \begin{pmatrix} a_0 & a_{-1} & \cdots & a_{-p} & & & \\ a_1 & a_0 & \ddots & \ddots & \ddots & & \\ \vdots & \ddots & \ddots & \ddots & \ddots & a_{-p} & \\ a_q & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ & \ddots & \ddots & \ddots & \ddots & a_0 & a_{-1} \\ & & a_q & \cdots & a_1 & a_0 & \end{pmatrix}$$

Außerdem sei $0 \leq p,q < N$. Weiterhin sei $x_n = h_n$ die Lösung der Differenzengleichung

$$a_{-p}x_{n+p} + \ldots + a_0x_n + \ldots + a_qx_{-q} = 0 \qquad (1)$$

mit $n \geq 1$ unter den Anfangsbedingungen $h_p = 1/a_{-p}$ und $h_n = 0$ für $n < p$, die sich etwa mit Hilfe der Operatorenrechnung ermitteln läßt (vgl. /1/). Dann besitzt die zu A_N gehörende

zweiseitig unendliche Matrix A die Inverse $A^{-1} = (h_{n-m})$, während die Elemente der Inversen $A_N^{-1} = (d_{nm})$ die Darstellung

$$d_{nm} = \frac{1}{\Delta} \begin{vmatrix} h_{n-m} & h_{N+1-m} & h_{N+2-m} & \cdots & h_{N+p-m} \\ h_n & h_{N+1} & h_{N+2} & \cdots & h_{N+p} \\ h_{n+1} & h_{N+2} & h_{N+3} & \cdots & h_{N+p+1} \\ \vdots & \vdots & \vdots & & \vdots \\ h_{n+p-1} & h_{N+p} & h_{N+p+1} & \cdots & h_{N+2p-1} \end{vmatrix} \qquad (2)$$

mit

$$\Delta = \begin{vmatrix} h_{N+1} & \cdots & h_{N+p} \\ \vdots & & \vdots \\ h_{N+p} & \cdots & h_{N+2p-1} \end{vmatrix}$$

besitzen. Die Elemente d_{nm} bilden zugleich die Greensche Funktion der Differenzengleichung (1) unter den Randbedingungen

$$x_o = x_{-1} = \cdots = x_{1-q} = 0, \quad x_{N+1} = \cdots = x_{N+p} = 0.$$

Für die Existenz von A_N^{-1} ist das Nichtverschwinden der Nennerdeterminante Δ notwendig und hinreichend. Die Formel (2) stammt in leicht abgeänderter Bezeichnungsweise von D. S. Meek /3/ und wurde dort mit Hilfe des Satzes von Jacobi bewiesen. Für p = 1 geht sie in die Formel (16) von I. Jagnow /2/ über, wo sie unabhängig von /3/ für q = 2 bewiesen wurde. Für p = 0 lautet sie $d_{nm} = h_{n-m}$ und beinhaltet die bekannte Tatsache, daß die Elemente der Inversen einer Dreiecksmatrix von der Ordnung N unabhängig sind. Anschließend folgt für den allgemeinen Fall ein neuer elementarer

Beweis: Die Inverse A^{-1} ist nicht eindeutig bestimmt. Man erhält alle nur möglichen Inversen, wenn man ihre Elemente durch

$$d_{nm} = h_{n-m} + \sum_{i,j=1}^{p+q} x_{ni} c_{ij} y_{jm} \qquad (3)$$

ersetzt, wobei $x_{n1}, \ldots, x_{n,p+q}$ ein Fundamentalsystem von (1) und $y_{1m}, \ldots, y_{p+q,m}$ ein Fundamentalsystem der zugehörigen adjungierten Gleichung

$$a_q y_{m+q} + \cdots + a_o y_m + \cdots + a_{-p} y_{m-p} = 0 \qquad (4)$$

ist. Die c_{ij} sind willkürliche Konstanten. Die Elemente (3) gehen für $1 \leq n,m \leq N$ in die Elemente von A_N^{-1} über, wenn die c_{ij} so bestimmt werden, daß die Randbedingungen

$$d_{no} = d_{n,-1} = \dots = d_{n,1-p} = 0, \quad d_{n,N+1} = \dots = d_{n,N+q} = 0 \tag{5}$$

oder die dazu äquivalenten Randbedingungen

$$d_{om} = d_{-1,m} = \dots = d_{1-q,m} = 0, \quad d_{N+1,m} = \dots = d_{N+p,m} = 0 \tag{6}$$

erfüllt sind. Da die h_{n-m} für $n-m > -q$ als Funktionen von n Lösungen von (1) und als Funktionen von m Lösungen von (4) sind, spezialisieren wir (3) zu

$$d_{nm} = h_{n-m} + \sum_{i,j=1}^{p} h_{n+i-1} b_{ij} h_{N+j-m} \tag{7}$$

mit willkürlichen Konstanten b_{ij}. Wegen $h_n = 0$ für $n < p$ sind dann bereits die letzten q Bedingungen von (5) sowie die ersten q Bedingungen von (6) erfüllt. Wie man durch Entwicklung der Zählerdeterminante von (2) nach der ersten Spalte und anschließenden Entwicklungen nach den ersten Zeilen sieht, hat (2) mit $\Delta \neq 0$ die Gestalt (7). Aus (2) geht weiterhin unmittelbar hervor, daß auch die ersten p Bedingungen von (5) sowie die letzten p Bedingungen von (6) erfüllt sind.
Verschwindet die Nennerdeterminante Δ in (2), so gibt es eine Linearkombination

$$x_n = \sum_{i=0}^{p-1} \beta_i h_{n+i}$$

mit $x_{N+1} = \dots = x_{N+p} = 0$ und natürlich $x_n = 0$ für $n \leq 0$. Da x_n eine Lösung von (1) ist, ist dann der Vektor $(x_1,\dots,x_N)^T$ eine Nullösung von A_N. Bei vorausgesetzter Regularität von A_N kann somit dieser Fall nicht eintreten.

2. Die reduzierte Wronskische Determinante

Ohne Beschränkung der Allgemeinheit wählen wir p = 1 und

$a_{-1} = -1$, so daß die Gleichungen (1) und (4)

$$x_{n+1} = a_o x_n + \ldots + a_q x_{n-q}, \quad y_{n-1} = a_o y_n + \ldots + a_q y_{n+q} \tag{8}$$

lauten. Es sei $q \geq 1$. Aus q Lösungen $y_{1n}, \ldots, y_{qn}$ der zweiten Gleichung bilden wir die q-dimensionale Vektorlösung $y_n = (y_{1n}, \ldots, y_{qn})$ und betrachten die Determinanten

$$D_n = \begin{vmatrix} y_{n+1} \\ y_{n+2} \\ \vdots \\ y_{n+q} \end{vmatrix} = \begin{vmatrix} y_{1,n+1} & \cdots & y_{q,n+1} \\ y_{1,n+2} & \cdots & y_{q,n+2} \\ \vdots & & \vdots \\ y_{1,n+q} & \cdots & y_{q,n+q} \end{vmatrix}, \quad D_n^{(m)} = \begin{vmatrix} y_n \\ \vdots \\ y_{n+m-1} \\ y_{n+m+1} \\ \vdots \\ y_{n+q} \end{vmatrix} \tag{9}$$

mit $0 \leq m \leq q$. Da die Gleichungen (8) die Ordnung q+1 besitzen, entsteht die Determinante $D_n = D_n^{(o)}$ aus der zugehörigen Wronskischen Determinante (vgl. /1/) durch Streichung der ersten Zeile und der letzten Spalte. Für später merken wir uns noch die Beziehung $D_n^{(q)} = D_{n-1}$. Wie wir jetzt zeigen wollen, genügt die aus D_n gebildete Folge

$$x_n = (-1)^{nq} a_q^n D_n \tag{10}$$

der ersten Gleichung in (8). Für $q = 2$ wurde dies bereits in /2/ gezeigt, wo auch auf eine Anwendung dieser Aussage hingewiesen wird.

Beweis: Ersetzen wir n in der Determinante $D_n^{(m)}$ aus (9) durch n-1 und anschließend y_{n-1} durch (8), so folgt für $m \geq 1$

$$D_{n-1}^{(m)} = \begin{vmatrix} y_{n-1} \\ y_n \\ \vdots \\ y_{n+m-2} \\ y_{n+m} \\ \vdots \\ y_{n+q-1} \end{vmatrix} = a_{m-1} \begin{vmatrix} y_{n-m-1} \\ y_n \\ \vdots \\ y_{n+m-2} \\ y_{n+m} \\ \vdots \\ y_{n+q-1} \end{vmatrix} + a_q \begin{vmatrix} y_{n+q} \\ y_n \\ \vdots \\ y_{n+m-2} \\ y_{n+m} \\ \vdots \\ y_{n+q-1} \end{vmatrix}.$$

Durch Vertauschung entsprechender Zeilen erkennt man, daß die erste Determinante auf der rechten Seite gleich $(-1)^{m-1}D_{n-1}$ und die zweite gleich $(-1)^{q-1}D_n^{(m-1)}$ ist, d. h., es gilt

$$D_{n-1}^{(m)} = (-1)^{m-1}a_{m-1}D_{n-1} + (-1)^{q-1}a_q D_n^{(m-1)}.$$

Nach Ersetzung von n durch n-m+2 entsteht hieraus

$$D_{n-m+1}^{(m)} = (-1)^{m-1}a_{m-1}D_{n-m+1} + (-1)^{q-1}a_q D_{n-m+2}^{(m-1)}$$

und nach der Substitution $z_m^{(n)} = (-1)^{mq-m}a_q^{-m}D_{n-m+1}^{(m)}$

$$z_m^{(n)} = (-1)^{mq-1}a_q^{-m}a_{m-1}D_{n-m+1} + z_{m-1}^{(n)}.$$

Durch Summation über m von 1 bis q folgt nun wegen $z_o^{(n)} = D_{n+1}$ und

$$z_q^{(n)} = a_q^{-q}D_{n-q+1}^{(q)} = a_q^{-q}D_{n-q}$$

die Gleichung

$$D_{n+1} = a_q^{-q}D_{n-q} + \sum_{m=1}^{q}(-1)^{mq}a_q^{-m}a_{m-1}D_{n-m+1}$$

oder nach einer Indexverschiebung

$$D_{n+1} = \sum_{m=o}^{q}(-1)^{(m+1)q}a_q^{-m-1}a_m D_{n-m}.$$

Dies ist aber nach Multiplikation mit $((-1)^q a_q)^{n+1}$ für die Folge (10) gerade die Behauptung (8).

3. Übertragung auf Differentialgleichungen

Das vorhergehende Ergebnis läßt sich leicht auf Differentialgleichungen mit sogar variablen Koeffizienten

$$y^{(n)}(x) + p_1(x)y^{(n-1)}(x) + \ldots + p_n(x)y(x) = 0 \qquad (11)$$

übertragen. Aus n-1 Lösungen $y_1(x),\ldots,y_{n-1}(x)$ dieser Gleichung bilden wir den Vektor $y(x) = (y_1(y),\ldots,y_{n-1}(x))^T$ und die redu-

zierte Wronskische Determinante

$$D(x) = \left| y(x),\ y'(x), \ldots, y^{(n-2)}(x) \right| .$$

Für die Funktion $z(x) = D(x)/W(x)$, wobei $W(x)$ die Wronskische Determinante mit der bekannten Eigenschaft $W'(x) = -p_1(x)W(x)$ ist, finden wir durch m-malige Differentiation mit $0 < m < n$ die Gleichung

$$z^{(m)} - (p_1 z)^{(m-1)} + (p_2 z)^{(m-2)} + \ldots + (-1)^m p_m z =$$

$$\left| y,\ y',\ \ldots,\ y^{(n-m-2)},\ y^{(n-m)},\ \ldots,\ y^{(n-1)} \right| / W$$

und für $m = n-1$ nach einer weiteren Differentiation die zu (11) adjungierte Differentialgleichung

$$z^{(n)}(x) - (p_1(x)z(x))^{(n-1)} + \ldots + (-1)^n p_n(x)z(x) = 0.$$

Es sei noch erwähnt, daß für Differentialgleichungen mit konstanten Koeffizienten $W = W_0 e^{-p_1 x}$ gilt und somit $z(x) = e^{p_1 x} D(x)$ gewählt werden kann.

Literatur

/1/ Berg, L.: Lineare Gleichungssysteme mit Bandstruktur und ihr asymptotisches Verhalten. Berlin (im Druck)

/2/ Jagnow, I.: Zur Invertierung tetradiagonaler Toeplitzscher Matrizen. Rostock. Math. Kolloq. 25, 31 - 36 (1984)

/3/ Meek, D. S.: The inverses of Toeplitz band matrices. Linear Algebra and its Appl. 49, 117 - 129 (1983)

eingereicht: 12. 06. 1984

Anschrift des Verfassers:

Prof. Dr. L. Berg
Wilhelm-Pieck-Universität Rostock
Sektion Mathematik
Universitätsplatz 1
DDR-2500 Rostock

Rostock. Math. Kolloq. 26, 51 - 54 (1984)

46F10
16A38

Lothar Berg

Distribution Algebras with Periodic Matrix Representations

According to /1/, a distribution algebra is an associative differential algebra with an element h satisfying

$$h^2 = h \tag{1}$$

and $\delta \neq 0$ with $\delta = h'$. The construction of distribution algebras can be realized by means of matrix representations, using especially twosided infinite matrices $f = (f_{ik})$, $-\infty < i,k < +\infty$, with the derivative

$$f' = (f_{i,k-1} - f_{i+1,k}). \tag{2}$$

cf. /1/ as well as Pho That Shay /3/. A generalization of (1) was investigated by G. L. N. Rao /4/. In what follows we restrict ourselves to s-periodic matrices with

$$f_{i+s,k+s} = f_{ik}$$

for all i,k, because with such matrices it is possible to calculate in a lucid way. As a consequence of this restriction the element δ satisfies new differential equations. The special case s = 2 was already treated in /2/.

Algebraic preliminaries

Let be σ the matrix with the entries $\sigma_{i,i+1} = 1$ and $\sigma_{ik} = 0$ else. Then the derivative (2) is the inner derivative with respect to $-\sigma$

$$f' = f\sigma - \sigma f. \tag{3}$$

For an s-periodic diagonal matrix f with the diagonal entries $f_{ii} = f_i$ and a fixed integer $s \geq 2$ we introduce the notation

$$f = (f_1, f_2, \ldots, f_s). \tag{4}$$

Then we have the rules

$$\sigma(f_1,f_2,\ldots,f_{s-1},f_s) = (f_2,f_3,\ldots,f_s,f_1)\sigma, \tag{5}$$

$$(f_1,f_2,\ldots,f_{s-1},f_s)' = (f_1-f_2,f_2-f_3,\ldots,f_{s-1}-f_s,f_s-f_1)\sigma, \tag{6}$$

and together with another s-periodic diagonal matrix
$g = (g_1,\ldots,g_s)$ and a number λ

$$f + g = (f_1+g_1,\ldots,f_s+g_s),$$

$$fg = (f_1g_1,\ldots,f_sg_s),$$

$$\lambda f = (\lambda f_1,\ldots,\lambda f_s).$$

The zero matrix O and the unit matrix I read respectively

$$O = (0,0,\ldots,0,0), \quad I = (1,1,\ldots,1,1).$$

The commuting rule (5) implies for all s-periodic matrices f

$$\sigma^s f = f\sigma^s. \tag{7}$$

From (3) it follows $\sigma' = O$.

A special choice for h

A diagonal matrix satisfies the relation (1) if the diagonal entries attain only the values 0 and 1. Hence the special s-periodic matrix

$$h = (1,1,\ldots,1,0) \tag{8}$$

generates a distribution algebra. According to (6) the first derivatives of h read

$$\left.\begin{aligned} \delta &= (0,\ldots,0,1,-1)\sigma, \\ \delta' &= (0,\ldots,0,-1,2,-1)\sigma^2, \\ \delta'' &= (0,\ldots,0,1,-3,3,-1)\sigma^3, \end{aligned}\right\} \tag{9}$$

and the relations

$$h\delta^{(ns-1)} = \delta^{(ns-1)}h, \quad \delta^{(ns-1)}\delta^{(ms-1)} = \delta^{(ms-1)}\delta^{(ns-1)} \tag{10}$$

with arbitrary natural numbers n,m are consequences of (7). From (5) and (9) we obtain

$$\delta^2 = (0,\ldots,0,-1,0)\sigma^2,$$

so that actually we have the nontrivial case $\delta^2 \neq 0$, but $\delta^3 = 0$ for $s > 2$.

The case of odd s

In the following we assume s to be an odd integer (greater than 1). Then in continuation of (9) we find

$$\delta^{(s-2)} = (-1, s-1, -\binom{s-1}{2}, \ldots, s-1, -1)6^{s-1},$$

$$\delta^{(s-1)} = (-s, \binom{s}{2}, \ldots, -\binom{s}{2}, s, 0)6^{s},$$

and the first equation of (10) allows for n = 1 the improvement

$$\delta^{(s-1)} = h\delta^{(s-1)} = \delta^{(s-1)}h. \tag{11}$$

Hence the well known consequence of (1)

$$\delta^{(n)} = h\delta^{(n)} + \delta^{(n)}h + \sum_{v=1}^{n} \binom{n+1}{v}\delta^{(v-1)}\delta^{(n-v)}$$

simplifies for n = s-1 to the differential equation

$$\delta^{(s-1)} = -\sum_{v=1}^{s-1} \binom{s}{v}\delta^{(v-1)}\delta^{(s-v-1)}. \tag{12}$$

Integrating (12) we obtain

$$\delta^{(s-2)} = -\sum_{v=1}^{s-2}\sum_{\mu=1}^{v} (-1)^{v+\mu}\binom{s}{\mu}\delta^{(v-1)}\delta^{(s-v-2)} - 6^{s-1}, \tag{13}$$

since according to $6' = 0$ equation (13) implies (12), and the correctness of the constant of integration we see as follows. All terms in (13) possess the form

$$(f_1, f_2, \ldots, f_s)6^{s-1}.$$

Since $f_1 = 0$ for all products $\delta^{(v-1)}\delta^{(s-v-2)}$, $f_1 = -1$ for $\delta^{(s-2)}$, and $f_1 = 1$ for 6^{s-1}, the factor of 6^{s-1} must be -1.

The case s = 3

Finally, let us consider the special case s = 3 in more detail. From

$$h = (1,1,0), \qquad \delta''' = 3(-2,1,1)6^4,$$

$$\delta = (0,1,-1)6, \qquad \delta^{(4)} = 9(-1,0,1)6^5,$$

$$\delta' = (-1,2,-1)6^2, \qquad \delta^{(5)} = 9(-1,-1,2)6^6,$$

$$\delta'' = 3(-1,1,0)6^3, \qquad \delta^{(6)} = 27(0,-1,1)6^7$$

we find $\delta^{(6)} = -27\delta\delta^6$ and therefore

$$\delta^{(n+6)} = -27\delta^{(n)}\delta^6$$

for all integers $n \geq 0$. But we also have $\delta'' = -3\delta\dot{\delta}$ and consequently

$$\delta^{(2n)} = (-3)^n\delta^n\dot{\delta}^n, \quad \delta^{(2n+1)} = (-3)^n\delta^n\dot{\delta}\cdot\delta^n.$$

Equations (12) and (13) read now respectively

$$\delta'' + 3\delta\dot{\delta} + 3\dot{\delta}\delta = 0, \quad \dot{\delta} + 3\delta^2 + \delta^2 = 0.$$

From further properties we only mention that in spite of $\delta^3 = 0$ no powers of the elements $\dot{\delta}$, δ'' and δ''' vanish.

References

/1/ Berg, L.: Representations for distribution algebras. Z. Angew. Math. Mech. 56, 177 - 181 (1976)

/2/ Berg, L.: Distributionenalgebren mit der Eigenschaft $h\dot{\delta} = \dot{\delta}h$. Rostock. Math. Kolloq. 1, 15 - 20 (1976)

/3/ Pho Thet Shay: Distribution Algebra of Infinite Matrices and Formal Sums. M. Sc. Thesis, Arts and Science University of Rangoon 1982

/4/ Rao, G. L. N.: Certain distribution algebras by matrix representation. Rostock. Math. Kolloq. 25, 45 - 52 (1984)

received: 16. 02. 1984

Author's address:

Prof. Dr. L. Berg
Wilhelm-Pieck-Universität Rostock
Sektion Mathematik
Universitätsplatz 1
DDR-2500 Rostock

Rostock. Math. Kolloq. 26, 55 - 62 (1984)

03G20
08A40
20M99

Dietlinde Lau

Unterhalbgruppen von $(P_3^1, \varkappa)$

Für eine Reihe von Untersuchungen in der 3-wertigen Logik P_3(/3/), insbesondere bei der Ermittlung gewisser "Überschauberer" Unterverbände des Verbandes der abgeschlossenen Teilmengen von P_3, sind Kenntnisse über die von einstelligen Funktionen erzeugten abgeschlossenen Teilmengen, d. h. über die Unterhalbgruppen von $(P_3^1, \varkappa)$, nützlich. So lassen sich z. B. mit Hilfe der weiter unten angegebenen Halbgruppen ohne Schwierigkeiten sämtliche abgeschlossenen Mengen von quasilinearen Funktionen der 3-wertigen Logik bestimmen (/2/, /1/).
Die Funktionen aus P_3^1 sind in Tabelle 1 definiert. Es sei

$C = \{c_0, c_1, c_2\}$, $J = \{c_0, c_1, j_0, j_1, j_2, j_3, j_4, j_5\}$,

$U = \{c_0, c_2, u_0, u_1, \ldots, u_5\}$, $V = \{c_1, c_2, v_0, v_1, \ldots, v_5\}$ und

$S = \{s_1, \ldots, s_6\}$. Wie üblich bezeichne $f \varkappa g$ die durch

$(f \varkappa g)(x) = f(g(x))$ definierte Funktion. Einen Überblick über $f \varkappa g$ für die "wesentlichen" f und g gibt Tabelle 2.

x	$c_0(x)$	$c_1(x)$	$c_2(x)$	$j_0(x)$	$j_1(x)$	$j_2(x)$	$j_3(x)$	$j_4(x)$	$j_5(x)$	$u_0(x)$
0	0	1	2	1	0	0	1	1	0	2
1	0	1	2	0	1	0	1	0	1	0
2	0	1	2	0	0	1	0	1	1	0

x	$u_1(x)$	$u_2(x)$	$u_3(x)$	$u_4(x)$	$u_5(x)$	$v_0(x)$	$v_1(x)$	$v_2(x)$	$v_3(x)$	$v_4(x)$
0	0	0	2	2	0	2	1	1	2	2
1	2	0	2	0	2	1	2	1	2	1
2	0	2	0	2	2	1	1	2	1	2

x	$v_5(x)$	$s_1(x)$	$s_2(x)$	$s_3(x)$	$s_4(x)$	$s_5(x)$	$s_6(x)$
0	1	0	0	1	1	2	2
1	2	1	2	0	2	0	1
2	2	2	1	2	0	1	0

Tabelle 1

f \ g	j_i	u_i	v_i	s_2	s_3	s_4	s_5	s_6
j_0	j_{5-i}	j_{5-i}	c_0	j_0	j_1	j_2	j_1	j_2
j_1	j_i	c_0	j_{5-i}	j_2	j_0	j_0	j_2	j_1
j_2	c_0	j_i	j_i	j_1	j_2	j_1	j_0	j_0
j_3	c_1	j_{5-i}	j_{5-i}	j_4	j_3	j_4	j_5	j_5
j_4	j_{5-i}	c_1	j_i	j_3	j_5	j_5	j_3	j_4
j_5	j_i	j_i	c_1	j_5	j_4	j_3	j_4	j_3
u_0	u_{5-i}	u_{5-i}	c_0	u_0	u_1	u_2	u_1	u_2
u_1	u_i	c_0	u_{5-i}	u_2	u_0	u_0	u_2	u_1
u_2	c_0	u_i	u_i	u_1	u_2	u_1	u_0	u_0
u_3	c_2	u_{5-i}	u_{5-i}	u_4	u_3	u_4	u_5	u_5
u_4	u_{5-i}	c_2	u_i	u_3	u_5	u_5	u_3	u_4
u_5	u_i	u_i	c_2	u_5	u_4	u_3	u_4	u_3
v_0	v_{5-i}	v_{5-i}	c_1	v_0	v_1	v_2	v_1	v_2
v_1	v_i	c_1	v_{5-i}	v_2	v_0	v_0	v_2	v_1
v_2	c_1	v_i	v_i	v_1	v_2	v_1	v_0	v_0
v_3	c_2	v_{5-i}	v_{5-i}	v_4	v_3	v_4	v_5	v_5
v_4	v_{5-i}	c_2	v_i	v_3	v_5	v_5	v_3	v_4
v_5	v_i	v_i	c_2	v_5	v_4	v_3	v_4	v_3
s_2	u_i	j_i	v_{5-i}	s_1	s_5	s_6	s_3	s_4
s_3	j_{5-i}	v_i	u_i	s_4	s_1	s_2	s_6	s_5
s_4	v_i	j_{5-i}	u_{5-i}	s_3	s_6	s_5	s_1	s_2
s_5	u_{5-i}	v_{5-i}	j_i	s_6	s_2	s_1	s_4	s_3
s_6	v_{5-i}	u_{5-i}	j_{5-i}	s_5	s_4	s_3	s_2	s_1

Tabelle 2

Bekanntlich sind die in Tabelle 3 angegebenen Abbildungen $\varepsilon_i : f \to s_i^{-1} \circ f \circ s_i$, $i = 1,2,\ldots,6$, Automorphismen von $(P_3^1, \circ)$. Mit Hilfe dieser Abbildungen lassen sich isomorphe Unterhalbgruppen ermitteln.

Bewiesen werden soll im folgenden, daß $(P_3^1, \circ)$ genau 1434 Unterhalbgruppen (einschließlich $\emptyset$) besitzt[1].

1 Nach einem Hinweis von B. Csákány ist dieses Ergebnis von Zoltán Székely (Szeged) mit Hilfe eines Computers ebenfalls erhalten worden. Auf Grund dieser Computerberechnung konnte die erste Fassung des vorliegenden Artikels im Fall 6.2 korrigiert werden.

f	s_2*f*s_2	s_3*f*s_3	s_4*f*s_5	s_5*f*s_4	s_6*f*s_6
c_0	c_0	c_1	c_1	c_2	c_2
c_1	c_2	c_0	c_2	c_0	c_1
c_2	c_1	c_2	c_0	c_1	c_0
j_0	u_0	j_4	v_1	u_3	v_3
j_1	u_2	j_5	v_2	u_5	v_4
j_2	u_1	j_3	v_0	u_4	v_5
j_3	u_4	j_2	v_5	u_1	v_0
j_4	u_3	j_0	v_3	u_0	v_1
j_5	u_5	j_1	v_4	u_2	v_2
u_0	j_0	v_1	j_4	v_3	u_3
u_1	j_2	v_0	j_3	v_5	u_4
u_2	j_1	v_2	j_5	v_4	u_5
u_3	j_4	v_3	j_0	v_1	u_0
u_4	j_3	v_5	j_2	v_0	u_1
u_5	j_5	v_4	j_1	v_2	u_2
v_0	v_5	u_1	u_4	j_2	j_3
v_1	v_3	u_0	u_3	j_0	j_4
v_2	v_4	u_2	u_5	j_1	j_5
v_3	v_1	u_3	u_0	j_4	j_0
v_4	v_2	u_5	u_2	j_5	j_1
v_5	v_0	u_4	u_1	j_3	j_2
s_1	s_1	s_1	s_1	s_1	s_1
s_2	s_2	s_6	s_6	s_3	s_3
s_3	s_6	s_3	s_2	s_6	s_2
s_4	s_5	s_5	s_4	s_4	s_5
s_5	s_4	s_4	s_5	s_5	s_4
s_6	s_3	s_2	s_3	s_2	s_6

Tabelle 3

Zum Beweis werden sämtliche Unterhalbruppen bestimmt. Dazu bezeichne H in diesem Artikel stets eine Unterhalbgruppe von $(P_3^1,\varkappa)$. Die Anzahl der Möglichkeiten für H in einem der betrachteten Fälle i für H sei n_i. In der sich jetzt anschließenden Fallunterscheidung für H werden jeweils n_i sowie die Möglichkeiten für H angegeben und nur, wenn das direkte Durchmustern der in Frage kommenden Mengen zu umfangreich ist, einige Anmerkungen gemacht, die das Durchmustern - darauf läuft letztlich der Beweis hinaus - erleichtern.

1. Fall: $H \subseteq C$.
Offensichtlich ist $n_1 = 8$ und H eine beliebige Teilmenge von C.

2. Fall: $H \nsubseteq C \wedge H \subseteq A \in \{J,U,V\}$.
Es ist nicht schwierig nachzuprüfen, daß in diesem Fall $n_2 = 123$ und H eine der Mengen J_i, $U_i = a_2(J_i)$ oder $V_i = a_6(J_i)$ aus Tabelle 4 ist, $i = 1,2,\ldots,41$.

3. Fall: $H \subseteq J \cup U \wedge H \nsubseteq J \wedge H \nsubseteq U$.

3.1: $H \subseteq C \cup \{j_1,j_4,u_2,u_3\} = J_{25} \cup U_{25}$.

Die Möglichkeiten für H sind $J_3 \cup \{c_0,c_2\}$, $U_3 \cup \{c_0,c_1\}$, $J_{12} \cup \{c_0,c_2\}$, $U_{12} \cup \{c_0,c_1\}$, $J_{25} \cup \{c_0,c_2\}$, $U_{25} \cup \{c_0,c_1\}$ und $J_p \cup U_q$, wobei $(p,q) \in \{(3,3),(3,12),(12,3),(12,12),(3,25),(25,3),(12,25),(25,12),(25,25)\}$, d. h., es gilt $n_{3.1} = 15$.

3.2: $H \cap J \nsubseteq \{c_0,c_1,j_1,j_4\} \wedge H \cap U \subseteq \{c_0,c_2,u_2,u_3\}$.

Für $H \cap U$ kommen in diesem Fall nur die Mengen $\{c_2\}$, $\{c_0,c_2\}$, $\{c_0,u_2\} = U_3$, $\{c_0,c_2,u_2\} = U_{12}$ und $\{c_0,c_2,u_2,u_3\} = U_{25}$ in Frage.

3.2.1: $H \cap U = \{c_2\}$.
Offensichtlich ist $H = J_8 \cup \{c_2\}$.

3.2.2: $H \cap U = \{c_0,c_2\}$.
$H \cap J$ ist dann eine beliebige Unterhalbgruppe von J, die $\{c_0,c_1\}$ enthält aber keine Teilmenge von J_{25} ist, d. h., es gilt $n_{3.2.2} = 19$ und $H = \{c_0,c_2\} \cup J_i$ mit $i \in \{13,14,15,22,23,24,26,27,28,29,33,34,35,36,37,38,39,40,41\}$.

3.2.3: $H \cap U = U_3$.
Es gilt $n_{3.2.3} = 17$ und $H = U_3 \cup J_i$ mit $i \in \{4,13,14,16,18,23,24,27,28,30,33,34,36,38,39,40,41\}$.

J_i										$U_i = s_2 * J_i * s_2$										$V_i = s_6 * J_i * s_6$									
i	c_0	c_1	J_0	J_1	J_2	J_3	J_4	J_5		i	c_0	c_2	u_0	u_1	u_2	u_3	u_4	u_5		i	c_1	c_2	v_0	v_1	v_2	v_3	v_4	v_5
1	-	-	-	+	-	-	-	-		1	-	-	-	-	+	-	-	-		1	-	-	-	-	-	-	+	-
2	-	-	-	-	-	-	-	+		2	-	-	-	-	-	-	-	+		2	-	-	-	-	+	-	-	-
3	+	-	-	+	-	-	-	-		3	+	-	-	-	+	-	-	-		3	-	+	-	-	-	-	+	-
4	+	-	-	-	+	-	-	-		4	+	-	-	+	-	-	-	-		4	-	+	-	-	-	-	-	+
5	+	-	-	-	-	-	-	+		5	+	-	-	-	-	-	-	+		5	-	+	-	-	+	-	-	-
6	-	+	-	+	-	-	-	-		6	-	+	-	-	+	-	-	-		6	+	-	-	-	-	-	+	-
7	-	+	-	-	+	-	-	-		7	-	+	-	-	-	+	-	-		7	+	-	+	-	-	-	-	-
8	-	+	-	-	-	-	-	+		8	-	+	-	-	-	-	-	+		8	+	-	-	-	+	-	-	-
9	-	-	+	-	-	-	-	+		9	-	-	+	-	-	-	-	+		9	-	-	-	-	+	+	-	-
10	-	-	-	+	-	-	+	-		10	-	-	-	-	+	+	-	-		10	-	-	-	+	-	-	+	-
11	-	-	-	+	-	-	-	+		11	-	-	-	-	+	-	-	+		11	-	-	-	-	+	-	+	-
12	+	+	-	+	-	-	-	-		12	+	+	-	-	+	-	-	-		12	+	+	-	-	-	-	+	-
13	+	+	-	-	+	-	-	-		13	+	+	-	+	-	-	-	-		13	+	+	-	-	-	-	-	+
14	+	+	-	-	-	+	-	-		14	+	+	-	-	-	-	+	-		14	+	+	+	-	-	-	-	-
15	+	+	-	-	-	-	-	+		15	+	+	-	-	-	-	-	+		15	+	+	-	-	+	-	-	-
16	+	-	-	+	+	-	-	-		16	+	-	-	+	+	-	-	-		16	-	+	-	-	-	-	+	+
17	+	-	-	+	-	-	-	+		17	+	-	-	-	+	-	-	+		17	-	+	-	-	+	-	+	-
18	+	-	-	-	+	-	-	+		18	+	-	-	+	-	-	-	+		18	-	+	-	-	+	-	-	+
19	-	+	-	+	-	+	-	-		19	-	+	-	-	+	-	+	-		19	+	-	+	-	-	-	+	-
20	-	+	-	+	-	-	-	+		20	-	+	-	-	+	-	-	+		20	+	-	-	-	+	-	+	-
21	-	+	-	-	-	+	-	+		21	-	+	-	-	-	-	+	+		21	+	-	+	-	+	-	-	-
22	+	+	+	-	-	-	-	+		22	+	+	+	-	-	-	-	+		22	+	+	-	-	+	+	-	-
23	+	+	-	+	+	-	-	-		23	+	+	-	+	+	-	-	-		23	+	+	-	-	-	-	+	+
24	+	+	-	+	-	+	-	-		24	+	+	-	-	+	-	+	-		24	+	+	+	-	-	-	+	-
25	+	+	-	+	-	-	+	-		25	+	+	-	-	+	+	-	-		25	+	+	-	+	-	-	+	-
26	+	+	-	+	-	-	-	+		26	+	+	-	-	+	-	-	+		26	+	+	-	-	+	-	+	-
27	+	+	-	-	+	+	-	-		27	+	+	-	+	-	-	+	-		27	+	+	+	-	-	-	-	+
28	+	+	-	-	+	-	-	+		28	+	+	-	+	-	-	-	+		28	+	+	-	-	+	-	-	+
29	+	+	-	-	-	+	-	+		29	+	+	-	-	-	-	+	+		28	+	+	+	-	+	-	-	-
30	+	-	-	+	+	-	-	+		30	+	-	-	+	+	-	-	+		30	-	+	-	-	+	-	+	+
31	-	+	-	+	-	+	-	+		31	-	+	-	-	+	-	+	+		31	+	-	+	-	+	-	+	-
32	-	-	+	+	-	-	+	+		32	-	-	+	-	+	+	-	+		32	-	-	-	+	+	+	+	-
33	+	+	-	+	+	+	-	-		33	+	+	-	+	+	-	+	-		33	+	+	+	-	-	-	+	+
34	+	+	-	+	+	-	-	+		34	+	+	-	-	+	+	-	+		34	+	+	-	-	+	-	+	+
35	+	+	-	+	-	+	-	+		35	+	+	-	-	+	-	+	+		35	+	+	+	-	+	-	+	-
36	+	+	-	-	+	+	-	+		36	+	+	-	+	-	-	+	+		36	+	+	+	-	+	-	-	+
37	+	+	+	+	-	-	+	+		37	+	+	+	-	+	+	-	+		37	+	+	-	+	+	+	+	-
38	+	+	+	-	+	+	-	+		38	+	+	+	+	-	-	+	+		38	+	+	+	-	+	+	-	+
39	+	+	-	+	+	+	+	-		39	+	+	-	+	+	+	+	-		39	+	+	+	+	-	-	+	+
40	+	+	-	+	+	-	+	-		40	+	+	-	+	-	+	-	+		40	+	+	+	-	+	-	+	+
41	+	+	+	+	+	+	+	+		41	+	+	+	+	+	+	+	+		41	+	+	+	+	+	+	+	+

(+ steht in der Tabelle für enthalten, - für nicht enthalten)

Tabelle 4

3.2.4: $H \cap U = U_{12}$.
Man prüft leicht nach, daß $n_{3.2.4} = 13$ und $H = U_{12} \cup J_i$ ist, $i \in \{13,14,23,24,27,28,33,34,36,38,39,40,41\}$.

3.2.5: $H \cap U = U_{25}$.
In diesem Fall ist $H \cap J$ eine beliebige Unterhalbgruppe von J, die $\{j_2, j_3\}$ enthält. Also ist $n_{3.2.5} = 7$ und $H = U_{25} \cup J_i$ für $i \in \{27,33,36,38,39,40,41\}$. Folglich gilt $n_{3.2} = 57$.

3.3: $H \cap J \subsetneq \{c_0, c_1, j_1, j_4\} \wedge H \cap U \not\subseteq \{c_0, c_2, u_2, u_3\}$.
In diesem Fall ist H isomorph zu einer die Bedingungen von 3.2 erfüllenden Unterhalbgruppe von $J \cup U$. Folglich ist $n_{3.3} = 57$ und $H = a_2(H')$, wobei H' die Bedingung 3.2 erfüllt.

3.4: $H \cap J \not\subseteq \{c_0, c_1, j_1, j_4\} \wedge H \cap U \not\subseteq \{c_0, c_2, u_2, u_3\}$.
Dann gibt es in H zwei Funktionen f und g mit $f\binom{0}{1} \in \binom{0\ 2}{2\ 0}$ und $g\binom{0}{2} \in \binom{0\ 1}{1\ 0}$. Hieraus folgt $|H \cap J| = |H \cap U|$ und, falls $(H \setminus C) \cap J = \{j_1 \mid i \in I\}$ ist, $(H \setminus C) \cap U \in \{\{u_i \mid i \in I\}, \{u_{5-i} \mid i \in I\}\}$.
Durchmustern der möglichen Fälle liefert $n_{3.4} = 19$ und $H = J_p \cup U_q$ mit $(p,q) \in \{(2,2),(5,5),(8,8),(9,9),(15,15),(16,16),(17,18),(18,17),(22,22),(23,23),(24,24),(26,28),(28,26),(30,30),(34,34),(37,38),(38,37),(39,39),(41,41)\}$.
Zusammenfassend ergibt sich $n_3 = 148$.

4. Fall: $H \subsetneq A \cup B \in \{J \cup V, U \cup V\} \wedge H \not\subseteq A \wedge H \not\subseteq B$.
Wegen $a_3(J \cup U) = J \cup V$ sowie $a_4(J \cup U) = U \cup V$ ist $n_4 = 296$ und H isomorph zu einer im Fall 3 bestimmten Unterhalbgruppe.

5. Fall: $H \subsetneq J \cup U \cup V \wedge H \not\subseteq J \cup U \wedge H \not\subseteq J \cup V \wedge H \not\subseteq U \cup V$.

5.1: $H \cap (U \cup V) \subsetneq C \cup \{u_2, u_3, v_2, v_3\}$.

Für $H \cap (U \cup V)$ kommen nur die Mengen $U_3 \cup V_8 = \{c_0, c_1, u_2, v_2\}$, $U_{12} \cup V_{15} = C \cup \{u_2, v_2\}$ und $U_{25} \cup V_{22} = C \cup \{u_2, u_3, v_2, v_3\}$ in Frage.

5.1.1: $H \cap (U \cup V) = U_3 \cup V_8$.
Möglich für $H \cap J$ sind dann alle Teilmengen J_i von J, die $\{c_0, c_1\}$ enthalten und für die gilt:
$$j_0 \in J_i \vee j_4 \in J_i \rightarrow \{j_2, j_3\} \subset J_i, \quad j_1 \in J_i \Rightarrow j_3 \in J_i, \quad j_5 \in J_i \Rightarrow j_2 \in J_i.$$
Also ist $n_{5.1.1} = 11$ und $H = J_i \cup U_3 \cup V_8$, $i \in \{13,14,24,27,28,33,36,38,39,40,41\}$.

5.1.2: $H \cap (U \cup V) = U_{12} \cup V_{15}$.

Mit Hilfe von 5.1.1 erhält man $n_{5.1.2} = 11$ und $H = J_i \cup U_{12} \cup V_{15}$, $i \in \{13,14,24,27,28,33,36,38,39,40,41\}$.

5.1.3: $H \cap (U \cup V) = U_{25} \cup V_{22}$.

$H \cap J$ ist in diesem Fall eine Unterhalbgruppe von J, die j_2 und j_3 enthält. Folglich ist $n_{5.1.3} = 7$ und $H = J_i \cup U_{25} \cup V_{22}$, $i \in \{27,33,36,38,39,40,41\}$.

5.2: $H \cap (J \cup V) \subsetneq C \cup \{j_1, j_4, v_1, v_4\}$.

Wegen $a_2(C \cup \{j_1, j_4, v_1, v_4\}) = U_{25} \cup V_{22}$ ist $n_{5.2} = 29$ und $H = a_2(A)$, wobei A eine der im Fall 5.1. bestimmten Mengen ist.

5.3: $H \cap (J \cup U) = C \cup \{j_0, j_5, u_0, u_5\}$.

Auch dieser Fall ist isomorph zu Fall 5.1. Sämtliche 29 Möglichkeiten für H erhält man mittels der Abbildung a_6 aus den in 5.1 angegebenen Mengen.

5.4: H erfüllt keine der in 5.1, 5.2. oder 5.3. genannten Bedingungen.

In diesem Fall gibt es in H Funktionen f, g und h mit $f\binom{0}{1} \in \binom{0\ 2\ 1\ 2}{2\ 0\ 2\ 1}$, $g\binom{0}{2} \in \binom{0\ 1\ 1\ 2}{1\ 0\ 2\ 1}$ und $h\binom{1}{2} \in \binom{0\ 1\ 0\ 2}{1\ 0\ 2\ 0}$.

Da zu H außerdem gewisse Funktionen j_a, u_b, v_c gehören, überzeugt man sich leicht davon, daß $C \subsetneq H$ und, falls $(H \setminus C) \cap J = \{j_1 \mid i \in I\}$ ist, $(H \setminus C) \cap U \in \{\{u_i \mid i \in I\}, \{u_{5-i} \mid i \in I\}\}$ sowie $(H \setminus C) \cap V \in \{\{v_i \mid i \in I\}, \{v_{5-i} \mid i \in I\}\}$ gilt. Durchmustern der möglichen Fälle unter Verwendung der Ergebnisse von Fall 2 liefert $n_{5.4} = 7$ und $H = J_p \cup U_q \cup V_r$ mit $(p,q,r) \in \{(24,24,26), (26,28,24), (28,26,28), (37,38,39), (38,37,38), (39,39,37), (41,41,41)\}$. Folglich ist $n_5 = 94$.

6. Fall: $S \cap H \neq \emptyset$.

Da S bekanntlich 6 Untergruppen besitzt, sind folgende Fälle möglich:

6.1: $S \cap H = \{s_1\}$.

Dann ist $H = A \cup \{s_1\}$, wobei A eine der 669 oben bestimmten Unterhalbgruppen von $J \cup U \cup V$ ist.

6.2: $S \cap H = \{s_1, s_3\}$.

$H \setminus S$ ist dann eine der folgenden 30 Mengen: $\emptyset$, $\{c_2\}$, $\{c_0, c_1\}$, C, J_{27}, J_{32}, J_{37}, J, $J_{27} \cup \{c_2\}$, $J_{37} \cup \{c_2\}$, $C \cup J$, $U_1 \cup V_2$, $U_6 \cup V_5$, $U_{12} \cup V_{15}$, $U_{10} \cup V_9$, $U_{25} \cup V_{22}$, $U_{21} \cup V_{16}$, $U_{29} \cup V_{23}$, $U_{31} \cup V_{30}$, $U_{35} \cup V_{34}$, $U_{38} \cup V_{39}$, $U \cup V$, $J_{27} \cup U_1 \cup V_2$, $J_{27} \cup U_6 \cup V_2$, $J_{27} \cup U_{25} \cup V_{22}$, $J_{37} \cup U_{38} \cup V_{39}$, $J \cup U_1 \cup V_2$, $J \cup U_6 \cup V_2$, $J \cup U_{25} \cup V_2$, $J \cup U \cup V$.

6.3: $S \cap H \in \{\{s_1, s_2\}, \{s_1, s_6\}\}$.

In diesem Fall ist $n_{6.3} = 60$, und die Möglichkeiten für H erhält man aus 6.2 mit Hilfe der Abbildungen a_2 und a_6.

6.4: $S \cap H \in \{\{s_1, s_4, s_5\}, S\}$.

Für $H \setminus S$ kommen nur die Mengen $\emptyset$, C und $J \cup U \cup V$ in Frage, d. h., es gilt $n_{6.4} = 6$. Folglich ist $n_6 = 765$, womit sich 1434 als Anzahl der möglichen Unterhalbgruppen von P_3^1 ergibt.

Literatur

/1/ Lau, D.: Abgeschlossene Mengen quasilinearer Funktionen von P_3. Preprint, Wilhelm-Pieck-Universität Rostock, Sektion Mathematik 1983

/2/ Мальцев, И. А.: Некоторые свойства клеток алгебр Поста. Дискрет. Анализ 23, 24 - 31 (1973)

/3/ Pöschel, R., und Kalužnin, L. A.: Funktionen- und Relationenalgebren, Berlin 1979

eingegangen: 15. 12. 1983

Anschrift des Verfassers:

Dr. D. Lau
Wilhelm-Pieck-Universität Rostock
Sektion Mathematik
Universitätsplatz 1
DDR-2500 Rostock

Rostock. Math. Kolloq. 26, 63 - 68 (1984)

06A10
05A05

Konrad Engel

Optimal representations, LYM posets, Peck posets, and the Ahlswede-Daykin inequality

In this nota (partly with survey character) we will show that results concerning LYM posets, Peck posets, the Ahlswede-Daykin inequality, and results concerning the asymptotic size of antichains in products of finite posets come together in the general notion of optimal representetions of posets.
Throughout we consider finite partially ordered sets.
A representation of a poset is a function $x: P \longrightarrow \mathbb{R}$ such that $a \geq_P b$ implies $x(a) - x(b) \geq 1$. The expected value μ_x end the variance σ_x^2 of the representation x of P are defined by

$$\mu_x := \frac{1}{|P|} \sum_{a \in P} x(a)$$

and

$$\sigma_x^2 := \frac{1}{|P|} \sum_{a \in P} x^2(a) - \mu_x^2,$$

respectively. The variance $\sigma^2(P)$ of the poset P is defined by $\sigma^2(P) := \inf \sigma_x^2$, where the infimum is extended over all representations x of P. A representation x of P is called optimal iff $\sigma_x^2 = \sigma^2(P)$.
Let d(P) be the meximum size of an antichain in P and let P^n be the (direct) product of n fectors P. The importance of optimal representations is given by a theorem of Alekseev /3/.

Theorem 1 (Alekseev): It holds

$$d(P^n) \sim (\sqrt{2\pi n}\,\sigma(P))^{-1} |P|^n \quad \text{as } n \to \infty.$$

This theorem was generalized by Engel and Kuzjurin /9/ to products of posets with boundad cardinalities which ere not necessarily equal each other end by Engel /8/ to weighted posets.

A poset P is ranked iff there is a function r: $P \longrightarrow \mathbb{N}$ such that $r(a) = 0$ if x is minimal in P and $r(b) = r(a) + 1$ if $b \geq a$ (b covers a if there is no element between a and b). In all what follows let $p := \max_{a \in P} r(a)$. The set $N_i := \{a \in P,\ r(a) = i\}$ and the number $W_i := |N_i|$ are called the i-th level and the i-th Whitney number of P, respectively. Let w(P) be the maximum Whitney number of P. Obviously, the rank function is a representation of the ranked poset P. Using the Local Limit Theorem of Gnedenko and Theorem 1 one can prove (see /6/) that for ranked posets P it holds $d(P^n) \sim w(P^n)$ as $n \to \infty$, if the rank function is an optimal representation (in that case one can call P^n asymptotically Sperner). Thus it is interesting to investigate in which ranked posets the rank function is an optimal representation.

Denote the edge set of the Hasse graph of P by E(P) and call $F \subseteq P$ an order filter if $a \in F$, $b > a$ imply $b \in F$. Finally for a subset S of P define $r(S) := \sum_{a \in S} r(a)$ (note that then r(P) is not the maximum rank in P).

Using the Theorem of Kuhn and Tucker in quadratic programming and the Duality Theorem in linear programming one can prove the following theorem (Engel /8/, for a more combinatorial approach see Alekseev /3/).

Theorem 2: Let P be a poset with rank function r. The following conditions are equivalent:

i) r is an optimal representation of P.

ii) There is a function $f: E(P) \longrightarrow \mathbb{R}_+$ such that

$$\sum_{a,a<b} f((a,b)) - \sum_{c,b<c} f((b,c)) = r(b) - \mu_r$$

for all $b \in P$.

iii) For all order filters F of P it holds

$$\frac{r(F)}{|F|} \geq \frac{r(P)}{|P|} \quad (= \mu_r).$$

In the following we show how Theorem 2 can be applied directly to certain posets without using Theorem 1 and other previously discovered product theorems /5, 12, 14, 17/.
A ranked poset P is called a LYM poset iff there is a function $g: E(P) \longrightarrow \mathbb{R}_+$ such that

$$\sum_{b,\, a \lessdot b} g((a,b)) = \frac{1}{W_o} \qquad \text{if } a \in N_o,$$

$$\sum_{a,\, a \lessdot b} g((a,b)) = \sum_{c,\, b \lessdot c} g((b,c)) = \frac{1}{W_i} \qquad \text{if } b \in N_i, \quad i = 1,\ldots,p-1,$$

$$\sum_{a,\, a \lessdot b} g((a,b)) = \frac{1}{W_p} \qquad \text{if } b \in N_p.$$

For equivalent conditions see Harper /12/ and Kleitman /16/.

Theorem 3 (Engel /8/): If P is a LYM poset, then the rank function r is an optimal representation of P.

Proof: The function f defined by

$$f((a,b)) := g((a,b)) \sum_{j=0}^{i} (\mu_r - j)\, W_j, \quad a \in N_i,\ a \lessdot b,$$

(i = 0,...,p-1) satisfies condition ii) of Theorem 2.

q.e.d.

Remark: In /8/ it was proved more generally that flow morphisms preserve variances (see /7/, /13/).

A subset T of P is called a k-family iff there are no k+1 elements of T which lie on any single chain in P. A ranked poset P is called a Peck poset if $W_o = W_p \leq W_1 = W_{p-1} \leq \ldots$ and the maximum size of a k-family in P equals the sum of the k largest Whitney numbers in P for all k.

Theorem 4 (Engel /8/): If P is a Peck poset, then the rank function r is an optimal representation.

Proof: From a result of Griggs /11/(for all $i < \frac{p}{2}$ there exist W_i disjoint chains with bottom in N_i and top in N_{p-i}) it follows that in Peck posets there is a bijection

$\varphi: \bigcup_{i<\frac{p}{2}} N_i \longrightarrow \bigcup_{j>\frac{p}{2}} N_j$ such that $x < \varphi(x)$ and

$r(x) + r(\varphi(x)) = p$ for all $x \in P$ with $r(x) < \frac{p}{2}$.
Obviously, we have

$$\frac{r(P)}{|P|} = \mu_r = \frac{p}{2}.$$

If F is any order filter in P, then there is no pair $(x, \varphi(x))$ such that $x \in F$ but $\varphi(x) \notin F$. Hence

$$\frac{r(F)}{|F|} \geq \frac{p}{2}$$

and condition iii) of Theorem 2 is satisfied.

q.e.d.

A lattice P is called distributive iff for all $x,y,z \in P$ it holds $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$. It is well-known (see Aigner /2/) that each distributive lattice possesses a rank function r. In Theorem 6 we will apply the inequality of Ahlswede and Daykin /1/ which is a generalization of inequalities of Kleitman /15/, Fortuin, Kasteleyn and Ginibre /10/ and others.

Theorem 5 (Ahlswede and Daykin /1/): Let P be a distributive lattice and let α, β, γ, δ be functions $P \to \mathbb{R}_+$ such that

$$\alpha(a)\beta(b) \leq \gamma(a \vee b)\, \delta(a \wedge b) \text{ for all } a,b \in P.$$

Then

$$\alpha(A)\beta(B) \leq \gamma(A \vee B)\, \delta(A \wedge B) \text{ for all } A,B \subseteq P,$$

where $\alpha(A) := \sum_{a \in A} \alpha(a)$ etc., $A \vee B := \{a \vee b,\ a \in A,\ b \in B\}$,

and $A \wedge B := \{a \wedge b,\ a \in A,\ b \in B\}$.

Theorem 6: If P is a distributive lattice, then its rank function r is an optimal representation of P.

Proof: Let F be any order filter of P. Put in Theorem 5 $\alpha := \gamma := r$, $\beta := \delta := 1$, $A := P$, and $B := F$. Then $A \vee B = F$ and $A \wedge B = P$. Obviously, it holds $r(a) \cdot 1 \leq r(a \vee b) \cdot 1$ for all $a, b \in P$. Hence by Theorem 5

$$r(P)\,|F| \leq r(F)\,|P|,$$

i.e.

$$\frac{r(F)}{|F|} \geq \frac{r(P)}{|P|}$$

and condition iii) of Theorem 2 is satisfied. q.e.d.

Finally let P be the partition lattice. Rota's problem /18/ to decide whether $d(P) = w(P)$ has a long history and influenced the development of the Sperner theory. Canfield /4/ proved that in general $d(P) > w(P)$. The following problem is a natural generalization of Rota's problem.

Problem: Decide whether the rank function of the partition lattice is an optimal representation of it.

References

/1/ Ahlswede, R., and Daykin, D. E.: An inequality for the weights of two families of sets, their unions and intersections. Z. Wahrsch. Verw. Gebiete 43, 183 - 185 (1978)

/2/ Aigner, M.: Kombinatorik I. Grundlagen und Zähltheorie. Berlin 1976

/3/ Alekseev, V. B.: O čisle monotonnych k-značnych funkcij. Problemy Kibernet. 28, 5 - 24 (1974)

/4/ Canfield, E. R.: On a problem of Rota. Adv. in Math. 29, 1 - 10 (1978)

/5/ Canfield, E. R.: A Sperner property preserved by products. Linear and Multilinear Algebra 9, 151 - 157 (1980)

/6/ Engel, K.: An asymptotic formula for maximal h-families in ranked product orders. Rostock. Math. Kolloq. 21, 11 - 14 (1982)

/7/ Engel, K.: A new proof of a theorem of Harper on the Sperner-Erdös problem. J. Combin. Theory Ser. A (to appear)

/8/ Engel, K.: Optimal representations of partially ordered sets and a limit Sperner theorem. Submitted to European J. Combin.

/9/ Engel, K., and Kuzjurin, N. N.: An asymptotic formula for the maximum size of an h-family in products of partially ordered sets. J. Combin. Theory Ser. A (to appear)

/10/ Fortuin, C. M., Kasteleyn, P. W., and Ginibre, J.: Correlation inequalities on some partially ordered sets. Comm. Math. Phys. 22, 89 - 103 (1971)

/11/ Griggs, J. R.: On chains and Sperner k-families in ranked posets. J. Combin. Theory Ser. A 28, 156 - 168 (1980)

/12/ Harper, L. H.: The morphology of partially ordered sets. J. Combin. Theory Ser. A 17, 44 - 58 (1974)

/13/ Harper, L. H.: The global theory of flows in networks. Adv. in Appl. Math. 1, 158 - 181 (1980)

/14/ Hsieh, W. N., and Kleitman, D. J.: Normalized matching in direct products of partial orders. Stud. Appl. Math. 52, 285 - 289 (1973)

/15/ Kleitman, D. J.: Families of non-disjoint subsets. J. Combin. Theory 1, 153 - 155 (1966)

/16/ Kleitman, D. J.: On an extremal property of antichains in partial orders. The LYM property and some of its implications and applications. In: Hall, M., and van Lint, J. H. (Eds.): Combinatorics, Part 2: Graph Theory; Foundations, Partitions and Combinatorial Geometry. Math. Centre Tracts 56, pp. 77 - 90. Amsterdam 1974

/17/ Proctor, R. A., Saks, M. E., and Sturtevant, D. G.: Product partial orders with the Sperner property. Discrete Math. 30, 173 - 180 (1980)

/18/ Rota, G.-C.: Research problem: A generalization of Sperner's theorem. J. Combin. Theory 2, 104 (1967)

received: June 7, 1984

Author's address:

Dr. K. Engel
Wilhelm-Pieck-Universität Rostock
Sektion Mathematik
Universitätsplatz 1
DDR-2500 Rostock

Rostock. Math. Kolloq. 26, 69 - 82 (1984)

65D10
03E72

Hans Bandemer

Zur Bestimmung funktionaler Beziehungen aus Fuzzy-Beobachtungen[1]

1. Einleitung

Üblicherweise wird die Ungenauigkeit, Unbestimmtheit und Unsicherheit von Beobachtungsergebnissen in einem mathematischen Modell dadurch berücksichtigt, daß man sie als Realisierungen von Zufallsvariablen auffaßt. Bei der Untersuchung funktionaler Beziehungen nimmt man dann gewöhnlich an, daß diese von den Erwartungswerten jener Zufallsvariablen gebildet werden. Um den Graph dieser Beziehung zu schätzen, wählt man dann in der Regel ein numerisches Approximationsverfahren, das den gegebenen Realisierungen den Graph einer Funktion zuordnet. Um nun zu weiterführenden Aussagen über diese Schätzung zu kommen, bedarf es einer Reihe von Annahmen wie über die Unabhängigkeit oder Unkorreliertheit der Zufallsvariablen bei verschiedenen Beobachtungen, über deren Varianz bis hin zu deren Verteilungstyp. Diese Annahmen, die in ihrer aktuellen Gesamtheit das statistische Modell bilden, können in gewissen Fällen gerechtfertigt werden, in anderen unwidersprochen bleiben, aber es gibt immer wieder Fälle, in denen sie recht willkürlich und unmotiviert erscheinen, z.B. wenn die Beobachtungen unterschiedlichen, kaum vergleichbaren Quellen entstammen (z.B. aus unterschiedlichen Situationen mit nicht vergleichbarer Meßtechnik, eventuell zusammen mit Expertenmeinungen).
In diesen oder ähnlichen Situationen sieht man sich nach einer Alternative um. Eine solche soll im folgenden angeboten werden. Sie benutzt die Denkweise der Fuzzy-Theorie, besitzt eine hohe Flexibilität in der Anpassung an reale Situationen, verlangt

1 Vortrag, gehalten am 13.4.1984 im Rostocker Mathematischen Kolloquium

keine vergleichbar einschneidenden Annahmen, erlaubt aber Aussagen ähnlicher Struktur (z.B. ähnlich zu Konfidenzschätzungen).
Nachdem im nächsten Abschnitt der Begriff der Fuzzy-Beobachtung eingeführt wurde, folgen einige Vorschläge für die Zuordnung von Graphen von Funktionen zu diesen Beobachtungen. Entsprechend dem Charakter der Beobachtungen werden diese ebenfalls zu unscharfen Mengen führen.

2. Fuzzy-Beobachtungen

2.1. Spezifizierung einer Fuzzy-Beobachtung

Gegeben sei eine Grundmenge G, z.B. als kompakte Untermenge des R^2. Jede Teilmenge $H \subset G$ läßt sich bekanntlich durch ihre Zugehörigkeitsfunktion μ darstellen:

$$\mu_H(x,y) = 1 \quad \text{falls} \quad (x,y) \in H\,, \qquad \mu_H(x,y) = 0 \quad \text{falls} \quad (x,y) \notin H\,. \tag{2.1}$$

Als Fuzzy-Menge M aus G bezeichnet man nun jede Abbildung von G in $[0,1]$, gegeben durch ihre Zugehörigkeitsfunktion μ_M

$$M : \mu_M \mid G \rightarrow [0,1]. \tag{2.2}$$

Man kann sie sich als Grautonbild auf G vorstellen (als Verallgemeinerung des Scherenschnitts H).
Zur Fuzzy-Beobachtung kommt man, indem man μ als Grad der Möglichkeit deutet, wo nach dem Vorwissen oder/und der durchgeführten konkreten Beobachtung Punkte der funktionalen Beziehung, die selbst als unbekannt, aber fest und determiniert (nicht fuzzy) angenommen wird, liegen können. (Der Weg, daß auch diese funktionale Beziehung als fuzzy angenommen wird, d.h., daß der "wahre" Parameter dieser Beziehung eine Fuzzy-Menge ist, wird z.B. von Tanaka et. al. /1/ beschritten.)
Die Möglichkeiten, solche Fuzzy-Beobachtungen aus den verschiedenen Quellen zu spezifizieren, sind überaus vielgestaltig und nur beschränkt durch den notwendigen Konsens der an der Lösung des Problems Beteiligten und Interessierten. Die spezielle analytische oder numerische Form der Zugehörigkeitsfunktion hat dabei wenig Einfluß auf das Ergebnis, wenn nur die Isotonie des Verlaufs gewahrt wird (z.B. Dreieck oder Parabelab-

schnitt, Pyramide oder Kugelkappe).
Wenn man es gewöhnt ist, Beobachtungspunkte mit den Variabilitätsgrenzen anzugeben, so fällt es nicht schwer, statt dessen einen Grautonsaum festzulegen, als Unsicherheitszone für die Lage der funktionalen Beziehung.
Ebenso scheint es nicht sehr kompliziert, in Expertengesprächen solche Grauton-Bilder zu erarbeiten, die dann dokumentieren, wo die einzelnen Experten die Lage des Graphen der funktionalen Beziehung erwarten.

Liegen "exakt" angegebene Beobachtungspunkte, z.B. (x_i,y_i) , $i = 1,\ldots,n$, vor, so kann man ein geeignetes Strukturelement B wählen (etwa einen kleinen Kreis), das den Unsicherheitsgrad der Beobachtungen repräsentiert. Es bezeichne $B_{(x,y)}$ das Strukturelement, wenn es den "Mittelpunkt" (x,y) hat. (Zur Rolle des Strukturelementes in der mathematischen Morphologie s. z.B. Serra /2/.)
Dann wäre eine Möglichkeit zur Festlegung der Fuzzy-Beobachtung M

$$M : \mu(x,y) = (1/c)\,\mathrm{card}\,\{(x_i,y_i) : (x_i,y_i) \in B_{(x,y)}\}\,, \qquad (2.3)$$

wobei die Normierungskonstante c so zu wählen wäre, daß μ für alle $(x,y) \in G$ in $[0,1]$ liegt.

Man könnte auch Thieles Astro-Cumulus-Methode verwenden (Thiele /3/), die das optische Verschmelzen der Elemente eines Sternhaufens nachgestaltet.

2.2. Zusammenführen von Beobachtungen

Es bezeichne

$$\mathrm{supp}\ M = \{(x,y) : \mu_M(x,y) > 0\}\,. \qquad (2.4)$$

Nun kann es vorkommen, daß es unter den n Fuzzy-Beobachtungen M_i, $i = 1,\ldots,n$, gewisse mit $j \neq k$ gibt, für die

$$\mathrm{supp}\ M_j \cap \mathrm{supp}\ M_k \neq \phi\,, \qquad (2.5)$$

und gewisse (x_m,y_m), für die

$$0 \neq \mu_{M_j}(x_m,y_m) \neq \mu_{M_k}(x_m,y_m) \neq 0 \qquad (2.6)$$

ausfällt. Für gewisse Verfahren (s. Abschnitte 3.1 und 3.2) ist es erforderlich, die Fuzzy-Beobachtungen vor deren Verarbeitung zu vereinigen. Dies kann wiederum auf verschiedene Art erfolgen. Am einfachsten zu motivieren ist:

$$M = \bigvee_i M_i \quad : \quad \mu_M(x,y) = \max_i \mu_{M_i}(x,y), \tag{2.7}$$

bei der jeweils der höchste Möglichkeitsgrad für jeden Punkt benutzt wird. Es tritt hierbei der bemerkenswerte Fall ein, daß zwei identische Fuzzy-Beobachtungen zur gleichen Ausgangssituation führen wie eine von ihnen. Beim Vorliegen zweier (als unabhängig angenommener!) identischer Realisierungen ist das bekanntlich nicht so.
Die Vereinigungsvorschrift (2.7) geht jedoch davon aus, daß zwei identische unscharfe Informationen ("Freiberg liegt in der Nähe von Dresden") so gut sind wie eine von ihnen.

Bei der Verwendung von Mittelwerten, die im Gegensatz zu (2.7) eine geringe Rolle spielen, muß man unterscheiden, ob $\mu_{M_i}(x_o,y_o) = 0$ bedeuten soll:

a) das Bestehen der funktionalen Beziehung wird in (x_o,y_o) für nicht möglich gehalten oder

b) über das Bestehen wird keine Aussage gemacht.

Im Fall a) wählt man dann

$$M = \mathop{A}_i M_i \quad : \quad \mu_M(x,y) = (1/n) \sum_i \mu_{M_i}(x,y) \tag{2.8}$$

und für b)

$$M = \mathop{a}_i M_i \quad : \quad \mu_M(x,y) = \frac{\sum_i \mu_{M_i}(x,y)}{\operatorname{card}\{i: \mu_{M_i}(x,y) > 0\}} \,. \tag{2.9}$$

Darüber hinaus wäre von Fall zu Fall zu entscheiden, ob noch andere Verknüpfungsmöglichkeiten (z.B. algebraische Summe, beschränkte Summe) sinnvoll interpretiert werden können.

3. Einige Verfahren zur Bestimmung funktionaler Beziehungen aus Fuzzy-Beobachtungen

Der Einfachheit halber nehmen wir die gesuchte funktionale Beziehung in expliziter Form gegeben an

$$y = f(x) \ , \quad x \in X \quad , \quad y \in Y \ , \tag{3.1}$$

und betrachten zuerst den Fall, daß wir ihre Struktur bis auf einen Vektor unbekannter Parameter kennen:

$$f(x) = \eta(x;a_0,\ldots,a_r). \tag{3.2}$$

Als Beispiel wollen wir den einfachsten Fall

$$f(x) = \eta(x;\ a_0,\ a_1) = a_0 + a_1 x \tag{3.3}$$

mitführen, wobei $x \in [0,1] = X$ und $f \in [0,1] = Y$ vorausgesetzt wird, was auf die Parametermenge

$$A = \left\{(a_0,a_1):\ \ a_0 \in [0,1] \ \ , \ a_1 \in [-a_0, 1-a_0]\right\} \tag{3.4}$$

führt. Ferner seien für das Beispiel die vier Fuzzy-Beobachtungen, $i = 1,2,3,4$,

$$M_i \ : \ \mu_i(x,y) = \left[1 - r_i^{-2}((x-x_i)^2 + (y-y_i)^2)\right]^+ \tag{3.5}$$

gegeben mit

$$\begin{array}{lll} x_1 = 0.2 \ , & y_1 = 0.2 \ , & r_1 = 0.1 \ , \\ x_2 = 0.4 \ , & y_2 = 0.45, & r_2 = 0.1 \ , \\ x_3 = 0.7 \ , & y_3 = 0.6 \ , & r_3 = 0.12, \\ x_4 = 0.7 \ , & y_4 = 0.8 \ , & r_4 = 0.08. \end{array} \tag{3.6}$$

Die Zugehörigkeitsfunktionen sind also Kugelkappen von verschiedenen Radien.

Im folgenden werden drei Verfahren zur "Fuzzy-Schätzung" vorgestellt und am Beispiel erläutert, wobei die beiden Verfahren in den Abschnitten 3.1 und 3.2 von Bandemer /4/ und das Verfahren in Abschnitt 3.3 von Bandemer und Schmerling /5/ vorgeschlagen wurden.

3.1. Erwartete Kardinalität

Hierzu werden die Fuzzy-Beobachtungen gemäß (2.7) zusammengefaßt ($M = \bigvee_i M_i$).

Betrachten wir eine spezielle Funktion, z.B. eine bestimmte Gerade $a_0 + a_1 x$, an einer festen Stelle x_0, so hat ihr

Funktionswert $a_o + a_1 x_o$ an dieser Stelle den Zugehörigkeitswert $\mu_M(x_o, a_o + a_1 x_o)$. Dieser wird sich gewöhnlich über dem Bereich X der Beziehung stark ändern, denn die Beobachtungen sind im allgemeinen nicht regelmäßig über $X \times Y$ angeordnet. Wenn wir annehmen können, daß die Werte $x \in X$, für die wir die funktionale Beziehung benutzen wollen, zufällig auftreten (als Beispiel sei an die Abhängigkeit des täglichen Wachstums von den klimatischen Bedingungen gedacht), dann können wir zu jeder Funktion $a_o + a_1 x$ ihren erwarteten Zugehörigkeitswert bestimmen. Sei P die Wahrscheinlichkeitsverteilung, nach der die x auftreten; dann hätten wir

$$E\,\mu_M(x, a_o + a_1 x) = \int_X \mu_M(x, a_o + a_1 x)\,dP(x) \qquad (3.7)$$

als sogenannte <u>erwartete Kardinalität</u> natürlich bezüglich P. (siehe Dubois und Prade /6/). Auch wenn wir die Vorstellung des zufälligen Auftretens von x nicht aufrecht erhalten können, aber eine normierte Gewichtsfunktion festlegen können, dann dürfen wir (3.7) als gewichtetes Mittel längs des Graphen der Funktion deuten und verwenden. Liegen keine Gründe für die Bevorzugung von Gebieten aus X vor, wird man die gleichmäßige Verteilung auf X wählen.

Betrachtet man nun (3.7) als Funktion der Parameter, so erhält man mit

$$\mu_E(a_o, a_1) = E\,\mu_M(x,\ a_o + a_1 x) \qquad (3.8)$$

eine Fuzzy-Menge über A, der Menge der möglichen Parameterwerte (a_o, a_1); siehe (3.4). Diese repräsentiert die in der Fuzzy-Beobachtung (3.5) enthaltene Information, transformiert in den Parameterbereich A und beurteilt durch das Wahrscheinlichkeitsmaß P.

Wenden wir uns dem konkreten Beispiel näher zu! Der Einfachheit halber wählen wir als Wahrscheinlichkeitsmaß P die Gleichverteilung $dP(x) = dx$. Um μ_E für ein spezielles (a_o^*, a_1^*) zu berechnen, prüfen wir zuerst, ob der Geradenabschnitt $y = a_o^* + a_1^* x$, $x \in [0,1]$, den Grundkreis supp μ_{M_i} einer Fuzzy-Beobachtung trifft, d.h., ob

$$(x - x_i)^2 + (a_o^* + a_1^* x - y_i)^2 = r_i^2 \qquad (3.9)$$

reelle Lösungen hat, z.B. $x_{1i} < x_{2i}$. Dann integrieren wir

$$k_i(a_o^*, a_1^*) = \int_{x_{1i}}^{x_{2i}} [1-r_i^{-2}((x-x_i)^2+(a_o^* + a_1^*x - y_i)^2)]dx \quad (3.10)$$

und erhalten

$$\mu_E(a_o^*, a_1^*) = \sum_{i=1}^{4} k_i(a_o^*, a_1^*), \quad (3.11)$$

wobei wir $k_i(a_o^*, a_1^*) = 0$ gesetzt haben, falls (3.9) keine oder nur eine reelle Lösung hat. Bild 1 zeigt die Höhenlinien für (3.11) für $\alpha = 0{,}25;\ 0{,}30;\ 0{,}35$.

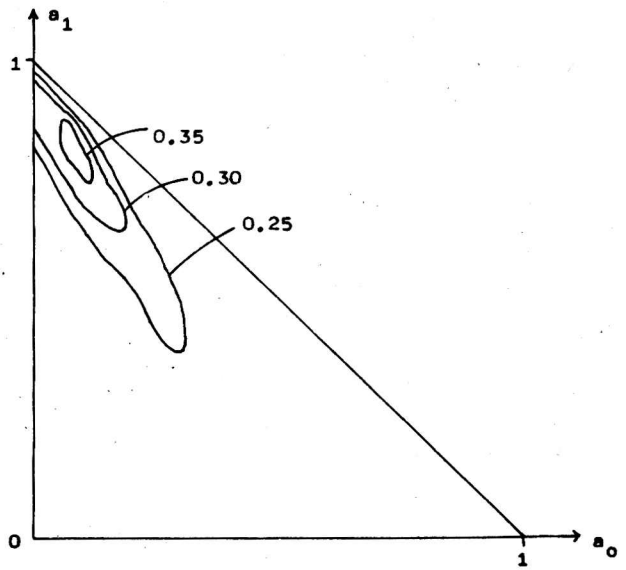


Bild 1

3.2. Fuzzy-Erwartungswert

Im Falle, daß es nicht möglich ist, eine Gewichtsfunktion über X festzulegen, wir aber trotzdem unterschiedlichen Teilmengen von X eine unterschiedliche Bedeutung zumessen müssen, können wir zuweilen ein Fuzzy-Maß g konstruieren. Dies ist eine Mengenfunktion, die nicht mehr die Additionseigenschaft hat, daß das Maß der Vereinigung zweier disjunkter Mengen die Summe der Einzelmaße ist. Statt dessen wird nur noch gefordert, daß das Maß jeder Obermenge nicht kleiner ist als das der Menge selbst. Die bei Wahrscheinlichkeitsmaßen übliche σ-Additivität wird durch die Stetigkeit bezüglich der Mengeninklusion ersetzt. Solche Eigenschaften für ein Maß können notwendig werden, wenn gewissen Teilgebieten subjektiv der Vorzug vor anderen gegeben werden soll, ohne daß der Vereinigung disjunkter Gebiete die Summe der Zahlen, die den Vorzug ausdrücken sollen, zugeschrieben werden kann oder darf.
Das Integral bezüglich eines Fuzzy-Maßes muß definiert werden, da die bekannten Integralbegriffe auf der Maßadditivität aufbauen. Sugeno /7/ gab eine Definition und zeigte, daß die folgende Darstellung mit seiner Definition äquivalent, aber handlicher ist.

Sei $h \mid X \rightarrow [0,1]$ eine Funktion und $X_0 \subseteq X$, dann ist das Fuzzy-Integral bezüglich des Fuzzy-Maßes g als

$$\int_{X_0} h(x) \circ g(\cdot) = \sup_{\alpha \in [0,1]} \left(\min \{\alpha,\ g(X_0 \cap F_\alpha)\}\right) \tag{3.12}$$

gegeben, wobei

$$F_\alpha := \{ x \in X \ : \ h(x) \geq \alpha \} \tag{3.13}$$

die α-Niveauüberschreitungsmenge von h bedeutet.
Analog zu (3.7) und (3.8) läßt sich nun

$$\mu_F (a_0, a_1) = \int_X \mu_M (x,\ a_0 + a_1 x) \circ g(\cdot) \tag{3.14}$$

als Fuzzy-Erwartungswert bezüglich g definieren.
Es ist eine Fuzzy-Menge über A und läßt sich analog zu μ_E interpretieren.

Um μ_F für ein $(a_0^*, a_1^*) \in A$ zu berechnen, haben wir die Mengen F_α zu betrachten, für die

$$\mu_M(x, a_0 + a_1 x) \tag{3.15}$$

ist. Ähnlich wie bei der Bestimmung von μ_E stellen wir fest, ob das Geradenstück $y = a_0^* + a_1^* x$, $x \in [0,1]$, den Kreis schneidet, in dem für ein bestimmtes M_i (3.15) gilt, d.h., ob

$$(x-x_i)^2 + (a_0^* + a_1^* x - y_i)^2 = r_i^2 \,(1-\alpha) \tag{3.16}$$

reelle Lösungen hat, die wir mit

$$x_{1i}^{(\alpha)} \,,\; x_{2i}^{(\alpha)} = \varkappa_i^{(\alpha)} \pm \sqrt{\lambda_i^{(\alpha)}(a_0^*, a_1^*)} \tag{3.17}$$

bezeichnen. So erhalten wir

$$g([0,1] \cap F_\alpha) = 2 \sum_{i=1}^{4} \sqrt{\left[\lambda_i^{(\alpha)}(a_0^*, a_1^*)\right]^+} \,, \tag{3.18}$$

wenn wir als Fuzzy-Maß zu Vergleichszwecken und der Einfachheit halber wieder die gleichmäßige Verteilung wählen. Wegen der Monotonie bezüglich α liefert die Lösung von

$$g([0,1] \cap F_\alpha) = \alpha \tag{3.19}$$

das gewünschte Integral (3.14). Im Bild 2 sehen wir die Niveaulinien von $\mu_F(a_0, a_1)$ für $\alpha = 0{,}30;\ 0{,}35;\ 0{,}40$.

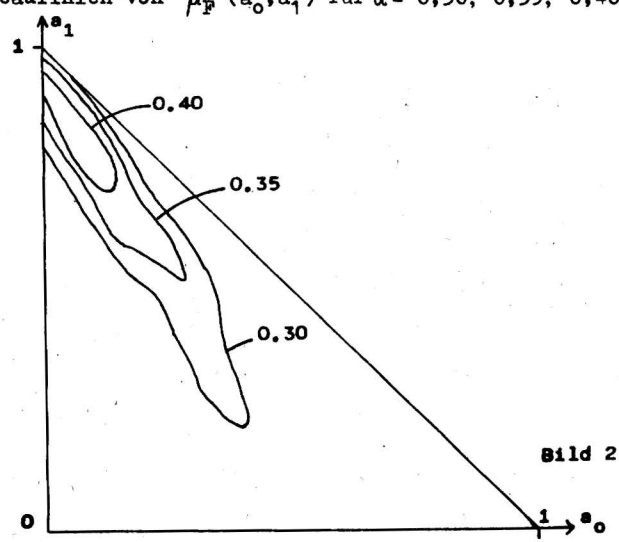


Bild 2

3.3. Fuzzyfizieren des Gültigkeitsbegriffes

Wenn wir sagen, daß eine funktionale Beziehung f in einer Menge M_i gilt, falls es einen Punkt gibt, in dem sie gilt, dann bedeutet dies, daß der Graph der funktionalen Beziehung die Menge M_i trifft. Diese Aussage läßt sich fuzzyfizieren, wenn M_i eine Fuzzy-Menge mit der Zugehörigkeitsfunktion μ_{M_i} ist, indem wir mit

$$\mu_G(f\ ,\ M_i) = \sup_{x \in X}\ \mu_{M_i}(x,\ f(x)) \tag{3.20}$$

den Fuzzy-Gültigkeitsgrad von f bezüglich M_i einführen. Im Unterschied zum Vorgehen in den Abschnitten 2.1 und 2.2 werden die Fuzzy-Beobachtungen M_i nicht vereinigt, sondern ihre Fuzzy-Gültigkeitsgrade getrennt betrachtet.

Entsprechend der Vorstellung, daß die funktionale Beziehung in jeder Beobachtung im Fuzzy-Sinne gelten soll, erhalten wir als sinnvolle Begriffsbildung den Fuzzy-Gültigkeitsgrad von f bezüglich $M_1,\ldots,M_n$ durch

$$\mu_{G,n}\ (f\ ;\ M_1,\ldots,M_n) = \min_i \sup_{x \in X}\ \mu_{M_i}(x, f(x)). \tag{3.21}$$

Betrachten wir unser Beispiel für $f(x) = a_0 + a_1 x$, $x \in [0,1]$! Nehmen wir die Gerade in ihrer Hesseschen Normalform

$$x \cos\varphi + y \sin\varphi - p = 0, \tag{3.22}$$

so liefert bekanntlich

$$x_0 \cos\varphi + y_0 \sin\varphi - p = r_0 \tag{3.23}$$

den Abstand r_0 eines Punktes (x_0, y_0) von dieser Geraden. Wegen der Kreissymmetrie von μ_{M_i} gilt

$$\mu_i(a_0, a_1) = \sup_{x \in X}\ \mu_{M_i}(x, a_0 + a_1 x) = [1 - r_0^2 / r_i^2]^+, \tag{3.24}$$

wobei r_0^2 das Abstandsquadrat der Geraden vom Kreismittelpunkt (x_i, y_i) gemäß (3.23) ist. Aus (3.24) erhält man dann

$$\mu_{G,4}(a_0, a_1;\ M_1,\ M_2,\ M_3,\ M_4) = \min_i\ \mu_i(a_0, a_1) \tag{3.25}$$

Im Bild 3 sehen wir die Niveaulinien von $\mu_{G,4}$ für $\alpha = 0.0;\ 0.5;\ 0.8$.

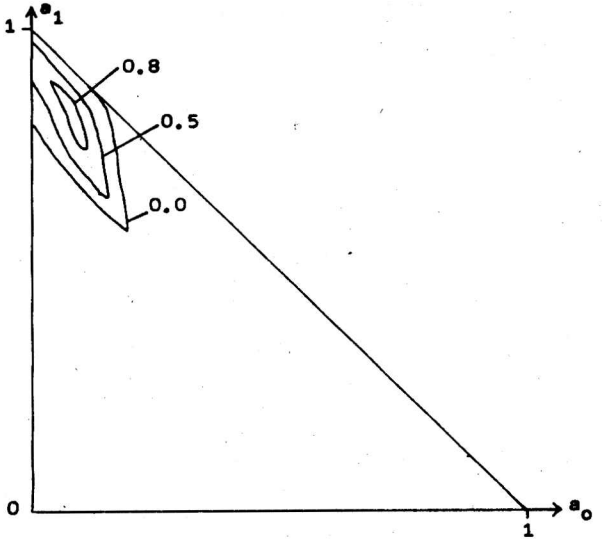


Bild 3

3.4. Einige Bemerkungen

1. In einer anderen Arbeit (Schmerling und Bandemer /8/) werden statt der Kugelkappen Pyramidenstümpfe verwendet. Die Ergebnisse sind ähnlich; zur Berechnung der Fuzzy-Mengen im Parameterraum lassen sich Methoden der Intervallarithmetik heranziehen.

2. Die Fuzzy-Mengen im Parameterraum haben Ähnlichkeit mit der Angabe von Konfidenzschätzungen zu verschiedenen Niveaus beim gleichen Datenmaterial. Die Deutung, im stochastischen Modell meist eine frequentistische, ist hier jedoch eine andere; sie entspricht dem Fuzzy-Begriff, der bei der Spezifizierung der Beobachtungen und bei deren Transformation verwendet wurde. Ein Vergleich von Fuzzy- und Konfidenzschätzung findet man in Bandemer /9/.

3. Den Unabhängigkeitsbegriff für die Beobachtungen, der in der Statistik eine so große Rolle spielt, haben wir bei der Einführung der Fuzzy-Beobachtungen nicht benutzt. Vorkenntnisse und Annahmen über gewisse Beziehungen zwischen den Beobachtungen (z.B. gleiche Quellen, gleichzeitige Spezifizierung) lassen sich gegebenenfalls durch die Art der Zusammenfassung berücksichtigen.

4. Die Annahme, daß die funktionale Beziehung explizit vorliegt, läßt sich umgehen. In einer Arbeit Bandemer und Gerlach /10/ werden die Methoden, die in der vorliegenden Arbeit vorgestellt wurden, auf implizite Beziehungen ausgedehnt. Die von einigen Methoden verlangten Maße müssen dann für $X \times Y$ spezifiziert werden, und der numerische Aufwand für die Berechnung der Fuzzy-Mengen im Parameterbereich ist verständlicherweise wesentlich höher.

4. Ein Analogon zur empirischen Regression

Im vorangehenden Abschnitt wurde davon ausgegangen, daß die Struktur der funktionalen Beziehung bis auf einen Vektor unbekannter Parameter bekannt sei. Halten wir allein die Voraussetzung aufrecht, daß die explizite funktionale Beziehung eindeutig und fest, aber unbekannt ist, dann läßt sich eine punktweise "Schätzung" für die Beziehung aus den Fuzzy-Beobachtungen in einfacher Weise angeben, die man als ein Analogon zur empirischen Regression ansehen kann. Als Funktionswert der Schätzung wählt man für jedes $x \in X$ den mit dem höchsten Zugehörigkeitsgrad der vereinigten Beobachtungen, d.h., sei μ_M die Zugehörigkeitsfunktion der Fuzzy-Beobachtung $M = \bigvee_i M_i$, dann ist

$$\hat{y} = \hat{f}(x) = \arg \left\{ \sup_{y \in Y} \mu_M(x,y) \right\} . \qquad (4.1)$$

Zu Fuzzy-Aussagen über diese Schätzung kann man dann z.B. durch eine der folgenden Methoden kommen:

a) Man bestimmt für eine Folge $d_1, d_2, \ldots, d_N$ die erwartete Kardinalität (oder den Fuzzy-Erwartungswert) bezüglich eines gegebenen Maßes längs der Graphen der folgenden Funktionen

$$\hat{f}(x) - d_N,\ \hat{f}(x) - d_{N-1},\ldots,\hat{f}(x)-d_1,\ \hat{f}(x),\ldots,\hat{f}(x)+d_N \qquad (4.2)$$

und erhält so Aussagen über eine Funktionsschicht mit dem Abstandsparameter d.

b) Man bestimmt für eine Folge $m_1,\ldots,m_L$

$$\hat{y}_{+i} = \hat{f}_{+i}(x) = \arg\sup_y \{\mu_M(x,y) := \mu_M(x,\hat{f}(x) - m_i)\}, \qquad (4.3)$$

$$\hat{y}_{-i} = \hat{f}_{-i}(x) = \arg\inf_y \{\mu_M(x,y) := \mu_M(x,\hat{f}(x) - m_i)\} \qquad (4.4)$$

und erhält so lokale Aussagen über die Schwankungsbreite der vermuteten Lage von f(x) durch

$$\hat{f}_{-i}(x) \leq f(x) \leq \hat{f}_{+i}(x). \qquad (4.5)$$

Damit dieses Vorgehen insgesamt sinnvoll wird, muß sich die Fuzzy-Beobachtung über den gesamten Bereich X erstrecken. Ist $\mu_M(x,y)$ für ein Teilintervall X_i identisch gleich Null, dann sind dafür keine Voraussagen möglich, wenn man sich nicht zu einer zusätzlichen Interpolationsregel entschließen will.

Es kann vorkommen, daß die Schätzung gemäß (4.1) nicht zu einem eindeutigen Ergebnis führt. Dann steht die Fuzzy-Beobachtung im Widerspruch zur Annahme einer eindeutigen und festen funktionalen Beziehung. Die Lösung dieses Widerspruchs, Änderung der Beobachtung oder Annahme einer impliziten Beziehung, muß im Zuge einer Reformulierung des Modells erfolgen.

Wenn man von einer hinreichend großen Anzahl "exakt" angegebener Beobachtungspunkte ausgeht und die Fuzzy-Beobachtungen gemäß (2.3) spezifiziert, erhält man ein mit der empirischen Regression nach Schmerling und Peil /11/ vergleichbares Verfahren. Der Radius des Strukturelementes B spielt dann eine ähnliche Rolle wie der Wirkungsradius h der Einzelbeobachtung dort.

Literatur

/1/ Tanaka, H., Uesima, S., and Asai, K.: Linear regression analysis with fuzzy model. IEEE Trans. Systems Man Cybernet. 12, 903 - 907 (1982)

/2/ Serra, J.: Image Analysis and Mathematical Morphology. London 1982

/3/ Thiele, H.: Die Astro-Cumulus-Methode. Persönliche Mitteilung, Dezember 1982

/4/ Bandemer, H.: Evaluating explicit functional relationships from fuzzy observations. Submitted to: Fuzzy Sets and Systems

/5/ Bandemer, H., and Schmerling, S.: Evaluating an explicit functional relationship by fuzzyfying the statement of its satisfying. Submitted to: Biometrical J.

/6/ Dubois, D., and Prade, H.: Fuzzy Sets and Systems: Theory and Application. New York 1980

/7/ Sugeno, M.: Theory of fuzzy integrals and its application. Dissertation, Tokyo Institute of Technology, Tokyo 1974

/8/ Schmerling, S., and Bandemer, H.: Transformation of fuzzy data of interval type into the parameter region of explicit functional relationships. Freiberger Forschungshefte Ser. D, Leipzig 1985 (in print)

/9/ Bandemer, H.: Einige Gedanken zur Regression mit Fuzzy-Daten. Vortrag auf der Fachtagung Qualitätsanalyse, Frankfurt/O 1983 (als Mikrofiche verfügbar bei KdT Frankfurt/O)

/10/ Bandemer, H., and Gerlach, W.: Evaluating implicit functional relationships from fuzzy observations. Freiberger Forschungshefte Ser. D, Leipzig 1985 (in print)

/11/ Schmerling, S., and Peil, J.: Some procedures of the empirical regression demonstrated by examples. Wiss. Z. Martin-Luther-Univ. Halle-Wittenberg Math.-Natur. Reihe 31, 119 - 129 (1982)

eingegangen: 13. 04. 1984

Anschrift des Verfassers:

Prof. Dr. H. Bandemer
Bergakademie Freiberg
Sektion Mathematik
PF 47

DDR-9200 Freiberg

Rostock. Math. Kolloq. 26, 83 - 88 (1984)

65007
65005
65010

Burkhard Meeß
Gerhard Meeß

Interpolating quadratic splines with norm-minimal curvature

Abstract

An interpolating polynomial spline of degree 2 is uniquely determined by interpolation requirements $p(x_n) = f_n$, $n = 0(1)N$, continuity conditions for the first derivative $p'(x_n-0) = p'(x_n+0)$, $n = 1(1)N-1$, and one boundary condition. In this paper we determine an initial condition $p'(x_0) =: d_0$ such that the resulting spline has a minimal second derivative or a minimal curvature in the L_2-norm. This spline is cheaper to compute and for practical purposes as good as the corresponding natural cubic spline (or even better).

1. Introduction

Let be given an ordered set of nodes x_n, $n = 0(1)N$,

$$x_0 < x_1 < \dots < x_N$$

and a set of corresponding values f_n, $n = 0(1)N$. By $p(x)$ we denote an interpolating quadratic spline function, i. e. the union

$$p(x) := \bigcup_{n=1}^{N} p_{2,n}(x) \tag{1}$$

of piecewise defined quadratic polynomials

$$p_{2,n}(x_{n-1}+\vartheta h_n) = f_{n-1} + h_n d_{n-1}\vartheta + (g_n - d_{n-1})h_n\vartheta^2, \quad n = 1(1)N, \quad 0 \leq \vartheta \leq 1. \tag{2}$$

Here we used the abbreviations

$$h_n := x_n - x_{n-1}, \quad g_n := (f_n - f_{n-1})/h_n.$$

The d_n are unknown parameters. The condition of smoothness at the inner nodes

$$p'_{2,n}(x_{n-1}+h_n) = p'_{2,n+1}(x_n), \quad n = 1(1)N-1,$$

yields N-1 equations for the N parameters d_n, $n = 0(1)N-1$,

$$d_{n-1} + d_n = 2g_n, \quad n = 1(1)N-1. \tag{3}$$

Usually there is added a boundary or a periodicity or an anti-periodicity condition:

$$\begin{aligned} &p'(x_o) = f'_o \quad \text{or} \quad p'(x_N) = f'_N \quad \text{or} \quad p'(x_o) = p'(x_N), \quad N \text{ odd}, \\ &\qquad\qquad \text{or} \quad p'(x_o) = -p'(x_N), \quad N \text{ even}. \end{aligned} \tag{4}$$

It is well-known (cf./1/) that in each of the four cases the d_n are determined uniquely.

If none of the four assumptions (4) is motivated by the physical background of the interpolation problem, one can use a suitable optimality condition instead. So one can try to minimize the second derivative $p''(x)$ or the curvature $p''(x)/(1+p'(x)^2)^{3/2}$ in a certain norm.

2. Quadratic splines with norm-minimal second derivative

If we use the L_2-norm, we get

$$Q := \int_{x_o}^{x_N} (p''(x))^2 dx = \sum_{n=1}^{N} h_n \int_0^1 (p''_{2,n}(x_{n-1}+sh_n))^2 ds. \tag{5}$$

The second derivatives of the polynomials $p_{2,n}(x)$ are constant,

$$p''_{2,n}(x) = 2t_n/h_n, \quad t_n := g_n - d_{n-1}, \quad n = 1(1)N.$$

If $t := t_1$ is a free parameter, the t_n are linear functions of t and can be determined recursively (cf. (3)):

$$t_{n+1} = t_n + g_{n+1} - g_n. \tag{6}$$

So Q is a quadratic function of t, the minimum of which is easily computed by

$$\frac{dQ}{dt} = 8 \sum_{n=1}^{N} \frac{t_n}{h_n} \frac{dt_n}{dt} = 8 \sum_{n=1}^{N} (-1)^{n-1} \frac{t_n}{h_n} = 0.$$

With the abbreviations

$$a_j := \sum_{n=j}^{N} 1/h_n, \quad j = 1(1)N, \tag{7}$$

we get the minimizer

$$t_{opt} = \frac{1}{a_1} \sum_{j=2}^{N} (-1)^j a_j (g_j - g_{j-1}) \tag{8}$$

and finally the initial slope of the interpolating spline

$$p'(x_o) = d_o = g_1 - t_{opt}. \tag{9}$$

The computing costs are about 3N multiplicative floating point operations for the determining of the d_n and two multiplications for each computed value p(x). (In the special case of equidistant nodes the parameters can be computed even without any multiplicative operation.) So quadratic splines with norm-minimal second derivative are much cheaper than cubic splines.

3. Quadratic splines with norm-minimal curvature

If we want to minimize the L_2-norm of the curvature, the integrands in the sum on the right hand side of (5) have to be replaced by $(p''(x))^2/(1+p'(x)^2)^3$ and are no longer constant. We therefore use the discretization $p'(x_{n-1}+sh_n) \approx g_n$ and get the function $\tilde{Q}$ to be minimized instead of Q:

$$\tilde{Q} := 4 \sum_{n=1}^{N} t_n^2/(h_n(1+g_n^2)^3). \tag{10}$$

So the result (8), (9) remains valid for this case too. We have only to replace (7) by

$$a_j := \sum_{n=j}^{N} 1/(h_n(1+g_n^2)^3). \tag{11}$$

For small g_n we have, of course, approximately the old result.

4. Two examples

Example 1: For a ship rib a designer wants a graph joining smoothly the following data

x_n	0	2	5	7	9	12	15	[m]
f_n	0.60	1.40	2.00	3.40	6.40	10.00	11.00	[m]

From formulae (11), (8) we get $t_{opt} = 0.25549$ and by (9) and (3) further $d_0 = 0.6555$, $d_1 = 0.1445$, $d_2 = 0.2555$, $d_3 = 1.1445$, $d_4 = 1,8555$, $d_5 = 0.5445$. So the components of the quadratic spline p(x) have the representation

$$p_{2,1}(x) = 0.60 + 0.6555\,x - 0.1277\,x^2, \quad 0 \le x \le 2,$$
$$p_{2,2}(x) = 1.40 + 0.1445\,(x-2) + 0.0185\,(x-2)^2, \quad 2 \le x \le 5,$$
$$p_{2,3}(x) = 2.00 + 0.2555\,(x-5) + 0.2223\,(x-5)^2, \quad 5 \le x \le 7,$$
$$p_{2,4}(x) = 3.40 + 1.1445\,(x-7) + 0.1777\,(x-7)^2, \quad 7 \le x \le 9,$$
$$p_{2,5}(x) = 6.40 + 1.8555\,(x-9) - 0.2185\,(x-9)^2, \quad 9 \le x \le 12,$$
$$p_{2,6}(x) = 10.00 + 0.5445\,(x-12) - 0.07039\,(x-12)^2, \quad 12 \le x \le 15.$$

With only two multiplications now the value p(x) can be computed for arbitrary arguments x (cf. Figure 1). There is no

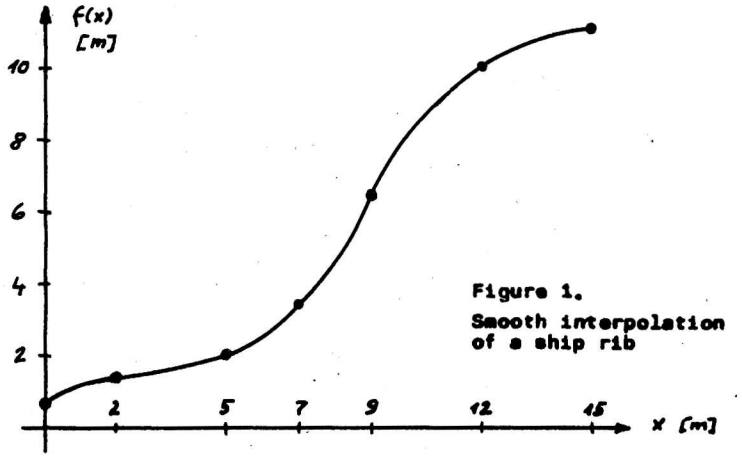


Figure 1.
Smooth interpolation of a ship rib

significant difference to the corresponding natural cubic spline.

Example 2: We interpolate the five values

x_n	0	1	2	3	4
f_n	1	1	2	6	24

of the function n factorial by quadratic splines $p_{2,n}(x)$ with norm-minimal curvature and by natural cubic splines $p_{3,n}(x)$. The results are

$$p_{2,1}(x) = 1 - 0.1107\,x + 0.1107\,x^2, \quad 0 \le x \le 1,$$
$$p_{2,2}(x) = 1 + 0.1107\,(x-1) + 0.8893\,(x-1)^2, \quad 1 \le x \le 2,$$
$$p_{2,3}(x) = 2 + 1.8893\,(x-2) + 2.1107\,(x-2)^2, \quad 2 \le x \le 3,$$
$$p_{2,4}(x) = 6 + 6.1107\,(x-3) + 11.8893\,(x-3)^2, \quad 3 \le x \le 4,$$

and

$$p_{3,1}(x) = 1 - 0.3036\,x + 0.3036\,x^3, \quad 0 \le x \le 1,$$
$$p_{3,2}(x) = 1 + 0.6071(x-1) + 0.9107(x-1)^2 - 0.5179(x-1)^3, \quad 1 \le x \le 2,$$
$$p_{3,3}(x) = 2 + 0.8750(x-2) - 0.6429(x-2)^2 + 3.7679(x-2)^3, \quad 2 \le x \le 3,$$
$$p_{3,4}(x) = 6 + 10.893\,(x-3) + 10.661\,(x-3)^2 - 3.5536(x-3)^3, \quad 3 \le x \le 4,$$

respectively. Mathematically the cubic spline is smoother since it has a continuous second derivative, but Figure 2 shows that the quadratic spline is "optically" smoother, its "total curvature" is smaller. This can be shown analytically too, if we use the approximative formula

$$\tilde{\tilde{Q}} := \sum_{n=1}^{N} \frac{1}{(1+g_n'^2)^3} \int_{x_{n-1}}^{x_n} (p''_{3,n}(x))^2\,dx \qquad (12)$$

for the total curvature of the cubic spline. For the quadratic spline $p_2(x)$ we get $\tilde{Q} = 0.448$ whereas (12) yields $\tilde{\tilde{Q}} = 1.245$.

5. Remarks

Quadratic splines are well suited to interpolate convex data. In the non-convex case, turning points lie at the nodes only. It is not difficult to use other norms, so as the L_1 or the

L_{∞}-norm. Quadratic splines with norm-minimal curvature are suitable to higher-dimensional interpolation too.

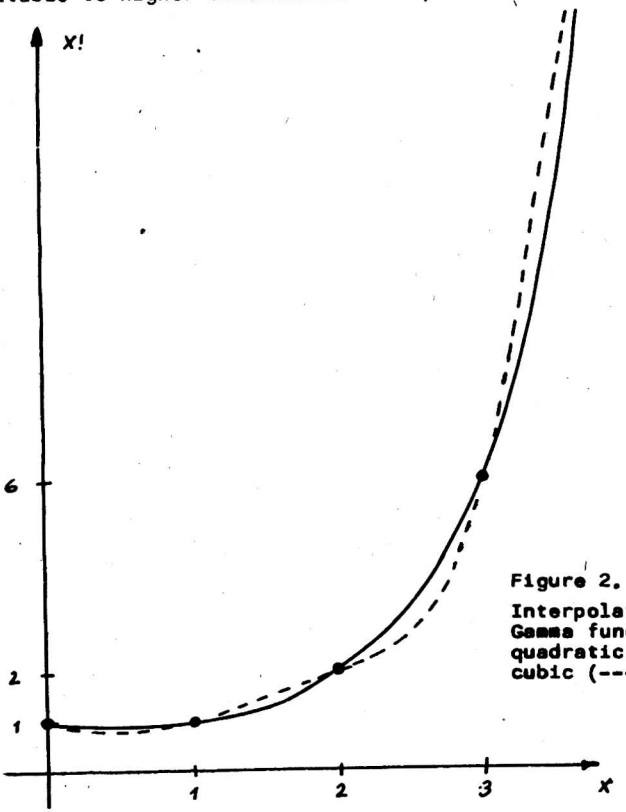


Figure 2.
Interpolation of the Gamma function by quadratic and cubic (---) splines

References

/1/ Karlin, S., and A. Pinkus: Interpolation by splines with mixed boundary conditions. In: Karlin, S., et al. (Eds.): Studies in Spline Functions and Approximation Theory. pp. 305 - 325. New York 1976

received: January 16, 1984

Authors addresses:

stud. phys. B. Maeß
Karl-Marx-Universität Leipzig
Sektion Physik
Linnéstraße 5
DDR-7010 Leipzig

Prof. Dr. G. Maeß
Wilhelm-Pieck-Universität Rostock
Sektion Mathematik
Universitätsplatz 1
DDR-2500 Rostock

Rostock. Math. Kolloq. 26, 89 - 102 (1984) 65N30

Wolfgang Moldenhauer

A k-Pascal triangle in the finite element method for the solution of elliptic boundary value problems of the second order[1]

1. Introduction

There are known a lot of triangular elements for the solution of boundary value problems with the finite element method. In this paper a so called k-Pascal triangle is constructed. Using the analysis of Ciarlet and Raviart /1/ estimates are generated for the approximation error. An elliptic boundary value problem of the second order is solved by means of k-Pascal triangles and error estimates in Sobolev spaces are derived. The numerical algorithm is described and the stiffness matrices are calculated. Numerical results are given in the testing example "torsion of a square".

2. Notations and preliminaries

Let k be a positive intager and let P_k denote the space of polynomials of degree $\leq k$ over R^2. It is $N = N(k) = \dim P_k = \binom{k+2}{2}$. Let l, m, n be integers such that

$$0 \leq l, m, n \quad \text{and} \quad l + m + n = k. \tag{0}$$

The conditions (0) are satisfied for exactly $\binom{k+2}{2}$ triples (l,m,n). Therefore N points of R^2 are collected such that the function values in these points determine uniquely a polynomial of P_k. This suggests the following definition.

Definition 1: N disjunct points $a_i \in R^2$ $(1 \leq i \leq N)$ form a k-unisolvant set $\{a_i\}_{i=1}^N$ iff for any given real numbers α_i, $1 \leq i \leq N$, there exists exactly one polynomial $p \in P_k$ such that $p(a_i) = \alpha_i$ $(1 \leq i \leq N)$ holds. (For a similar definition see /1/.)

---

1 The work on this paper was done during my stay at the Wilhelm-Pieck-University. I would like to thank all colleagues of the department of mathematics for their support.

For any k-unisolvent set a fixed reference k-unisolvent set is considered. For a triangle T with the nodes $P_i(x_i,y_i)$, i=1,2,3, the union triangle T_0 with nodes (0,0), (1,0) and (0,1) will be the reference triangle such that $T = \Phi(T_0)$.
The affine transformation Φ is given by:

$x = x_0(\xi,\eta) = x_1 + (x_2-x_1)\xi + (x_3-x_1)\eta,$

$y = y_0(\xi,\eta) = y_1 + (y_2-y_1)\xi + (y_3-y_1)\eta$ or shorter

$X = A\overline{X} + B$ with $X = (x,y)^T$, $\overline{X} = (\xi,\eta)^T$, $B = (x_1,y_1)^T$, and

$$A = \begin{bmatrix} x_2 - x_1 & x_3 - x_1 \\ y_2 - y_1 & y_3 - y_1 \end{bmatrix}.$$

The Jacobian of Φ satisfies $A = J(\Phi) = 2mes(T) \neq 0$. Let be h_T the greatest side and ϑ_0 the smallest angle of the triangle T. Now it is possible to give sharper bounds for the Jacobian.

Lemma 1: There are constants $c_1, c_2 > 0$ such that

$$c_1 h_T^2 \leq J(\Phi) \leq c_2 h_T^2 \quad (see/2/).$$

Remarks:

a) The best possible constants are $c_1 = \frac{1}{2}\sin\vartheta_0$ and $c_2 = \frac{1}{2}\sqrt{3}$.

b) From Lemma 1 and the explanations stated above it follows that the inverse transformation Φ^{-1} exists.

3. The k-Pascal triangle

Let $P_i(x_i,y_i)$, i = 1,2,3, be three given points of R^2 not lying on a straight line. The points $P_{l,m,n}$ are constructed such that $P_{l,m,n} = (k^{-1}(lx_1+mx_2+nx_3); k^{-1}(ly_1+my_2+ny_3))$ with l, m, n explained in conditions (0). The following definition is suggested from the geometrical point of view.

Definition 2: The set $P_{l,m,n}$, $0 \leq l,m,n$, $l+m+n = k$ is called a k-Pascal triangle (or k-Pascal element[2]).

---

2 The notation "element" is used for the triangle and for the basic functions defined on the triangle as well.

Remark: For k = 1 the linear triangular element, for k = 2 the quadratic triangular element, and for k = 3 a cubic triangular element are contained in the definition (cf. /3/).

The following result allows to apply the techniques of Ciarlet and Raviart /1/.

Theorem 1: A k-Pascal triangle is k-unisolvent.

Proof: It is sufficient to prove that the function values α_i, $1 \leq i \leq \binom{k+2}{2}$, in the points $P_{l,m,n}$ determine uniquely the N coefficients of a polynomial $p \in P_k$. Therefore it is sufficient to show that the homogeneous system $p(P_{l,m,n}) = 0$, $0 \leq l,m,n$, $l + m + n = k$, has only the trivial solution $p \equiv 0$. Let $\Phi^{-1}(P_{l,m,n}) = R_{l,m,n}$ and $r(\xi,\eta) = p[x_o(\xi,\eta), y_o(\xi,\eta)]$. Then it holds $r \in P_k$ and $r(R_{l,m,n}) = 0$ for all $0 \leq l,m,n, l + m + n = k$.

Now $r(0,\eta)$ is a polynomial of degree k in η which has k+1 roots in $R_{i,o,k-i}$, $0 \leq i \leq k$. It follows $r(0,\eta) \equiv 0$ and therefore we have $r = \xi q_{k-1}(\xi,\eta)$ with $q_{k-1} \in P_{k-1}$ and $q_{k-1}(R_{l,m,n}) = 0$ for all $0 \leq l,n$, $1 \leq m$, $l+m+n = k$. Now $q_{k-1}(k^{-1},\eta)$ is a polynomial of degree k-1 in η with k roots in the points $R_{i,1,k-1-i}$, $0 \leq i \leq k-1$. Hence it holds $r(k^{-1},\eta) = 0$, and $r = \xi q_{k-1} = \xi(\xi - k^{-1})q_{k-2}$ with $q_{k-2} \in P_{k-2}$, and $q_{k-2}(R_{l,m,n}) = 0$ for all $0 \leq l,n$, $2 \leq m$, $l+m+n = k$. Similarly it follows

$$r = \xi(\xi-k^{-1})(\xi-2k^{-1}) \ldots (\xi-jk^{-1})q_{k-j-1} \qquad (1)$$

with $q_{k-j-1} \in P_{k-j-1}$ and $q_{k-j-1}(R_{l,m,n}) = 0$ for all $0 \leq l,n$, $j+1 \leq m$, $l+m+n = k$. However, for $j = k-1$ it follows $q_o \in P_o$ and $q_o(R_{o,k,o}) = 0$. Hence, it is $q_o = 0$. By means of relation (1) one concludes $r \equiv 0$.

Let $\Omega \subset R^2$ be a bounded convex domain with the boundary Γ. As usual in the finite element method one has to triangulate the domain Ω.

Definition 3: A domain Ω is called triangulated iff $\bar{\Omega} = \Omega \cup \Gamma$ is divided in a finite number of triangles with the following properties: 1. The union of all triangles is Ω. 2. Any two triangles are either disjunct or have a common node or a common side.

Each triangulation has two significant parameters: h will denote the greatest side and ϑ the smallest angle of all triangles of the triangulation.

Lemma 2: A k-Pascal triangle generates functions of $C^0(\bar{\Omega})$, but not of $C^1(\bar{\Omega})$.

Remark: Since the functions defined over a k-Pascal triangle are piecewise polynomials over Ω they do not belong to $H^2 = W_2^{(2)}(\Omega)$. The norm $\|\cdot\|_{H^m} = \|\cdot\|_m = \|\cdot\|_{m,2}$ is given by $\|v\|_m^2 = \sum_{|\alpha| \leq m} \int_\Omega (D^\alpha v)^2 dx$. Therefore a k-Pascal triangle cannot be used for the solution of problems of higher than the second order.

To receive estimates for the approximation error the following notations are introduced.
Let the function u in $a \in R^2$ be k-times differentiable. The k-th (Fréchet) derivative is denoted by $D^k u(a)$, and let $\mathcal{L}_k(R^2,R)$ be the space of the k-linear mappings $A_k: \overset{k}{\underset{1=1}{X}} R^2 \to R$. Obviously $D^k u(a)$ is a symmetrical element of the space $\mathcal{L}_k(R^2,R)$. Its norm is given by

$$\|D^k u(a)\| = \sup_{\substack{\|h_i\| \leq 1 \\ 1 \leq i \leq k}} \|D^k u(a)\cdot(h_1,h_2,\ldots,h_k)\|.$$

For any partial derivative $\partial^\alpha u(a)$ with $|\alpha| = k$ and any function u there are constants C(k) such that

$$|\partial^\alpha u(a)| \leq \|D^k u(a)\| \leq C(k) \max_{|\alpha|=k} |\partial^\alpha u(a)|$$

holds, if the R^2 is equipped with the Euclidian norm. Let u be

defined on a k-unisolvent set $\sum = \{a_i\}_{i=1}^{N}$.

Definition 4: The function $\tilde{u}$ is called an interpolating polynomial of the function u iff $\tilde{u} \in P_k$ and $\tilde{u}(a_i) = u(a_i)$ for all i with $1 \leq i \leq N$.

The closed convex hull of the set $\sum$ is denoted by $K = K(\sum)$. The following theorem will be helpful to get estimates for the approximation error of the interpolating polynomial $\tilde{u}$.

Theorem 2: Let be $k \in N$, $k \geq 1$, and $\sum$ a k-unisolvent set of R^2. Further let $u \in C^k(\bar{K}) \cap C^{k+1}(K)$ be given. Then for any $x \in K$ and any m with $0 \leq m \leq k$ it holds

$$D^m[\tilde{u}(x)-u(x)] = \frac{1}{(k+1)!} \sum_{i=1}^{N} \{D^{k+1}u[\eta_i(x)](a_i-x)^{k+1}\} D^m p_i(x)$$

with $\eta_i(x) = \Theta_i x + (1-\Theta_i)a_i$, $0 < \Theta_i < 1$,

and the polynomials $p_i \in P_k$ are uniquely determined by

$p_i(a_j) = \delta_{ij}$, $1 \leq i,j \leq N$.

This theorem follows from Theorem 1 in /1/.

Now it is possible to give error estimates, since

$|a_i - x|^{k+1} \leq h_T^{k+1} \leq h^{k+1}$ holds.

Theorem 3: Let T be a given k-Pascal triangle and $u \in C^k(\bar{T}) \cap C^{k+1}(T)$ a given function. It is assumed that $M_{k+1} = \sup\{\|D^{k+1}u\|,\ u \in T\} < \infty$. Further let be $\tilde{u} \in P_k$ the interpolating polynomial of u. Then for any m, $0 \leq m \leq k$, it holds

$$\sup\{\|D^m[u(x)-\tilde{u}(x)]\|,\ x \in T\} \leq \frac{2^{\frac{3}{2}m}}{(k+1)!} \sin^{-m}\vartheta_o\, h_T^{k+1-m} M_{k+1} C_o$$

with $C_o = \sum_{i=1}^{N} \sup\{\|D^m \tilde{p}_i(\tilde{x})\|,\ \tilde{x} \in T_o\}$ and $\tilde{p} = \Phi^{-1}p$.

The proof is realizable with the methods contained in /1/. Using a generalized Markov inequality (cf. /4/) it is possible to establish estimates for the constants C_o.

For $p \in P_k$ and any compact convex subset $K \subset R^2$ it holds

$$\sup\{\|Dp(x)\|,\ x \in K\} \leq 2Ck^2 \varrho^{-1}(K) \sup\{|p(x)|,\ x \in K\}$$

in which $\varrho(K)$ denotes the supremum of the diameters of all K inscribed spheres. By means of this fact the following lemma is established.

Lemma 3: Let be $p \in P_k$ defined over T_o. Then for any m, $0 \leq m < k$, it holds $\sup\{\|D^m p(x)\|,\ x \in T_o\} \leq C_1 \sup\{|p(x)|,\ x \in T_o\}$ with $C_1 = 2^m(2+\sqrt{2})^m C^m k^2(k-1)^2 \ldots (k-m+1)^2$.

4. Some applications

Let V be a Hilbert space with $H_o^m \subset V \subset H^m$. The norm is induced by H^m. We consider a continuous bilinear form a(u,v) on $V \times V$, that means that the mapping $a: V \times V \to R$ is linear in u and v and bounded such that $|a(u,v)| \leq M \|u\|_m \|v\|_m$ $(u,v \in V, M = \text{const.} > 0)$.

L(v) is a continuous functional on V. Now one concludes the following problem:

Determine a function $u \in V$, such that

$$a(u,v) = L(v) \qquad (2)$$

holds for all $v \in V$.

Let a(u,v) be V-elliptic, that means $a(u,v) \geq \alpha \|v\|_m^2$ $(\forall v \in V, \alpha = \text{const.} > 0)$, then the problem (2) has a unique solution (cf. /5/). If the additional assumption a(u,v) = a(v,u) for all $u,v \in V$ is fulfilled then u is the solution of (2) iff u minimizes the energy functional

$$\Phi(v) = \frac{1}{2} a(u,v) - L(v) \qquad (3)$$

over V (cf. /3/).

Let $\Omega \subset R^2$ be a bounded domain with the convex boundary $\Gamma = \partial\Omega$. Now the following problem is considered:

$$-\Delta u = f \text{ in } \Omega,$$

$$\sigma u + \frac{\partial u}{\partial \nu} = g \text{ on } \Gamma \text{ with } \sigma \not\equiv 0 \text{ but } \sigma \geq 0.$$

We choose $V = H^1$, $a(u,v) = \int_\Omega (u_x v_x + u_y v_y)dxdy + \int_\Gamma \sigma uv d\Gamma$ and $L(v) = (f,v)_o + \int_\Gamma gv d\Gamma$.

Lemma 4: $a(u,v)$ is V-elliptic.

Proof: Because of $\sigma \not\equiv 0$, $\sigma \geq 0$ there is $\Gamma' \subset \Gamma$ with $\sigma > 0$ on Γ'. Choosing Γ'' such that $\Gamma'' \subset \Gamma'$ it follows $\inf_{\Gamma''} \sigma > 0$. Then

$$a(u,v) = \int_\Omega (u_x^2+u_y^2)dxdy + \int_\Gamma \sigma u^2 d\Gamma \geq \int_\Omega (u_x^2+u_y^2)dxdy$$

$$+ \inf_{\Gamma''} \sigma \int_{\Gamma''} u^2 d\Gamma'' \geq \min(1, \inf_{\Gamma''} \sigma)\left(\int_\Omega (u_x^2+u_y^2)dxdy + \int_{\Gamma''} u^2 d\Gamma''\right)$$

$$\geq c^{-2}\min(1, \inf_{\Gamma''} \sigma)\, |u|^2_{W_2^{(1)}(\Omega)} = \alpha \|u\|_1^2$$

where the last estimate follows from the generalized Friedrich's inequality (cf. /5/). Let h and ϑ be the parameters of the triangulation of Ω and let V_h^k denote the finite dimensional subspace of V belonging to piecewise polynomials of k-th degree. Now the following discretization problem of (2) is considered:

Find a function $u_h \in V_h^k$, such that

$$a(u_h,v) = L(v) \qquad (4)$$

holds for all $v \in V_h$.

Under the assumptions stated above the approximate solution u_h satisfies $\|u - u_h\|_1 = o(h)$. If the additional assumption $u \in H^{r+1}$ is fulfilled then it follows $\|u - u_h\|_1 \leq Ch^r \|u\|_{r+1}$ in which the constant C does not depend on h and u (cf. /2/, /3/). (This assumption is unrealistic in general.)
Attract our attention to the following Dirichlet problem (see the numerical example):

$$-\Delta u = f \text{ in } \Omega,\ u = 0 \text{ on } \Gamma. \qquad (5)$$

It is assumed $f \in L_2$. Then it follows $\|u\|_2 \leq C\|f\|_0$ (/6/). Therefore for the subspaces V_h^1 (i. e. the 1-Pascal triangle) we have the estimate $\|u - u_h\|_1 \leq Ch\|f\|_0$. Using an argument of Nitsche /7/ it is possible to give an error estimate in the L_2-norm. Let be $-\Delta v = u - u_h$, $v = 0$ on Γ. Then

$$\|u - u_h\|_0^2 = (u-u_h, -\Delta v)_0 = a(u-u_h, v) = a(u-u_h, v-\bar{v}),$$

since (2) and (4) imply $a(u-u_h, w) = 0$ for all $w \in V_h$. Thereby $\bar{v} \in V_h^1$ denotes the function which equals to u in the nodes.

Now $|a(u-u_h, v-\bar{v})| \leq M\|u-u_h\|_1 \|v-\bar{v}\|_1 \leq Ch\|u-u_h\|_1\|v\|_2$ $\leq Ch\|u-u_h\|_1 \|u-u_h\|_0$ and a division by $\|u-u_h\|_1$ shows $\|u-u_h\|_0 \leq Ch^2\|f\|_0$.

5. A numerical algorithm

The algorithm is similar to the algorithm described in /8/. Therefore only some remarks are necessary. The problem

$$-\Delta u + c_0 u = f \text{ in } \Omega,\ u = 0 \text{ on } \Gamma$$

is considered in a bounded domain $\Omega \subset R^2$ with a convex polygonal boundary Γ. An equivalent formulation reads:
Minimize the energy functional

$$F(u) = \int_\Omega (u_x^2 + u_y^2 + c_0 u^2 - 2fu)\,d\Omega \qquad (6)$$

in the class $V \subset W_2^{(1)}(\Omega)$, these elements fulfil the Dirichlet condition $u = 0$ on Γ. A triangulation of Ω is established and a function in V_h is a k-Pascal triangle. Over the subspace V_h the integral (6) is a sum of the integrals corresponding to each triangle. Now we are giving an algorithm for computing each of such integrals. The basic idea is to compute each of such integrals once for all on T_0. Approximating the functional F(u) by

$$F(p) = \sum_T I(p,T) = \sum_T \int_T (p_x^2+p_y^2+c_0p^2-2fp)dxdy, \qquad (7)$$

where the sum is taken over all triangles of the triangulation and $p|_T \in P_k$, the functional (6) is a quadratic function in all values in the nodes of the k-Pascal triangles. Any integral of (7) is expressed in the form $w^TKw - 2w^T\delta$, where w^T is the vector of unknowns. $K + K^T$ is the element stiffness matrix and the components of the vector δ are determined by $f(x,y)$. The determination of the minimum of (7) leads to the linear algebraic system of equations

$$(K+K^T)w = 2\delta. \qquad (8)$$

Now an algorithm for computation of K and δ is established.

Transforming a triangle T with the nodes $P_i(x_i,y_i)$, i = 1,2,3, by Φ to T_o, it follows

$$I(p,T) = \int_{T_o} (ar_\xi^2+2br_\xi r_\eta+cr_\eta^2+dr^2-2gr)d\xi\,d\eta \qquad (9)$$

with

$a = |J(\Phi)|^{-1}[(x_3-x_1)^2+(y_3-y_1)^2]$, $c = |J(\Phi)|^{-1}[(x_2-x_1)^2+(y_2-y_1)^2]$,

$b = -|J(\Phi)|^{-1}[(x_2-x_1)(x_3-x_1)+(y_2-y_1)(y_3-y_1)]$, $d = |J(\Phi)|c_o$,

$g(\xi,\eta) = |J(\Phi)|f[x_1+(x_2-x_1)\xi+(x_3-x_1)\eta,\ y_1+(y_2-y_1)\xi+(y_3-y_1)\eta]$.

Let be $\alpha^T = (\alpha_i)$, $1 \le i \le N$, then $r(\xi,\eta)$ is defined as $r(\xi,\eta) = \alpha^T z(k)$, where the vector z is given in the following table:

k	$z(k)^T$
1	$(1, \xi, \eta)$
2	$(1, \xi, \eta, \xi^2, \xi\eta, \eta^2)$
3	$(1, \xi, \xi^2, \xi^3, \eta, \eta^2, \eta^3, \xi^2\eta, \xi\eta^2, \xi\eta)$

and so on. Then $r(\xi,\eta)$ is uniquely determined by the following conditions at the points $P_{l,m,n}$:

k = 1: $u_1 = r(0,0)$, $u_2 = r(1,0)$, $u_3 = r(0,1)$.

k = 2: $u_1 = r(0,0)$, $u_2 = r(\frac{1}{2},0)$, $u_3 = r(1,0)$,
$u_4 = r(\frac{1}{2},\frac{1}{2})$, $u_5 = r(0,1)$, $u_6 = r(0,\frac{1}{2})$.

k = 3: $u_1 = r(0,0)$, $u_2 = r(\frac{1}{3},0)$, $u_3 = r(\frac{2}{3},0)$,
$u_4 = r(1,0)$, $u_5 = r(0,\frac{1}{3})$, $u_6 = r(0,\frac{2}{3})$,
$u_7 = r(0,1)$, $u_8 = r(\frac{1}{3},\frac{1}{3})$, $u_9 = r(\frac{2}{3},\frac{1}{3})$,
$u_{10} = r(\frac{1}{3},\frac{2}{3})$.

Let be $w^T = (u_i)$, $1 \le i \le N$, then we have

$$w = S(k)\alpha \quad \text{and} \quad \alpha = S(k)^{-1}w. \qquad (10)$$

Let be $l^T = \frac{\partial}{\partial\xi} z(k)^T$, $m^T = \frac{\partial}{\partial\eta} z(k)^T$ then it follows

$$\int_{T_o} r_\xi^2 d\xi d\eta = w^T A w, \quad \int_{T_o} 2 r_\xi r_\eta d\xi d\eta = w^T (B+B^T) w,$$

$$\int_{T_o} r_\eta^2 d\xi d\eta = w^T C w, \quad \int_{T_o} r^2 d\xi d\eta = w^T D w$$

with $A = (S^{-1})^T L S^{-1}$, $B = (S^{-1})^T M S^{-1}$, $C = (S^{-1})^T N S^{-1}$, $D = (S^{-1})^T O S^{-1}$ $(S=S(k))$ and $L = (\int_{T_o} l_i l_j d\xi d\eta)$,

$M = (\int_{T_o} l_i m_j d\xi d\eta)$, $N = (\int_{T_o} m_i m_j d\xi d\eta)$, $O = (\int_{T_o} z_i z_j d\xi d\eta)$,

where l_i, m_i, z_i are the components of l, m, z = z(k), respectively. The sum of the first four terms of I(p,T) is therefore given as $w^T K w$ with

$$K = aA + b(B+B^T) + cC + dD. \tag{11}$$

For the last term we find $-2 \int_{T_o} grd\xi d\eta = -2 w^T \delta$ with

$$\delta = (S^{-1})^T \beta \quad \text{and} \tag{12}$$

$$\beta = (\int_{T_o} g z_i d\xi d\eta)^T. \tag{13}$$

The integrals can be calculated by quadrature formulas. Here we only remark the formula: For $\alpha, \beta \in N$ it holds $\int_{T_o} \xi^\alpha \eta^\beta d\xi d\eta = \Gamma(\alpha+1)\Gamma(\beta+1)\Gamma(\alpha+\beta+3)^{-1}$. The matrices S^{-1}, A, B, C, D for k = 1 and k = 3 read (for k = 2 see /8/):

<u>k = 1:</u>

$$S^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}, \quad A = \frac{1}{2}\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad B = \frac{1}{2}\begin{bmatrix} 1 & 0 & -1 \\ -1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$

$$C = \frac{1}{2}\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix} \quad D = \frac{1}{24}\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}.$$

k = 3:

$$S^{-1} = \frac{1}{2}\begin{bmatrix}
2 & & & & & & & & & \\
-11 & 18 & -9 & 2 & & & & & & \\
18 & -45 & 36 & -9 & & & & & & \\
-9 & 27 & -27 & 9 & & & & & & \\
-11 & & & & 18 & -9 & 2 & & & \\
18 & & & & -45 & 36 & -9 & & & \\
-9 & & & & 27 & -27 & 9 & & & \\
-27 & 54 & -27 & & 27 & & & -54 & 27 & \\
-27 & 27 & & & 54 & -27 & & -54 & & 27 \\
36 & -45 & 9 & & -45 & 9 & & 54 & -9 & -9
\end{bmatrix}$$

$$A = \frac{1}{720}\begin{bmatrix}
306 & -486 & 243 & -63 & 27 & -27 & 0 & & -27 & -27 \\
-486 & 1215 & -972 & 243 & & & 0 & & & \\
243 & -972 & 1215 & -486 & & & 0 & & & \\
-63 & 243 & -486 & 306 & -27 & -27 & 0 & & 27 & 27 \\
27 & & & -27 & 1215 & -243 & 0 & -1458 & 243 & 243 \\
27 & & & -27 & -243 & 1215 & 0 & & 243 & -1215 \\
 & & & & & & 0 & & & \\
 & & & & -1458 & & 0 & 2916 & -1458 & \\
-27 & & & 27 & 243 & 243 & 0 & -1458 & 1215 & -243 \\
-27 & & & 27 & 243 & -1215 & 0 & & -243 & 1215
\end{bmatrix}$$

$$B' = \begin{bmatrix}
22482 & 54 & 54 & 0 & -1046 & 560 & -21996 & & -54 & -54 \\
-22768 & 1215 & -243 & 0 & -10861 & -12319 & -21796 & -1458 & 243 & 243 \\
22356 & -1701 & 1215 & 0 & -317 & -169 & -21870 & 1458 & -1215 & 243 \\
-126 & 432 & -1026 & 0 & & & 126 & & 1026 & -432 \\
-21816 & 1215 & -243 & 0 & 1289 & -1775 & 22302 & -1458 & 243 & 243 \\
54 & -243 & -243 & 0 & -243 & 1215 & -1026 & 1458 & 243 & -1215 \\
 & & & 0 & & & & & & \\
-21796 & -1458 & 1458 & 0 & -13534 & -10618 & -21796 & 2916 & -1458 & -1458 \\
-21924 & 243 & -1215 & 0 & 317 & 169 & 21438 & -1458 & 1215 & 1215 \\
-54 & 243 & 243 & 0 & 243 & -1215 & 1026 & -1458 & -243 & 1215
\end{bmatrix}$$

$$C = \frac{1}{720}\begin{bmatrix}
306 & 27 & -27 & 0 & -486 & 243 & -63 & & -27 & -27 \\
27 & 1215 & -243 & 0 & & & -27 & -1458 & 243 & 243 \\
27 & -243 & 1215 & 0 & & & -27 & & -1215 & 243 \\
 & & & 0 & & & & & & \\
-486 & & & 0 & 1215 & -972 & 243 & & & \\
243 & & & 0 & -972 & 1215 & -486 & & & \\
-63 & -27 & -27 & 0 & 243 & -486 & 306 & & 27 & 27 \\
 & -1458 & & 0 & & & & 2916 & & -1458 \\
-27 & 243 & -1215 & 0 & & & 27 & & 1215 & -243 \\
-27 & 243 & 243 & 0 & & & 27 & -1458 & -243 & 1215
\end{bmatrix}$$

$$D = \frac{1}{40320}\begin{bmatrix} 228 & 54 & & 33 & 54 & & 33 & 108 & 81 & 81 \\ 20182 & 1620 & -567 & & 810 & -405 & 81 & 486 & -405 & -162 \\ & -567 & 1620 & 54 & -405 & -162 & 81 & 486 & 810 & -405 \\ 33 & & 54 & 228 & 81 & 81 & 33 & 108 & 54 & \\ 20182 & 810 & -405 & 81 & 1620 & -567 & & 486 & -162 & -405 \\ & -405 & -162 & 81 & -567 & 1620 & 54 & 486 & -405 & 810 \\ 33 & 81 & 81 & 33 & & 54 & 228 & 108 & & 54 \\ 108 & 486 & 486 & 108 & 486 & 486 & 108 & 5832 & 486 & 486 \\ 81 & -405 & 810 & 54 & -162 & -405 & & 486 & 1620 & -567 \\ 81 & -162 & -405 & & -405 & 810 & 54 & 486 & -567 & 1620 \end{bmatrix}$$

It holds $B' = 1440\ B$. Now it is possible to compute K and δ. From K and δ one can compute K^* and δ^* and after this one can solve the linear algebraic system

$$K^* w^* = \delta^*.$$

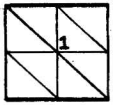
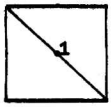
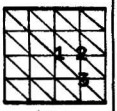
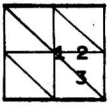
K^* is the stiffness matrix and w^* is a vector, in which any unknown of the region Ω appears only once. An algorithm for computation of K^* and δ^* from K and δ is given in /8/, p. 105.

6. A numerical example

Let be $\Omega = \{(x,y): -0{,}25 < x,\ y < 0{,}25\}$. The boundary value problem $-\Delta u = 1$ in Ω, $u = 0$ on Γ is considered. The exact solution of this problem ist

$$u(x,y) = -0{,}5\left(y^2 - \frac{1}{16}\right) + \sum_{n=0}^{\infty} \frac{(-1)^{n+1}}{(2n+1)^3 \pi^3 \cosh\left[(2n+1)\frac{\pi}{2}\right]} \cos[2(2n+1)\pi y] \cosh[2(2n+1)\pi x].$$

The form of triangulation and the numerical results are seen in the following table (ε denotes the relative error):

No. of point	exact value	 8 linear elements	ε in %	 2 quadratic elements	ε in %	 32 linear elements	ε in %	 8 quadratic elements	ε in %
1	.018418	.015625	15	.015625	15	.017578	4.6	.018747	1.8
2	.014334					.013672	4.6	.014061	1.9
3	.011322					.010742	5.1	.010936	3.4

References

/1/ Ciarlet, P. G., and Raviart, P. A.: General Lagrange and Hermite interpolation in R^n with applications to finite element methods. Arch. Rational Mech. Anal. 46, 177 - 199 (1972)

/2/ Zlámal, M.: Lectures on the finite element method. Skripte, Technische Hochschule Karl-Marx-Stadt 1974

/3/ Ciarlet, P. G.: The Finite Element Method for Elliptic Problems. Amsterdam 1978

/4/ Coatmélec, C.: Approximation et interpolation des fonctions différentiables de plusieurs variables. Ann. Sci. École Norm. Sup. (3), 83, 271 - 341 (1966)

/5/ Necas, J.: Les méthodes directes en théorie des équations elliptiques. Prague 1967

/6/ Kadlec, J.: On the regularity of the solution of the Poisson problem on a domain with boundary locally similar to the boundary of a convex open set (in Russian). Czechoslovak Math. J. 14 (89), 386 - 393 (1964)

/7/ Nitsche, J.: Ein Kriterium für die Quasioptimalität des Ritzschen Verfahrens. Numer. Math. 11, 346-348 (1968)

/8/ Moldenhauer, W., und Strauß, R.: Zur Lösung der Helmholtz-Gleichung in einem polygonberandeten Gebiet mittels des Finite-Elemente-Verfahrens. Rostock. Math. Kolloq. 12, 101 - 111 (1979)

/9/ Zlámal, M.: On some finite element procedures for solving second order boundary value problems. Numer. Math. 14, 42 - 48 (1969)

/10/ Schwarz, H. R.: Methode der finiten Elemente. Stuttgart 1980

received: January 26, 1984

Author's address:

Dr. W. Moldenhauer
Pädagogische Hochschule
"Dr. Theodor Neubauer"
Erfurt/Mühlhausen
Nordhäuser Str. 63
DDR-5060 Erfurt

Rostock. Math. Kolloq. 26, 103 - 109 (1984)

10A10
94A11

Reiner Creutzburg
Manfred Tasche

Zahlentheoretische Transformationen und primitive Einheitswurzeln in einem Restklassenring modulo m, II

Eine wesentliche Rolle bei der Definition zahlentheoretischer Transformationen spielen die primitiven Einheitswurzeln modulo m. In /1/ wurde für einen gegebenen Modul m und eine geeignet gewählte Ordnung μ ein Algorithmus zur Berechnung aller primitiven μ-ten Einheitswurzeln modulo m beschrieben. Im Unterschied zu dieser Konstruktionsmethode von primitiven Einheitswurzeln modulo m werden jetzt passende Moduln m bei vorgegebenen Zahlen a und $\mu > 1$ derart gesucht, daß a eine primitive μ-te Einheitswurzel modulo m ist. Als Spezialfälle dieser zweiten Konstruktionsmethode, die die Eigenschaften der Kreisteilungspolynome wesentlich ausnutzt, bekommt man bekannte Aussagen von P. J. Erdelsky (/3/). Als Moduln erhält man u. a. Fermat-, Mersenne-, Pseudo-Fermat- und Pseudo-Mersenne-Zahlen, für die primitive Einheitswurzeln modulo m angegeben werden.

1. Konstruktion passender Moduln m

Im folgenden werden die in /1/ eingeführten Bezeichnungen weiterhin verwendet. In /1/ wurden bei vorgegebenem ganzzahligen Modul $m > 1$ die möglichen Ordnungen μ und die primitiven μ-ten Einheitswurzeln modulo m bestimmt. Wie bereits einfache Beispiele zeigen, können die primitiven Einheitswurzeln modulo m "recht groß" ausfallen. Deshalb interessiert man sich für die folgende Frage: Wie muß ein Modul m bei vorgegebenen ganzen Zahlen $\mu > 2$ und $a \geq 2$ lauten, damit a eine primitive μ-te Einheitswurzel modulo m ist?
Nach /1/, Satz 4.7 muß m ein Teiler von $\chi_\mu(a)$ sein, so daß es zu vorgegebenen Zahlen μ und a nur endlich viele Moduln m geben kann.
Im folgenden seien μ, $a \in G$ mit $\mu > 2$, $a \geq 2$, und p sei der größte

Primfaktor von μ mit $p^{\beta} \mid \mu$, aber $p^{\beta+1} \nmid \mu$ für gewisses $\beta \geq 1$. Ein bereits auf L. Kronecker (/2/) zurückgehendes Resultat, daß der Wert $\chi_\mu(a)$ des μ-ten Kreisteilungspolynoms χ_μ nur p oder Primzahlen $\equiv 1 \bmod \mu$ als Primfaktoren besitzen kann, wurde zu folgender Aussage verschärft:

Satz 1 (/4/): Es seien μ, $a \in G$ mit $\mu > 2$ und $a \geq 2$ gegeben, wobei der Fall $\mu = 6$ und $a = 2$ ausgeschlossen ist. Ferner sei

$$M = \begin{cases} \chi_\mu(a)/p, & \text{falls a zum Exponenten } \mu/p^{\beta} \text{ modulo p gehört,} \\ \chi_\mu(a) & \text{andernfalls.} \end{cases} \quad (1)$$

Dann besitzt M lauter Primfaktoren $\equiv 1 \bmod \mu$.

Hieraus ergibt sich folgende einfache Konstruktionsmethode für einen Modul m:

Satz 2: Es seien μ, $a \in G$ mit $\mu > 2$ und $a \geq 2$ gegeben, wobei der Fall $\mu = 6$ und $a = 2$ ausgeschlossen ist. Ferner sei M durch (1) erklärt und $m > 1$ ein Teiler von M.
Dann ist a eine primitive μ-te Einheitswurzel modulo m. Ist μ ungerade, so ist -a eine primitive (2μ)-te Einheitswurzel modulo m. Die Teiler $m > 1$ von M sind sämtliche möglichen Moduln, für die a eine primitive μ-te Einheitswurzel modulo m ist.

Beweis: Wegen /1/, Satz 4.7 muß ein Modul m stets ein Teiler von $\chi_\mu(a)$ sein. Nach Satz 1 genügt μ der Bedingung $\mu \mid p_i - 1$ für jeden Primfaktor p_i $(i=1,\ldots,s)$ von m. Offensichtlich gilt

$$\chi_\mu(a) \equiv 0 \bmod m, \quad (2)$$

so daß a nach /1/, Satz 4.7 eine primitive μ-te Einheitswurzel modulo m ist.
Ist μ ungerade, so besitzt m nach Satz 1 nur Primfaktoren $\equiv 1 \bmod 2\mu$. Denn sämtliche Primteiler von m sind ungerade, da andernfalls m einen Primfaktor $\equiv 2 \bmod \mu$ hätte. Wegen $\chi_{2\mu}(-a) = \chi_\mu(a)$ für ungerades μ gilt nach (2)

$$\chi_{2\mu}(-a) \equiv 0 \bmod m.$$

Somit ist -a nach /1/, Satz 4.7 eine primitive (2μ)-te Einheitswurzel modulo m. □

Als Folgerungen aus Satz 2 ergeben sich zahlreiche, bisher isoliert betrachtete Ergebnisse:

Folgerung 3 (/3/): Es sei p eine Primzahl, $t \geq 1$, $\mu = p^t$ und $a \geq 2$. Ferner sei m die Pseudo-Mersenne-Zahl

$$m = \begin{cases} \chi_\mu(a)/p, & \text{falls } a \equiv 1 \bmod p, \\ \chi_\mu(a) & \text{sonst} \end{cases}$$

mit $\chi_\mu(a) = (a^\mu - 1)/(a^{\mu/p} - 1)$.
Dann ist a eine primitive μ-te Einheitswurzel modulo m. Im Fall $p > 2$ ist -a eine primitive (2μ)-te Einheitswurzel modulo m.

Folgerung 4 (/3/): Es sei $p > 2$ eine Primzahl und m die Mersenne-Zahl

$$m = \chi_p(2) = 2^p - 1.$$

Dann ist 2 eine primitive p-te Einheitswurzel modulo m und -2 eine primitive (2p)-te Einheitswurzel modulo m.

Folgerung 5 (/3/): Es sei $d \geq 0$, $\mu = 2^{d+1}$ und m die Fermat-Zahl

$$m = \chi_\mu(2) = 2^{2^d} + 1.$$

Dann ist 2 eine primitive μ-te Einheitswurzel modulo m. Im Fall $d \geq 2$ ist

$$a = 2^{2^{d-2}} (2^{2^{d-1}} - 1) \qquad (3)$$

eine primitive (2μ)-te Einheitswurzel modulo m.

Beweis: Es bleibt zu zeigen, daß im Fall $d \geq 2$ die Zahl (3) eine primitive (2μ)-te Einheitswurzel modulo m ist. Zunächst gilt $a^2 \equiv 2 \bmod m$ und damit

$$a^{2^{d+2}} \equiv 2^{2^{d+1}} \equiv 1 \bmod m,$$

$$a^{2^{d+1}} - 1 \equiv 2^{2^d} - 1 \equiv -2 \bmod m.$$

Ferner ist m ungerade. Folglich ist a nach /1/, Satz 1.2 eine primitive (2μ)-te Einheitswurzel modulo m. □

Bemerkung: Nach Satz 2 und Folgerung 5 kommen im Fall $\mu = 2^{d+1}$ ($d \geq 0$) und $a = 2$ nur die Fermat-Zahl $\chi_\mu(2)$ und deren mögliche

Teiler >1 als Moduln m in Frage, so daß 2 eine primitive (2^{d+1})-te Einheitswurzel modulo m ist.
Nach Satz 2 gibt es beispielsweise nur einen einzigen Modul m, für den 2 eine primitive 120-te Einheitswurzel modulo m ist, nämlich die Primzahl $m = \chi_{120}(2) = 4\,562\,284\,561$.

Folgerung 6 (/3/): Es sei $p > 2$ eine Primzahl, $t \geq 1$, $\mu = 2p^t$ und $a \geq 2$, wobei der Fall $\mu = 6$ und $a = 2$ ausgeschlossen ist. Ferner sei m die Pseudo-Fermat-Zahl

$$m = \begin{cases} \chi_\mu(a)/p, & \text{falls } a \equiv -1 \bmod p, \\ \chi_\mu(a) & \text{sonst} \end{cases}$$

mit $\chi_\mu(a) = (a^{\mu/2} + 1)/(a^{\mu/(2p)} + 1)$.
Dann ist a eine primitive μ-te Einheitswurzel modulo m.

Folgerung 7: Es sei $p > 2$ eine Primzahl, $d \geq 0$ und $\mu = 2^{d+1}p$, wobei der Fall $\mu = 6$ ausgeschlossen ist. Ferner sei m die Pseudo-Fermat-Zahl

$$m = \begin{cases} \chi_\mu(2)/p, & \text{falls 2 zum Exponenten } 2^{d+1} \text{ modulo p gehört,} \\ \chi_\mu(2) & \text{sonst, d. h. } 2^{\mu/(2p)} \not\equiv -1 \bmod p, \end{cases}$$

mit $\chi_\mu(2) = (2^{\mu/2} + 1)/(2^{\mu/(2p)} + 1)$.
Dann ist 2 eine primitive μ-te Einheitswurzel modulo m. Im Fall $d \geq 2$ ist

$$a = 2^{p2^{d-2}}(2^{p2^{d-1}} - 1) \qquad (4)$$

eine primitive (2μ)-te Einheitswurzel modulo m.

Beweis: Es ist zu zeigen, daß im Fall $d \geq 2$ die Zahl (4) eine primitive (2μ)-te Einheitswurzel modulo m ist. Wegen $a^2 \equiv 2 \bmod m$ gilt

$$a^{2\mu} \equiv 2^\mu \equiv 1 \bmod m,$$
$$a^\mu - 1 \equiv 2^{\mu/2} - 1 \equiv -2 \bmod m,$$
$$a^{2\mu/p} - 1 \equiv 2^{\mu/p} - 1 \bmod m.$$

Da 2 eine primitive μ-te Einheitswurzel modulo m ist, sind $2^{\mu/p} - 1$ und m nach /1/, Satz 1.2 teilerfremd. Ferner ist m ungerade, so daß folgt

$$(a^\mu - 1, m) = (a^{2\mu/p} - 1, m) = 1.$$

Nach /1/, Satz 1.2 ist a eine primitive (2μ)-te Einheitswurzel modulo m. □

Bemerkung: Nach /1/, Beispiel 4.1 sind 57 und 307 primitive vierte Einheitswurzeln modulo 1625. Wendet man nun für $\mu = 4$ und $a = 57$ Folgerung 3 an, so erhält man erneut $m = 1625$. Im Fall $\mu = 4$, $a = 307$ bekommt man dagegen $m = 47125 = 29 \cdot 1625$. Nach Satz 2 ist die Zahl (1) der größtmögliche Modul m, für die a eine primitive μ-te Einheitswurzel modulo m ist.

2. Anwendungen

Die voranstehenden Ergebnisse gestatten eine Reihe von wichtigen Aussagen über primitive μ-te Einheitswurzeln a modulo m, die in der folgenden Tabelle zusammengestellt sind. Dabei bezeichnet p stets eine Primzahl.

Modul m	Primitive μ-te Einheitswurzel a modulo m	
	a	μ
$2^p - 1$	2	p
$2^p - 1$ $p > 2$	-2	2p
$2^{2^d} + 1$ $d \geq 0$	2	2^{d+1}
$2^{2^d} + 1$ $d \geq 2$	$2^{2^{d-2}(2^{2^{d-1}} - 1)}$	2^{d+2}
$2^{q2^d} + 1$ q > 2 ungerade, $d \geq 0$	2^q	2^{d+1}
$2^{q2^d} + 1$ q > 2 ungerade, $d \geq 2$	$2^{q2^{d-2}(2^{q2^{d-1}} - 1)}$	2^{d+2}
$(2^{p^2} - 1)/(2^p - 1)$	2	p^2

$(2^{p^2}-1)/(2^p-1)$ $p>2$	-2	$2p^2$
$(2^{pq}-1)/(2^q-1)$ $q \geq 2$, $2^q \not\equiv 1 \bmod p$	2^q	p
$(2^{pq}-1)/(2^q-1)$ $p>2$, $q \geq 2$ $2^q \not\equiv 1 \bmod p$	-2^q	$2p$
$(2^p+1)/3$ $p>3$	2	$2p$
$(2^{p^2}+1)/(2^p+1)$ $p>3$	2	$2p^2$
$(2^{pq}+1)/(2^q+1)$ $p>2$, $q \geq 2$ $2^q \not\equiv -1 \bmod p$	2^q	$2p$
$(2^{p2^d}+1)/(2^{2^d}+1)$ $p>2$, $d \geq 0$ $2^{2^d} \not\equiv -1 \bmod p$	2	$p2^{d+1}$
$(2^{p2^d}+1)/(2^{2^d}+1)$ $p>2$, $d \geq 2$ $2^{2^d} \not\equiv -1 \bmod p$	$2^{p2^{d-2}}(2^{p2^{d-1}}-1)$	$p2^{d+2}$

Literatur

/1/ Creutzburg, R., und Tasche, M.: Zahlentheoretische Transformationen und primitive Einheitswurzeln in einem Restklassenring modulo m. Rostock. Math. Kolloq. 25, 4 - 22 (1984)

/2/ Kronecker, L.: Über die arithmetischen Sätze, welche Lejeune Dirichlet in seiner Breslauer Habilitationsschrift entwickelt hat. Monatsberichte der Königlich

Preußischen Akademie der Wissenschaften zu Berlin 1888, 417 - 423 (= Leopold Kronecker's Werke 3, 1, 281 - 292. Leipzig 1899)

/3/ Nussbaumer, H. J.: Fast Fourier Transform and Convolution Algorithms. Berlin 1981

/4/ Richter, B.: Die Primfaktorzerlegung der Werte der Kreisteilungspolynome. J. Reine Angew. Math. 254, 123 - 132 (1972)

eingegangen: 18. 06. 1984

Anschrift der Verfasser:

Dipl.-Lehrer R. Creutzburg
Doz. Dr. sc. nat. M. Tasche
Wilhelm-Pieck-Universität Rostock
Sektion Mathematik
Universitätsplatz 1
DDR-2500 Rostock

Hinweise für Autoren

Manuskripte (in deutscher, ggf. auch in russischer oder englischer Sprache) bitten wir, an die Schriftleitung zu schicken. Die gesamte Arbeit ist linksbündig zu schreiben. Eine Ausnahme hiervon bilden hervorzuhebende Formeln und das Literaturverzeichnis. Der Kopf der Arbeit soll folgende Form haben: Rostock. Math. Kolloq. /Leerzeile/ Vorname Name/ Leerzeile/ Titel der Arbeit/ 1 Zeilenumschaltung/ Unterstreichung/ Leerzeile. Der Text der Arbeit ist einzeilig (= 3 Zeilenumschaltungen) zu schreiben mit maximal 63 Anschlägen je Zeile und maximal 37 Zeilen je Seite. Zwischenüberschriften sind wie folgt einzuordnen: 6 Zeilenumschaltungen/ Zwischenüberschrift/ Unterstreichung (ohne Zeilenumschaltung)/ 5 Zeilenumschaltungen. Hervorhebungen sind durch Unterstreichen und Sperren möglich. Ankündigungen wie Satz, Definition, Bemerkung, Beweis u. ä. sind zu unterstreichen und mit einem Doppelpunkt abzuschließen. Vor und nach Sätzen, Definitionen u. ä. ist ein Zeilenabstand von 5 Umschaltungen zu lassen. Fußnoten sind möglichst zu vermeiden. Sollte doch davon Gebrauch gemacht werden, so sind sie durch eine hochgestellte Ziffer im Text zu kennzeichnen und innerhalb des oben angegebenen Satzspiegels unten auf der gleichen Seite anzugeben. Formeln und Bezeichnungen sollen möglichst mit der Schreibmaschine zu schreiben sein. Hervorzuhebende Formeln sind drei Leerzeichen einzurücken und mit 6 Umschaltungen zum übrigen Text zu schreiben. Formelzähler sollen am rechten Rand stehen. Der Platz für Abbildungen ist beim Schreiben auszusparen; die Abbildungen selbst sind in der dem ausgesparten Platz entsprechenden Größe gesondert nach TGL-Vorschrift auf Transparentpapier beizufügen. Der zugehörige Begleittext ist im Manuskript mitzuschreiben. Sein Abstand nach unten beträgt 5 Umschaltungen. Literaturzitate im Text sind durch laufende Nummern in Schrägstrichen (vgl. /8/, /9/ und /10/) zu kennzeichnen und am Schluß der Arbeit unter der Zwischenüberschrift Literatur zusammenzustellen.

Beispiele: (Zeitschriftenabkürzungen nach Math. Reviews)

/8/ Zariski, O., and Samuel, P.: Commutative Algebra. Princeton 1958

/9/ Steinitz, E.: Algebraische Theorie der Körper. J. Reine Angew. Math. 137, 167 - 309 (1920)

/10/ Gnedenko, B. W.: Über die Arbeiten von C. F. Gauß zur Wahrscheinlichkeitsrechnung. In: Reichardt, H. (Ed.): C. F. Gauß, Gedenkband anläßlich des 100. Todestages. S. 193 - 204, Leipzig 1967

Die Angaben sollen in Originalsprache erfolgen; bei kyrillischen Buchstaben soll die bibliothekarische Transkription (Duden) verwendet werden.

Am Ende der Arbeit stehen folgende Angaben zum Autor und zur Arbeit: eingegangen: Datum/ Leerzeile/ Anschrift des Verfassers:/ Titel Initialen der Vornamen Name/ Institution/ Struktureinheit/ Straße Hausnummer/ Land Postleitzahl Ort.

Der Autor wird gebeten, eine Korrektur des Durchschlags vom Offsetmanuskript zu lesen und dabei die mathematischen Symbole einzutragen. Ferner sollte er 1 - 2 Klassifizierungsnummern (entsprechend der "1980 Mathematics Subject Classification" der Math. Reviews) zur inhaltlichen Einordnung seiner Arbeit angeben.

ODR II-15-14 C 70/84 4722