Die Methode der Finiten Elemente

Hans-Jürgen Reinhardt[1)]

Vorlesungsskriptum
im WS 1995/96

Fachbereich Mathematik
Universität – GH Siegen

[1)]Email: hjreinhardt@numerik.math.uni-siegen.de

Vorwort

Das vorliegende Skriptum ist aus einer Vorlesung entstanden, die der Autor mehrfach — zuletzt im WS 1995/96 — an der Univ. – GH Siegen gehalten hat. Zu der Vorlesung wurden theoretische und praktische Übungen gestellt; die Übungsaufgaben finden sich im Anhang C zusammen mit numerischen Lösungen der praktischen Aufgaben.

Die Darstellung des Vorlesungsstoffes im Rahmen einer verallgemeinerten Interpolationstheorie wurde wesentlich geprägt durch eine Vorlesung von Herrn Professor F. Stummel, an der der Autor als Hörer selbst teilnehmen durfte. Seinem verehrten akademischen Lehrer ist der Autor zu besonderem Dank verpflichtet, auch dafür, Teile seiner Vorlesung verwenden zu dürfen. Auch Herrn Prof. Dr. R. Rannacher sei gedankt, dessen Vorlesungsskriptum vom WS 1976/77 den Kapiteln 6 und 7 zugrunde liegt.

Dieses Skriptum wurde von Frau M. Beier sowie Herrn cand. math. M. Schüler mit LaTeX erstellt. Herr Dipl.–Math. F. Seiffarth hat die Übungsaufgaben zusammengestellt und betreut. Ihnen und meinen Studenten sei für Ihren Einsatz und Ihre Mitarbeit gedankt.

Siegen, im April 1996 H.–J. Reinhardt

Inhaltsverzeichnis

1 Einleitung **1**

1.1 Variationelle Formulierung eines 1–dimensionalen Modellproblems . . 2

1.2 Die einfachste Finite Elemente Methode für das Modellproblem . . . 7

1.3 Erste Fehlerabschätzungen im eindimensionalen Fall 10

1.4 FEM für die Poisson–Gleichung 14

1.5 Die Hilberträume $L^2(G)$, $H^1(G)$, $H_0^1(G)$ 27

1.6 Die FEM für ein Neumann Problem 30

2 Finite–Element–Approximationen elliptischer Differentialgleichungen **35**

2.1 Schwache Lösungen elliptischer Differentialgleichungen 35

2.2 Ritzsches Projektionsverfahren, Fehlerabschätzungen 44

2.3 Beispiele für elliptische RWP's in einer Veränderlichen 47

3 Interpolation von Funktionen einer Veränderlichen **51**

3.1 Finite Elemente für Funktionen einer Veränderlichen 51

3.2 Die verallgemeinerte Interpolationsaufgabe 55

3.3 Klassische Interpolationsaufgaben 63

3.4 Interpolation durch stückweise quadratische Funktionen einer Veränderlichen 70

3.4.1 Lagrange–Interpolation in $\mathcal{P}_2^1(\Delta)$ 70

3.4.2 Quadratische Spline–Funktionen 73

3.4.3 Weitere Interpolation und Differenzenquotienten 75

4 Finite Elemente in mehreren Veränderlichen 81

4.1 Räume von Polynomen in mehreren Veränderlichen 82

4.2 Lagrange–Interpolation in $\mathcal{P}_1(\mathbb{R}^n)$ 86

4.3 Lagrange–Interpolation in $\mathcal{P}_r(\mathbb{R}^n)$ 95

4.4 Lagrange–Interpolation in $\mathcal{Q}_{r_1,\ldots,r_n}$ 99

4.5 Hermite–Interpolation 109

5 Interpolationsabschätzungen 121

5.1 Interpolatorische Approximation und Restglieder 121

5.2 Approximation durch Polynome in einer Veränderlichen 130

5.3 Approximation durch Polynome in mehreren Veränderlichen 146

5.4 Beste Approximation von Linearformen 163

6 Numerische Aspekte der FEM 167

6.1 Bezeichnungen und Wiederholung 167

6.2 Exakte Bestimmung der Matrix K 170

6.3 Näherungsweise Bestimmung von K und F 172

6.3.1 Interpolation mit exakter Polynomintegration 174

6.3.2 Integration mit Quadraturformeln 175

7 FEM auf nicht polygonalen Gebieten; Isoparametrische Finite Elemente 183

7.1 Fall: $G \subset \mathbb{R}^2$ konvex und ∂G aus der Klasse C^2. 183

7.2 Fall: $G \subset \mathbb{R}^2$ nicht konvex und ∂G aus der Klasse C^2. 189

7.3 Fall: $G \subset \mathbb{R}^2$ nicht notwendig konvex und ∂G aus der Klasse C^{m+1}, $m \geq 1$. 191

8 Finite–Element–Verfahren für zeitabhängige Probleme 195

8.1 Parabolische Probleme 195

8.2 Hyperbolische Probleme 200

A Theorie gewöhnlicher Differentialgleichungen **205**

A.1 Existenz und Eindeutigkeit von Lösungen von AWPs 206

A.2 Lineare Differentialgleichungen 207

A.3 Systeme mit konstanten Koeffizienten 212

A.4 Lineare Differentialgleichungen n-ter Ordnung 214

A.5 Lineare Differentialgleichungen n-ter Ordnung mit konstanten Koeffizienten 218

A.6 Lineare Randwertaufgaben 2. Ordnung 221

B Lösungen partieller Differentialgleichungen **227**

B.1 Modellbeispiele (partielle Differentialgleichungen) 227

B.2 Methode der Separation der Variablen 230

C Übungen **235**

Literaturverzeichnis **247**

Index **251**

Abbildungsverzeichnis

1.1 Elastischer Stab 3
1.2 Beispiel einer Funktion $v \in V_h$ 7
1.3 Die Basisfunktionen φ_j 8
1.4 Gleichförmige Triangulierung 25
1.5 Die Basisfunktionen w^{ij} bei Triangulierung 26

3.1 Basisfunktionen $p_0, \ldots, p_k, \ldots, p_n$ in $\mathcal{P}_1^1(\Delta)$ 58
3.2 Lagrange–Polynome aus $\mathcal{P}_2$ 71
3.3 Basisfunktionen in $\mathcal{P}_2^1(\Delta)$ 72
3.4 Basisfunktionen p_k für $k = 2, \ldots, n-1$ 78

4.1 Polynomräume $\mathcal{P}_r(\mathbb{R}^2)$ und $\mathcal{Q}_r(\mathbb{R}^2)$ 84
4.2 Schiefwinkliges Koordinatensystem in $\mathbb{R}^2$ 88
4.3 Einfache gleichförmige Triangulierung 91
4.4 Lagrange-Interpolation in $\mathcal{P}_1(\mathbb{R}^2)$ 93
4.5 Stützstellen für Lagrange–Interpolation in $\mathcal{P}_2(\mathbb{R}^2)$ 96
4.6 Stützstellen der Interpolation in $\mathcal{Q}_1(\mathbb{R}^2)$ 102
4.7 Differenzenquotienten in $\mathcal{Q}_1(\mathbb{R}^2)$ 103
4.8 Stützstellen der Interpolation in $\mathcal{Q}_{1,r}(\mathbb{R}^2)$ 105
4.9 Differenzenquotienten in $\mathcal{Q}_{22}(\mathbb{R}^2)$ 106
4.10 Stützstellen für Dreieck von Morley 113
4.11 Stützstellen für Rechteck von Adini 116
4.12 Basisfunktion P_1 (Rechteck von Adini) 120

5.1 Affine Transformation finiter Elemente 147

5.2 In– und Umkreis finiter Elemente 156
5.3 Quasi–uniforme Triangulierung 159
5.4 Transformation auf Einheitselement 160
5.5 Knotenpunkte der Lagrange—Interpolation in $\mathcal{P}_2(G_\iota)$ 162

6.1 Quasi–uniforme Triangulierung 168
6.2 Träger von w_i 171

7.1 Approximation von $\partial G \in C^2$ durch Streckenzüge 184
7.2 Randabstand $\mathrm{O}(h^2)$ 184
7.3 Segmenteigenschaft 187
7.4 Fall $G_h \not\subset \overline{G}$ 189
7.5 Triangulierung mit krummlinigen Seiten 192

8.1 Linienmethode 197

Kapitel 1

Einleitung

Die Methode der finiten Elemente (Abk.: FEM) beschäftigt sich mit Verfahren zur Approximation von Funktionen einer und mehrerer Veränderlicher. Typische Anwendungen findet diese Methode in der näherungsweisen Berechnung von Funktionen, der Interpolation und Quadratur für Funktionen mehrerer Veränderlicher, der Approximation der Lösungen von Differential- und Integralgleichungen, von Anfangs-, Rand- und Eigenwertaufgaben bei gewöhnlichen und partiellen Differentialgleichungen. Die Methode der finiten Elemente ist verhältnismäßig neu, der weit überwiegende Teil der Literatur zu diesem Gebiet ist seit 1967 entstanden (s. [Whi73, NdV73, ZT89]). Der größte Teil dieser Literatur behandelt sehr anwendungs- und techniknahe Fragestellungen.

Die obengenannten Aufgabenstellungen treten in zahlreichen Anwendungsgebieten auf: Mathematische Darstellung von Oberflächen, Rumpfformen und Strukturen von Schiffen, Flugzeugen, Autokarosserien und Fahrwerken, Brücken, Staudämmen, Turbinenrädern, Platten und Schalen usw., im Zusammenhang mit der Berechnung der interessierenden physikalischen und technischen Zustandsgrößen. Durch dieses Verfahren wird die näherungsweise Lösung auch komplexer Probleme in Statik, Elastizitätstheorie, Wärmeleitung, Strömungsmechanik und vielen anderen ingenieurwissenschaftlichen Gebieten auf Computern ermöglicht.

Die Methode der finiten Elemente wird wesentlich dadurch geprägt, daß sie zur numerischen Realisierung auf Computern dienen soll. Die betrachteten Funktionenräume und Gebietsapproximationen sind geeignet, auch sehr komplexe mathematische Problemstellungen numerisch stabil zu lösen. Im Hinblick auf diese numerischen Anwendungen arbeitet die Methode der finiten Elemente vor allem mit Approximationen durch Polynome.

Die finiten Elemente in der Ebene haben im allgemeinen die geometrische Form von Dreiecken, Rechtecken oder Vierecken. Dreieckszerlegungen führen auf sogenannte

Triangulationen des Grundgebiets. Dreieckszerlegungen gestatten bei allgemeineren Gebieten eine genauere Approximation des Grundgebiets sowie eine bessere Anpassung der Feinheit der Überdeckung an das Verhalten der zu approximierenden Funktionen als dies bei Rechteckzerlegungen der Fall ist. In dreidimensionalen Problemen arbeitet man entsprechend mit Zerlegungen in Tetraeder oder Simplexe, dreidimensionale Intervalle, Parallelepipede und ähnliches. Oberbegriff für die eben beschriebenen geometrischen Mengen ist der beschränkte abgeschlossene konvexe Polyeder. Im eindimensionalen Fall reduzieren sich derartige Zerlegungen in finite Elemente einfach auf Intervallteilungen.

Das zentrale Approximationsverfahren der Methode finiter Elemente ist die Interpolation, genauer die verallgemeinerte Interpolation. Unter diesem Aspekt läßt sich die mathematische Theorie in drei Schritten aufbauen, die zugleich drei verschiedenen Ebenen der Komplexität des Approximationsproblems entsprechen:

1. Mikroskopische Theorie oder lokale Approximationstheorie,
2. Makroskopische Theorie oder globale Approximationstheorie,
3. Approximationstheorie für Funktionalgleichungen, insbesondere in der Form von Variationsgleichungen.

Auf jeder dieser Ebenen umfaßt die verallgemeinerte Interpolation Taylor–Entwicklungen, Lagrange– und Hermite–Interpolation sowie optimale Approximationen nach der Methode des kürzesten Abstands oder des kleinsten Fehlerquadrats in Hilbertschen Räumen.

Die skizzierte Gliederung läßt bereits erkennen, daß sich hier ein relativ umfangreiches und sehr detailliertes Gebiet eröffnet.

Zur Motivation wird in diesem einleitenden Kapitel Punkt 3. aufgegriffen und die Idee der FEM anhand von Modellproblemen für 1- und 2-dimensionale Randwertprobleme erläutert. Die Punkte 1. und 2. werden in den späteren Kapiteln behandelt.

1.1 Variationelle Formulierung eines 1–dimensionalen Modellproblems

Die 2–Punkt–Randwertaufgabe

$$\begin{aligned} -u''(x) &= f(x)\,, \qquad 0 < x < 1 \\ u(0) &= u(1) = 0 \end{aligned} \tag{D}$$

stellt zu gegebener stetiger Funktion f ein Modellproblem für verschiedene Anwendungsbeispiele aus der Mechanik dar. Aus der Theorie gewöhnlicher Differentialgleichungen weiß man außerdem, daß das Problem (D) eine eindeutig bestimmte Lösung besitzt.

Beispiel 1.1 Elastischer Stab

Auf einen elastischen Stab, der an beiden Enden aufliegt, wirke eine tangentiale Kraft mit Intensität f. Es bezeichne $\sigma(x)$ bzw. $u(x)$ die Spannung bzw. tangentiale Auslenkung im Punkt x unter f. Nimmt man an, daß die Auslenkung klein bleibt und das elastische Material sich linear verhält, dann gelten die folgenden (physikalischen) Beziehungen,

$$\begin{aligned} \sigma &= Eu' && \text{(Hookesches Gesetz)} \\ -\sigma' &= f && \text{(Gleichgewichtszustand)} \\ u(0) &= u(1) = 0 && \text{(Randbedingungen)} , \end{aligned}$$

wobei E den Elastizitätsmodul beschreibt. Setzt man $E = 1$ und eliminiert σ, dann erhält man (D). □

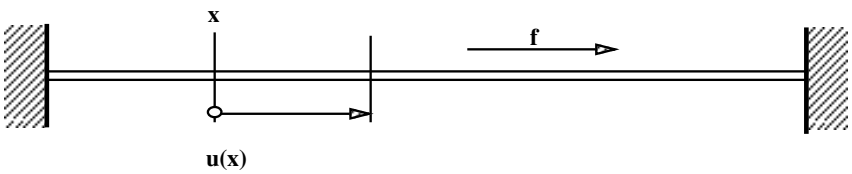


Abbildung 1.1: Elastischer Stab

Beispiel 1.2 Wärmeleitung

Es bezeichne u die Temperatur und q den Wärmefluß in einem leitenden Stab, verursacht durch eine Wärmequelle der Intensität f. Nimmt man an, daß die Temperatur an beiden Enden des Stabes Null ist, dann gelten im stationären Fall folgende Beziehungen,

$$\begin{aligned} -q &= ku' && \text{(Fouriersches Gesetz)} \\ q' &= f && \text{(Energieerhaltung)} \\ u(0) &= u(1) = 0 && \text{(Randbedingungen)} . \end{aligned}$$

Hierbei bezeichnet noch k die Wärmeleitfähigkeit, die auf 1 normiert, gerade wieder das Randwertproblem (D) ergibt. □

Als nächstes wird es unser Ziel sein zu zeigen, daß die Lösung des Randwertproblems auch die Lösung eines Minimierungsproblems (M) und eines Variationsproblems (V) ist. Dazu bezeichne

$$(v, w) = \int_0^1 v(x)w(x)\,dx$$

das L^2–Skalarprodukt für reellwertige Funktionen, die wir als stückweise stetig mit höchstens endlich vielen Sprungstellen annehmen. Außerdem sei

$$\begin{aligned} V \;:= \{v|\; & v \text{ stetig in } [0,1]\,,\; v' \text{ stückweise stetig und beschränkt,} \\ & v(0) = v(1) = 0\,\} \end{aligned}$$

ein entsprechender linearer Raum von Funktionen und

$$F(v) := \frac{1}{2}(v', v') - (f, v)$$

definiere ein quadratisches Funktional. Wir zeigen nun, daß die oben erwähnten Probleme die folgende Gestalt haben,

$$\begin{array}{llllll} \textit{Gesucht} & u \in V & \textit{mit} & F(u) & \leq & F(v) \quad \forall v \in V\,. \qquad (M) \\ \textit{Gesucht} & u \in V & \textit{mit} & (u', v') & = & (f, v) \quad \forall v \in V\,. \qquad (V) \end{array}$$

Im obigen Beispiel 1.1 repräsentiert $F(v)$ die *gesamte potentielle Energie* des Systems bei gegebener Auslenkung v. Hierbei stellt $\frac{1}{2}(v', v')$ die *innere elastische Energie* und (f, v) die *aufgebrachte Energie* dar. Das *Minimierungsproblem* (M) entspricht daher dem physikalischen *Prinzip der minimalen potentiellen Energie*; das *Variationsproblem* (V) wird in der Physik als *Prinzip der virtuellen Arbeit* bezeichnet.

Behauptung 1.3 *Eine Lösung von (D) löst auch (V).*

Beweis: Durch Multiplikation der Differentialgleichung in (D) mit einer Testfunktion aus V und Integration erhält man

$$-(u'', v) = (f, v)\,, \qquad v \in V\,.$$

Partielle Integration und Einsetzen der Randbedingungen liefert

$$-(u'', v) = -u'(1)v(1) + u'(0)v(0) + (u', v') = (u', v')\,,$$

woraus die Gleichung in (V) folgt. □

Behauptung 1.4 *Eine Lösung von (M) ist auch Lösung von (V) und umgekehrt.*

Beweis: (i) Mit einer Lösung u von (V) sei $w := v - u$ für beliebiges $v \in V$. Dann ist auch $w \in V$ und $v = u + w$, und man erhält

$$\begin{aligned} F(v) &= F(u+w) = \frac{1}{2}(u'+w', u'+w') - (f, u+w) \\ &= \underbrace{\frac{1}{2}(u', u') - (f, u)}_{=F(u)} + \underbrace{(u', w') - (f, w)}_{=0} + \frac{1}{2}\underbrace{(w', w')}_{\geq 0} \geq F(u)\,; \end{aligned}$$

damit ist u auch Lösung von (M).

(ii) Für eine Lösung u von (M), ein beliebiges $v \in V$ und eine reelle Zahl ε hat man

$$F(u) \leq F(u + \varepsilon v)\,,$$

da auch $u + \varepsilon v \in V$. Definiert man

$$\begin{aligned} g(\varepsilon) &:= F(u + \varepsilon v) \\ &= \frac{1}{2}(u', u') + \varepsilon(u', v') + \frac{\varepsilon^2}{2}(v', v') - (f, u) - \varepsilon(f, v)\,, \end{aligned}$$

dann ist $g(.)$ differenzierbar mit der Ableitung

$$g'(\varepsilon) = (u', v') + \varepsilon(v', v') - (f, v)$$

und hat ein Minimum bei $\varepsilon = 0$. Notwendigerweise muß daher $g'(0) = 0$ sein, was auf die Gleichung

$$(u', v') - (f, v) = 0$$

führt. □

Behauptung 1.5 *Eine Lösung von (V) ist eindeutig bestimmt.*

Beweis: Für zwei Lösungen u_1, u_2 von (V) hat man

$$(u_1', v') = 0 \qquad \text{und} \qquad (u_2', v') = 0 \qquad \forall v \in V\,.$$

Durch Subtraktion und die Wahl $v := u_1 - u_2 \in V$ folgt

$$\int_0^1 (u_1' - u_2')^2 \, dx = 0 \; .$$

Damit stimmen u_1' und u_2' fast überall in $[0,1]$ überein. Nach dem Hauptsatz der Differential- und Integralrechnung gilt auch für die stückweise differenzierbare Funktion $v = u_1 - u_2$, daß

$$v(x) = v(0) + \int_0^x v'(s) \, ds \; , \qquad x \in [0,1] \; .$$

Mit Hilfe der Cauchy–Schwarzschen Ungleichung für Integrale folgt daraus

$$|v(x) - v(0)| \leq \int_0^1 |v'(s)|^2 \, ds \; , \qquad x \in [0,1] \; .$$

Wie oben gezeigt, verschwindet die rechte Seite dieser Ungleichung; wegen der Randbedingungen ist $v(0) = 0$, was $v(x) = 0$, $x \in [0,1]$, beweist. □

Die drei letzten Behauptungen lassen sich symbolisch wie folgt darstellen,

$$(D) \Rightarrow (V) \Leftrightarrow (M) \, .$$

Wir zeigen nun abschließend, daß die Lösungen von (V) unter gewissen Regularitätsvoraussetzungen auch klassische Lösungen des Randwertproblems liefern.

Behauptung 1.6 *Eine Lösung u von (V) ist auch Lösung des Randwertproblems (D), falls u'' existiert und stetig ist, und falls f stetig ist.*

Beweis: Für $u \in V$ gelte

$$\int_0^1 u'v' \, dx - \int_0^1 fv \, dx = 0 \qquad \forall v \in V \; .$$

Durch partielle Integration und die in der Definition von V integrierten Randbedingungen $v(0) = v(1) = 0$ folgt

$$-\int_0^1 (u'' + f)v \, dx = 0 \qquad \forall v \in V \; .$$

Man überlegt sich leicht mit Hilfe eines indirekten Beweises, daß wegen der Stetigkeit von $u'' + f$ dann

$$(u'' + f)(x) = 0 \qquad \forall x \in [0,1]$$

gelten muß. Damit erfüllt u die Differentialgleichung in (D); die homogenen Randbedingungen sind erfüllt, da $u \in V$ angenommen wurde. □

1.2 Die einfachste Finite Elemente Methode für das Modellproblem

Zur Approximation des Modellproblems (D) aus Abschnitt 1.1 legen wir den endlich–dimensionalen Teilraum V_h von V aller stetigen, stückweise linearen Funktionen zugrunde. Dazu bezeichne

$$\Delta : 0 = x_0 < \cdots < x_{n-1} < x_n = 1$$

eine *Unterteilung* von $[0,1]$ in Teilintervalle $I_j = (x_{j-1}, x_j)$ der Länge $h_j = x_j - x_{j-1}$, $j = 1, \ldots, n$; weiter sei $h = \max_{1 \le j \le n} h_j$. Der Parameter h gibt also an, wie fein das Intervall unterteilt ist. Wie angekündigt, sei

$$V_h := \{v \in C[0,1] \,|\, v|I_j = \text{ linear}, \quad v(0) = v(1) = 0\} \, .$$

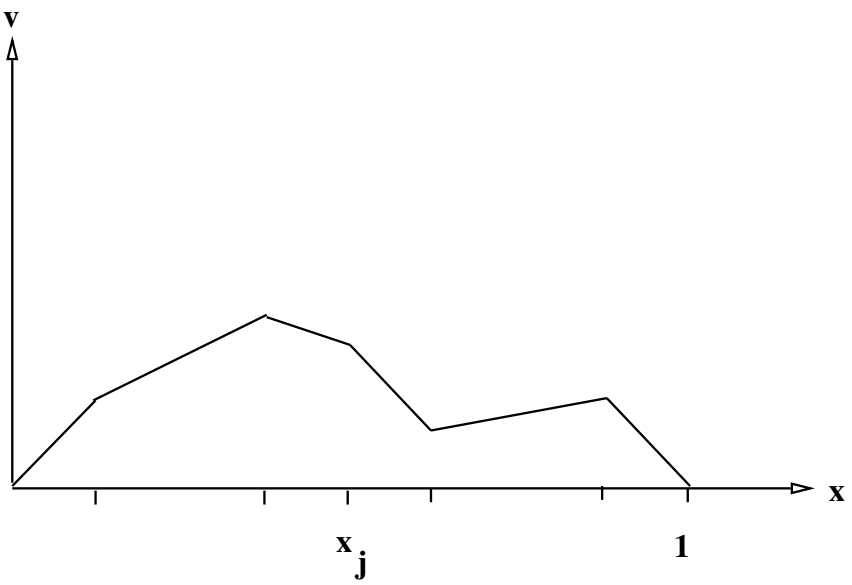


Abbildung 1.2: Beispiel einer Funktion $v \in V_h$

Man erhält so also einen Teilraum $V_h \subset V$ der Dimension $n-1$; Funktionen aus V_h sind eindeutig bestimmt durch deren Werte $\eta_j = v(x_j)$, $j = 1, \ldots, n-1$, an den *Gitterpunkten* x_j. Basisfunktionen in V_h erhält man durch die sogenannten „Dachfunktionen“ (engl. ‘roof functions’), die durch die Eigenschaften

$$\varphi_i(x_j) = \begin{cases} 1 & , \quad i = j \\ 0 & , \quad i \neq j \, , \end{cases} \qquad j = 1, \ldots, n-1$$

charakterisiert sind. Diese ergeben sich als

$$\varphi_i(x) = \begin{cases} 0 & , \quad 0 \le x \le x_{i-1} \ , \\ \dfrac{x - x_{i-1}}{h_i} & , \quad x_{i-1} \le x \le x_i \ , \\ \dfrac{x_{i+1} - x}{h_{i+1}} & , \quad x_i \le x \le x_{i+1} \ , \\ 0 & , \quad x_{i+1} \le x \le 1 \, , \qquad i = 1, \ldots, n-1 \ . \end{cases}$$

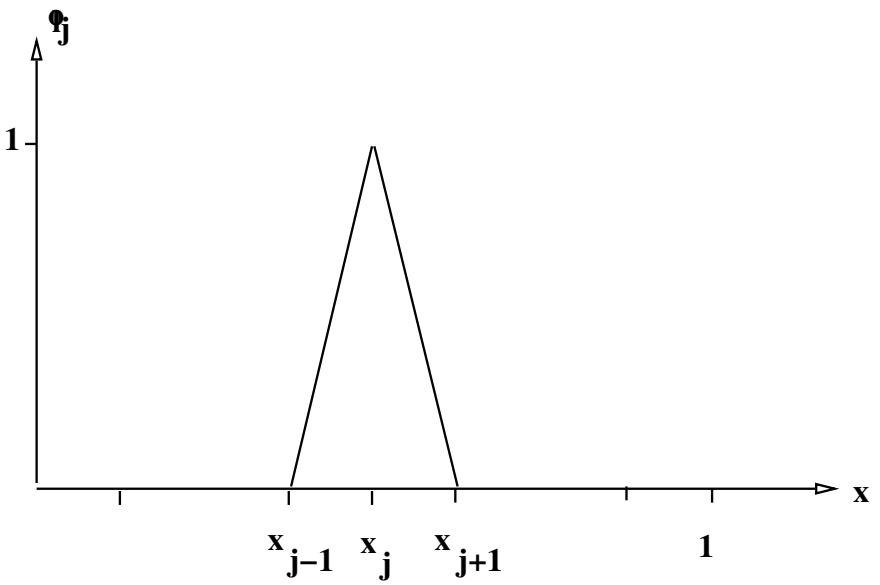


Abbildung 1.3: Die Basisfunktionen φ_j

Jede Funktion $v \in V_h$ läßt sich offenbar durch

$$v(x) = \sum_{i=1}^{n-1} \eta_i \varphi_i(x) \ , \qquad x \in [0,1] \ ,$$

darstellen mit $\eta_i = v(x_i)$.

Die *Methode der Finiten Elemente* für das Randwertproblem (D) und den ausgewählten endlichdimensionalen Raum V_h besteht nun darin, die folgende Aufgabe zu lösen:

$$\textit{Gesucht} \quad u_h \in V_h \quad \textit{mit} \quad F(u_h) \ \le \ F(v) \ \ \forall v \in V_h \ . \tag{M_h}$$

Wie in Abschnitt 1.1, Behauptung 1.4, macht man sich klar, daß diese Aufgabenstellung äquivalent dazu ist, das folgende *endlichdimensionale Variationsproblem* zu lösen:

$$\textit{Gesucht} \quad u_h \in V_h \quad \textit{mit} \quad (u_h', v') \;=\; (f, v) \quad \forall v \in V_h \; . \tag{V_h}$$

Das Problem (V_h) heißt üblicherweise auch *Galerkin Methode* ; das Minimierungsproblem wird auch *Ritz Methode* genannt. Setzt man

$$u_h(x) = \sum_{i=1}^{n-1} \xi_i \varphi_i(x) \; , \qquad \xi_i = u_h(x_i) \; , \tag{1.1}$$

und in (V_h) nacheinander $v = \varphi_j$, $j = 1, \ldots, n-1$, dann ist die Galerkin Methode (V_h) äquivalent zur Lösung des linearen Gleichungssystems

$$\sum_{i=1}^{n-1} \xi_i (\varphi_i', \varphi_j') = (f, \varphi_j) \; , \qquad j = 1, \ldots, n-1 \; , \tag{1.2}$$

zur Bestimmung der Koeffizienten ξ_i in der Darstellung (1.1) der *Galerkin Approximation* u_h. Die zugehörige Matrix $A = (a_{ij})$ mit den Koeffizienten $a_{ij} = (\varphi_i', \varphi_j')$ bezeichnet man auch als *(Gesamt-)Steifigkeitsmatrix* ; der Vektor der rechten Seiten $b = (b_1, \ldots, b_{n-1})^\top$, mit $b_j = (f, \varphi_j)$, heißt auch *Lastvektor*. Diese Bezeichnungen stammen aus der ursprünglichen Entwicklung und Anwendung der FEM in der Strukturmechanik.

Im vorliegenden Beispiel der Dachfunktionen können die Koeffizienten der Steifigkeitsmatrix A leicht berechnet werden. Zunächst beobachtet man, daß $(\varphi_i', \varphi_j') = 0$ ist, falls $|i - j| > 1$. Dies zeichnet eine *lokale Basis* aus, bei der im vorliegenden eindimensionalen Fall nur benachbarte Basisfunktionen einen gemeinsamen *Träger* [1] haben. Dadurch wird A tridiagonal, und man rechnet aus, daß

$$\begin{aligned}
(\varphi_j', \varphi_j') &= \int_{x_{j-1}}^{x_j} \frac{1}{h_j^2}\, dx + \int_{x_j}^{x_{j+1}} \frac{1}{h_{j+1}^2}\, dx = \frac{1}{h_j} + \frac{1}{h_{j+1}} \; , \; j = 1, \ldots, n-1 \; , \\
(\varphi_j', \varphi_{j-1}') &= (\varphi_{j-1}', \varphi_j') = -\int_{x_{j-1}}^{x_j} \frac{1}{h_j^2}\, dx = -\frac{1}{h_j} \; , \qquad j = 2, \ldots, n-1 \, , \\
(\varphi_j', \varphi_{j+1}') &= (\varphi_{j+1}', \varphi_j') = -\frac{1}{h_{j+1}}, \quad j = 1, \ldots, n-2.
\end{aligned}$$

Offenbar ist A symmetrisch und auch positiv definit, denn für einen Vektor $(\eta_1, \ldots, \eta_{n-1})^\top \in \mathbb{R}^n$ hat man mit $v = \sum_{j=1}^{n-1} \eta_j \varphi_j(x)$, daß

[1] Bezeichnung und Definition: $\operatorname{supp} \varphi := \{x | \varphi(x) \neq 0\}$

$$\sum_{i,j=1}^{n-1} \eta_i(\varphi_i', \varphi_j')\eta_j = \left(\sum_{i=1}^{n-1} \eta_i\varphi_i', \sum_{j=1}^{n-1} \eta_j\varphi_j'\right) = (v', v') \geq 0 \ .$$

In der letzten Ungleichung gilt Gleichheit nur, wenn $v' \equiv 0$ ist, was in Verbindung mit $v(0) = 0$ nur für $v \equiv 0$ — oder $\eta_j = 0\,,\ j = 1, \ldots, n-1$ — richtig ist.

Für eine positiv definite Matrix ist bekanntlich das zugehörige Gleichungssystem für jede rechte Seite eindeutig lösbar, und man kann hier das Gaußsche Eliminationsverfahren für tridiagonale Matrizen — ohne Zeilenvertauschung — als ein direktes Verfahren zur Berechnung der Lösung einsetzen (vgl. [SH82, §6]).

Im Spezialfall einer gleichförmigen Unterteilung, i. e. $h_j = h = \frac{1}{n}$, hat das Gleichungssystem (1.2) die Form

$$\begin{aligned} -\frac{1}{h^2}\left(\xi_{j-1} - 2\xi_j + \xi_{j+1}\right) &= \frac{1}{h}(f, \varphi_j)\ , \qquad j = 1, \ldots, n-1\ , \\ \xi_0 &= \xi_n = 0\ . \end{aligned}$$

Dies ist auf der linken Seite offenbar genau die Approximation der zweiten Ableitung durch den zentralen Differenzenquotienten zweiter Ordnung . Im Gegensatz zu einem entsprechenden Differenzenverfahren hat man bei der Galerkin Methode rechts eine Art Mittelung von f über das Interval (x_{j-1}, x_{j+1}) zu nehmen; eine Approximation z. B. durch die Simpson–Formel in I_j und I_{j+1} hat die Gestalt

$$\frac{1}{h}(f, \varphi_j) \approx \frac{1}{3}(f(x_{j-1/2}) + f(x_j) + f(x_{j+1/2}))\ ,$$

wobei $x_{j\pm1/2} := \frac{1}{2}(x_j + x_{j\pm1})$.

1.3 Erste Fehlerabschätzungen im eindimensionalen Fall

Wir betrachten zunächst wieder das Modellproblem (D) und die zugehörige Galerkin–Approximation (V_h). Fehlerabschätzungen für die Differenz der Lösung u von (D) bzw. u_h von (V_h) beruhen auf der folgenden *Fehlergleichung*

$$((u - u_h)', v') = 0 \qquad \forall v \in V_h\ , \tag{1.3}$$

die durch Subtraktion der Gleichung in (V) und (V_h) unter Beachtung von $V_h \subset V$ folgt.

Für die folgenden Abschätzungen benötigen wir die L^2–Norm

$$\|w\| = (\|w\|_0 =)(w,w)^{1/2} = \left(\int_0^1 w^2\,dx\right)^{1/2} ,$$

für die bekannntlich die Cauchy–Schwarzsche Ungleichung

$$|(v,w)| \le \|v\|\,\|w\| \tag{1.4}$$

gilt.

Der folgende Satz zeigt, daß u_h in gewissem Sinne die beste Approximation von u in Teilraum V_h darstellt; die zugrundeliegende Norm ist dabei

$$|v|_1 = \|v'\| , \qquad v \in V . \tag{1.5}$$

Wir werden später sehen, daß die so definierte Halbnorm in V — wegen der Nullrandbedingungen — sogar eine Norm darstellt.

Satz 1.7 *Für die Lösungen u von (V) bzw. u_h von (V_h) und jedes $v \in V_h$ gilt die Ungleichung*

$$|u - u_h|_1 \le |u - v|_1 . \tag{1.6}$$

Beweis: Für beliebiges $v \in V_h$ sei $w = u_h - v$. Damit ist $w \in V_h$ und mit w anstelle von v in der Fehlergleichung (1.3) erhält man

$$\begin{aligned} \|(u-u_h)'\|^2 &= ((u-u_h)', (u-u_h)') + ((u-u_h)', w') \\ &= ((u-u_h)', (u-u_h+w)') = ((u-u_h)', (u-v)') \\ &\le \|(u-u_h)'\|\,\|(u-v)'\| . \end{aligned}$$

Division durch $\|(u-u_h)'\|$ liefert die behauptete Abschätzung. □

Als Konsequenz des letzten Satzes kann man eine Abschätzung des Fehlers $|u-u_h|_1$ dadurch erhalten, daß man z. B. den Interpolationsfehler $|u-\tilde{u}_h|_1$ abschätzt, wobei $\tilde{u}_h$ die stetige, stückweise lineare Interpolierende von u bezeichnet,

$$\tilde{u}_h \in V_h : \tilde{u}_h(x_j) = u(x_j) , \qquad j = 0, \ldots, n .$$

Nach bekannten Ergebnissen eines Grundkurses in Numerik erhält man durch Abschätzung des Interpolationsfehlers auf jedem Teilintervall

$$|u'(x) - \tilde{u}_h'(x)| \le h \max_{0\le y\le 1} |u''(y)| , \tag{1.7}$$

$$|u(x) - \tilde{u}_h(x)| \le \frac{1}{2} h^2 \max_{0\le y\le 1} |u''(y)| . \tag{1.8}$$

Für die L^2–Norm der Ableitung folgt aus (1.7) durch Integration

$$\|(u-u_h)'\| \le h \max_{0\le y\le 1} |u''(y)| \; . \tag{1.9}$$

Für den maximalen Fehler der Galerkin–Approximation kann man mit Hilfe von (1.7) und der Cauchy–Schwarzschen Ungleichung zeigen, daß

$$|(u-u_h)(x)| \le h \max_{0\le y\le 1} |u''(y)| \; , \qquad x \in [0,1] \; , \tag{1.10}$$

denn

$$\begin{aligned} |u(x)-u_h(x)| &\le \int_0^1 |(u-u_h)'(y)|\, dy \\ &\le \left(\int_0^1 |(u-u_h)'(y)|^2 dy\right)^{1/2} \left(\int_0^1 1 dy\right)^{1/2} \\ &\le h \max_{0\le y\le 1} |u''(y)| \; , \qquad x \in [0,1] \; . \end{aligned}$$

Im Gegensatz zu (1.9) ist die letzte Abschätzung nicht optimal in dem Sinne, daß der kleinste Abstand von u zu einer Funktion in V_h bzgl. der Maximumnorm ja wegen (1.8) mindestens $\mathrm{O}(h^2)$ — und nicht wie $\mathrm{O}(h)$ in (1.10) — gegen Null strebt für $h \to 0$.

Die FEM mit einer zugehörigen ersten Fehlerabschätzung wird jetzt noch für eine allgemeinere Klasse von 2–Punkt–Randwertaufgaben bei gewöhnlichen Differentialgleichungen hergeleitet. Mit stetigen Funktionen p, q auf $[a,b]$, die den Voraussetzungen

$$0 < p_0 \le p(x) \; , \quad 0 \le q(x) \; , \quad a \le x \le b \; , \tag{1.11}$$

genügen, betrachten wir die Randwertaufgabe

$$\begin{aligned} -(pu')' + qu &= f \quad in \quad [a,b] \\ u(a) = u(b) &= 0 \; . \end{aligned} \tag{D}$$

Das zugehörige Variationsproblem hat dann die Form

$$u \in V : (pu', v') + (qu, v) = (f, v) \qquad \forall v \in V \; , \tag{V}$$

mit demselben Raum V wie oben. Wir bezeichnen mit

$$a(v,w) := (pv', w') + (qv, w) \tag{1.12}$$

die zugehörige *Bilinearform* und wählen als zugrundeliegende Norm jetzt

$$\|v\|_1 = (\|v\|^2 + \|v'\|^2)^{1/2} , \qquad v \in V . \tag{1.13}$$

Mit Hilfe der Cauchy–Schwarzschen Ungleichung und der Hölderungleichung für Summen erhält man die folgende Beschränktheitsbedingung für die Bilinearform $a(.,.)$,

$$|a(v,w)| \le \alpha_1 \|v\|_1 \, \|w\|_1 , \qquad v, w \in V , \tag{1.14}$$

wobei $\alpha_1 = \max(p_1, q_1)$, $p_1 = \max_{a \le x \le b} |p(x)|$, $q_1 = \max_{a \le x \le b} |q(x)|$. Weiter sieht man leicht, daß

$$\|v\| \le (b-a)|v|_1 \tag{1.15}$$

für alle $v \in C^1(a,b)$ mit $v(a) = 0$. Damit gilt (1.15) auch für alle $v \in V$. Aus (1.15) folgt unmittelbar

$$\|v\|_1 \le (1 + (b-a)^2)^{1/2} |v|_1 , \qquad v \in V , \tag{1.16}$$

was beweist, daß $|v|_1$ und $\|v\|_1$ äquivalente Normen auf V sind. Schließlich beweist man über die Voraussetzungen in (1.11) die folgende Abschätzung nach unten für die Bilinearform $a(.,.)$,

$$a(v,v) \ge \alpha_0 \|v\|_1^2 , \qquad v \in V , \tag{1.17}$$

mit $\alpha_0 = p_0(1 + (b-a)^2)^{-1}$, denn wegen (1.11) und (1.16) gilt

$$\begin{aligned} a(v,v) &\ge p_0 |v|_1^2 + \underbrace{q_{\min}}_{\ge 0} \|v\|^2 \ge p_0 |v|_1^2 \\ &\ge p_0 (1 + (b-a)^2)^{-1} \|v\|_1^2 , \qquad v \in V . \end{aligned}$$

Als Konsequenz von (1.17) sei noch bemerkt, daß mit $a(v,w)$ auch ein Skalarprodukt erklärt wird, und die zugehörige Norm äquivalent zu $\|.\|_1$ — und damit auch zu $|.|_1$ — ist,

$$\alpha_0^{1/2} \|v\|_1 \le a(v,v)^{1/2} \le \alpha_1^{1/2} \|v\|_1 , \qquad v \in V .$$

Die Norm $\|v\|_a = a(v,v)^{1/2}$ heißt auch die zu (D) bzw. (V) gehörige *Energienorm.*

Die *Galerkin Methode* besteht nun darin, zu einem gegebenen endlichdimensionalen Teilraum $V_h \subset V$ eine Lösung $u_h \in V_h$ zu suchen von

$$a(u_h, v) = (f, v) \qquad \forall v \in V_h \ . \tag{1.18}$$

Diese Gleichung nennt man wie oben die zugehörige *Variationsgleichung*, die äquivalent dazu ist, ein lineares Gleichungssystem zur Bestimmung der Koeffizienten von $u_h = \sum_{j=1}^{n-1} \xi_j \varphi_j$ zu lösen, wobei die zugehörige Matrix $A = (a_{ij})$, $a_{ij} = a(\varphi_i, \varphi_j)$, wegen der obigen Voraussetzungen — insbesondere wegen (1.17) — wieder symmetrisch und positiv definit ist.

Der Fehler erfüllt nun die *Fehlergleichung*

$$a(u - u_h, v) = 0 \qquad \forall v \in V_h \ , \tag{1.19}$$

woraus man durch einen analogen Schluß wie oben unter Benutzung von (1.14) und (1.17) zeigen kann, daß

$$\|u - u_h\|_1 \leq \left(\frac{\alpha_1}{\alpha_0}\right)^{1/2} \|u - v\|_1 \qquad \forall v \in V_h \ . \tag{1.20}$$

Man hat nämlich

$$\begin{aligned} a(u - u_h, u - u_h) &\overset{(1.19)}{=} a(u - u_h, u) \overset{(1.19)}{=} a(u - u_h, u - v) \\ &\leq a(u - u_h, u - u_h)^{1/2} a(u - v, u - v)^{1/2} \ , \quad v \in V_h \ , \end{aligned}$$

wobei die letzte Ungleichung die Cauchy–Schwarzsche Ungleichung für das durch $a(.,.)$ definierte Skalarprodukt darstellt. Division durch $a(u - u_h, u - u_h)^{1/2}$ und Anwendung von (1.14), (1.17) liefert

$$\begin{aligned} \alpha_0^{1/2} \|u - u_h\|_1 &\leq a(u - u_h, u - u_h)^{1/2} \\ &\leq a(u - v, u - v)^{1/2} \leq \alpha_1^{1/2} \|u - v\|_1 \qquad \forall v \in V_h \ . \end{aligned}$$

Man nennt die Abschätzung (1.20) „quasi–optimal“, da die Galerkin–Approximation bis auf einen Faktor gleich dem minimalen Abstand von u zu dem Teilraum V_h ist, wobei als Norm $\|.\|_1$ zugrundeliegt. Es sei bemerkt, daß außerdem eine „optimale“Abschätzung bzgl. der durch $a(.,.)$ induzierten Norm vorliegt, da dann der Faktor gerade Eins ist. Es wird nun zu klären sein, wie man den minimalen Abstand von u zu V_h bzgl. $\|.\|_1$ — oder bzgl. äquivalenter Normen – abschätzen kann.

1.4 FEM für die Poisson–Gleichung

Als einfachsten Modellfall einer partiellen Differentialgleichung 2. Ordnung betrachten wir das sog. *erste Randwertproblem der Potentialtheorie:*

Gesucht eine Funktion $u \in C(\overline{G}) \cap C^2(G)$ *mit*

$$\begin{aligned} u &= 0 \quad \text{auf} \quad \partial G\,, \\ -\Delta u &= f \quad \text{in} \quad G\,. \end{aligned} \tag{D}$$

Dabei sind G ein offenes, beschränktes Gebiet im $\mathbb{R}^2$, hier der Einfachheit halber das Einheitsquadrat $G = \{(x,y) \in \mathbb{R}^2 | 0 < x, y < 1\}$, mit dem Rand ∂G und $\Delta = (\partial^2/\partial x^2) + (\partial^2/\partial y^2)$ der *Laplace–Operator*. Das Problem (D) ist das mathematische Modell einer über G eingespannten elastischen Membran mit Massedichte $\rho = 1$, welche durch eine äußere Kraft mit Kraftdichte $-f(x,y)$ eine Auslenkung $u(x,y)$ erfährt. Die Randbedingung $u|_{\partial G} = 0$ entspricht der Forderung, daß die Membran am Rand eingespannt ist. Die physikalische Grundannahme für die elastische Membran ist die Proportionalität ihrer *elastischen Energie* F_E zur Flächenänderung (o.B.d.A. Proportionalitätsfaktor $\mu = 1$; $u_x = \partial u/\partial x$, $u_y = \partial u/\partial y$)

$$F_E(u) := \int_G \sqrt{1 + u_x^2 + u_y^2}\, dx\, dy - \int_G dx\, dy\ .$$

Da wir nur kleine Auslenkungen u, u_x, u_y betrachten wollen, entnehmen wir der Taylorentwicklung um $(u_x, u_y) = (0,0)$

$$\sqrt{1 + u_x^2 + u_y^2} = 1 + \frac{1}{2}\left(u_x^2 + u_y^2\right) + \mathrm{O}\left((u_x^2 + u_y^2)^2\right)$$

die Näherungsformel für die elastische Energie

$$F_E(u) \sim \frac{1}{2}\int_G \left(u_x^2 + u_y^2\right)\, dx\, dy\ .$$

Der äußeren Kraft entspricht ferner die *potentielle Energie*

$$F_A(u) := -\int_G fu\, dx\, dy\ .$$

Die stabile Gleichgewichtslage $u(x,y)$ des physikalischen Systems wird nun dadurch charakterisiert (Hamiltonsches Prinzip) , daß sie der gesamten Energie

$$F(u) = \frac{1}{2}\int_G \left\{u_x^2 + u_y^2\right\}\, dx\, dy - \int_G fu\, dx\, dy$$

ein Minimum erteilt, d. h. es ist das folgende *Minimierungsproblem* zu lösen,

$$\min_{v \in V} F(v) = F(u) \; . \tag{M}$$

Dabei sind als Vergleichsfunktionen $v \in V$ alle stetigen auf ∂G verschwindende Funktionen mit stückweise stetigen beschränkten ersten Ableitungen zugelassen. Folglich ist für jedes in diesem Sinne zulässige v

$$\frac{d}{d\varepsilon} F(u + \varepsilon v)\Big|_{\varepsilon=0} = 0 \; .$$

Wir werten dies aus und gelangen zu

$$\int_G \{u_x v_x + u_y v_y\} \, dx \, dy - \int_G f v \, dx \, dy = 0$$

bzw. abgekürzt geschrieben

$$\int_G \nabla u \cdot \nabla v \, dx \, dy = \int_G f v \, dx \, dy \quad \forall v \in V \; . \tag{V}$$

Dies ist das zugehörige *Variationsproblem* bzw. die *Variationsgleichung*. Ist die stabile Gleichgewichtslage u sogar aus $C^2(G)$, so ergibt die Anwendung der Greenschen Formel

$$-\int_G (\Delta u) v \, dx \, dy = \int_G f v \, dx \, dy$$

für alle zulässigen Funktionen v. Hieraus folgt dann notwendig das Bestehen der Beziehung (D) in G. Man nennt (D) die *Eulersche Differentialgleichung* zum Variationsproblem (V).

Wir werden nun an Hand des Modellproblems (D) die Vorgehensweise der *Differenzenverfahren* und der *Methode der finiten Elemente* beschreiben.

Differenzenverfahren

Mit einer natürlichen Zahl N und zugehöriger Schrittweite $h = \frac{1}{N}$ werden die Gitterpunktmengen erklärt

$$\begin{aligned} G_h &= \{(ih, jh) | i, j = 0, \ldots, N\} \cap G \\ \partial G_h &= \{(ih, jh) | i, j = 0, \ldots N\} \cap \partial G \; . \end{aligned}$$

Für eine Funktion $v \in C(\overline{G})$ bezeichnet $r_h v$ die *Restriktion* auf das Gitter G_h. Ist $v \in C^3(\overline{G})$, so gilt mit den üblichen zentralen Differenzenquotienten 2. Ordnung in Punkten $(x,y) \in G_h$

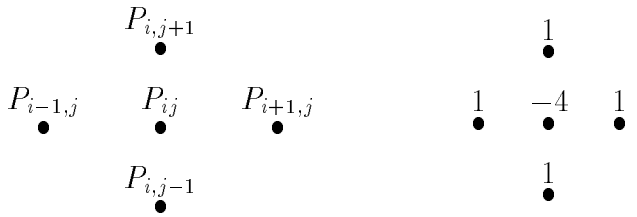
$$\begin{aligned} v_{xx}(x,y) &= h^{-2}\{v(x+h,y) - 2v(x,y) + v(x-h,y)\} + \mathrm{O}(h)\,, \\ v_{yy}(x,y) &= h^{-2}\{v(x,y+h) - 2v(x,y) + v(x,y-h)\} + \mathrm{O}(h)\,, \end{aligned}$$

was man leicht mit Hilfe der entsprechenden Taylorentwicklungen ableitet. Die einfachste Differenzenapproximation der Randwertaufgabe (D) lautet dann:

Gesucht eine Funktion u_h *auf* $\overline{G}_h = G_h \cup \partial G_h$ *mit*

$$\begin{aligned} -\Delta_h u_h(x,y) &:= h^{-2}\{-u_h(x+h,y) - u_h(x,y+h) - u_h(x-h,y) \\ &\quad -u_h(x,y-h) + 4u_h(x,y)\} = r_h f(x,y) \;\; in \;\; G_h \\ und \;\; u_h(x,y) &= 0 \;\; auf \;\; \partial G_h\,. \end{aligned} \tag{D_h}$$

Wir bemerken, daß dabei der Wert von Δu im Punkt $P_{i,j} = (ih, jh) \in G$ durch einen Mittelwert $\Delta_h u_h$ von u_h über die Punkte aus folgendem Punktestern approximiert wird (5–Punkte–Formel):



Dies führt zu der Bezeichnung *Mittelwertoperator* für Δ_h. Ein allgemeinerer Mittelwertoperator zur Approximation des Laplace–Operators Δ ist gegeben durch

$$\Delta_h u_h(x,y) = h^{-2} \sum_{i,j} b_{ij} u_h(x+ih, y+jh)$$

mit den Eigenschaften

$$\begin{aligned} \sum_{i,j} b_{ij} &= \sum_{i,j} i b_{ij} = \sum_{i,j} j b_{ij} = \sum_{i,j} ij\, b_{ij} = 0\,, \\ \frac{1}{2}\sum_{i,j} i^2 b_{ij} &= \frac{1}{2}\sum_{i,j} j^2 b_{ij} = 1\,. \end{aligned}$$

Diese Differenzenoperatoren haben bzgl. des Laplace–Operators die Konsistenzordnung $\nu = 1$; d. h. für jede Funktion $v \in C^3(\overline{G})$ ist bzgl. der Maximumnorm $\|.\|_\infty$ im $\mathbb{R}^{N\times N}$

$$\|\Delta_h r_h v - r_h \Delta v\|_\infty = \mathrm{O}(h) \ .$$

Dies liest man unter Verwendung der obigen Bedingungen direkt aus der durch Taylorentwicklung gewonnenen Beziehung ab:

$$\begin{aligned}
\Delta_h r_h v(x,y) &= h^{-2} \sum_{i,j} b_{ij} r_h v(x+ih, y+jh) \\
&= h^{-2} r_h v(x,y) \sum_{i,j} b_{ij} + h^{-1} \sum_{i,j} b_{ij} \{ i r_h v_x + j r_h v_y \}(x,y) + \\
&\quad + \frac{1}{2} \sum_{i,j} b_{ij} \left\{ i^2 r_h v_{xx} + 2ij r_h v_{xy} + j^2 r_h v_{yy} \right\}(x,y) + \mathrm{O}(h) \\
&= r_h(v_{xx} + v_{yy}) + \mathrm{O}(h) \ , \qquad (x,y) \in G_h \ .
\end{aligned}$$

Gilt darüberhinaus noch

$$b_{ij} \geq 0 \ , \quad |i| + |j| \neq 0 \ , \quad b_{00} < 0 \ ,$$

so heißt der Differenzenoperator Δ_h von *positivem Typ* . Dies trifft für die obige einfache 5–Punkte–Formel offenbar zu. Differenzenverfahren dieser Art gestatten eine besonders vollständige mathematische Analyse. Für sie gilt ein sogenanntes *diskretes Maximumprinzip*, mit dessen Hilfe man für jede Gitterfunktion v_h mit $v_h|_{\partial G_h} = 0$ die *a–priori–Abschätzung* beweisen kann

$$\|v_h\|_\infty \leq C \|\Delta_h v_h\|_\infty$$

(C unabhängig von der Schrittweite h).

Hieraus folgt sofort die eindeutige Lösbarkeit der endlich dimensionalen Probleme (D_h), denn die zugehörigen homogenen Aufgaben mit $r_h f \equiv 0$ können offenbar nur die triviale Lösung $u_h \equiv 0$ besitzen. Ferner erhält man für die Lösungen $u \in C^3(\overline{G})$ von (D) und u_h von (D_h) die Fehlerabschätzung

$$\begin{aligned}
\|r_h u - u_h\|_\infty &\leq C \|\Delta_h (r_h u - u_h)\|_\infty \\
&\leq C \|\Delta_h r_h u - r_h \Delta u\|_\infty + C \|r_h \Delta u - \Delta_h u_h\|_\infty \\
&\leq \mathrm{O}(h) + C \|r_h f - r_h f\|_\infty = \mathrm{O}(h) \ .
\end{aligned}$$

Die Konvergenzordnung eines Differenzenverfahrens von positivem Typ für die Randwertaufgabe (D) ist gleich seiner (lokalen) Konsistenzordnung. Werden statt der Randbedingung $u|_{\partial G} = 0$ jedoch kompliziertere Randbedingungen, wie z. B. $(u + \partial u/\partial n)|_{\partial G} = g$, gestellt und ist G ein allgemeineres Gebiet im $\mathbb{R}^2$, so sind die zugehörigen Differenzenapproximationen Δ_h, insbesondere solche höherer Konsistenzordnung, meist nicht von positivem Typ. In diesen Fällen ist die Frage nach der Konvergenz und der Konvergenzordnung der Verfahren wesentlich schwerer zu beantworten.

Methode der Finiten Elemente

Im Gegensatz zu den Differenzenverfahren, die direkt an der Differentialgleichung (D) ansetzen, geht die Methode der finiten Elemente als Spezialfall des allgemeinen *Ritzschen Projektionsverfahrens* von der ursprünglichen variationellen Formulierung (M) des Problems aus:

$$\begin{aligned} &\textit{Minimiere} \quad F(u) = \int_G \left(\frac{1}{2} |\nabla u|^2 - fu \right) dx\, dy \quad \textit{für} \quad u \in V \ , \\ &\textit{mit} \qquad\quad V = \{ u \in C(\overline{G}) \quad \textit{mit stückweise stetigen (beschränkten) ersten Ableitungen und } u|_{\partial G} = 0 \} \end{aligned} \tag{M}$$

Es ist naheliegend, das Problem dadurch zu approximieren, daß lediglich über endlichdimensionale Teilräume $V_h \subset V$ minimiert wird:

$$\textit{Minimiere} \quad F(v_h) \quad \textit{für} \quad v_h \in V_h \ . \tag{M_h}$$

Dabei beschreibt der Parameter h die „Feinheit“ von V_h, d. h. $\dim(V_h)$ ist eine Funktion von h. Läßt man $h \to 0$ gehen, so besteht die Hoffnung, daß auch der Approximationsfehler $u - u_h \to 0$ konvergiert (in einem geeigneten Sinne).

Mit dieser Fragestellung werden wir uns im folgenden hauptsächlich beschäftigen.

Offensichtlich ist eine sehr vielfältige Wahl der Räume $V_h \subset V$ möglich, z. B.:

1. Eigenfunktionen des kontinuierlichen Problems (D),
2. Polynomräume (Legendre, Tchebycheff usw.),
3. trigonometrische Polynome,
4. rationale Funktionen,
5. Wavelets, usw.

Dies ist vor allem ein Problem der Approximationstheorie. Vom numerischen Standpunkt aus müssen jedoch einige spezielle Grundforderungen an die Räume V_h gestellt werden. Hierauf soll etwas näher eingegangen werden.

Ist $N_h = \dim(V_h)$ und $\{w_h^i \ , \ i = 1, \ldots, N_h\}$ eine Basis von V_h, so gestattet jedes $v_h \in V_h$ eine Darstellung

$$v_h = \sum_{i=1}^{N_h} a^i w_h^i \ .$$

Da V_h endlichdimensionaler Vektorraum ist, erhält man durch Nullsetzen der 1. Variation von $F(.)$ bei $u_h \in V_h$ das zu (M_h) äquivalente Problem

$$\int_G \nabla u_h \nabla v_h dx\, dy = \int_G f v_h dx\, dy\ , \qquad \forall v_h \in V_h\ , \tag{V_h}$$

bzw.

$$\sum_{i=1}^{N_h} a^i \int_G \nabla w_h^i \nabla w_h^j dx\, dy = \int_G f w_h^j dx\, dy\ , \qquad j = 1, \ldots, N_h\ .$$

Dies ist ein Matrix–Problem im $\mathbb{R}^{N_h}$,

$$Aa = b \tag{V_h}$$

mit

$$A := \left(\int_G \nabla w_h^i \nabla w_h^j\, dx\, dy \right)_{i,j=1}^{N_h}\ , \qquad a := \left(a^1, \ldots, a^{N_h}\right)$$

und

$$b := \left(\int_G f w_h^j dx\, dy \right)_{j=1}^{N_h}\ .$$

Das Problem (V_h) ist eindeutig lösbar, da A Gramsche Matrix der Basis $\{w_h^i\ ,\ i = 1, \ldots, N_h\}$ bzgl. des Skalarproduktes $\int_G \nabla v \nabla w\, dx\, dy$ ist. Insbesondere ist A symmetrisch und positiv definit. Durch Subtraktion der Gleichungen (V) und (V_h) erhalten wir die *Fehlergleichung*

$$\int_G \nabla(u - u_h) \nabla v_h\, dx\, dy = 0\ , \qquad \forall v_h \in V_h\ .$$

Dies ergibt dann, wie später in Abschnitt 2.2 gezeigt wird, die erste Fehlerabschätzung

$$\int_G |\nabla(u - u_h)|^2 dx\, dy \le \inf_{v_h \in V_h} \int_G |\nabla(u - v_h)|^2 dx\, dy\ .$$

Hieran schließen sich nun vom numerischen Standpunkt die folgenden *drei zentralen Forderungen* an die Teilräume $V_h \subset V$ an:

1. Die Lösung u von (D) soll für $h \to 0$ gut durch Funktionen aus V_h approximierbar sein:

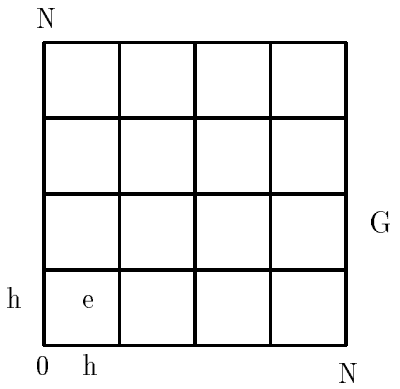
$$\inf_{v_h \in V_h} \int_G |\nabla(u - v_h)|^2 dx\, dy \to 0 \qquad (h \to 0)\,.$$

2. Die Elemente $\int_G \nabla w_h^i \nabla w_h^j \, dx\, dy$ der Matrix A sollen einfach berechenbar sein.

3. Die Matrix A darf für $h \to 0$ nicht zu schlecht konditioniert werden (wird unten noch näher erläutert).

Es hat sich gezeigt, daß unter allen möglichen Klassen von Teilräumen $V_h \subset V$ die sog. Finite–Elemente–Räume diesen drei Anforderungen besonders gut genügen. Zu ihrer Definition zerlegen wir das Gebiet G in Quadrate e der Seitenlänge $h = \frac{1}{N}$.

(Später werden wir vor allem auch Zerlegungen in Dreiecke betrachten.)
Die Teilräume $V_h \subset V$ sind dann definiert durch



$V_h := \{v_h \in C(\overline{G}) |\ v_h$ auf jedem Quadrat e Polynom vom Grade $\leq m$; $v_h = 0$ auf $\partial G\}$.

Wir betrachten jetzt den einfachsten Fall für $m = 1$, bei dem die Funktionen $v_h \in V_h$ auf jedem Quadrat e bilinear sind:

$$v_h|_e(x, y) = \alpha xy + \beta x + \gamma y + \delta\ .$$

Diese Setzung ist sinnvoll, denn durch Vorgabe der Funktionswerte von v_h in den Eckpunkten aller Quadrate e ist v_h offenbar eindeutig bestimmt und stetig.

Eine Basis von V_h erhalten wir, wenn wir alle Eckpunkte der Quadrate durchnumerieren zu $P_{i,j} = (x_i, y_j)\,,\ x_i = ih\,,\ y_j = jh\,,\ i, j = 1, \ldots, N-1$, und dann setzen:

$$w_h^{ij} \in V_h : w_h^{ij}(P_{l,k}) = \delta_{il}\delta_{jk}\ , \quad \begin{matrix} i, j = 1, \ldots, N-1\ , \\ l, k = 1, \ldots, N-1\ . \end{matrix}$$

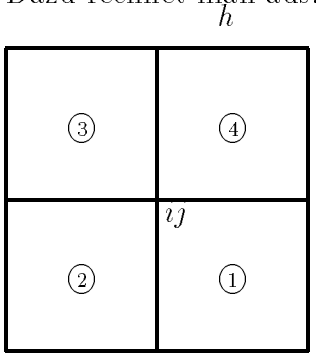
Jede Funktion $v_h \in V_h$ besitzt dann offenbar die Darstellung

$$v_h = \sum_{i,j=1}^{N_h} v_h(P_{i,j}) w_h^{ij} \, .$$

Die Elemente der Matrix A ergeben sich wie folgt:

$$\int_G \nabla w_h^{ij} \nabla w_h^{lk} dx\, dy = \begin{cases} 0 & , \quad 2 \le \max\{|i-l|\,,\ |j-k|\} \, , \\ 8/3 & , \quad 0 = \max\{|i-l|\,,\ |j-k|\} \, , \\ -1/3 & , \quad 0 < \max\{|i-l|\,,\ |j-k|\} \le 1 \, . \end{cases}$$

Dazu rechnet man aus:



$$a_{ij}^{kl} = \iint_G \nabla w_h^{ij} \nabla w_h^{lk} dx\, dy =$$

$$\begin{cases} = \underset{①}{\iint} + \underset{②}{\iint} + \underset{③}{\iint} + \underset{④}{\iint} \\ \quad \text{für} \ \ |i-k| \le 1 \, , \ |j-l| \le 1 \\ = 0 \qquad \text{sonst} \end{cases}$$

$$w_h^{ij}(x,y) = X_i(x) Y_j(y) \ : \qquad w_h^{ij}(x_l, y_k) = \delta_{il}\delta_{jk}$$

$$X_i(x) = \begin{cases} \frac{1}{h}(x - x_{i-1}) & , \quad x_{i-1} \le x \le x_i \, , \\ \frac{1}{h}(x_{i+1} - x) & , \quad x_i \le x \le x_{i+1} \, , \\ 0 & , \quad \text{sonst} \, , \end{cases}$$

$$Y_j(y) = \begin{cases} \frac{1}{h}(y - y_{j-1}) & , \quad y_{j-1} \le y \le y_j \, , \\ \frac{1}{h}(y_{j+1} - y) & , \quad y_j \le y \le y_{j+1} \, , \\ 0 & , \quad \text{sonst} \, . \end{cases}$$

$$\begin{aligned} a_{ij}^{lk} &= \iint_G \left(\frac{\partial w_h^{ij}}{\partial x} \frac{\partial w_h^{lk}}{\partial x} + \frac{\partial w_h^{ij}}{\partial y} \frac{\partial w_h^{lk}}{\partial y} \right) dx\, dy \\ &= \iint_G \left(Y_j Y_k \frac{dX_i}{dx} \frac{dX_l}{dx} + X_i X_l \frac{dY_j}{\partial y} \frac{dY_k}{dy} \right) dx\, dy \end{aligned}$$

wobei

$$\frac{dX_i}{dx} = \pm\frac{1}{h}\,, \quad \frac{dY_j}{dy} = \pm\frac{1}{h} \quad \text{auf} \quad ① — ④$$

$i = l\,,\ j = k$:

$$① \qquad \frac{dX_i}{dx} = -\frac{1}{h}\,, \quad \frac{dY_j}{dy} = \frac{1}{h}\,,$$

$$\begin{aligned} \iint_{①} &= \frac{1}{h^2} h \int_{y_{j-1}}^{y_j} \frac{1}{h^2}(y - y_{j-1})^2 dy + \frac{1}{h^2} h \int_{x_i}^{x_{i+1}} \frac{1}{h^2}(x_{i+1} - x)^2 dx \\ &= \frac{1}{h^3}\int_0^h z^2 dz + \frac{-1}{h^3}\int_h^0 z^2 dz = \frac{2}{3}\,, \end{aligned}$$

analog für ②, ③, ④. Insgesamt: $a_{ij}^{ij} = \frac{8}{3}$ („Assemblierung“) .
$l = i + 1\,,\ k = j$:

$$① \qquad \frac{dX_{i+1}}{dx} = \frac{1}{h}\,, \quad \frac{dX_i}{dx} = -\frac{1}{h}\,,$$

$$\begin{aligned} \iint_{①} &= \frac{-1}{h^2} h \int_{y_{j-1}}^{y_j} \frac{1}{h^2}(y - y_{j-1})^2 dy + \frac{1}{h^2} h \int_{x_i}^{x_{i+1}} \frac{1}{h^2}(x_{i+1} - x)(x - x_i)\, dx \\ &= \frac{-1}{h^3}\int_0^h z^2 dz + \frac{1}{h^3}\underbrace{\int_0^h z(h - z)\, dz}_{+\frac{1}{2}h^3 - \frac{1}{3}h^3} = -\frac{1}{3} + \frac{1}{2} - \frac{1}{3} = -\frac{1}{6} \end{aligned}$$

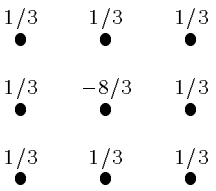
$$④ \qquad \frac{dX_{i+1}}{dx} = \frac{1}{h}\,,$$

$$\begin{aligned} \iint_{④} &= \frac{-1}{h^2} h \int_{y_j}^{y_{j+1}} \frac{1}{h^2}(y_{j+1} - y)^2 dy + \frac{1}{h^2} h \int_{x_i}^{x_{i+1}} \frac{1}{h^2}(x_{i+1} - x)(x - x_i)\, dx \\ &= -\frac{1}{3} + \frac{1}{2} - \frac{1}{3} = -\frac{1}{6} \Rightarrow a_{ij}^{i+1,j} = -\frac{1}{3}\,. \end{aligned}$$

Die Gleichung (V_h) bekommt somit die Gestalt $(a = (u_h(P_{1,1}), \ldots, u_h(P_{N,N}))$

$$\begin{aligned}(Aa)_{i,j} &= h^{-2}\Big\{\frac{8}{3}u_h(P_{i,j}) - \frac{1}{3}u_h(P_{i+1,j}) - \frac{1}{3}u_h(P_{i-1,j}) - \frac{1}{3}u_h(P_{i,j+1}) \\ &\quad - \frac{1}{3}u_h(P_{i,j-1}) - \frac{1}{3}u_h(P_{i+1,j+1}) - \frac{1}{3}u_h(P_{i+1,j-1}) \\ &\quad - \frac{1}{3}u_h(P_{i-1,j+1}) - \frac{1}{3}u_h(P_{i-1,j-1})\Big\} \\ &= h^{-2}\int_G f w_h^{ij}\,dx\,dy\ , \qquad i,j = 1,\ldots,N-1\ .\end{aligned}$$

Offenbar ist dies gerade das Gleichungssystem, welches bei Anwendung eines Differenzenverfahrens mit einem 9–Punkte–Mittelwertoperator Δ_h mit dem zugehörigen Punkteschema



entsteht. In diesem einfachen Spezialfall entspricht die Methode der finiten Elemente also einem Differenzenverfahren. Bei allgemeineren Räumen V_h (insbesondere bei Dreieckzerlegungen von G) ist eine solche Zuordnung jedoch nicht mehr möglich.

Es sei hier bemerkt, daß bei einer gleichförmigen Zerlegung in Dreiecke — man spricht dann von einer *gleichförmigen Triangulierung* — wieder eine besonders einfache Gestalt der Gesamtsteifigkeitsmatrix A entsteht:

$$\begin{pmatrix}
4 & -1 & 0 & 0 & -1 & 0 & & & \\
-1 & 4 & -1 & 0 & 0 & -1 & 0 & & \\
0 & -1 & 4 & -1 & 0 & 0 & -1 & 0 & \\
0 & 0 & -1 & 4 & -1 & 0 & 0 & -1 & 0 \\
-1 & 0 & 0 & -1 & 4 & -1 & 0 & 0 & -1 \\
 & & & & \ddots & & & & \\
 & & & -1 & 0 & 0 & -1 & 4 & -1 \\
 & & & 0 & -1 & 0 & 0 & -1 & 4
\end{pmatrix}$$

Der zugehörige Differenzenstern ist gerade der der einfachsten 5–Punkt–Differenzenapproximation des Laplace–Operators, den wir schon oben im Zusammenhang mit

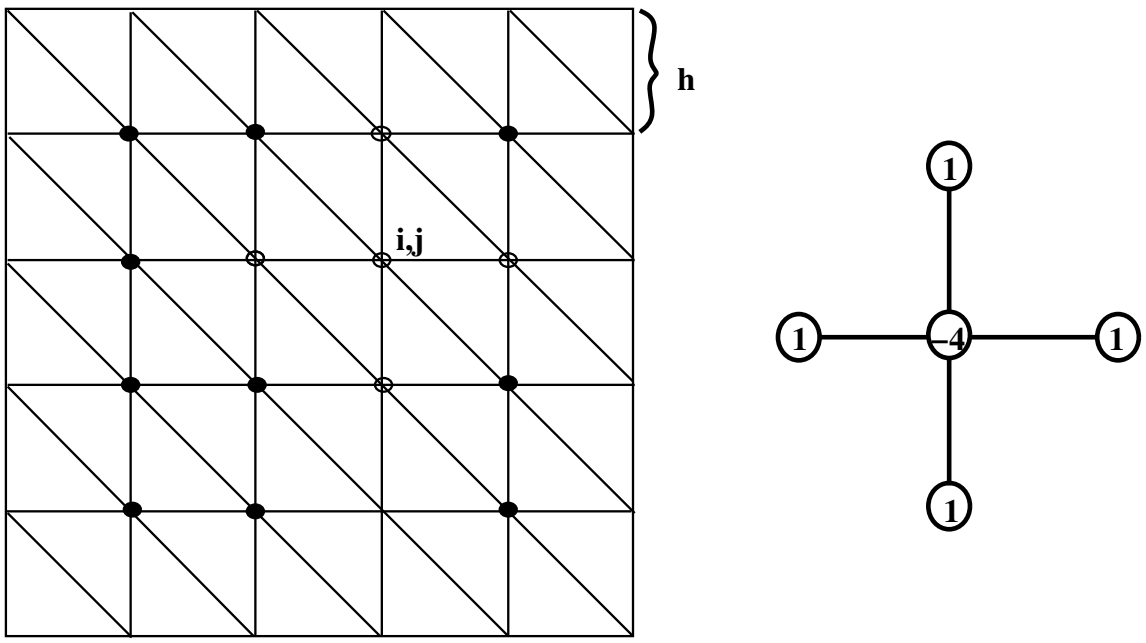


Abbildung 1.4: Gleichförmige Triangulierung

den Differenzenapproximationen analysiert haben. Die rechte Seite des entstehenden Gleichungssystems unterscheidet sich allerdings von dem des Differenzenverfahrens, da dort für jeden Gitterpunkt $P_{i,j}$ Mittelungen (nach Division durch h^2)

$$\frac{1}{h^2}\int_G f w_h^{ij}\,dx\,dy$$

stehen, wobei die Integration nur über die höchstens 6 Dreiecke erfolgt, die den Eckpunkt $P_{i,j}$ gemeinsam haben.

Die Forderung 2. nach der einfachen Berechenbarkeit der Matrixelemente von A ist bei Verwendung stückweise polynomialer Funktionen meist erfüllt. Später werden wir die Approximationsbeziehung

$$\inf_{v_h\in V_h}\int_G |\nabla(u-v_h)|^2\,dx\,dy = \mathrm{O}(h^2)$$

zeigen, so daß auch die erste Forderung 1. erfüllt ist.

Die dritte Forderung zielt auf eine gute Auflösbarkeit der linearen Gleichungssysteme (V_h). Die *Kondition* κ der Matrix A kann definiert werden als Quotient von größtem und kleinstem Eigenwert von A:

$$\kappa := \frac{\lambda_{\max}}{\lambda_{\min}}\,.$$

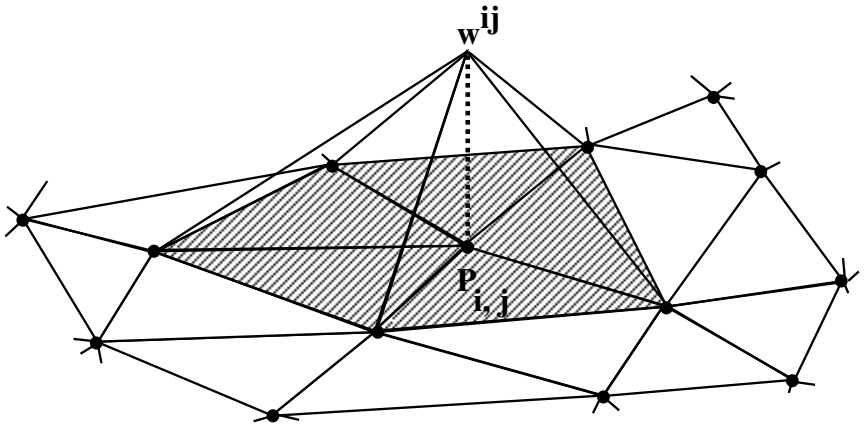


Abbildung 1.5: Die Basisfunktionen w^{ij} bei Triangulierung

Ist bei der Auflösung des linearen Gleichungssystems (V_h) die Information über die rechte Seite b mit einem Fehler δb behaftet, so gelten für die Lösungen der Gleichungssysteme

$$Aa = b\ , \quad A(a+\delta a) = b + \delta b$$

die Abschätzungen ($\|.\|$ Euklidische Vektornorm)

$$\begin{aligned} \|b\| &= \|Aa\| &\leq \lambda_{\max}\|a\| \\ \lambda_{\min}\|\delta a\| &\leq \|A\delta a\| &= \|\delta b\| \end{aligned}$$

wegen

$$\|A^{-1}b\| \leq \lambda_{\max}(A^{-1})\|b\| = \lambda_{\min}^{-1}\|b\| \Rightarrow \|Aa\| \geq \lambda_{\min}\|a\|,$$

und somit

$$\frac{\|\delta a\|}{\|a\|} \leq \frac{\|\delta b\|}{\lambda_{\min}} \cdot \frac{\lambda_{\max}}{\|b\|} = \kappa \frac{\|\delta b\|}{\|b\|}\ ,$$

mit der *Konditionszahl* $\kappa = \kappa(A) := \lambda_{\max}/\lambda_{\min}$. Ist nun $\kappa \sim 10^8$, so muß zur sicheren Erzielung einer brauchbaren Güte der Lösung $a + \delta a$ von (V_h) schon mit doppelter Rechnergenauigkeit gearbeitet werden. Es sei hier bemerkt, daß für die obigen Beispiele sowie die in Kapitel 6 betrachteten Verfahren gilt $\kappa(A) = \mathrm{O}(h^{-2})$. Damit hängt die $\mathrm{O}(h)$–Ordnung der Konditionszahl $\kappa(A)$ nicht von der Wahl der speziell gewählten Basisfunktionen sondern von der Ordnung des zugrunde liegenden Differentialoperators ab.

Abschließend sei bemerkt, daß sich die Methode der Finiten Elemente besonders bei der Behandlung komplizierter Randwertprobleme auf allgemeineren Gebieten $G \subset \mathbb{R}^n$ den Differenzenverfahren als überlegen erwiesen hat.

1.5 Die Hilberträume $L^2(G)$, $H^1(G)$, $H_0^1(G)$

Die variationellen Formulierungen der ein- bzw. zweidimensionalen Randwertprobleme, wie sie in den vorigen Abschnitten aufgestellt wurden, reichen insoweit nicht aus, als eine mathematische Theorie zur Sicherung von Existenz und Eindeutigkeit von Lösungen gewisse größere Räume benötigt, die insbesondere vollständig sein müssen. Zu diesem Zweck führen wir hier kurz die benötigten Räume ein. Im Rahmen der in Abschnitt 2.1 dargestellten Hilbertraum–Theorie elliptischer Differentialgleichungen werden diese Räume dann verwendet. In diesem Abschnitt geben wir dann noch die geeigneten variationellen Formulierungen der betrachteten Modellprobleme in den größeren Räumen an.

Im eindimensionalen Fall, d. h. für ein Intervall (a, b), bezeichnet $L^2(a, b)$ den Raum der quadratisch integrierbaren Funktionen, d. h. die L^2–Norm

$$\|v\| = \|v\|_0 = \int_a^b v(x)^2 \, dx$$

bleibt endlich. $L^2(a, b)$ ist ein Hilbertraum mit Skalarprodukt

$$(v, w) = \int_a^b v(x) w(x) \, dx \; ;$$

wir nehmen hier an, daß die Funktionen reellwertig sind. Zum Nachweis der Vollständigkeit von $L^2(a, b)$ benötigt man Kenntnisse aus der Theorie Lebesguescher Integral, wozu z. B. auf [For84, Wal92] verwiesen sei. Für unsere Zwecke reicht es aus, sich das übliche Riemann–Integral vorzustellen und typischerweise Funktionen zu nehmen, die stückweise stetig, möglicherweise unbeschränkt sind, aber endliche Norm $\int_a^b v^2 dx$ besitzen.

Solch eine Funktion ist zum **Beispiel** $v(x) = x^{-\beta}$, $x \in (0, 1)$, die für $\beta < \frac{1}{2}$ in $L^2(0, 1)$ liegt. Dazu rechnet man aus, daß

$$\|v\|_0^2 = \int_0^1 x^{-2\beta} dx = \frac{1}{1-2\beta} x^{1-2\beta} \Big|_0^1 = \frac{1}{1-2\beta} < \infty \, , \text{ falls } \beta < \frac{1}{2} \, .$$

Für eine beschränkte offene Menge G im $\mathbb{R}^n$, $n = 1, 2, 3$, benutzen wir die folgenden Bezeichnungen und Definitionen:

$$
\begin{aligned}
L^p(G) &= \left\{ u | \int_G |u(x)|^p dx < \infty \right\} \\
D^\alpha u &= D_1^{\alpha_i} \cdots D_n^{\alpha_n} \ , \quad D_i^{\alpha_i} = \frac{\partial^{\alpha_i}}{\partial x_i^{\alpha_i}} \ , \\
\alpha &= (\alpha_1, \ldots, \alpha_n) = \text{ Multiindex}, \qquad \alpha_i \in \{0\} \cup \mathbb{N} \ , \\
|\alpha| &= \alpha_1 + \ldots + \alpha_n = \text{ Ordnung von } \alpha \\
C^m(G) &= \{ u : G \to \mathbb{R} \ , \ m\text{-mal stetig differenzierbar auf } G \} \\
C^{m,p}(G) &= \{ u \in C^m(G) | D^\alpha u \in L^p(G) \quad \forall |\alpha| \le m \} \\
C^\infty(G) &= \cap_{m=0}^\infty C^m(G) \\
C_0^\infty(G) &= \{ u \in C^\infty(G) | \text{ supp } u \subset G \text{ kompakt} \} \ .
\end{aligned}
$$

Für den zuletzt definierten Raum hat man auch die Darstellung

$$
C_0^\infty(G) = \{ u \in C^\infty(G) | D^\alpha u = 0 \ \text{ auf } \ \partial G \quad \forall \alpha \} \, .
$$

Grundlage für die variationellen Formulierungen ist die *Greensche Formel*

$$
\begin{aligned}
\int_G u D_i v \, dx \ &= \ - \int_G (D_i u) v \, dx + \int_{\partial G} u v n_i ds \ , \quad u, v \in C^{1,p}(G) \, , \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad i = 1, \ldots, n \, ,
\end{aligned}
\tag{1.21}
$$

wobei $n = (n_1, \ldots, n_n)$ den Vektor der *äußeren Normalen* an ∂G bezeichnet. Wendet man die Greensche Formel $|\alpha|$-mal an und beachtet, daß die Ableitungen einer Funktion aus $C_0^\infty(G)$ am Rand verschwinden, so erhält man

$$
\int_G (D^\alpha u) \varphi \, dx = (-1)^{|\alpha|} \int_G u D^\alpha \varphi \, dx \quad \forall \varphi \in C_0^\infty(G) \, . \tag{1.22}
$$

Diese Beziehung ist gleichzeitig die Definitionsgleichung für die *schwache Ableitung* $D^\alpha u$ einer Funktion $u \in L^p(G)$.

Im Falle $n = 1$ schreiben wir auch $v' = Dv$ und definieren

$$
\begin{aligned}
H^1(a,b) \ &= \ H^{1,2}(a,b) = \{ v \in L^2(a,b) | v' \in L^2(a,b) \} \, , \\
H_0^1(a,b) \ &= \ \{ v \in H^1(a,b) | v(a) = v(b) = 0 \} \, .
\end{aligned}
$$

Als **Beispiel** betrachten wir $u = |x|^\gamma$, $x \in (-1,1)$; für $u'(x) = \gamma x^{\gamma - a}$, $x > 0$, bzw. $u'(x) = -\gamma x^{\gamma - 1}$, $x < 0$, ist die Beziehung (1.22) zu überprüfen, was offenbar nach Aufteilung des Intervalls $(-1,1)$ in $(-1,0)$ und $(0,1)$ zutrifft. Außerdem

überlegt man sich wie im vorigen Beispiel, daß $u' \in L^2(a,b)$, falls $\beta = 1-\gamma < \frac{1}{2}$ ist, d. h. $\gamma > \frac{1}{2}$. Unter dieser Annahme ist also $u \in H^1(a,b)$.

In $H^1(a,b)$ legen wir die in Abschnitt 1.3 eingeführten Normen bzw. Halbnormen $\|.\|_1$ bzw. $|.|_1$ zugrunde. Wir werden später noch ausführen, daß die Gleichung (1.15) auch in $H_0^1(a,b)$ gilt — sie heißt auch *Ungleichung von Poincaré–Friedrichs* — und damit stellen $|.|_1$ und $\|.\|_1$ auf $H_0^1(a,b)$ äquivalente Normen dar.

Mit $H_0^1(a,b)$ anstelle von V lautet jetzt die „richtige“ variationelle Formulierung des Modellproblems (D) von Abschnitt 1.3:

$$\textit{Gesucht} \quad u \in H_0^1(a,b) \quad \textit{mit} \quad (u',v') \;=\; (f,v) \quad \forall v \in H_0^1(a,b)\,. \tag{V}$$

Man sucht also eine Lösung in einem etwas größeren Raum $H_0^1(a,b)$ — entsprechendes gilt für das äquivalente Minimierungsproblem —, der bzgl. der durch

$$(v,w)_1 = \int_a^b (vw + v'w')dx \quad \text{oder} \quad (u',w')$$

definierten Skalarprodukte ein Hilbertraum ist.

In höheren Dimensionen bezeichnen wir analog

$$\begin{aligned} H^1(G) &= \left\{ v \in L^2(G) \mid \frac{\partial v}{\partial x_i} \in L^2(G)\,, \qquad i = 1,\dots,n \right\}\,, \\ H_0^1(G) &= \left\{ v \in H^1(G) \mid v = 0 \quad \text{auf} \quad \partial G \right\}\,. \end{aligned}$$

Mit dem *Gradienten* $\nabla v = \left(\frac{\partial v}{\partial x_1},\dots,\frac{\partial v}{\partial x_n}\right)^\top = (D_1 v,\dots,D_n v)^\top$ können wir ein Skalarprodukt und zugehörige Normen bzw. Halbnormen definieren,

$$\begin{aligned} (v,w)_1 &= \int_G (vw + \nabla v \cdot \nabla w)\,dx \qquad v,w \in H^1(G)\,, \\ \|v\|_1 &= (v,v)_1^{1/2} = \left(\int_G (|v|^2 + |\nabla v|^2)\,dx \right)^{1/2}\,, \qquad v \in H^1(G)\,, \\ |v|_1 &= \left(\int_G \left(\left(\frac{\partial v}{\partial x_1}\right)^2 + \left(\frac{\partial v}{\partial x_2}\right)^2 + \cdots + \left(\frac{\partial v}{\partial x_n}\right)^2 \right) dx \right)^{1/2} \\ &= \left(\int_G |\nabla v|^2 dx \right)^{1/2}\,. \end{aligned}$$

Im Gegensatz zum 1-dimensionalen Beispiel hat man in 2-d. mit $u(x) = |x|^\gamma$, $G = \{x_1^2 + x_2^2 \le 1\}$ eine $H^1(G)$–Funktion für alle $\gamma > 0$ (vgl. Übungsaufg. 4).

Die zur Poisson–Gleichung (vgl. (D), in Abschnitt 1.4) gehörige variationelle Formulierung lautet folgendermaßen:

$$\textit{Gesucht} \quad u \in H_0^1(G) \quad \textit{mit} \quad (\nabla u, \nabla v) \;=\; (f, v) \quad \forall v \in H_0^1(G)\ . \tag{V}$$

Man sucht also wieder Lösungen in einem größeren Raum; entsprechend beim dazu äquivalenten Minimierungsproblem,

$$\textit{Gesucht} \quad u \in H_0^1(G) \quad \textit{mit} \quad F(u) \;\leq\; F(v) \quad \forall v \in H_0^1(G)\ . \tag{M}$$

Man nennt die Lösung von (V) bzw. (M) auch die *schwache Lösung* und (V) bzw. (M) die *schwache Formulierung* der Randwertaufgabe (D). Existenz und Eindeutigkeit von schwachen Lösungen werden in Abschnitt 2.1 behandelt.

Für spätere Zwecke sei an dieser Stelle noch bemerkt, daß man unter gewissen Voraussetzungen an das Gebiet G, auf die wir hier nicht weiter eingehen wollen, die Räume $H^1(G)$ und $H_0^1(G)$ auch als *Vervollständigung* von Räumen stetig differenzierbarer Funktionen erhalten kann (vgl. z. B. [Rei95], Kap. 2.4, [Ada75], [Agm65]). Bei Vervollständigungen ist es wichtig zu sagen, bzgl. welcher Normen man vervollständigt — hier bzgl. $\|.\|_1$. Man hat die Beziehungen $H^1(G) = \overline{C^{1,2}(G)}$ und $H_0^1(G) = \overline{C_0^\infty(G)}$, wobei die Vervollständigungen wie folgt definiert sind,

$$\begin{aligned}
\overline{C^{1,2}(G)} \;&=\; \Big\{ u \in L^2(G) | \exists u_k \in C^{1,2}(G)\ ,\ (u_k) \text{ Cauchy–Folge bzgl. } \|.\|_1 \\
&\qquad \text{und } \|u_k - u\|_0 \to 0 \Big\}\ , \\
\overline{C_0^\infty(G)} \;&=\; \Big\{ u \in L^2(G) | \exists u_k \in C_0^\infty(G)\ ,\ (u_k) \text{ Cauchy–Folge bzgl. } \|.\|_1 \\
&\qquad \text{und } \|u_k - u\|_0 \to 0 \Big\}\ .
\end{aligned}$$

Wegen der Vollständigkeit von $L^2(G)$ gibt es für die genannten Folgen bzw. deren Ableitungen $D^\alpha u_k$, $|\alpha| = 1$, Limites in $L^2(G)$, die die Gleichung (1.22) einer verallgemeinerten Ableitung erfüllen, also

$$\|D^\alpha u_k - D^\alpha u\|_0 \to 0 \ (k \to \infty)\,,\ |\alpha| = 1\ .$$

1.6 Die FEM für ein Neumann Problem

In diesem Abschnitt betrachten wir eine andere partielle Differentialgleichung mit einer anderen Art von Randbedingung,

$$\begin{aligned} -\Delta u + u &= f \quad \text{in } G \subset \mathbb{R}^n \\ \frac{\partial u}{\partial n} &= g \quad \text{auf } \partial G \end{aligned} \tag{D}$$

wobei wieder G ein beschränktes offenes Gebiet und $\frac{\partial u}{\partial n} = \nabla u \cdot n$ die Ableitung in Richtung der äußeren Normalen bezeichnet. Man nennt die obige Bedingung am Rand eine *Neumann–Bedingung*; im Gegensatz hierzu wird $u = g$ auf ∂G als *Dirichletsche Randbedingung* bezeichnet. In der Mechanik bedeutet eine Neumann–Randbedingung einen Fluß in das oder aus dem Gebiet.

Aufgrund der *Greenschen Formel* (vgl. (1.21)) erhält man

$$\int_G \nabla u \cdot \nabla v \, dx = \int_{\partial G} \frac{\partial u}{\partial n} v \, dx - \int_G (\Delta u) v \, dx$$

und man sieht sofort, daß die (klassische) Lösung des obigen Randwertproblems auch das folgende *Variationsproblem* löst:

$$\begin{aligned} &\textit{Gesucht} \quad u \in H^1(G) \quad \textit{mit} \quad a(u,v) = (f,v) + < g, v > \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall v \in H^1(G) \, , \end{aligned} \tag{V}$$

wobei

$$a(u,v) = \int_G (\nabla u \cdot \nabla v + uv) \, dx \, , \quad < g, v > = \int_{\partial G} gv \, ds \, . \tag{1.23}$$

Wie in Abschnitt 1.4 macht man sich klar, daß das Variationsproblem wieder äquivalent zu einem Minimierungsproblem ist, nämlich:

$$\textit{Gesucht} \quad u \in H^1(G) \quad \textit{mit} \quad F(u) \leq F(v) \quad \forall v \in H^1(G) \, , \tag{M}$$

wobei

$$F(v) = \frac{1}{2} a(v,v) - (f,v) - < g, v > \, .$$

Wir zeigen nun, daß eine Lösung u von (V) auch das Randwertproblem (D) löst, falls $u \in C^2(G)$ und $f \in C(G)$, $g \in C(\partial G)$. Benutzt man wieder die Greensche Formel, so folgt nämlich, daß

$$(f,v)+ < g,v >= a(u,v) = \int_{\partial G} \frac{\partial u}{\partial n} v \, dx + \int_G (-\Delta u + u) v \, dx \ , \quad v \in H^1(G) \ ,$$

also

$$\int_G (-\Delta u + u - f) v \, dx + \int_{\partial G} \left(\frac{\partial u}{\partial n} - g \right) v \, ds = 0 \quad \forall v \in H^1(G) \ . \tag{1.24}$$

Die letzte Beziehung gilt insbesondere für Funktionen $v \in H_0^1(G)$, und da man weiß, daß $H_0^1(G)$ dicht in $L^2(G)$ liegt, folgt

$$-\Delta u + u - f = 0 \quad \text{in } G \ .$$

(Die letzte Beziehung gilt zunächst nur fast überall in G; wegen der Stetigkeit der auftretenden Funktionen folgt daraus die punktweise Identität.) Als Konsequenz erhält man daher aus (1.24), daß

$$\int_{\partial G} \left(\frac{\partial u}{\partial n} - g \right) v \, ds = 0 \quad \forall v \in H^1(G) \ .$$

Mit derselben Argumentation wie oben — $H^1(\partial G)$ liegt dicht in $L^2(\partial G)$ — folgt daraus punktweise

$$\frac{\partial u}{\partial n} - g = 0 \quad \text{auf } \partial G \ ,$$

was beweist, daß u auch Lösung von (D) ist.

Man beachte, daß die Neumannsche Randbedingung nicht explizit in der Variationsformulierung auftaucht, während (homogene) Dirichletsche Randbedingungen im Raum $H_0^1(G)$ explizit integriert sind. Man spricht deshalb von Neumann–Bedingungen als den *natürlichen Randbedingungen* und von Dirichlet–Bedingungen als den *erzwungenen* oder *wesentlichen* Randbedingungen.

Eine entsprechende *Finite Elemente Approximation* für das Neumann–Problem erhält man nun durch einen endlichdimensionalen Teilraum V_h von $H^1(G)$ und der Lösung des folgenden endlichdimensionalen Variationsproblems:

$$\textit{Gesucht} \quad u_h \in V_h \quad \textit{mit} \quad a(u_h, v) \ = \ (f,v)+ < g,v > \ \forall v \in V_h \ . \tag{V_h}$$

Es sei darauf hingewiesen, daß keinerlei Randbedingungen in V_h enthalten sind und eine Gitterfunktion $u_h \in V_h$ z. B. über einer Triangulierung durch die Werte an den Gitterpunkten inklusive den Randgitterpunkten erklärt ist.

Ähnlich wie in Abschnitt 1.4 kann man zeigen, daß die Finite Elemente Methode eine eindeutige Lösung u_h besitzt, da das zugehörige Gleichungssystem wieder symmetrisch und positiv definit ist. Über eine entsprechende Fehlerbeziehung kann man zeigen, daß der Fehler bzgl. der Norm $\|.\|_1$ *optimal* ist, d. h. es gilt wieder — hier bzgl. $\|.\|_1$ — daß

$$\|u - u_h\|_1 \leq \|u - v\|_1 \quad \forall v \in V_h \ .$$

Die Tatsache, daß hier $\|.\|_1$ als Norm in der optimalen Fehlerabschätzung auftaucht, liegt daran, daß hier $\|.\|_1$ auch die zugehörige Energienorm darstellt,

$$\|v\|_a = a(v,v)^{1/2} = \|v\|_1 \ , \qquad v \in H^1(G) \ ,$$

während für die Poisson–Gleichung gerade $|.|_1$ die zugehörige Energienorm (in $H_0^1(G)$) ist.

Kapitel 2

Finite–Element–Approximationen elliptischer Differentialgleichungen

In diesem Kapitel betrachten wir allgemein elliptische Differentialgleichungen 2. Ordnung bzw. entsprechende Randwertprobleme und leiten die zugehörigen Variationsprobleme her.

Die Beispiele aus Kapitel 1 sind hierin enthalten. Wesentlich ist die Betrachtung der Probleme in (größeren) vervollständigten Räumen, wo über die Vollständigkeit mit Hilfe des Satzes von Fréchet–Riesz (oder Lax–Milgram) die Existenz und Eindeutigkeit von (schwachen) Lösungen gesichert ist. Für elliptische Probleme sind die zugehörigen endlichdimensionalen Minimierungs- bzw. Variationsprobleme immer eindeutig lösbar; deren Lösungen konvergieren unter einer Approximierbarkeitsforderung (an die endlichdimensionalen Räume) gegen die (schwache) Lösung des Randwertproblems, und es gelten quasi–optimale Fehlerabschätzungen. Im letzen Abschnitt dieses Kapitels werden die gewonnenen Ergebnisse für eindimensionale Randwertprobleme konkretisiert.

2.1 Schwache Lösungen elliptischer Differentialgleichungen

Wir betrachten ein Gebiet $G \subset \mathbb{R}^n$, $n \geq 1$, das beschränkt, offen und zusammenhängend, und das die sog. Kegelbedingung erfüllt (zu letzterem vgl. z. B. [Agm65]). In Verallgemeinerung der Aufgabenstellungen aus den Abschnitten 1.3 und 1.4 wollen wir jetzt das folgende *Randwertproblem* untersuchen:

$$
\begin{aligned}
&\textit{Gesucht } u \in C(\overline{G}) \cap C^{2,2}(G) \textit{ mit} \\
&(Au :=) - \sum_{i,j=1}^{n} D_i(a^{ij} D_j u) + au = f \quad \text{in } G\,, \\
&u = g \quad \text{auf } \partial G\,.
\end{aligned}
\tag{D}
$$

Wir machen die folgenden Voraussetzungen:

$$
\begin{aligned}
&a^{ij} = a^{ji} \in C^1(\overline{G})\,, \quad i,j = 1,\ldots,n\,, \quad a \in C(\overline{G})\,, \\
&\lambda\,|\xi|_2^2 \le \sum_{i,j=1}^{n} a^{ij}\xi^i\xi^j \text{ in } G\,, \quad \xi = (\xi^1,\ldots,\xi^n) \in \mathbb{R}^n (\lambda > 0)\,, \\
&f \in L^2(G)\,, \quad g \in C(\overline{G}) \cap C^{2,2}(G)\,.
\end{aligned}
\tag{2.1}
$$

Die Aufgabenstellung (D) heißt auch das „1. Randwertproblem für einen elliptischen Differentialoperator 2. Ordnung“.

Zu den Vorausssetzungen sei bemerkt, daß keine klassische Lösung vorliegt, falls z. B. f unstetig ist.

Das Randwertproblem (D) stellt die sogenannte „Eulersche Gleichung“ des folgenden *Minimierungsproblems* dar:

$$
\begin{aligned}
&\textit{Minimiere } F(u) := \frac{1}{2}a(u) - (f,u) \\
&\textit{für } u \in C(\overline{G}) \cap C^{1,2}(G) \textit{ mit } u = g \textit{ auf } \partial G\,.
\end{aligned}
\tag{M}
$$

Hierbei bezeichnet

$$
a(u,v) = \int_G \left\{ \sum_{ij=1}^{n} a^{ij} D_i u D_j v + auv \right\} dx\,, \qquad u,v \in C^{1,2}(G)\,,
$$

die zum Differentialoperator A gehörige *Bilinearform* und $a(u) := a(u,u)$ die zugehörige *quadratische Form*.

Für eine Bilinearform $a : E \times E \to \mathbb{R}$, E ein normierter Raum und V ein abgeschlossener Teilraum von E, hat man die folgenden Bezeichnungen und *Definitionen*:

$a(.,.)$ *symmetrisch* $\Longleftrightarrow a(u,v) = a(v,u)\,, \quad u,v \in E\,;$

a(.,.) *beschränkt* (oder *stetig*)
$\Longleftrightarrow |a(u,v)| \le \alpha_1 \|u\|\,\|v\|\,, \quad u,v \in E\,;$

$a(.,.)$ *positiv semidefinit auf* V
$\Longleftrightarrow a$ symmetrisch und $a(u) \ge 0\,, \quad u \in V\,;$

$a(.,.)$ *positiv definit auf* V
$\Longleftrightarrow a$ symmetrisch und $a(u) > 0\,, \quad 0 \ne u \in V\,;$

$a(.,.)$ *elliptisch auf* V (oder V*–elliptisch*)
$\Longleftrightarrow |a(u)| \ge \alpha_0 \|u\|^2\,, \quad u \in V\,.$

Wir werden nun zeigen, daß die oben definierte Bilinearform unter den in (A) angegebenen Voraussetzungen insbesondere elliptisch auf $H_0^1(G)$ ist. Hierzu benötigen wir noch die *Ungleichung von Poincaré–Friedrichs*,

$$\|u\|_0 \le C|u|_1\,, \qquad u \in H_0^1(G)\,, \tag{2.2}$$

die wir für eindimensionale Probleme im Raum aller stetig differenzierbaren Funktionen bereits in Kapital 1 benutzt haben; die Konstante C hängt hierbei nur vom Gebiet G ab. Eine Konsequenz der Ungleichung von Poincaré–Friedrichs ist die folgende zweiseitige Ungleichung für Funktionen aus $H_0^1(G)$,

$$|u|_1 \le \|u\|_1 \le \tilde{C}|u|_1\,, \qquad u \in H_0^1(G)\,,$$

die zeigt, daß $|.|_1$ und $\|.\|_1$ äquivalente Normen für $H_0^1(G)$ sind.

Lemma 2.1 *Unter den Voraussetzungen in (2.1) kann die Bilinearform* $a(.,.)$ *eindeutig auf* $H^1(G) \times H^1(G)$ *fortgesetzt werden; sie ist symmetrisch, beschränkt, positiv definit und elliptisch auf* $H_0^1(G)$*, denn es gelten die Ungleichungen*

$$|a(u,v)| \;\le\; \alpha_1 \|u\|_1 \|v\|_1\,, \qquad u,v \in H^1(G)\,, \tag{2.3}$$

$$\alpha_0 \|u\|_1^2 \;\le\; a(u)\,, \qquad u \in H_0^1(G)\,. \tag{2.4}$$

Beweis: (i) Da die a^{ij} und a beschränkt sind, folgt mit Hilfe der Schwarzschen Ungleichung (für Summen)[1)]

[1)] $|.|$ oder $|.|_2$ bezeichnet auch die Euklidische Norm im $\mathbb{R}^n$.

$$\left|\sum_{i,j=1}^{n} a^{ij} D_j u D_i v + auv\right| \le C\left(\sum_{i,j} |D_j u|\,|D_i v| + |u|\,|v|\right)$$
$$\le Cn\left\{\underbrace{\left(\sum_j |D_j u|^2\right)^{1/2}}_{|\nabla u|_2}\underbrace{\left(\sum_i |D_i v|^2\right)^{1/2}}_{|\nabla v|_2} + |u|\,|v|\right\}, \quad u, v \in C^{1,2}(G),$$

$$\Longrightarrow\ |a(u,v)| \ \le\ Cn\left\{\int_G |\nabla u|_2 |\nabla v|_2\, dx + \int_G |u|\,|v|\, dx\right\}$$
$$\overset{\text{(Schwarz/Integrale)}}{\le} Cn\left\{\left(\int_G |\nabla u|_2^2\, dx\right)^{1/2}\left(\int_G |\nabla v|_2^2\, dx\right)^{1/2} + \right.$$
$$\left. + \left(\int_G |u|^2\, dx\right)^{1/2}\left(\int_G |v|^2\, dx\right)^{1/2}\right\}$$

Daraus folgt (2.3) für Funktionen aus $C^{1,2}(G)$.

ii) Unter (2.1) ist

$$a(v) \ = \ \int_G \left\{\sum_{i,j=1}^{n} a^{ij} D_j v D_i v + \underbrace{a}_{\ge 0} v^2\right\} dx$$
$$\ge\ \lambda \int_G |\nabla v|_2^2\, dx\,, \qquad v \in C^{1,2}(G)\,,$$

und mit Hilfe der Ungleichung von Poincaré–Friedrichs folgt

$$|u|_1^2 \le \|u\|_1^2 = |u|_0^2 + |u|_1^2 \le (1 + C^2)|u|_1^2 \le (1 + C)^2 |u|_1^2$$

und

$$a(u) \ge \lambda(1 + C)^{-2} \|u\|_{1,2}^2\,, \qquad u \in C_0^\infty(G)\,.$$

Seien $u, v \in H^1(G)$. Dann ex. $u_\nu, v_\nu \in C^{1,2}(G)$, $\nu \in \mathbb{N}$, mit

$$\|u_\nu - u\|_1 \longrightarrow 0\,, \quad \|v_\nu - v\|_1 \longrightarrow 0 \ (\nu \to \infty)\,.$$

Definiert man

$$\hat{a}(u, v) := \lim_{\nu \to \infty} a(u_\nu, v_\nu)\,,$$

dann ist $\hat{a}$ unabhängig von der Wahl der $(u_\nu), (v_\nu)$ und stimmt auf $C^{1,2}(G)$ mit $a(.,.)$ überein (setze: $a = \hat{a}$). Im Limes folgen dann (2.3) und (2.4). □

Satz 2.2 *Ist $a(.,.)$ positiv semidefinit auf $V \subset E$ und ℓ eine stetige Linearform auf E, dann ist $u \in V$ eine Lösung des Minimierungsproblems*

$$u \in V : F(u) = \min_{v \in V} F(v) \quad (\textit{ mit } \; F(v) := \frac{1}{2} a(v) - \ell(v))$$

d. u. n. d. wenn u Lösung der folgenden (Variations–)Gleichung ist:

$$u \in V : a(u, v) = \ell(v) , \qquad v \in V .$$

Ist $a(.,.)$ positiv definit auf V, dann gibt es höchstens eine Lösung.

Beweis: Für $u, v \in E$ hat man

$$\begin{aligned} 2(F(v) - F(u)) &= a(v,v) - 2\ell(v) - a(u,u) + 2\ell(u) \\ &= a(v, v-u) - \ell(v-u) + a(v-u, u) - \ell(v-u) \\ &\underset{a \text{ symm.}}{=} a(v-u, v-u) + 2(a(u, v-u) - \ell(v-u)) . \end{aligned}$$

Wenn $a(u, v) = \ell(v) \;\; \forall v \in V$, dann gilt (wegen a pos. semidef.)

$$a(u, v-u) - \ell(v-u) = \ell(v) - \underbrace{a(u)}_{\ell(u)} - \ell(v-u) = 0$$

und $F(v) - F(u) \geq 0 \;\; \forall v \in V$. Also ist u Minimallösung.

Umgekehrt sei u Lösung des Minimierungsproblems, dann gilt mit $v := u + tz$, $t > 0$, $z \in V \, (v - u = tz)$, daß

$$0 \leq 2(F(v) - F(u)) = t^2 a(z) + 2t\{a(u, z) - \ell(z)\} .$$

Im Limes $t \to 0$ folgt

$$0 \leq a(u, z) - \ell(z) \quad \forall z \in V .$$

Mit $-z$ anstelle von z erhält man

$$0 \geq a(u, z) - \ell(z) \quad \forall z \in V .$$

Also ist $a(u, z) = \ell(z) \;\; \forall z \in V$.

Zur Eindeutigkeit: Seien $\tilde{u} \neq u$ Lösungen der Variationsgleichung. Dann ist

$$a(u-\tilde{u},v)=0 \quad \forall v \in V \text{ und } a(u-\tilde{u},u-\tilde{u})=0$$

(Widerspruch zu $u \neq \tilde{u}$, a pos. def.) □

Man nennt nun die folgende Aufgabenstellung die „schwache Form“ des elliptischen RWP's (D):

$$\begin{aligned} &\textit{Gesucht} \quad u \in H^1(G) \quad \textit{mit} \quad u-g \in H_0^1(G) \quad \textit{und} \\ &a(u,v)=\underbrace{(f,v)}_{\ell(v)} \quad \forall v \in H_0^1(G)\ . \end{aligned} \tag{$\hat{V}$}$$

Setzt man $\hat{f}=f-Ag(\in L^2(G))$ und betrachtet die Aufgabe

$$\textit{Gesucht} \quad u \in H_0^1(G) \quad \textit{mit} \quad a(u,v)=(\hat{f},v)\ , \quad v \in H_0^1(G)\,, \tag{V}$$

dann ist u Lösung von (V) d. u. n. d. wenn $\hat{u}=u+g$ Lösung von $(\hat{V})$ ist.

Nach Satz 2.2 ist $u \in H_0^1(G)$ Lösung von (V) d. u. n. d. wenn u Lösung des *Minimierungsproblems*

$$u \in H_0^1(G) : F(u)=\min_{v \in H_0^1(G)} F(v)\ . \tag{M}$$

Mit der Lösung u von (D) (und damit der Lösung des zuerst angegebenen Minimierungsproblems) löst $u-g$ dann (M). In (M) sucht man jedoch das Minimum in einem größeren Raum, so daß die Lösung von (M) — und die Lösung von (V) — (bzw. $u-g$) die *schwache Lösung* des RWP (D) heißt. Man nennt (V) (bzw. $(\hat{V})$) das *Variationsproblem* zu (D) (oder die *Variationsformulierung* von (D)) und die zugehörige Gleichung die *Variationsgleichung* . Die „erste Variation“ von F , i. e. die Fréchet–Ableitung

$$F'(u)v=a(u,v)-\ell(v)\ , \qquad u,v \in V\ ,$$

verschwindet nämlich bei einem Extremum von F, und dies ist offenbar äquivalent zur Lösung der Variationsgleichung (o. B. d. A. im folgenden: $g=0$).

Die Existenz der Lösung des Variationsproblems läßt sich hier mit dem Darstellungssatz von Riesz erledigen. Man hat nämlich mit

$$(u,v)_a := a(u,v)\ , \qquad u,v \in H_0^1(G)\ ,$$

ein Skalarprodukt auf dem Hilbertraum $V = H_0^1(G)$, und $\ell(v) = (f,v)$ ist ein beschränktes lineares Funktional auf $E = H^1(G)$ (bzgl. $\|.\|_1$). Nach dem Satz von Riesz gibt es daher ein eindeutig bestimmtes Element $u \in E$ mit

$$(u,v)_a = \ell(v) \quad \forall v \in V\ .$$

Satz 2.3 *Unter den Voraussetzungen von (2.1) besitzt das RWP (D) für jedes $f \in L^2(G)$ genau eine schwache Lösung $u \in H_0^1(G)$, und es gilt die folgende Stabilitätsungleichung*

$$\|u\|_1 \le \frac{1}{\alpha_0}\|f\|_0\ .$$

Beweis: Existenz und Eindeutigkeit ist mit Hilfe des Satzes von Fréchet–Riesz gesichert. Die Stabilitätsungleichung folgt aus der Variationsgleichung (mit $v = u$) und (2.4),

$$\alpha_0\|u\|_1^2 \le a(u,u) = \ell(u) \le \|f\|_0\|u\|_0 \le \|f\|_0\|u\|_1\ .$$

□

Variationsgleichungen sind auch für nicht notwendig symmetrische Bilinearformen sinnvoll. Die Existenz von deren Lösungen (für elliptische Bilinearformen) sichert das folgende *Lemma von Lax–Milgram.*

Satz 2.4 (LEMMA von Lax–Milgram) *Sei V ein Hilbert–Raum und $a : V \times V \to \mathbb{R}$ eine beschränkte, auf V elliptische Bilinearform. Dann gibt es zu jedem beschränkten linearen Funktional ℓ auf V eine eindeutige Lösung $u \in V$ des Variationsproblems*

$$a(u,v) = \ell(v)\ , \qquad v \in V\ ,$$

und es gilt die Abschätzung

$$\alpha_0\|u\| \le \sup_{0\neq v\in V} |\ell(v)|/\|v\| \ (=: \|\ell\|)\ .$$

Beweis: Sei $(u_\nu)_{\nu\in\mathbb{N}}$ in V eine Folge mit der Eigenschaft $\lim_{\nu\to\infty} F(u_\nu) = d := \inf_{v\in V} F(v)$ (Minimalfolge). Offenbar ist wegen $2F(v) = a(v) - 2\ell(v) \geq \alpha_0\|v\|^2 - \|\ell\|\,\|v\|$ das Infimum $d > -\infty$.

Unter Anwendung der Parallelogrammidentität

$$a(u-v) + a(u+v) = 2a(u) + 2a(v)$$

erhalten wir

$$\begin{aligned}
\alpha_0\|u_\nu - u_\mu\|^2 &\leq a(u_\nu - u_\mu) \;=\; 2a(u_\nu) + 2a(u_\mu) - a(u_\nu + u_\mu) \\
&= 2a(u_\nu) + 2a(u_\mu) - 4a\left(\frac{u_\nu + u_\mu}{2}\right) \\
&\quad - 4\ell(u_\nu) - 4\ell(u_\mu) + 8\ell\left(\frac{u_\nu + u_\mu}{2}\right) \\
&= 2F(u_\nu) + 2F(u_\mu) - 4F\left(\frac{u_\nu + u_\mu}{2}\right) .
\end{aligned}$$

Wegen $F\left(\frac{u_\nu+u_\mu}{2}\right) \geq d$ und $\lim_{\mu,\nu\to\infty} F(u_\nu) + F(u_\mu) = 2d$
ergibt sich

$$\|u_\nu - u_\mu\| \longrightarrow 0 \qquad (\nu,\mu \to \infty) ,$$

d. h. $(u_\nu)_{\nu\in\mathbb{N}}$ ist eine Cauchyfolge und besitzt folglich im Hilbertraum V einen Limes $u : \|u - u_\nu\| \longrightarrow 0$ $(\nu \to \infty)$.

Aufgrund der Stetigkeit von $F(.)$, welche direkt aus der Beschränktheit von a und ℓ folgt, sehen wir

$$F(u) = \lim_{\nu\to\infty} F(u_\nu) = d .$$

Durch Nullsetzen der ersten Variation von $F(.)$ bei u

$$\frac{d}{d\varepsilon} F(u + \varepsilon v)\bigg|_{\varepsilon=0} = 0 , \qquad \forall v \in V ,$$

erhalten wir die gewünschte Beziehung (vgl. (V))

$$a(u, v) = \ell(v) , \qquad \forall v \in V .$$

Es bleibt noch, die Eindeutigkeit der Lösung u zu zeigen. Für zwei Lösungen u, u' der Variationsgleichung gilt

$$a(u-u',v)=0\ , \qquad \forall v\in V\ .$$

Setzen wir speziell $v:=u-u'$, so folgt mit Hilfe der Elliptizität

$$\|u-u'\|^2\le\frac{1}{\alpha_0}|a(u-u')|=0\ .$$

□

Zur Regularität der Lösung des Variationsproblems gilt das folgende Ergebnis (vgl. z. B. [Fri69, Lad85])

Satz 2.5 *Gilt (2.1), ist G konvex und $\partial G\in C^2$, dann ist die Lösung $u\in H_0^1(G)$ von (V) aus $H^2(G)$ und genügt der a–priori Abschätzung*

$$\|u\|_2\le C\|f\|_0$$

(mit C unabhängig von f).

Unter etwas stärkeren Voraussetzungen an die Koeffizienten a^{ij},a, den Rand ∂G und die rechte Seite f läßt sich sogar zeigen, daß u im L^2–Sinne äquivalent zu einer Funktion aus $C(\overline{G})\cap C^{2,2}(G)$ ist, die wiederum klassische Lösung des Randwertproblems (D) ist.

Wir fassen noch einmal die vorliegende *Situation* zusammen, die dann unseren Approximationsmethoden zugrundeliegt:

$E=H^1(G)$ Hilbertraum, $\ell(.)=(f,.)$ stetige Linearform auf E ,
$V=H_0^1(G)$ abgeschl. Teilraum (= Hilbertraum) ,
$a(.,.)$ symm., beschr., ellipt. (auf V) Bilinearform ,
$(.,.)_a=a(.,.)$ Skalarprodukt auf V ,
Normen (äquiv.):

$$\alpha_0\|u\|_{1,2}^2\le a(u)=\|u\|_a^2\le\alpha_1\|u\|_{1,2}^2\ , \qquad u\in V\ .$$

Minimierungsproblem: $\quad u\in V: F(u)=\min_{v\in V}F(v)$

äquivalent zur

Variationsgleichung: $\quad u\in V:(u,v)_a=\ell(v)\ , \qquad v\in V.$

Existenz und Eindeutigkeit der Lösung gesichert (nach Fréchet–Riesz oder Lax–Milgram).

2.2 Ritzsches Projektionsverfahren, Fehlerabschätzungen

In diesem Abschnitt wird angenommen, daß die Voraussetzungen von Satz 2.5 gelten.

Auf Ritz[2)] geht das folgende direkte[3)] Verfahren zur Approximation der Lösung u des Variationsproblems (V) zurück. Für einen reellen Parameter h, $0 < h \leq h_0$, seien $V_h \subset H_0^1(G)$ endlichdimensionale Teilräume mit gegen ∞ wachsenden Dimensionen

$$\dim(V_h) \longrightarrow \infty \qquad (h \to 0) \; .$$

Wir betrachten die approximierenden Aufgaben:

$$\textit{Gesucht} \quad u_h \in V_h \quad \textit{mit} \quad a(u_h, v_h) = (f, v_h) \; , \quad \forall v_h \in V_h \; . \tag{V_h}$$

Da die Einschränkungen der Bilinearform a auf die Teilräume V_h offenbar den Bedingungen von Satz 2.4 (mit $V = V_h$) genügen, besitzen die Aufgaben(V_h) eindeutige Lösungen $u_h \in V_h$, welche zugleich auch Lösungen der folgenden Variationsprobleme sind:

$$\textit{Minimiere} \quad F(v_h) = \frac{1}{2} a(v_h) - (f, v_h) \quad \textit{für} \quad v_h \in V_h \; . \tag{M_h}$$

Bei Zugrundelegung einer Basis $\left\{w_h^1, \ldots, w_h^N\right\}$, $N = \dim(V_h)$, von V_h wird (V_h) äquivalent zu dem linearen Gleichungssystem $\left(u_h = \sum_{i=1}^N \eta^i w_h^i\right)$

$$\sum_{i=1}^{N} \eta^i \, a\left(w_h^i, \, w_h^j\right) = \left(f, w_h^j\right) \; , \qquad j = 1, \ldots, N \; ,$$

mit symmetrischer, positiv definiter Koeffizientenmatrix

$$K := \left(a\left(w_h^i, w_h^j\right) \right)_{i,j=1}^{N} \; .$$

Durch Kombination der Gleichungen in (V) und (V_h) erhalten wir die *Fehlergleichung*

[2)] W. Ritz: Über eine neue Methode zur Lösung gewisser Variationsprobleme der mathematischen Physik. J. f. reine u. angew. Math. 135,1–61 (1909)

[3)] „direkt“ im Gegensatz zu den „indirekten“ iterativen Verfahren.

$$a(u-u_h, v_h) = 0\ , \qquad \forall v_h \in V_h\ . \tag{2.5}$$

Die Bilinearform a definiert nun , wie man leicht nachprüft, durch $(.,.)_a := a(.,.)$ ein neues Skalarprodukt auf $H_0^1(G)$. Die Beziehung (2.2) besagt dann gerade, daß der Fehlervektor $u - u_h$ bzgl. $(.,.)_a$ senkrecht zu dem Teilraum V_h steht, bzw. daß u_h die „orthogonale Projektion" der Lösung u auf V_h ist. Daher resultiert der Name „Ritzsches Projektionsverfahren" .

Satz 2.6 *Ist für die Lösung $u \in H_0^1(G)$ des Variationsproblems (V)*

$$\delta_h(u) := \inf_{\varphi_h \in V_h} \|u - \varphi_h\|_1 \longrightarrow 0 \qquad (h \to 0)\ , \tag{2.6}$$

so konvergieren die Lösungen des Ritzschen Verfahrens,

$$\|u - u_h\|_1 \longrightarrow 0 \qquad (h \to 0)\ .$$

Ferner gelten mit den Größen

$$\Delta_h^m := \sup\left\{\frac{\delta_h(v)}{\|v\|_m}\ , \qquad v \in H_0^1(G) \cap H^m(G)\right\} \tag{2.7}$$

die (asymptotischen) Fehlerabschätzungen

$$\|u - u_h\|_1 \ \leq\ c\Delta_h^m \|u\|_m\ , \tag{2.8}$$

$$\|u - u_h\|_0 \ \leq\ c\Delta_h^2 \Delta_h^m \|u\|_m\ . \tag{2.9}$$

Beweis: Unter Verwendung der Elliptizität (Lemma 2.1) der Bilinearform a und der Beziehung (2.5) erhalten wir für beliebiges $\varphi_h \in V_h$

$$\begin{aligned}\|u - u_h\|_1^2 \ &\leq\ \frac{1}{\alpha_0} a(u - u_h) = \frac{1}{\alpha_0} a(u - u_h, u - \varphi_h) \\ &\leq\ \frac{\alpha_1}{\alpha_0} \|u - u_h\|_1 \|u - \varphi_h\|_1\ .\end{aligned}$$

Dies beweist die erste Behauptung

$$\|u - u_h\|_1 \leq c \inf_{\varphi_h \in V_h} \|u - \varphi_h\|_1 \longrightarrow 0 \qquad (h \to 0)$$

und gleichzeitig auch die Abschätzung (2.8),

$$\|u - u_h\|_1 \leq c\Delta_h^m \|u\|_m\ .$$

Sei nun $w \in H_0^1(G) \cap H^2(G)$ die eindeutige Lösung des Hilfsproblems (bei festgehaltenem h)

$$a(w,v) = (u - u_h, v) \ , \quad \forall v \in H_0^1(G) \ , \tag{2.10}$$

welche dann der Ungleichung genügt

$$\|w\|_2 \leq c\|u - u_h\|_0$$

(vgl. Satz 2.5 zur Regularität).

Durch Einsetzen von $u - u_h \in H_0^1(G)$ in (2.10) erhält man

$$\|u - u_h\|_0^2 = a(w, u - u_h) \ .$$

Unter Ausnutzung der Orthogonalitätsbeziehung (2.5) folgt weiter für beliebiges $\varphi_h \in V_h$

$$\|u - u_h\|_0^2 = a(w - \varphi_h, u - u_h) \leq c\|u - u_h\|_1 \|w - \varphi_h\|_1 \ .$$

Nun ist nach Definition von Δ_h^2 und wegen Satz 2.5

$$\inf \|w - \varphi_h\|_1 \leq \Delta_h^2 \|w\|_2 \leq c\Delta_h^2 \|u - u_h\|_0 \ ,$$

woraus zusammen mit (2.8) schließlich die Behauptung (2.9) folgt,

$$\|u - u_h\|_0 \leq c\Delta_h^2 \Delta_h^m \|u\|_m \ .$$

□

Die Technik zur Herleitung von (2.9) nennt man auch den „Aubin–Nitsche–Trick“; ein Vergleich mit (2.8) legt die Bezeichnung „stärkere Abschätzung in schwächeren Normen“ für (2.9) nahe. Ein Ziel wird es in späteren Abschnitten sein, die Abstände Δ_h^m abzuschätzen.

2.3 Beispiele für elliptische RWP's in einer Veränderlichen

Beispiel 2.7

$$-u'' = f \quad \text{in } (0,1)\,, \quad u(0) = u(1) = 0 \tag{D}$$

$$V = H_0^1(I)\,, \quad I = (0,1)$$

$$a(v,w) = \int_I v'w'\,dx\,, \quad \ell(v) = \int_I fv\,dx\,, \qquad f \in L^2(I)$$

$$u \in H_0^1(I) : a(u,v) = \ell(v) \quad \forall v \in H_0^1(I) \tag{V}$$

$a(.,.)$ symm., bilinear, beschränkt, denn

$$|a(v,w)| \le |v|_1|w|_1 \le \|v\|_1\|w\|_1\ ;$$

$\ell(.)$ linear und beschränkt, denn $|\ell(v)| \le \|f\|_0\|v\|_0 \le \|f\|_0\|v\|_1$;
$a(.,.)$ elliptisch auf V, denn

$$|v|_0 \le |v|_1 \Rightarrow \|v\|_1^2 = |v|_0^2 + |v|_1^2 \le 2|v|_1^2$$

$$\Rightarrow \quad a(v,v) = |v|_1^2 \ge \frac{1}{2}\|v\|_1^2$$

d. h. $\alpha_0 = \frac{1}{2},\ \alpha_1 = 1$.[4)]
$V_h = \{$ stetig, stückw. linear $\} \cap H_0^1(0,1)\,,\ 0 = x_0 < x_1 < \cdots < x_n = 1$
(vgl. Abb. 1.2 und 1.3)

$$u_h \in V_h : a(u_h,v) = \ell(v)\,, \qquad v \in V_h \tag{V_h}$$

[4)] Es gilt sogar $v_0 \le (1/\pi)|v|_1$ und deshalb $\|v\|_1^2 \le (1 + 1/\pi^2)|v|_1^2$.

Fehlerabschätzung: $|v|_1 = a(v,v)^{1/2} (= \|v\|_a)$ Energienorm

$$\begin{aligned} |u-u_h|_1 &\leq |u-v|_1 \quad \forall v \in V_h \quad \text{(optimal)} \\ \|u-u_h\|_1 &\leq \frac{\alpha_1}{\alpha_0}\|u-v\|_1 \;=\; 2\|u-v\|_1\,, \quad v \in V_h \\ & \qquad\qquad\qquad\qquad \text{(quasi–optimal)} \end{aligned}$$

$v = \tilde{u}_h : \tilde{u}_h(x_j) = u(x_j) \overset{\text{Abschnitt 1.3}}{\Longrightarrow} |u-\tilde{u}_h|_1 \leq h \max_{s\in(0,1)} |u''(s)|.$ □

Beispiel 2.8

$$\begin{aligned} &-u'' + u = f \quad \text{in } (0,1) \\ &-u'(0) = \gamma_0\,, \quad u'(1) = \gamma_1 \end{aligned} \tag{D}$$

$$\begin{aligned} &\textit{Gesucht} \quad u \in H^1(I) (= V)\,: \\ &a(u,v) = (f,v) + < g, v >\,, \qquad v \in H^1(I)\,. \end{aligned} \tag{V}$$

wobei

$$\begin{aligned} a(u,v) &= \int_0^1 (u'v' + uv)\,dx\,, \\ < g, v > &= \gamma_0 v(0) + \gamma_1 v(1)\,, \\ g(x) &= (1-x)\gamma_0 + x\gamma_1\,, \end{aligned}$$

denn

$$(-u'' + u, v) = (u', v') \underbrace{-[u'v|_0 - u'v|_1]}_{\gamma_0 v(0) + \gamma_1 v(1)} + (u,v)\,.$$

$a(.,.)$ symm., bilinear und beschränkt, denn ($\alpha_1 = 1$)

$$|a(v,w)| \leq |v|_1|w|_1 + \|v\|_0\|w\|_0 \leq \underbrace{\left(|v|_1^2 + \|v\|_0^2\right)^{1/2}}_{\|v\|_1} \left(|w|_1^2 + \|w\|_0^2\right)^{1/2}\;;$$

$a(.,.)$ elliptisch auf $V = H^1(I)$, denn

$$a(v,v) = |v|_1^2 + |v|_0^2 = \|v\|_1^2 \quad \text{(trivial: } \alpha_0 = 1)\,.$$

$\ell(v) = (f,v) + < g, v >$ linear und beschränkt, denn

$$|\ell(v)| \le \|f\|_0 \|v\|_0 + |\gamma_0|\,|v(0)| + |\gamma_1|\,|v(1)|\,;$$

es ist offenbar $v(0)^2 = \int_0^1 \frac{d}{ds}\{(s-1)v(s)^2\}\,ds$ und

$$\frac{d}{ds}\Big\{(s-1)v(s)^2\Big\} = v(s)^2 + 2(s-1)v(s)v'(s)$$

$$\begin{aligned} \Rightarrow |v(0)|^2 &\le \int_0^1 |v(s)|^2\,ds + 2\int_0^1 |v(s)|\,|v'(s)|\,ds \\ &\le \|v\|_0^2 + 2\|v\|_0|v|_1 \le 2\|v\|_0^2 + |v|_1^2\,; \end{aligned}$$

analog $|v(1)|^2 \le \ldots$

$$\begin{aligned} \Rightarrow |\ell(v)| &\le \|f\|_0 \underbrace{\|v\|_0}_{\le\|v\|_1} + (|\gamma_0| + |\gamma_1|) \underbrace{(2\|v\|_0^2 + |v|_1^2)^{1/2}}_{\le\sqrt{2}\|v\|_1} \\ &\le (\|f\|_0 + \sqrt{2}\,(|\gamma_0| + |\gamma_1|))\,\|v\|_1\,. \end{aligned}$$

$V_h = \{\text{stetig, stückweise linear}\} \subset H^1(I)$ (vgl. Abb. 3.1 in Abschnitt 3.2)

$$u_h \in V_h : a(u_h, v) = (f, v) + \gamma_0 v(0) + \gamma_1 v(1) \qquad (V_h)$$

Fehlerabschätzung: $\|v\|_1 = a(v,v)^{1/2} =$ Energienorm

$$\|u - u_h\|_1 \le \|u - v\|_1 \quad \forall v \in V_h \quad \text{(optimal)}$$

□

Beispiel 2.9 Balkenbiegung

$$\begin{aligned} \frac{d^4u}{dx^4} &= f \quad \text{in } (0,1) \\ u(0) &= u'(0) = u(1) = u'(1) = 0 \end{aligned} \qquad (D)$$

(vgl. [Joh87, Fig. 1.5])

$H^2(I) = \{v \in L^2(I) : v', v'' \in L^2(I)\}$,

$\underbrace{H_0^2(I)}_{=V} = \{v \in H^2(I) : v(0) = v'(0) = v(1) = v'(1) = 0\}$,

$$\|v\|_2 = \left(\|v\|_0^2 + |v|_1^2 + |v|_2^2\right)^{1/2} .$$

$$\textit{Gesucht} \quad u \in V \;:\; a(u,v) = (f,v)\ \forall v \in V, \tag{V}$$

wobei

$$a(u,v) = (u'', v'') ,$$

denn

$$\begin{aligned}(u^{(4)}, v) &= -(u''', v') + (u''' \underbrace{v)(1)}_{=0} - (u''' \underbrace{v)(0)}_{=0} \\ &= +(u'', v'') - (u'' \underbrace{v')(1)}_{=0} + (u'' \underbrace{v')(0)}_{=0} , \quad v \in H_0^2(I) .\end{aligned}$$

$a(.,.)$ symm., bilinear und beschränkt (bzgl. $\|.\|_2$);

$a(.,.)$ elliptisch auf V, denn

$$\begin{aligned}\|v\|_0 &\le \tfrac{1}{\pi}|v|_1 \le \tfrac{1}{\pi^2}|v|_2 \quad \forall v \in H_0^2(I) \\ \Rightarrow \quad \|v\|_2^2 &= \|v\|_0^2 + |v|_1^2 + |v|_2^2 \le \underbrace{\left(\frac{1}{\pi^4} + \frac{1}{\pi^2} + 1\right)}_{1/\alpha_0} \underbrace{|v|_2^2}_{a(v,v)} \\ |\ell(v)| &= |(f,v)| \le \|f\|_0 \|v\|_0 \le \|f\|_0 \|v\|_2 .\end{aligned}$$

□

Kapitel 3

Interpolation von Funktionen einer Veränderlichen

Instruktive erste Beispiele für die Methode finiter Elemente erhält man mit Räumen stückweise polynomialer Funktionen einer Veränderlichen. Im ersten Abschnitt dieses Kapitels werden zunächst einige allgemeine Eigenschaften dieser Räume hergeleitet. Mit Hilfe von Stetigkeitsbedingungen definieren wir anschließend die Teilräume $\mathcal{P}_r^m(\Delta)$ und bestimmen die Dimensionen dieser Räume. Als erstes Beispiel wird dann die Lagrangesche Interpolationsaufgabe für lineare Spline-Funktionen behandelt. Mit Hilfe der zugehörigen kanonischen Darstellung von Linearformen wird schließlich die Herleitung von Quadraturformeln und Differenzenquotienten erläutert. Weitere Beispiele für Interpolationsaufgaben in Räumen stückweise polynomialer Funktionen findet man in den folgenden Abschnitten. Zusätzlich wird in Abschnitt 3.2 eine verallgemeinerte Interpolationsaufgabe formuliert und Aussagen zur Lösbarkeit bewiesen, in deren Rahmen die konkreten Aufgabenstellungen von Abschnitt 3.1 und 3.3 eingeordnet werden können.

3.1 Finite Elemente für Funktionen einer Veränderlichen

Gegeben sei ein beschränktes Intervall $[a, b]$ der reellen Zahlengeraden und eine Zerlegung von $[a, b]$ durch *Gitterpunkte*

$$\Delta : a = x_0 < x_1 < \ldots < x_n = b$$

in Teilintervalle

$$I_j = (a_j, b_j)\ ,\ a_j = x_{j-1}\ ,\ b_j = x_j\ ,\quad j = 1, \ldots, n\ .$$

Die Intervalle I_j sind die *finiten Elemente* dieser Zerlegung. Weiter sei $\mathcal{P}_r(\Delta)$ der Vektorraum aller *stückweise polynomialen* Funktionen u vom Grade r auf $\cup_j I_j \subset [a, b]$. Das heißt, zu jedem Teilintervall I_j der Zerlegung existiert ein Polynom $u^j \in \mathcal{P}_r = \mathcal{P}_r(\mathbb{K})$ mit der Eigenschaft

$$u(x) = u^j(x)\ ,\quad x \in I_j = (a_j, b_j)\ ,\qquad j = 1, \ldots, n\ . \tag{3.1}$$

Diese Polynome u^j sind eindeutig durch u bestimmt. Denn für $u = u_1^j = u_2^j$ auf I_j ist $u_1^j(x) = u_2^j(x)$ für alle x aus $x_{j-1} < x < x_j$. Bekanntlich sind dann u_1^j und u_2^j identisch. Umgekehrt erhält man zu jedem n–tupel von Polynomen $u^j \in \mathcal{P}_r\ ,\ j = 1, \ldots, n$, durch die Vorschrift (3.1) eine stückweise polynomiale Funktion $u \in \mathcal{P}_r(\Delta)$. Mit Hilfe dieser Eigenschaften zeigt man leicht, daß der Vektorraum $\mathcal{P}_r(\Delta)$ isomorph zum kartesischen Produkt von n Räumen $\mathcal{P}_r(\mathbb{K})$ ist,

$$\mathcal{P}_r(\Delta) \cong (\mathcal{P}_r(\mathbb{K}))^n\ . \tag{3.2}$$

Eine wichtige Folgerung aus dieser Beziehung ist die Dimensionsaussage

$$\dim \mathcal{P}_r(\Delta) = n(r+1)\ . \tag{3.3}$$

Der Raum $\mathcal{P}_0(\Delta)$ besteht aus den stückweise konstanten Funktionen oder *Treppenfunktionen* auf der Zerlegung Δ von $[a, b]$. $\mathcal{P}_1(\Delta)$ ist der Raum der *stückweise linearen* Funktionen, $\mathcal{P}_2(\Delta)$ der *stückweise quadratischen*, $\mathcal{P}_3(\Delta)$ der *stückweise kubischen* Funktionen auf $\cup I_j$. Funktionen u aus $\mathcal{P}_r(\Delta)$ sind im allgemeinen in den Intervallendpunkten $x_0, \ldots, x_n$ oder a_j, b_j nicht definiert. An den inneren Gitterpunkten $x_1, \ldots, x_{n-1}$ können darüber hinaus Unstetigkeiten auftreten. Für jedes Teilintervall läßt sich jedoch die Einschränkung von Funktionen $u \in \mathcal{P}_r(\Delta)$ und ihrer Ableitungen $\frac{d^s u}{dx^s}$ in die Intervallendpunkte fortsetzen.

$$\begin{aligned} \frac{d^s u}{dx^s}(a_j + 0) = \lim_{\substack{x \to a_j \\ x \in I_j}} \frac{d^s u}{dx^s}(x)\ ,\quad \frac{d^s u}{dx^s}(b_j - 0) = \lim_{\substack{x \to b_j \\ x \in I_j}} \frac{d^s u}{dx^s}(x)\ , \\ s = 0, \ldots, r\ ,\quad j = 1, \ldots, n\ . \end{aligned} \tag{3.4}$$

Mit Hilfe der in (3.1) erklärten zugehörigen Polynome $u^j \in \mathcal{P}_r$ und $a_j = x_{j-1}\ ,\ b_j = x_j$ ist dann offenbar

$$\frac{d^s u}{dx^s}(a_j + 0) = \frac{d^s u^j}{dx^s}(x_{j-1})\,,\quad \frac{d^s u}{dx^s}(b_j - 0) = \frac{d^s u^j}{dx^s}(x_j)\ . \tag{3.5}$$

Durch (3.4) oder (3.5) sind Linearformen auf $\mathcal{P}_r(\Delta)$ erklärt, die wir mit δ^s_{j-1+0} , δ^s_{j-0} bezeichnen für $j = 1, \ldots, n,\ s = 0, 1, 2, \ldots$

In einer Reihe von Approximationsproblemen ist man daran interessiert, daß die stückweise polynomialen Funktionen aus $\mathcal{P}_r(\Delta)$ oder auch alle Ableitungen bis zur Ordnung $m-1$ einschließlich stetig sind. Damit werden lineare Teilräume von $\mathcal{P}_r(\Delta)$ definiert, die wir mit $\mathcal{P}^m_r(\Delta) = \{u \in \mathcal{P}_r(\Delta) | u \in C^{m-1}(a,b)\}$ bezeichnen wollen. Mit Hilfe der oben definierten Linearformen sei noch

$$f_{js} := \frac{1}{2}\left(\delta^s_{j+0} + \delta^s_{j-0}\right) \ , \quad g_{js} := \frac{1}{2}\left(\delta^s_{j+0} - \delta^s_{j-0}\right) \ , \tag{3.6}$$

für $s = 0, \ldots, m-1\,,\ j = 1, \ldots, n-1$. Umgekehrt ist dann

$$\delta^s_{j+0} = f_{js} + g_{js}\,,\ \delta^s_{j-0} = f_{js} - g_{js}\ .$$

Die Linearformen g_{js}, angewandt auf stückweise polynomiale Funktionen $u \in \mathcal{P}_r(\Delta)$, messen die halbe *Sprunghöhe* der s-ten Ableitung von u an der Stelle x_j. Eine Funktion u ist genau dann mit ihren Ableitungen bis zur Ordnung $m-1$ stetig an der Stelle x_j, wenn die *Stetigkeitsbedingung*

$$g_{js}(u) = 0\,, \qquad s = 0, \ldots, m-1\ , \tag{3.7}$$

gültig ist. Unter dieser Bedingung ist

$$f_{js}(u) = \frac{d^s u}{dx^s}(x_j)\ , \qquad s = 0, \ldots, m-1\ ,$$

im Sinne der stetigen Fortsetzung von $\frac{d^s u}{dx^s}$ in den Punkt x_j.

Die Linearformen g_{js} sind linear unabhängig auf $\mathcal{P}_r(\Delta)$, solange $s \le r$ bleibt. Betrachtet man nämlich die stückweise polynomialen Funktionen

$$u_{kt}(x) = \frac{2}{t!}(x - x_k)^t_+ \ , \quad x \in \bigcup_{j=1}^{n} I_j \ , \tag{3.8}$$

wobei $\quad (x-x_k)_+ := \max(x - x_k, 0) = \begin{cases} x - x_k\ , & x \ge x_k \\ 0 & \text{sonst}\ , \end{cases}$

dann ist $u_{kt} \in \mathcal{P}_r(\Delta)$ und

$$g_{js}(u_{kt}) = \delta_{jk}\delta_{st}\ , \qquad j,k = 1, \ldots, n-1\ , \quad s,t = 0, \ldots, r\ . \tag{3.9}$$

Letzteres ergibt sich z. B. für $s = 0$ und $t > 0$, d. h. $g_{j0}(u_{kt}) = 0$, da u_{kt} selbst stetig ist für $t \geq 1$; für $s = t = 0$ und $j \neq k$ ist offenbar $g_{j0}(u_{k0}) = 0$ und $g_{j0}(u_{j0}) = 1$, da die halbe Sprunghöhe von u_{k0} bei x_k gerade Eins beträgt. Für beliebiges s hat man entsprechend $g_{js}(u_{kt}) = 0$ für $t > s$, da $u_{kt}^{(s)}$ für $t > s$ stetig ist; weiter ist $u_{ks}^{(s)} = 0$ für $x < x_k$ und $u_{ks}^{(s)} = 2$ für $x > x_k$, so daß $g_{js}(u_{ks}) = \delta_{kj}$. Im folgenden Abschnitt 3.2 werden wir sehen, daß aufgrund der Bedingung (3.9) ein Paar biorthonormaler Systeme vorliegt, was notwendigerweise die lineare Unabhängigkeit der g_{js} nach sich zieht.

Mit Hilfe dieser Linearformen haben die Räume $\mathcal{P}_r^m(\Delta)$ die Darstellung

$$\begin{aligned} \mathcal{P}_r^m(\Delta) \;=\; \{u \in \mathcal{P}_r(\Delta) | g_{js}(u) = 0\,, \quad & j = 1, \ldots, n-1\,, \\ & s = 0, \ldots, m-1\} \;. \end{aligned} \tag{3.10}$$

Die natürliche Zahl m bezeichnet also die Anzahl der Stetigkeitsbedingungen an jedem Gitterpunkt im Innern des Intervalls $[a, b]$. Der Vektorraum $\mathcal{P}_r(\Delta)$ hat gemäß (3.2) die Dimension $n(r+1)$. Aus der Darstellung (3.10) folgt dann die Dimensionsbeziehung

$$\begin{aligned} \dim \mathcal{P}_r^m(\Delta) = n(r+1) - (n-1)m = \; & n(r+1-m) + m \;, \\ & m = 0, \ldots, r+1 \;. \end{aligned} \tag{3.11}$$

Die Dimension von $\mathcal{P}_r^{r+1}(\Delta)$ ist gleich $r+1$. In diesem Falle ist jede Funktion $u \in \mathcal{P}_r^{r+1}(\Delta)$ an jedem Gitterpunkt $x_1, \ldots, x_{n-1}$ zusammen mit ihren Ableitungen bis zur Ordnung r stetig. Folglich ist dann u ein Polynom aus $\mathcal{P}_r(\mathbb{K})$ und damit beliebig oft stetig differenzierbar auf $[a, b]$. Hieraus folgt

$$\dim \mathcal{P}_r^m(\Delta) = r+1 \;, \qquad m = r+1, r+2, \ldots \tag{3.12}$$

Die Tabelle 3.1, Seite 55, veranschaulicht die Abhängigkeit der Dimensionen der Räume $\mathcal{P}_r^m(\Delta)$ von den Parametern m und r.

Nach den Polynomen auf $[a, b]$ besitzen die stückweise polynomialen Funktionen aus den Räumen $\mathcal{P}_r^r(\Delta)$ die weitestgehenden Stetigkeitseigenschaften. Diese Funktionen nennt man *Spline–Funktionen* oder *Polynom–Splines*. Die Dimension dieser Räume ist bei festem Polynomgrad r im wesentlichen gleich der Anzahl n der Intervalle in der Zerlegung Δ (genauer: $n + r$).

In den folgenden Abschnitten sollen Interpolationsaufgaben für stückweise polynomiale Funktionen aus $\mathcal{P}_r(\Delta)$, $\mathcal{P}_r^m(\Delta)$ oder aus Teilräumen dieser Räume betrachtet werden. Dabei werden nicht nur Existenz und Eindeutigkeit von Lösungen dieser Aufgaben gezeigt, sondern auch Darstellungen dieser Lösungen und biorthonormale Basissysteme hergeleitet. Ein einfaches aber wichtiges Beispiel bildet die Interpolationsaufgabe für *lineare Spline–Funktionen:*

$$u \in \mathcal{P}_1^1(\Delta) \; : \quad u(x_j) = y_j \; , \qquad j = 0, \ldots, n \; . \tag{3.13}$$

Man kann sehr einfach eine Lösung dieser Aufgabe angeben durch

$$u(x) = y_{j-1} \frac{x_j - x}{x_j - x_{j-1}} + y_j \frac{x - x_{j-1}}{x_j - x_{j-1}} \; , \qquad x \in I_j \; , \quad j = 1, \ldots, n \; . \tag{3.14}$$

Diese Lösung ist eindeutig bestimmt. Sei nämlich v eine Lösung der zugehörigen homogenen Aufgabe

$$v \in \mathcal{P}_1^1(\Delta); \quad v(x_j) = 0 \; , \quad j = 0, \ldots, n \; . \tag{3.15}$$

Dann ist v in jedem Teilintervall $I_j = (x_{j-1},\ x_j)$ ein Polynom ersten Grades, das in den beiden Endpunkten $x_{j-1},\ x_j$ verschwindet. Folglich ist $v = 0$ in I_j für jedes $j = 1, \ldots, n$, also $v = 0$.

$r \backslash m$	0	1	2	3	4
0	n	1	1	1	1
1	$2n$	$n+1$	2	2	2
2	$3n$	$2n+1$	$n+2$	3	3
3	$4n$	$3n+1$	$2n+2$	$n+3$	4

Tabelle 3.1: $\dim \mathcal{P}_r^m(\Delta)$

3.2 Die verallgemeinerte Interpolationsaufgabe

Die Grundaufgabe der Interpolation läßt sich in einfacher Weise im Rahmen abstrakter Vektorräume formulieren. Mit bekannten Begriffsbildungen der linearen Algebra kann man Bedingungen für die eindeutige Lösbarkeit und geeignete Darstellungen der Lösungen angeben. Die in diesem Abschnitt dargelegten allgemeinen Prinzipien werden anschließend auf die Behandlung zahlreicher konkreter Interpolationsaufgaben angewandt. Die Interpolationsaufgaben nach Taylor und Lagrange in Abschnitt 3.3 liefern klassische Beispiele für die Entwicklung nach Biorthonormalsystemen.

Sei M ein endlichdimensionaler Vektorraum über dem Körper $\mathbb{K}$ der reellen Zahlen $\mathbb{R}$ oder der komplexen Zahlen $\mathbb{C}$. Sei $\{q_1, \ldots, q_l\}$ eine Menge von Linearformen auf M, die wir *Knotenfunktionale* nennen. Die Grundaufgabe der *verallgemeinerten Interpolation* besteht darin, zu beliebigen Zahlen $\lambda_1, \ldots, \lambda_l$ aus $\mathbb{K}$ ein Element u aus dem Teilraum M zu bestimmen, das den Bedingungen

$$q_j(u) = \lambda_j \ , \qquad j = 1, \ldots, l \ , \tag{3.16}$$

genügt. Wir bezeichnen mit M' den Vektorraum aller Linearformen auf M. Man nennt M' auch den *algebraischen Dualraum* von M. Da M endlichdimensional ist, gilt hier $M' = M^* := \{\lambda \in M' \mid \lambda \text{ stetig}\}$; M^* heißt der *topologische Dualraum* von M. Die Elemente u aus M kann man bekanntlich als Linearformen auf dem Raum M' auffassen, wobei $u(q) = q(u)$ gesetzt wird für alle $u \in M$ und $q \in M'$. In diesem Sinne wird in Aufgabe (3.16) also eine Funktion u gesucht, die an den *Knotenpunkten* $q_1, \ldots, q_l$ die *Knotenwerte* $\lambda_1, \ldots, \lambda_l$ annimmt.

Für das Studium der verallgemeinerten Interpolationsaufgabe werden eine Reihe von Begriffsbildungen benötigt, die wir jetzt einführen wollen. Zwei Systeme $\{p_1, \ldots, p_l\} \subset M$ und $\{q_1, \ldots, q_l\} \subset M'$ mit der Eigenschaft

$$q_j(p_k) = \delta_{jk} \ , \qquad j, k = 1, \ldots, l \ , \tag{3.17}$$

heißen *biorthonormal*. Aus dieser Eigenschaft folgt unmittelbar, daß biorthonormale Systeme linear unabhängig sind. Offenbar gilt nämlich die Beziehung

$$0 = \sum_{k=1}^{l} \alpha_k p_k \Longrightarrow 0 = \sum_{k=1}^{l} \alpha_k q_j(p_k) = \alpha_j \ , \qquad j = 1, \ldots, l \ .$$

Entsprechend schließt man für das System $\{q_1, \ldots, q_l\}$. Ein Paar biorthonormaler Vektorsysteme bildet daher ein Paar *dualer Basen* der von diesen Elementen aufgespannten Teilräume. Wenn

$$M = [p_1, \ldots, p_l] \ , \quad L = [q_1, \ldots, q_l] \tag{3.18}$$

ist, dann haben die Elemente $u \in M$ und die Linearformen $q \in M'$ bezüglich der dualen Basissysteme stets die *kanonischen Darstellungen*

$$u = \sum_{k=1}^{l} q_k(u) p_k \ , \quad q = \sum_{k=1}^{l} q(p_k) q_k \ , \tag{3.19}$$

wie man mit Hilfe von (3.17) sofort zeigt. Man sieht, daß ein System von Knotenfunktionalen $\{q_1, \ldots q_l\}$ aus M', das biorthornormal zu einer Basis von M ist, eine Basis des Dualraums M' ergibt. Da L ein Teilraum von M' ist, folgt also $L = M'$. Die kanonischen Darstellungen spielen eine zentrale Rolle in der Approximation durch Interpolation, die wir in den folgenden Abschnitten studieren wollen.

Jede Basis des Teilraums M definiert umgekehrt eine zugehörige verallgemeinerte Interpolationsaufgabe und eine biorthonormale Basis in M'. Sei $\{u_1, \ldots, u_m\}$ eine Basis des Vektorraums M. Jedes Element $u \in M$ läßt sich dann schreiben in der Form

$$u = \sum_{k=1}^{m} \lambda_k u_k \tag{3.20}$$

mit eindeutig bestimmten Koeffizienten $\lambda_1, \ldots, \lambda_m$. Damit werden durch die Vorschrift

$$f_j(u) = \lambda_j \ , \qquad j = 1, \ldots, m \ , \tag{3.21}$$

Abbildungen f_j von M in $\mathbb{K}$ erklärt. Wie man leicht sieht, sind diese Abbildungen linear, also f_j Linearformen auf M. Weiter folgt aus der obigen Definition unmittelbar die Eigenschaft

$$f_j(u_k) = \delta_{jk} \ , \qquad j, k = 1, \ldots, m \ . \tag{3.22}$$

Daher bilden $\{u_1, \ldots, u_m\}$, $\{f_1, \ldots, f_m\}$ ein biorthonormales System, und es gelten die kanonischen Darstellungen

$$u = \sum_{k=1}^{m} f_k(u) u_k \ , \quad f = \sum_{k=1}^{m} f(u_k) f_k$$

für jedes $u \in M$, $f \in M'$.

Ein *erstes Beispiel* für die verallgemeinerte Interpolationsaufgabe stellt die Aufgabe (3.13) dar, wobei der Raum M und die Knotenfunktionale durch

$$M = \mathcal{P}_1^1(\Delta) \ , \quad q_j = \delta_j \ , \qquad j = 0, \ldots, n \ ,$$

definiert sind. Die zu $\{q_0, \ldots, q_n\}$ biorthonormale Basis $\{p_0, \ldots, p_n\}$ von $\mathcal{P}_1^1(\Delta)$ erhält man aus (3.13) mit den rechten Seiten $y_j = \delta_{jk}$, d. h. $p_k \in \mathcal{P}_1^1(\Delta) : p_k(x_j) = \delta_{jk}$. Die zugehörigen Lösungen p_k veranschaulicht die Abb. 3.1.

Jede Funktion $u \in \mathcal{P}_1^1(\Delta)$ hat die kanonische Darstellung

$$u(x) = \sum_{k=0}^{n} u(x_k) p_k(x) \ , \qquad x \in [a, b] \ .$$

Das bestimmte Integral über $[a, b]$ definiert die Linearform

$$q(u) = \int_a^b u(x)\, dx \ , \qquad u \in \mathcal{P}_1^1(\Delta) \ .$$

Diese Linearform hat die kanonische Darstellung

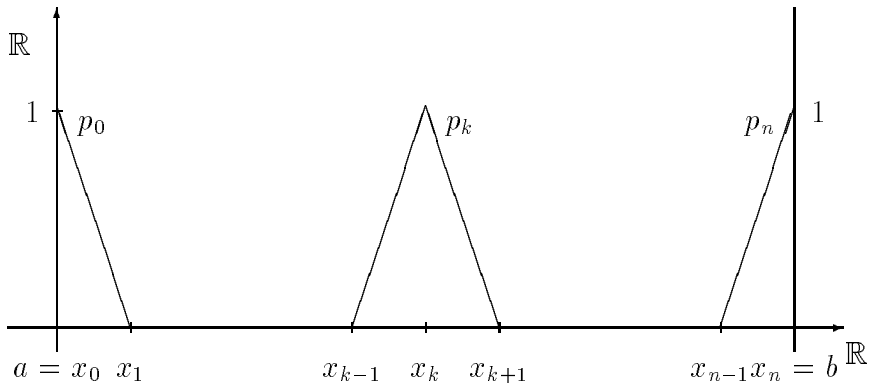


Abbildung 3.1: Basisfunktionen $p_0, \ldots, p_k, \ldots, p_n$ in $\mathcal{P}_1^1(\Delta)$

$$\int_a^b u(x)\,dx = \sum_{k=0}^{n} \beta_k u(x_k)\;, \qquad u \in \mathcal{P}_1^1(\Delta)\;, \tag{3.23}$$

mit den Koeffizienten

$$\beta_k = \int_a^b p_k(x)\,dx\;, \qquad k = 0, \ldots, n\;.$$

Für ein äquidistantes Gitter mit der Maschenweite $h = x_j - x_{j-1}\,,\ j = 1, \ldots, n$, ist offenbar

$$\beta_0 = \beta_n = \frac{h}{2}\;, \quad \beta_1 = \ldots = \beta_{n-1} = h\;,$$

also ist (3.23) die *summierte Sehnentrapezformel* .

Eine weitere wichtige Klasse von Linearformen erhält man durch die *erste Ableitung* der Funktionen $u \in \mathcal{P}_1^1(\Delta)$ an beliebigen, aber festen Stellen c in $[a, b]$. Betrachten wir die Linearform $f(u) = \frac{du}{dx}(c)$ für $u \in \mathcal{P}_1^1(\Delta)$. Die kanonische Darstellung ergibt dann

$$\frac{du}{dx}(c) = \frac{1}{h_j}(u(x_j) - u(x_{j-1}))\;, \qquad u \in \mathcal{P}_1^1(\Delta)\;, \tag{3.24}$$

falls $x_{j-1} < c < x_j$ ist. Wenn c gleich einer der Stützstellen x_j ist, dann setzen wir

$$f(u) := f_{j1}(u) = \frac{1}{2}\left(\frac{du}{dx}(c+0) + \frac{du}{dx}(c-0)\right)\;. \tag{3.25}$$

Wie in (3.24) ergibt sich

$$f_{j1}(u) = \frac{1}{2h_{j+1}} \left(u(x_{j+1}) - u(x_j)\right) + \frac{1}{2h_j} \left(u(x_j) - u(x_{j-1})\right) \ . \tag{3.26}$$

Bei äquidistanten Stützstellen mit der Schrittweite $h_j = h$ ist daher

$$f_{j1}(u) = \frac{1}{2h} \left(u(x_{j+1}) - u(x_{j-1})\right) \ , \qquad u \in \mathcal{P}_1^1(\Delta) \ .$$

Auf diesem Wege erhält man also Darstellungen der „Ableitungen“ von stückweise linearen Funktionen durch *Differenzenquotienten* .

Eine erste Anwendung der obigen Begriffe ist das folgende Kriterium:

Satz 3.1 *Die verallgemeinerte Interpolationsaufgabe (3.16) ist dann und nur dann eindeutig lösbar für jede rechte Seite* $(\lambda_1, \ldots, \lambda_l) \in \mathbb{K}^l$, *wenn eine zu* $\{q_1, \ldots, q_l\}$ *biorthonormale Basis* $\{p_1, \ldots, p_l\}$ *von* M *existiert. In diesem Falle hat die Lösung* u *von (3.16) die Darstellung*

$$u = \sum_{k=1}^{l} \lambda_k p_k \ . \tag{3.27}$$

Beweis: (i) Mit einem Paar dualer, biorthonormaler Basissysteme $\{p_1, \ldots, p_l\}$ von M und $\{q_1, \ldots, q_l\}$ von L erhält man durch (3.27) eine Lösung u der Interpolationsaufgabe (3.16), wie man mit Hilfe von (3.17) unmittelbar zeigt. Diese Lösung ist eindeutig bestimmt, weil $v \in M$ und $q_j(v) = 0$ für $j = 1, \ldots, l$ mit Hilfe von (3.19) notwendig $v = 0$ ergibt.

(ii) Umgekehrt sei jetzt die Interpolationsaufgabe (3.16) lösbar für jede rechte Seite. Dann existieren offenbar für jedes $k = 1, \ldots, l$ Lösungen $p_k \in M$ der obigen Gleichungen (3.17). Das System $\{p_1, \ldots, p_l\}$ ist linear unabhängig, wie wir bereits festgestellt haben. Sei u ein beliebiges Element aus M und setzen wir

$$v = u - \sum_{k=1}^{l} q_k(u) p_k \ ,$$

dann ist $v \in M$ und $q_j(v) = 0$ für $j = 1, \ldots, l$, also v Lösung der Interpolationsaufgabe (3.16) mit Nullen auf der rechten Seite. Nach Voraussetzung ist die Interpolationsaufgabe eindeutig lösbar, so daß $v = 0$ sein muß. Folglich hat jedes Element $u \in M$ die Darstellung (3.19), so daß $\{p_1, \ldots, p_l\}$ eine Basis von M ist. □

Die Menge $\{q_1, \ldots, q_l\}$ aus M' heißt *M–unisolvent* , wenn die verallgemeinerte Interpolationsaufgabe (3.16) für jede rechte Seite eindeutig lösbar ist. Der obige Satz 3.1 gibt also eine notwendige und hinreichende Bedingung für die M–Unisolvenz der Menge $\{q_1, \ldots, q_l\}$. In Anwendungen und Beispielen läßt sich die M–Unisolvenz damit oft sehr leicht folgendermaßen nachweisen. Als erstes stellt man fest, daß der Teilraum M von l Elementen aufgespannt wird, also $M = [u_1, \ldots, u_l]$ ist. Damit ist M höchstens l–dimensional. Dann gibt man explizit ein zu $\{q_1, \ldots, q_l\}$ biorthonormales System $\{p_1, \ldots, p_l\}$ in M an. Wie wir oben gezeigt haben, ist in diesem Fall $\{p_1, \ldots, p_l\}$ linear unabhängig und folglich eine Basis von M.

In einigen Fällen ist es jedoch nicht ohne weiteres möglich, eine duale Basis explizit anzugeben, so daß man weitere Kriterien benötigt. Ein wichtiges Hilfsmittel dazu bildet die *Gramsche Determinante*

$$\det(q_j(u_k))$$

zweier Systeme $\{u_1, \ldots, u_l\} \subset M$, $\{q_1, \ldots, q_l\} \subset M'$. Mit Hilfe der Gramschen Determinante erhält man ein einfaches hinreichendes Kriterium für die lineare Unabhängigkeit der beiden Vektorsysteme.

Satz 3.2 *Ist die Gramsche Determinante* $\det(q_j(u_k))$ *ungleich Null, dann sind die Systeme* $\{u_1, \ldots, u_l\}$ *und* $\{q_1, \ldots, q_l\}$ *linear unabhängig.*

Beweis: Es gilt die Aussage

$$0 = \sum_{k=1}^{l} \alpha_k u_k \Longrightarrow 0 = \sum_{k=1}^{l} \alpha_k q_j(u_k) \; , \qquad j = 1, \ldots, l \; .$$

Da $\det(q_j(u_k)) \neq 0$ ist, besitzt das homogene Gleichungssystem nur die triviale Lösung $\alpha_1 = \ldots = \alpha_l = 0$. Entsprechend hat man

$$0 = \sum_{k=1}^{l} \beta_k q_k \Longrightarrow 0 = \sum_{k=1}^{l} \beta_k q_k(u_j) \; , \qquad j = 1, \ldots, l \; .$$

Es ist $\det(q_k(u_j)) = \det(q_j(u_k)) \neq 0$, also folgt auch in diesem Falle $\beta_1 = \ldots = \beta_l = 0$. □

Die Gramsche Determinante gestattet die folgende Charakterisierung der eindeutigen Lösbarkeit der Aufgabe (3.16).

Satz 3.3 *Die verallgemeinerte Interpolationsaufgabe (3.16) ist dann und nur dann eindeutig lösbar für jede rechte Seite, wenn ein System* $\{u_1, \ldots, u_l\}$ *in* M *existiert mit der Eigenschaft*

$$M = [u_1, \ldots, u_l] , \qquad \det(q_j(u_k)) \neq 0 . \tag{3.28}$$

Beweis: (i) Ist die Gramsche Determinante ungleich Null, so sind die Vektoren $u_1, \ldots, u_l$ eine Basis von M. Zu jedem System $(\lambda_1, \ldots, \lambda_l) \in \mathbb{K}^l$ besitzt das inhomogene Gleichungssystem

$$\sum_{k=1}^{l} \alpha_k q_j(u_k) = \lambda_j , \qquad j = 1, \ldots, l ,$$

eine Lösung $\alpha_1, \ldots, \alpha_l$. Hiermit erhält man eine Lösung von (3.16) durch $u = \sum_{k=1}^{l} \alpha_k u_k$. Diese Lösung ist eindeutig bestimmt, weil für jedes $v = \sum_{k=1}^{l} \beta_k u_k$ aus M das homogene Gleichungssystem

$$q_j(v) = 0 \Longleftrightarrow \sum_{k=1}^{l} \beta_k q_j(u_k) = 0$$

für $j = 1, \ldots, l$ nur die triviale Lösung $\beta_1 = \ldots = \beta_l = 0$ hat.

(ii) Ist umgekehrt die verallgemeinerte Interpolationsaufgabe eindeutig lösbar, dann hat die zu $\{q_1, \ldots, q_l\}$ biorthonormale Basis $\{p_1, \ldots, p_l\}$ von M offenbar die Eigenschaft (3.28). □

Für die späteren Untersuchungen wollen wir schließlich noch eine weitere Charakterisierung der eindeutigen Lösbarkeit der verallgemeinerten Interpolationsaufgabe angeben. Sei $\{u_1, \ldots, u_m\}$ eine Basis von M und L der von $\{q_1, \ldots, q_l\}$ aufgespannte Teilraum,

$$M = [u_1, \ldots, u_m] , \quad L = [q_1, \ldots, q_l] . \tag{3.29}$$

Dann läßt sich die Interpolationsaufgabe (3.16) in der äquivalenten Form des *zugehörigen inhomogenen Gleichungssystems*

$$\sum_{k=1}^{m} \alpha_k q_j(u_k) = \lambda_j , \qquad j = 1, \ldots, l , \tag{3.30}$$

schreiben. Die $l \times m$ Matrix $(q_j(u_k))$ dieses Gleichungssystems bezeichnen wir als *Gramsche Matrix* der Systeme $\{u_1, \ldots, u_m\}$, $\{q_1, \ldots, q_l\}$. Dieses Gleichungssystem besitzt bekanntlich höchstens eine Lösung, wenn die m Spaltenvektoren der Gramschen Matrix im Zahlenraum $\mathbb{K}^l$ linear unabhängig sind, also die Aussage

$$\sum_{k=1}^{m} \alpha_k q_j(u_k) = 0 \ , \qquad j = 1, \ldots, l \Longrightarrow \alpha_1 = \ldots = \alpha_m = 0 \tag{3.31}$$

gilt. Diese Aussage ist gleichbedeutend damit, daß die Gramsche Matrix den Rang

$$Rg(q_j(u_k)) = m \tag{3.32}$$

hat. Insbesondere muß in diesem Falle $m \leq l$ sein. Weiter sind die obigen Aussagen offenbar äquivalent zur *Eindeutigkeitsbedingung*

$$u \in M \ ; \ q_j(u) = 0 \ , \qquad j = 1, \ldots, l \Longrightarrow u = 0 \ . \tag{3.33}$$

Das inhomogene Gleichungssystem (3.30) ist genau dann lösbar für jede rechte Seite, wenn die Spaltenvektoren der Gramschen Matrix den Zahlenraum aufspannen, also die Matrix den Rang

$$Rg(q_j(u_k)) = l \tag{3.34}$$

besitzt. Diese Bedingung ist gleichbedeutend damit, daß die l Zeilen der Gramschen Matrix linear unabhängig in $\mathbb{K}^m$ sind, weil der Spaltenrang einer Matrix bekanntlich gleich ihrem Zeilenrang ist. Das heißt,

$$\sum_{j=1}^{l} \beta_j q_j(u_k) = 0 \ , \quad k = 1, \ldots, m \Longrightarrow \beta_1 = \ldots = \beta_l = 0 \ . \tag{3.35}$$

Die Elemente $u_1, \ldots, u_m$ spannen den Raum M auf. Die letzte Bedingung ist daher äquivalent zu

$$\sum_{j=1}^{l} \beta_j q_j(u) = 0 \ , \quad u \in M \Longrightarrow \beta_1 = \ldots = \beta_l = 0 \ ,$$

also zur *Lösbarkeitsbedingung:*

$$q_1, \ldots, q_l \ \textit{sind linear unabhängig} \tag{3.36}$$

Insgesamt kann man also die Aussage treffen, daß die verallgemeinerte Interpolationsaufgabe (3.16) bzw. das Gleichungssystem (3.30) für jede rechte Seite eindeutig lösbar ist, wenn für die Gramsche Matrix $Rg(q_j(u_k)) = l = m$ gilt.

3.3 Klassische Interpolationsaufgaben

Die bekannten klassischen Interpolationsaufgaben liefern erste Beispiele für die allgemeinen Begriffe und Sätze des vorangegangenen Abschnitts. Gleichzeitig erhält man hiermit wichtige Hilfsmittel zur Behandlung weiterer Interpolationsprobleme für Polynome einer und mehrerer Veränderlicher. Zusammen mit der kanonischen Darstellung des Interpolationsproblems werden in der Regel die zugehörigen interpolatorischen Quadraturformeln für bestimmte Integrale angegeben. Eine weitere, für die Anwendungen interessante Klasse von Linearformen erhält man durch Differenzenquotienten, mit denen Ableitungen der interpolierenden Polynome dargestellt werden.

Beispiel 3.4 Taylor–Interpolation

Sei $\mathcal{P} = \mathcal{P}(\mathbb{K})$ der Vektorraum aller Polynome in einer Veränderlichen, aufgefaßt als Funktionen auf $\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$. Sei $\mathcal{P}_{r-1} = \mathcal{P}_{r-1}(\mathbb{K})$ der Teilraum aller Polynome höchstens $(r-1)$-ten Grades. Die Taylor–Interpolation bestimmt ein Polynom $u \in \mathcal{P}_{r-1}$, das an einer Stelle x_0 den Bedingungen

$$\frac{d^j u}{dx^j}(x_0) = \lambda_j \ , \qquad j = 0, \ldots, r-1 \ , \tag{3.37}$$

genügt. Durch die Vorschrift

$$\delta_0^{(j)}(v) = \frac{d^j v}{dx^j}(x_0) \ , \qquad j = 0, 1, 2, \ldots , \tag{3.38}$$

für Polynome $v \in \mathcal{P}$ werden offenbar Linearformen auf $\mathcal{P}$ definiert. Wir schreiben auch

$$\delta_0 = \delta_0^{(0)} \ , \quad \delta_0' = \delta_0^{(1)} \ , \quad \delta_0'' = \delta_0^{(2)}, \ldots \ .$$

Diese Linearformen sind bis auf das Vorzeichen gleich den Ableitungen des Diracschen δ–Funktionals an der Stelle x_0. Die Interpolationsaufgabe (3.37) hat die Form der verallgemeinerten Interpolationsaufgabe mit den Knotenfunktionalen $q_j = \delta_0^{(j)}$ und den Räumen

$$M = [1, x, \ldots, x^{r-1}] = \mathcal{P}_{r-1} \ , \quad L = [\delta_0, \delta_0', \ldots, \delta_0^{(r-1)}] \ . \tag{3.39}$$

Man kann leicht ein zu $\delta_0, \ldots, \delta_0^{(r-1)}$ biorthonormales System von Polynomen durch

$$p_k(x) = \frac{1}{k!}(x - x_0)^k \ , \qquad k = 0, \ldots, r-1 \ , \tag{3.40}$$

explizit angeben. Da $\mathcal{P}_{r-1}$ von höchstens r Elementen aufgespannt wird, ist $\{p_0, \ldots, p_{r-1}\}$ eine Basis von $\mathcal{P}_{r-1}$. Nach Satz 3.1 ist somit die Interpolationsaufgabe (3.37) für jede rechte Seite eindeutig lösbar durch

$$u(x) = \sum_{k=0}^{r-1} \lambda_k (x - x_0)^k \frac{1}{k!} \ . \tag{3.41}$$

Die kanonische Darstellung (3.19) bezüglich der biorthonormalen Systeme $\{p_0, \ldots, p_{r-1}\}$, $\{\delta_0, \ldots, \delta_0^{(r-1)}\}$ lautet

$$p(x) = \sum_{k=0}^{r-1} \frac{d^k p}{dx^k}(x_0) \frac{1}{k!} (x - x_0)^k \tag{3.42}$$

für jedes $p \in \mathcal{P}_{r-1}$. Dieses ist die bekannte *Taylorsche Formel* zur Darstellung von Polynomen. □

Beispiel 3.5 Lagrange–Interpolation

Sei eine Menge von $m+1$ paarweise verschiedenen Punkten $x_0, \ldots, x_m$ aus $\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$ gegeben. Dann lautet die Grundaufgabe der *Lagrange–Interpolation: Gegeben* $\lambda_0, \ldots, \lambda_m$, *gesucht* $u \in \mathcal{P}_m$ *mit*

$$u(x_j) = \lambda_j \ , \qquad j = 0, \ldots, m \ . \tag{3.43}$$

Die Knotenfunktionale q_j sind also die δ–Funktionale

$$\delta_j(v) = v(x_j) \ , \qquad j = 0, \ldots, m \ , \tag{3.44}$$

die für alle stetigen Funktionen v erklärt sind. Die Räume M, L haben hier die Form

$$M = [1, x, \ldots, x^m] = \mathcal{P}_m \ , \quad L = [\delta_0, \ldots, \delta_m] \ . \tag{3.45}$$

Mit den bekannten *Lagrange Polynomen*

$$L_k(x) = \prod_{\substack{l=0 \\ l \neq k}}^{m} \frac{x - x_l}{x_k - x_l} \ , \qquad k = 0, \ldots, m \ ,$$

erhält man offenbar das zu $\{\delta_0, \ldots, \delta_m\}$ biorthonormale System von Polynomen in $\mathcal{P}_m$. Nach Satz 3.1 ist die Interpolationsaufgabe (3.43) für jede rechte Seite eindeutig lösbar. Die kanonische Darstellung (3.19) führt auf die *Lagrange–Darstellung*

$$p(x) = \sum_{k=0}^{m} p(x_k) L_k(x) \tag{3.46}$$

der Polynome $p \in \mathcal{P}_m$.

Sei $[a, b]$ ein beschränktes Intervall der reellen Zahlengeraden und die Linearform q durch das bestimmte Integral

$$q(p) = \int_a^b p(x)\, dx \ , \qquad p \in \mathcal{P}_m \ ,$$

definiert. Dann ergibt die kanonische Darstellung (3.19) für q die bekannte *interpolatorische Quadraturformel*

$$\int_a^b p(x)\, dx = \sum_{k=0}^{m} \beta_k p(x_k) \ , \qquad p \in \mathcal{P}_m \ , \tag{3.47}$$

mit den Koeffizienten

$$\beta_k = \int_a^b L_k(x)\, dx \ , \qquad k = 0, \ldots, m \ .$$

Die Newtonsche Darstellung von Interpolationspolynomen gewinnt man durch Übergang zu anderen Basissystemen in $M = \mathcal{P}_m$, $L = [\delta_0, \ldots, \delta_m] = M'$. Wir bezeichnen mit $\delta[x_j, \ldots, x_{j+l}]$ diejenigen Linearformen, die jeder stetigen Funktion v die zugehörigen *Differenzenquotienten* l-ter Ordnung zuordnen, das heißt,

$$\delta[x_j, \ldots, x_{j+l}](v) = v[x_j, \ldots, x_{j+l}] \tag{3.48}$$

für $j = 0, \ldots, m - l$, $l = 1, \ldots, m$. Damit ist

$$\begin{aligned} \delta[x_j] &= \delta_j \\ \delta[x_j, \ldots, x_{j+l}] &= \frac{1}{x_j - x_{j+l}} \left\{ \delta[x_j, \ldots, x_{j+l-1}] - \delta[x_{j+1}, \ldots, x_{j+l}] \right\} \end{aligned} \tag{3.49}$$

Diese Linearformen sind Linearkombinationen der δ–Funktionale $\delta_0, \ldots, \delta_m$ und folglich wieder aus L. Hiermit setzen wir jetzt

$$q_j = \delta[x_0, \ldots, x_j] \ , \qquad j = 0, \ldots, m \ , \tag{3.50}$$

und definieren die Polynome

$$p_0(x) = 1 \ , \quad p_k(x) = (x - x_0) \cdots (x - x_{k-1}) \ , \qquad k = 1, \ldots, m \ . \tag{3.51}$$

Die Newtonsche Darstellung von Interpolationspolynomen lautet in diesen Bezeichnungen bekanntlich

$$u(x) = u[x_0] + \sum_{k=1}^{m} u[x_0, \ldots, x_k](x - x_0) \cdots (x - x_{k-1}) = \sum_{k=0}^{m} q_k(u) p_k(x) \ . \quad (3.52)$$

Das System $\{p_0, \ldots, p_m\}$ ist linear unabhängig und daher eine Basis von $\mathcal{P}_m$. Setzt man nämlich $\sum_{k=0}^{m} \alpha_k p_k = 0$, dann muß $\alpha_m = 0$ sein, da der führende Koeffizient von p_m (bei x^m) 1 ist und die Monome bekanntlich linear unabhängig sind — usw. für $\alpha_{m-1},\ \alpha_{m-2}, \ldots, \alpha_0$.

Aus der Darstellung (3.52) gewinnt man

$$\delta[x_0, \ldots, x_j](p_k) = \delta_{jk} \ , \qquad j, k = 0, \ldots, m \ ,$$

wie wir in (3.22) im allgemeinen Rahmen ausgeführt haben. Folglich sind die beiden Systeme $\{p_0, \ldots, p_m\}$, $\{\delta[x_0], \ldots, \delta[x_0, \ldots, x_m]\}$ ein Paar biorthonormaler Basen von $M = \mathcal{P}_m$ und $L = [\delta_0, \ldots, \delta_m] = M'$. Die zugehörige kanonische Darstellung ist gerade die Newtonsche Darstellung (3.52) für Polynome aus $\mathcal{P}_m$. □

Beispiel 3.6 Hermite–Interpolation

Gegeben seien $m+1$ paarweise verschiedene Punkte $x_0, \ldots, x_m \in \mathbb{K}$ und $m+1$ nichtnegative ganze Zahlen $r_0, \ldots, r_m$. Die zugehörige *Hermitesche Interpolationsaufgabe* lautet:

$$\frac{d^s u}{dx^s}(x_j) = \lambda_{js} \ , \qquad s = 0, \ldots, r_j - 1 \ ,\ j = 0, \ldots, m \ . \quad (3.53)$$

An jeder Stelle x_j sind r_j Bedingungen an die Funktion u gestellt, insgesamt also

$$r = r_0 + \cdots + r_m$$

Bedingungen. Daher wird ein Interpolationspolynom u vom Grade $r-1$ als Lösung der Aufgabe (3.53) gesucht. Die zugehörige *Knotenfunktionale* q_{js} bezeichnen wir mit $\delta_j^{(s)}$, definiert durch

$$\delta_j^{(s)}(v) = \frac{d^s v}{dx^s}(x_j)$$

für alle Polynome v. Dementsprechend wird jetzt

$$\begin{aligned} M &= [1, x, \ldots, x^{r-1}] = \mathcal{P}_{r-1} \ , \\ L &= [\delta_j^{(s)} | s = 0, \ldots, r_j - 1 \ ,\ j = 0, \ldots, m] \ . \end{aligned} \quad (3.54)$$

Offensichtlich sind die oben bereits betrachteten Interpolationsaufgaben von Taylor und Lagrange hier als Spezialfall enthalten. Die Taylor–Interpolation erhält man für $m = 0$ und die Lagrange–Interpolation für $r_j = 1\,,\ j = 0, \ldots, m$.

Wir wollen zeigen, daß die Hermitesche Interpolationsaufgabe eindeutig lösbar ist. Hierzu geben wir ein geeignetes System von Polynomen u_{kt} in $\mathcal{P}_{r-1}$ an, das mit den Knotenfunktionalen $\delta_j^{(s)}$ eine von Null verschiedene Gramsche Determinante besitzt. Sei

$$u_{kt}(x) = v_{kt}(x) w_k(x)\,, \quad v_{kt}(x) = \frac{1}{t!}(x - x_k)^t\,, \quad w_k(x) = \prod_{\substack{\ell=0 \\ \ell \neq k}}^{m} (x - x_\ell)^{r_\ell}\,,$$

für $t = 0, \ldots, r_k - 1\,,\ k = 0, \ldots, m$. Dann ist w_k ein Polynom höchstens $(r - r_k)$-ten Grades und

$$\delta_j^{(s)}(w_k) = \frac{d^s w_k}{dx^s}(x_j) = 0\,, \qquad s = 0, \ldots, r_j - 1\,,\ j \neq k\,,\ j = 0, \ldots, m\,.$$

Weiter hat man $w_j(x_j) \neq 0$, weil nach Voraussetzung $x_j \neq x_\ell$ ist für $j \neq \ell$. Offenbar wird

$$\frac{d^s v_{kt}}{dx^s}(x_k) = \delta_{st}\,, \qquad s, t = 0, \ldots, r_k - 1\,,$$

denn man hat

$$\begin{aligned}
s < t \quad &: \quad (x - x_k)^{t-s}\Big|_{x_k} = 0\,; \\
s = t \quad &: \quad \frac{t!}{t!} = 1\,; \\
s > t \quad &: \quad v_{kt}^{(s)} = 0\,.
\end{aligned}$$

Damit erhält man

$$\begin{aligned}
\delta_j^{(s)}(u_{kt}) &= \frac{d^s u_{kt}}{dx^s}(x_j) = \sum_{\ell=0}^{s} \binom{s}{\ell} \underbrace{v_{kt}^{(\ell)}(x_j)}_{=\delta_{\ell t}\ (j=k)} \underbrace{w_k^{(s-\ell)}(x_j)}_{=0,\, j \neq k} \\
&= \binom{s}{t} \underbrace{v_{kt}^{(t)}(x_k)}_{=1}\, w_k^{(s-t)}(x_k) \\
&= \delta_{jk} \binom{s}{t} w_k^{(s-t)}(x_j)\,.
\end{aligned}$$

Wie üblich ist dabei $\binom{s}{t} = 0$ für $t > s$. Die Gramsche Determinante ist genau dann ungleich Null, wenn das zugehörige homogene Gleichungssystem nur die triviale Lösung hat. Dieses Gleichungssystem hat auf Grund der letzten Beziehung die Form

$$\begin{aligned} 0 &= \sum_{k=0}^{m} \sum_{t=0}^{r_k-1} \gamma_{kt} \delta_j^{(s)}(u_{kt}) = \sum_{t=0}^{r_j-1} \gamma_{jt} \delta_j^{(s)}(u_{jt}) \\ &= \sum_{t=0}^{r_j-1} \gamma_{jt} \binom{s}{t} w_j^{(s-t)}(x_j) \ , \qquad s = 0, \ldots, r_j - 1 \ , \ j = 0, \ldots, m \ . \end{aligned}$$

Für jedes j hat das System dieser r_j homogenen Gleichungen für die Unbekannten $\gamma_{j0}, \ldots, \gamma_{jr_j-1}$ Dreiecksgestalt,

$$\begin{aligned} 0 &= \gamma_{j0} w_j(x_j) \ , \\ 0 &= \gamma_{j0} w_j'(x_j) + \gamma_{j1} w_j(x_j) \ , \\ &\vdots \\ 0 &= \gamma_{j0} w_j^{(r_j-1)}(x_j) + \binom{r_j - 1}{1} \gamma_{j1} w_j^{(r_j-2)}(x_j) + \cdots + \gamma_{jr_j-1} w_j(x_j) \ . \end{aligned}$$

Da $w_j(x_j) \neq 0$ ist, folgt hieraus

$$\gamma_{j0} = \gamma_{j1} = \cdots = \gamma_{jr_j-1} = 0 \ .$$

Also genügt die Gramsche Determinante der Bedingung

$$\det \left(\delta_j^{(s)}(u_{kt}) \right) \neq 0 \ .$$

Auf Grund der Sätze 3.2 und 3.3 sind dann die beiden Systeme u_{js} in $\mathcal{P}_{r-1}$ und $\delta_j^{(s)}$ in $\mathcal{P}_{r-1}'$ für $s = 0, \ldots, r_j - 1$, $j = 0, \ldots, m$, linear unabhängig und die Hermitesche Interpolationsaufgabe (3.53) ist für jede rechte Seite eindeutig lösbar durch ein Polynom $u \in \mathcal{P}_{r-1}$. □

Ein wichtiges *Beispiel* für die obigen Überlegungen erhält man mit Hermite–Interpolationen der Ordnungen $r_j = 2$ für alle $j = 0, \ldots, m$. Wir modifizieren w_j und definieren

$$w_j(x) := \prod_{\substack{k=0 \\ j \neq k}}^{m} \left(\frac{x - x_k}{x_j - x_k} \right)^2 \ , \qquad j = 0, \ldots, m \ , \tag{3.55}$$

und somit

$$w_j(x_k) = \delta_{jk} \ , \qquad j,k = 0,\ldots,m \ .$$

Die Funktionen w_j haben doppelte Nullstellen für $x_k \neq x_j$. Dann wird durch

$$\begin{aligned} p_{j0}(x) &= (1 - w_j'(x_j)(x - x_j)\,)w_j(x) \ , \\ p_{j1}(x) &= (x - x_j)w_j(x) \ , \qquad j = 0,\ldots,m \ , \end{aligned} \tag{3.56}$$

die biorthonormale Basis in $\mathcal{P}_{2m+1}[a,b]$ erklärt. Es ist nämlich

$$\begin{aligned} p_{j0}'(x) &= (1 - w_j'(x_j)(x - x_j)\,)w_j'(x) - w_j'(x_j)w_j(x) \ , \\ p_{j1}'(x) &= (x - x_j)w_j'(x) + w_j(x) \ . \end{aligned}$$

Für $k \neq j$ erhält man wegen $w_j'(x_k) = 0$

$$p_{j0}(x_k) = p_{j0}'(x_k) = p_{j1}(x_k) = p_{j1}'(x_k) = 0 \ ,$$

und für $k = j$ ist

$$p_{j0}(x_j) = 1 \ , \ p_{j0}'(x_j) = 0 \ ; \quad p_{j1}(x_j) = 0 \ , \ p_{j1}'(x_j) = 1 \ .$$

Jedes Polynom $u \in \mathcal{P}_{2m+1}$ hat also die kanonische Darstellung

$$u(x) = \sum_{k=0}^{m} \left[u(x_k)p_{k0}(x) + \frac{du}{dx}(x_k)p_{k1}(x) \right] \ . \tag{3.57}$$

Das bestimmte Integral über ein beschränktes Intervall $[a,b]$ definiert die Linearform

$$q(u) = \int_a^b u(x)\,dx \ , \qquad u \in \mathcal{P}_{2m+1} \ .$$

Die zugehörige kanonische Darstellung ergibt

$$\int_a^b u(x)\,dx = \sum_{k=0}^{m} \left[\beta_k u(x_k) + \beta_k' \frac{du}{dx}(x_k) \right] \tag{3.58}$$

mit den Koeffizienten

$$\beta_k = \int_a^b p_{k0}(x)\,dx \ , \quad \beta_k' = \int_a^b p_{k1}(x)\,dx \ , \qquad k = 0,\ldots,m \ .$$

Wählt man als Stützstellen $x_0, \ldots, x_m$ die Wurzeln des Legendre–Polynoms P_{m+1} zum Intervall $[a, b]$, so folgt aus der Orthogonalität von P_{m+1} zu allen Polynomen aus $\mathcal{P}_m$ die Bedingung

$$\begin{aligned}\int_a^b \underbrace{(x-x_j)w_j(x)}_{p_{j1}(x)}\,dx &= \int_a^b (x-x_0)\cdots(x-x_m)\prod_{\substack{k=0\\k\neq j}}^{m}\frac{x-x_k}{(x_j-x_k)^2}\,dx \\ &= \int_a^b P_{m+1}(x)\prod\cdots dx = 0\ .\end{aligned}$$

Daher ist in diesem Falle

$$\begin{aligned}\beta_j' &= \int_a^b p_{j1}(x)\,dx = 0\ , \qquad j=0,\ldots,m\ , \\ \beta_j &= \int_a^b p_{j0}(x)\,dx \overset{(3.56)}{=} \int_a^b w_j(x)\,dx - w_j'(x_j)\underbrace{\int_a^b (x-x_j)w_j(x)\,dx}_{=0}\end{aligned}$$

und (3.58) wird die bekannte *Gaußsche Quadraturformel* (vgl. z.B. [SH82, §4])

$$\int_a^b u(x)\,dx = \sum_{k=0}^{m}\beta_k u(x_k)\ , \qquad u\in\mathcal{P}_{2m+1}\ .$$

3.4 Interpolation durch stückweise quadratische Funktionen einer Veränderlichen

Dieser Abschnitt untersucht Interpolationsaufgaben für stückweise quadratische Funktionen aus Teilräumen von $\mathcal{P}_2(\Delta)$. Zur Bestimmung von Lagrange–Interpolationen bieten sich hier drei verschiedene Wege an. Zuerst wird interpoliert durch stetige, stückweise quadratische Funktionen aus $\mathcal{P}_2^1(\Delta)$. Im zweiten Abschnitt wird mit Stetigkeitsbedingungen gearbeitet, die der klassischen Vorgehensweise zur Herleitung von Differenzenapproximationen entsprechen. Der dritte Abschnitt zeigt, daß die Lagrangesche Interpolationsaufgabe zusammen mit einer Anfangsbedingung für die Ableitung stets durch quadratische Spline–Funktionen aus $\mathcal{P}_2^2(\Delta)$ eindeutig gelöst werden kann. Man erhält hiermit besonders glatte interpolierende Funktionen.

3.4.1 Lagrange–Interpolation in $\mathcal{P}_2^1(\Delta)$

Zu jedem Teilintervall $I_j = (a_j, b_j) = (x_{j-1}, x_j)$ sei noch ein weiterer Punkt $c_j = x_{j-\frac{1}{2}} \in I_j$ definiert für $j = 1, \ldots, n$. Damit ist die Interpolationsaufgabe

$$u \in \mathcal{P}_2^1(\Delta):\ u(x_j) = y_j, \quad j = 0, \frac{1}{2}, 1, \ldots, n - \frac{1}{2}, n, \tag{3.59}$$

gestellt. Zu jedem Intervall I_j und den Stützstellen $x_{j-1}, x_{j-\frac{1}{2}}, x_j$ sind die zugehörigen Lagrangepolynome $L_0^j, L_1^j, L_2^j \in \mathcal{P}_2$ erklärt:

$$\begin{aligned} L_0^j(x) &= \frac{x - x_{j-\frac{1}{2}}}{x_{j-1} - x_{j-\frac{1}{2}}} \cdot \frac{x - x_j}{x_{j-1} - x_j}, \\ L_1^j(x) &= \frac{x - x_{j-1}}{x_{j-\frac{1}{2}} - x_{j-1}} \cdot \frac{x - x_j}{x_{j-\frac{1}{2}} - x_j}, \\ L_2^j(x) &= \frac{x - x_{j-1}}{x_j - x_{j-1}} \cdot \frac{x - x_{j-\frac{1}{2}}}{x_j - x_{j-\frac{1}{2}}}. \end{aligned}$$

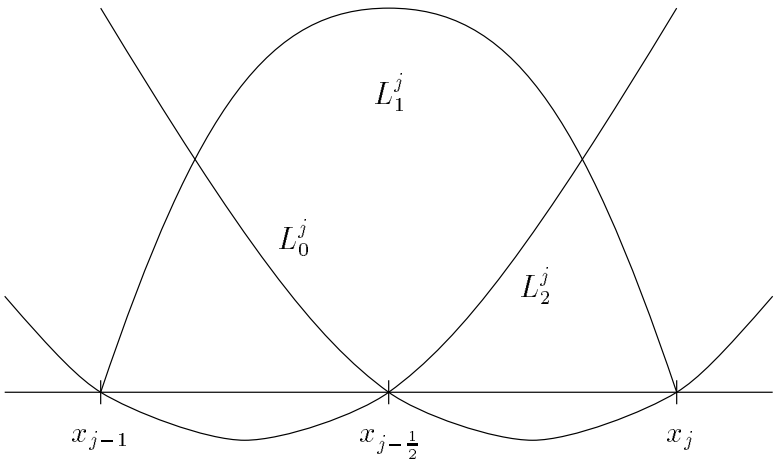


Abbildung 3.2: Lagrange–Polynome aus $\mathcal{P}_2$

Hiermit erhält man eine Lösung von (3.59) durch

$$u(x) = y_{j-1} L_0^j(x) + y_{j-\frac{1}{2}} L_1^j(x) + y_j L_2^j(x), \quad x \in I_j, \tag{3.60}$$

für $j = 1, \ldots, n$. Diese Lösung ist eindeutig bestimmt, weil die zugehörige homogene Interpolationsaufgabe

$$v \in \mathcal{P}_2^1(\Delta):\ v(x_j) = 0, \quad j = 0, \frac{1}{2}, 1, \ldots, n - \frac{1}{2}, n,$$

nur die triviale Lösung hat. Das System von Knotenfunktionalen

$$\{\delta_0, \delta_{\frac{1}{2}}, \delta_1, \ldots, \delta_{n-\frac{1}{2}}, \delta_n\}$$

ist daher $\mathcal{P}_2^1(\Delta)$–unisolvent.

Die zugehörige biorthonormale Basis $\{p_0, p_{\frac{1}{2}}, p_1, \ldots, p_{n-\frac{1}{2}}, p_n\}$ von $\mathcal{P}_2^1(\Delta)$ ist in der Abb. 3.3 veranschaulicht,

$$p_k(x) = \begin{cases} L_2^k(x), & x \in I_k, \\ L_0^{k+1}(x), & x \in I_{k+1}, \\ 0 & \text{sonst} \end{cases} \qquad p_{k-\frac{1}{2}}(x) = \begin{cases} L_1^k(x), & x \in I_k, \\ 0 & \text{sonst.} \end{cases}$$

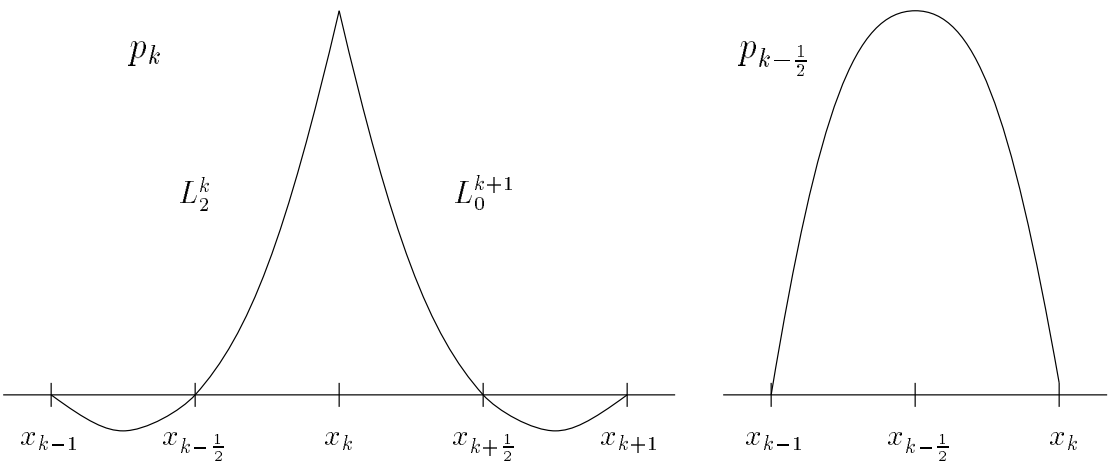


Abbildung 3.3: Basisfunktionen in $\mathcal{P}_2^1(\Delta)$

Anwendung (Quadraturformel)

Sei jetzt ein äquidistantes Gitter gegeben und

$$x_j = x_0 + jh,\ c_j = x_0 + \left(j - \frac{1}{2}\right) h, \quad j \text{ ganz.}$$

Die Quadraturformel (aus kanon. Darstellung von q)

$$\int_a^b u(x)\,dx = \sum_{k=0}^{n} \beta_k u(x_k) + \sum_{k=1}^{n} \beta_{k-\frac{1}{2}} u(x_{k-\frac{1}{2}}), \quad u \in \mathcal{P}_2^1(\Delta), \tag{3.61}$$

zu dieser Interpolationsaufgabe hat dann die Koeffizienten

$$\begin{aligned} \beta_0 = \beta_n &= \frac{h}{6},\ \beta_k = \int_a^b p_k(x)\,dx = \frac{h}{3}, \quad k = 1, \ldots, n-1, \\ \beta_{k-\frac{1}{2}} &= \int_a^b p_{k-\frac{1}{2}}(x)\,dx = \frac{2h}{3}, \quad k = 1, \ldots, n, \end{aligned} \tag{3.62}$$

wie man z. B. mit Hilfe der Simpsonschen Formel leicht sieht. Also ist (3.61) die summierte Simpsonsche Formel.

Beweis von (3.62):

$$\begin{aligned}
\beta_k &= \int_a^b p_k(x)\,dx = \int_{x_{k-1}}^{x_k} + \int_{x_k}^{x_{k+1}} \\
&\overset{\substack{\text{Simpson}\\ \text{exakt}}}{=} \frac{h}{6}\Big[\underbrace{p_k(x_{k-1})}_{0} + 4\underbrace{p_k(x_{k-\frac{1}{2}})}_{0} + \underbrace{p_k(x_k)}_{1}\Big] \\
&\quad + \frac{h}{6}\left[p_k(x_k) + 4p_k(x_{k+\frac{1}{2}}) + p_k(x_{k+1})\right] \\
&= \frac{h}{6} + \frac{h}{6} = \frac{h}{3} \\
\beta_{k-\frac{1}{2}} &= \int_a^b p_{k-\frac{1}{2}}\,dx = \int_{x_{k-1}}^{x_k} p_{k-\frac{1}{2}}\,dx \\
&= \frac{h}{6}[0+4+0] = \frac{2}{3}h, \quad k = 1,\ldots,n.
\end{aligned}$$

□

3.4.2 Quadratische Spline–Funktionen

Der Raum $\mathcal{P}_2^2(\Delta)$ der quadratischen Splines besteht aus stückweise quadratischen Funktionen, die zusammen mit ihren ersten Ableitungen an jedem Gitterpunkt $x_1,\ldots,x_{n-1}$ stetig sind. In Abschnitt 3.1 haben wir gezeigt, daß der Raum $\mathcal{P}_2^2(\Delta)$ die Dimension $n+2$ hat. Daher müssen hier nicht nur die Funktionswerte von u an den $n+1$ Stellen $x_0,\ldots,x_n$, sondern noch eine weitere Bedingung vorgeschrieben werden. Wählt man hier für die Ableitung von u an der Stelle x_0, dann lautet die zugehörige Interpolationsaufgabe:

Gegeben $y_0', y_0, \ldots, y_n$*; gesucht*

$$u \in \mathcal{P}_2^2(\Delta);\ u'(x_0) = y_0', u(x_j) = y_j, \quad j = 0,\ldots,n. \tag{3.63}$$

Bevor die eindeutige Lösbarkeit der Interpolationsaufgabe gezeigt wird, müssen wir noch einige einfache Hilfsmittel bereitstellen. Sei u eine beliebige Funktoine aus $\mathcal{P}_2^2(\Delta)$ und

$$y_j = u(x_j),\ y_j' = \frac{du}{dx}(x_j), \quad j = 0,\ldots,n.$$

Auf jedem Teilintervall $I_j = (x_{j-1}, x_j)$ ist u ein Polynom zweiten Grades, das als Lösung der Hermiteschen Interpolationsaufgabe

$$u(x_{j-1}) = y_{j-1},\ u'(x_{j-1}) = y_{j-1}',\ u(x_j) = y_j$$

aufgefaßt werden kann. Dementsprechend hat u auf I_j die Darstellung

$$u(x) = y_j z + y_{j-1}(1-z) + z(1-z)h_j(y'_{j-1} - \frac{1}{h_j}(y_j - y_{j-1})), \quad x \in I_j, \tag{3.64}$$

mit den Abkürzungen

$$z = \frac{x - x_{j-1}}{h_j},\ h_j = x_j - x_{j-1}, \quad j = 1, \ldots, n.$$

Die Ableitung von u auf I_j gestattet folglich die Darstellung

$$u'(x) = \frac{1}{h_j}(y_j - y_{j-1}) + (1-2z)(y'_{j-1} - \frac{1}{h_j}(y_j - y_{j-1})), \quad x \in I_j. \tag{3.65}$$

Setzt man hier $x = x_j$, so folgen unmittelbar die *charakteristischen Relationen*

$$y'_j + y'_{j-1} = \frac{2}{h_j}(y_j - y_{j-1}), \quad j = 1, \ldots, n, \tag{3.66}$$

zwischen den Werten von u und u' an den Gitterpunkten x_{j-1}, x_j.

Nach diesen Vorbereitungen können wir jetzt den folgenden Satz beweisen:

Satz 3.7 *Die Interpolationsaufgabe (3.63) für quadratische Spline-Funktionen ist eindeutig lösbar für jede rechte Seite $y'_0, y_0, \ldots, y_n$. Die Lösung u hat im Intervall I_j die Darstellung (3.64) mit den Lösungen y'_j der rekursiven Beziehung (3.66), wobei der Anfangswert y'_0 in (3.63) gegeben ist.*

Beweis: (Existenz) Die durch (3.64) definierte Funktion u ist auf jedem Teilintervall I_j ein Polynom zweiten Grades, das in den Endpunkten x_{j-1}, x_j die Werte y_{j-1}, y_j annimmt. Die Ableitung u' von u in I_j hat die Form (3.65). Auf Grund von (3.66) ist daher

$$u'(x_{j-1}) = y'_{j-1},\ u'(x_j) \overset{(3.66)}{=} y'_j, j = 1, \ldots, n.$$

Also ist $u \in \mathcal{P}_2(\Delta)$, die Funktion u ist zusammen mit ihren ersten Ableitungen stetig an den Stellen $x_1, \ldots, x_{n-1}$ und Lösung der Interpolationsaufgabe (3.63).

(Eindeutigkeit) Sei $v \in \mathcal{P}_2^2(\Delta)$ eine Lösung der homogenen Interpolationsaufgabe

$$v'(x_0) = 0,\ v(x_j) = 0, \quad j = 0, \ldots, n.$$

Das System $\{\delta_0', \delta_0, \delta_1\}$ beschreibt eine eindeutig lösbare Hermite–Interpolationsaufgabe auf I_1. Folglich ist $v|I_1 = 0$. Da die Ableitung v' stetig ist bei x_1, folgt $v'(x_1) = 0$. Durch fortgesetzte Anwendung der eben beschriebenen Schlußweise erhält man dann

$$0 = v|I_1 = v|I_2 = \ldots = v|I_n,$$

so daß $v = 0$ ist. □

Die Ableitung u' der Lösung u der Interpolationsaufgabe hat auf Grund von (3.65), (3.66) im Intervall I_j die Form

$$u'(x) = y_j' z + y_{j-1}'(1 - z). \tag{3.67}$$

Also wird

$$u'(x_{j-1}) = y_{j-1}',\ u'(x_j) = y_j', \quad j = 1, \ldots, n.$$

Beweis von (3.67):

$$\begin{aligned} u'(x) &\overset{(3.65)}{=} (1-2z)y_{j-1}' + 2z\frac{1}{h_j}(y_j - y_{j-1}) \\ &\overset{(3.66)}{=} (1-2z)y_{j-1}' + z(y_j' + y_{j-1}') \\ &= z y_j' + (1-z)y_{j-1}', \quad x \in I_j. \end{aligned}$$

□

Somit ist u' diejenige lineare Spline–Funktion, welche an den Stellen x_j die Werte y_j' annimmt.

3.4.3 Weitere Interpolation und Differenzenquotienten

Die bekannten Differenzenquotienten zur Approximation von Ableitungen erhält man mit Hilfe geeigneter Interpolationspolynome. Diese Polynome definieren stückweise quadratische Funktionen, die einem speziellen Typ von Stetigkeitsbedingungen genügen. Auf diesem Wege gewinnt man eine weitere Klasse von Interpolationsaufgaben in Teilräumen von $\mathcal{P}_2(\Delta)$.

Zu jedem Teilintervall $I_j = (a_j, b_j) = (x_{j-1}, x_j)$ sei wie in Abschnitt 3.4.1 ein weiterer Punkt $c_j \in I_j$ gegeben. Zu jeder Funktion $u \in \mathcal{P}_2(\Delta)$ und jedem Element $I_j \in \Delta$

bezeichnen wir mit u^j das eindeutig bestimmte Polynom aus $\mathcal{P}_2$, $u|I_j = u^j|I_j$. Hiermit wird durch

$$\begin{aligned} \ell_{j0}(u) &= u^j(c_{j-1}) - u(c_{j-1}), \quad j = 2, \ldots, n, \\ \ell_{j1}(u) &= u(c_{j+1}) - u^j(c_{j+1}), \quad j = 1, \ldots, n-1, \end{aligned} \tag{3.68}$$

ein System von $2n-2$ Linearformen auf $\mathcal{P}_2(\Delta)$ erklärt. Offenbar ist

$$\ell_{j0}(u) = \ell_{j1}(u) = 0 \tag{3.69}$$

eine Stetigkeitsbedingung, welche die Fortsetzung u^j von $u|I_j$ in die beiden Nachbar–Elemente I_{j-1}, I_{j+1} mit der Funktion u an den Stellen c_{j-1}, c_{j+1} zusammenheftet. Durch

$$\{u \in \mathcal{P}_2(\Delta),\ \ell_{11}(u) = \ell_{n0}(u) = 0,\ \ell_{j0}(u) = \ell_{j1}(u), \quad j = 2, \ldots, n-1\} \tag{3.70}$$

wird dann ein linearer Teilraum von $\mathcal{P}_2(\Delta)$ erklärt, der mit Hilfe von $2n-2$ Stetigkeitsbedingungen aus $\mathcal{P}_2(\Delta)$ erzeugt wird. Dementsprechend hat dieser Teilraum mindestens die Dimension $3n - (2n-2) = n+2$. Wir führen noch zwei weitere Gitterpunkte $c_0 < a_1, c_{n+1} > b_n$ ein und definieren

$$u(c_0) = u^1(c_0),\ u(c_{n+1}) = u^n(c_{n+1}), \tag{3.71}$$

das heißt, u wird mit Hilfe von u^1 und u^n extrapoliert in die Punkte c_0, c_{n+1}. Diese Bedingung kann man auch in der Form

$$\ell_{10}(u) = 0,\ \ell_{n1}(u) = 0, \tag{3.72}$$

schreiben. Hiermit setzen wir jetzt

$$M := \{u \in \mathcal{P}_2(\Delta) | \ell_{j0}(u) = \ell_{j1}(u) = 0, \quad j = 1, \ldots, n\}. \tag{3.73}$$

Die zugehörige Interpolationsaufgabe lautet

$$u \in M:\ u(c_j) = y_j, \quad j = 0, \ldots, n+1. \tag{3.74}$$

Satz 3.8 *Die Interpolationsaufgabe (3.74) ist für jede rechte Seite $y_0, \ldots, y_{n+1}$ eindeutig lösbar.*

Beweis: (Existenz) Zu jedem Intervall I_j seien L^j_{-1}, L^j_0, L^j_1 aus $\mathcal{P}_2$ die Lagrange–Polynome zu den Stützstellen c_{j-1}, c_j, c_{j+1} für $j = 1, \ldots, n$. Damit wird durch

$$u^j(x) = y_{j-1}L^j_{-1}(x) + y_jL^j_0(x) + y_{j+1}L^j_1(x),\ x \in \mathbb{R}, \tag{3.75}$$

ein Polynom u^j zweiten Grades definiert, das an den Stellen c_{j-1}, c_j, c_{j+1} die Werte y_{j-1}, y_j, y_{j+1} annimmt. Die zugehörige Funktion

$$u(x) = u^j(x),\ x \in I_j, \quad j = 1, \ldots, n,$$

genügt dann den Bedingungen $u(c_j) = y_j$ für $j = 0, \ldots, n+1$. Weiter ist

$$\ell_{j0}(u) = y_{j-1} - y_{j-1} = 0,\ \ell_{j1}(u) = y_{j+1} - y_{j+1} = 0,$$

also $u \in M$ und damit u eine Lösung der obigen Interpolationsaufgabe.

(Eindeutigkeit) Betrachten wir das zugehörige homogene Interpolationsproblem

$$v \in M :\ v(c_j) = 0, \quad j = 0, \ldots, n+1.$$

Sei I_j ein beliebiges Teilintervall und $v^j \in \mathcal{P}_2$ die Fortsetzung von $v|I_j$. Da v aus M ist, hat man $v^j(c_{j-1}) = 0,\ v^j(c_{j+1}) = 0$. Mit $v(c_j) = v^j(c_j)$ hat v_j Nullstellen bei c_{j-1}, c_j, c_{j+1}, so daß $v^j = 0$ sein muß. Also ist $v = 0$ und die Aufgabe (3.74) eindeutig lösbar.

□

Die Knotenfunktionale q_j in der obigen Interpolationsaufgabe sind die δ–Funktionale zu den Punkten c_j,

$$\delta_j(u) = u(c_j), \quad j = 0, \ldots, n+1. \tag{3.76}$$

Dieses System $\{\delta_0, \ldots, \delta_{n+1}\}$ ist also M–unisolvent und es existiert eine dazu biorthonormale Basis $\{p_0, \ldots, p_{n+1}\}$ von M. Die Basisfunktion $p_k,\ k = 2, \ldots n-1$, ist gegeben durch

$$p_k = \begin{cases} L^k_0 & \text{in } I_k, \\ L^{k-1}_1 & \text{in } I_{k-1}, \\ L^{k+1}_{-1} & \text{in } I_{k+1}; \end{cases}$$

für $k = 0, 1$, sowie $k = n, n+1$ hat man entsprechende Definitionen. Die Abb. 3.4 zeigt die typische Form dieser Basis–Funktionen.

Der Einfachheit halber betrachten wir jetzt ein äquidistantes Gitter und c_j als Mittelpunkt:

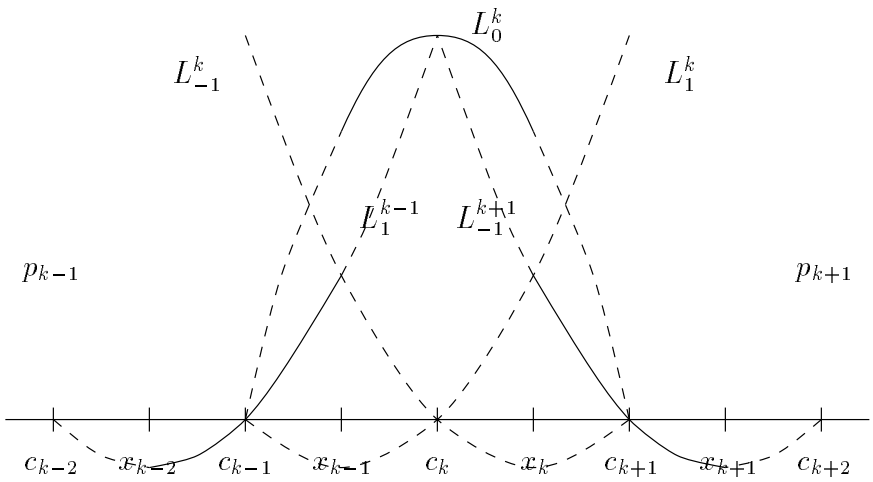


Abbildung 3.4: Basisfunktionen p_k für $k = 2, \ldots, n-1$

$$x_j = x_0 + jh,\ c_j = x_0 + (j - \frac{1}{2})h, \quad j \text{ ganz.}$$

Die Ableitungen der stückweise quadratischen Funktionen $u \in M$ lassen sich dann an den Gitterpunkten c_j darstellen durch

$$\begin{aligned} \frac{du}{dx}(c_j) &= \frac{1}{2h}(u(c_{j+1}) - u(c_{j-1})), \\ \frac{d^2u}{dx^2}(c_j) &= \frac{1}{h^2}(u(c_{j+1}) + u(c_{j-1}) - 2u(c_j)), \quad j = 1, \ldots, n. \end{aligned} \tag{3.77}$$

Die Funktionswerte von u kann man somit einfach aus der Taylorformel

$$u(x) = u(c_j) + (x - c_j)\frac{du}{dx}(c_j) + \frac{1}{2}(x - c_j)^2\frac{d^2u}{dx^2}(c_j), \quad x \in I_j, \tag{3.78}$$

berechnen. An den Gitterpunkten x_j sind die Funktionen $u \in M$ im allgemeinen unstetig.

Behauptung 3.9 *Die erste Ableitung von $u \in M$ ist stetig bei x_j $\forall j$.*

Beweis: Es ist allgemein

$$\frac{du}{dx}(c) = \frac{1}{h}\left(u\left(c + \frac{h}{2}\right) - u\left(c - \frac{h}{2}\right)\right) + \mathrm{O}(h^2 M_3)$$

mit $M_3 := \max |u^{(3)}|$, und daher ist für $u \in \mathcal{P}_2$ der zentrale Differenzenquotient exakt.

$$\Rightarrow \frac{du}{dx}(x_j + 0) = \frac{du^{j+1}}{dx}(x_j) = \frac{1}{h}(\underbrace{u^{j+1}(c_{j+1})}_{u(c_{j+1})} - \underbrace{u^{j+1}(c_j)}_{\overset{(3.69)}{=} u(c_j)} = \frac{1}{h}(u(c_{j+1}) - u(c_j))$$

und

$$\frac{du}{dx}(x_j - 0) = \frac{du^j}{dx}(x_j) = \frac{1}{h}(\underbrace{u^j(c_{j+1})}_{\overset{(3.69)}{=} u(c_{j+1})} - \underbrace{u^j(c_j)}_{u(c_j)}) = \frac{1}{h}(u(c_{j+1}) - u(c_j))$$

□

Die zweite Ableitung von $u \in M$ ist stückweise konstant auf den Intervallen I_j. Hiermit wird

$$\begin{aligned} u''|I_j &= \frac{d^2u}{dx^2}(x_j - 0) &&= \frac{1}{(h/2)^2}(u^j(c_j) - 2u^j(x_j) + u^j(c_{j+1})) \\ &&&= \frac{4}{h^2}(u(c_{j+1}) + u(c_j) - 2u(x_j - 0)) \\ u''|I_{j+1} &= \frac{d^2u}{dx^2}(x_j + 0) &&= \frac{1}{(h/2)^2}(u^{j+1}(c_j) - 2u^{j+1}(x_j) + u^{j+1}(c_{j+1})) \\ &&&= \frac{4}{h^2}(u(c_{j+1}) + u(c_j) - 2u(x_j + 0)). \end{aligned}$$

Aus diesen Beziehungen ergibt sich für die halbe Sprunghöhe von u an der Stelle x_j die Darstellung

$$\frac{1}{2}(u(x_j + 0) - u(x_j - 0)) = \frac{h^2}{16}(u''|I_j - u''|I_{j+1}) \tag{3.79}$$

für $j = 1, \ldots, n-1$. Dieser Ausdruck gibt bis auf das Vorzeichen den Fehler an, mit dem $u(x_j + 0)$ und $u(x_j - 0)$ vom Mittelwert dieser beiden Funktionswerte abweichen.

Kapitel 4

Finite Elemente in mehreren Veränderlichen

Analog zu den klassischen Interpolationsaufgaben für Polynome in einer Veränderlichen sollen jetzt entsprechende Interpolationsprobleme für Polynome mehrerer Veränderlicher studiert werden. Diese Aufgabenstellung ist wesentlich komplexer als die entsprechende Aufgabe für Polynome einer Veränderlichen. Zum Beispiel genügt es in der Lagrange–Interpolation bei einer Veränderlichen, daß die Stützstellen paarweise verschieden sind. Im Falle mehrerer Veränderlicher benötigt man dagegen Systeme affin unabhängiger Punkte und damit definierte Gitter. Weiter ergibt sich eine Differenzierung im Falle mehrerer Veränderlicher durch die geometrische Struktur von Definitionsgebieten und Stützstellenmengen in Form von Dreiecken und Simplexen, Rechtecken oder Parallelepipeden usw. Dementsprechend arbeitet man mit den Polynomräumen $\mathcal{P}_r$, $\mathcal{Q}_r$ oder $\mathcal{Q}_{r_1,\ldots,r_n}$ oder noch allgemeineren Räumen, auf die wir hier nicht eingehen können. Das Ziel dieses Paragraphen ist es, geeignete Systeme von Knotenfunktionalen anzugeben und die M–Unisolvenz, also die eindeutige Lösbarkeit der zugehörigen Interpolationsaufgaben zu zeigen. Die ersten Anwendungen der Interpolation durch Polynome erhält man in der interpolatorischen Approximation von Funktionen mehrerer Veränderlicher und durch Quadraturformeln.

4.1 Räume von Polynomen in mehreren Veränderlichen

Polynome in n reellen Veränderlichen sind Funktionen auf dem n–dimensionalen Zahlenraum $\mathbb{R}^n$ der Form

$$p(x) = \sum_{s_1,\ldots,s_n} \alpha_{s_1\ldots s_n} x_1^{s_1} \ldots x_n^{s_n} \ , \quad x = (x_1, \ldots, x_n) \in \mathbb{R}^n \ ,$$

mit höchstens endlich vielen von Null verschiedenen Koeffizienten $\alpha_{s_1\ldots s_n}$ aus dem Körper $\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$. Es bezeichnet $\mathcal{P}(\mathbb{R}^n)$ oder kürzer $\mathcal{P}^n$ den Vektorraum aller Polynome in n reellen Veränderlichen. Zur Vereinfachung der Schreibweise verwendet man Multiindizes $s = (s_1, \ldots, s_n)$. Dieses sind Vektoren des $\mathbb{R}^n$ mit nichtnegativen ganzzahligen Komponenten $s_1, \ldots, s_n$. Die *Ordnung* des Multiindex ist erklärt durch $|s| = s_1 + \cdots + s_n$, also durch die ℓ^1–Norm $|s|_1$ von s aus $\mathbb{R}^n$. Wir bezeichnen mit $D_j = \frac{\partial}{\partial x_j}$ die partiellen Ableitungen nach $x_1, \ldots, x_n$ und setzen

$$D^s = D_1^{s_1} \ldots D_n^{s_n} \ , \ s! = s_1! \ldots s_n! \ ,$$

für jeden Multiindex $s = (s_1, \ldots, s_n)$. Jedem Multiindex ist eindeutig das *Monom*

$$x^s = x_1^{s_1} \cdots x_n^{s_n} \ , \ x = (x_1, \ldots, x_n) \in \mathbb{R}^n \ ,$$

zugeordnet. Dieses ist ein Polynom vom Grade $|s|$. Jedes Polynom ist also eine endliche Linearkombination von Monomen. Jedes endliche System von Monomen ist linear unabhängig. Wenn nämlich

$$0 = \sum_t \alpha_t x^t \ , \qquad x \in \mathbb{R}^n \ ,$$

ist, dann wird insbesondere jede partielle Ableitung dieses Polynoms gleich Null, also

$$0 = D^s \left(\sum_t \alpha_t x^t \right) \Big|_{x=0} = s!\alpha_s + \sum_{t>s} \tilde{\alpha}_t x^{t-s} \Big|_{x=0} = s!\alpha_s \quad \forall s \ .$$

Sei $\mathcal{P}_r = \mathcal{P}_r(\mathbb{R}^n)$ der endlichdimensionale Vektorraum aller Polynome höchstens r–ten Grades,

$$\mathcal{P}_r := \left\{ p \Big| p(x) = \sum_{|s|\le r} \alpha_s x^s \ , \qquad x \in \mathbb{R}^n \right\} \ . \tag{4.1}$$

Polynome oder Funktionen aus $\mathcal{P}_r$ nennt man *linear* für $r = 1$, *quadratisch* für $r = 2$, *kubisch* für $r = 3$, *quartisch* für $r = 4$, *quintisch* für $r = 5$. Für $n = 2$ und $x_1 = x$, $x_2 = y$ ist zum Beispiel:

$$\begin{aligned} \mathcal{P}_1(\mathbb{R}^2) &= [1, x, y] , \quad \mathcal{P}_2(\mathbb{R}^2) = [1, x, y, x^2, xy, y^2] , \\ \mathcal{P}_3(\mathbb{R}^2) &= [1, x, y, x^2, xy, y^2, x^3, x^2y, xy^2, y^3], \dots . \end{aligned}$$

Diese Räume haben die Dimensionen

$$\dim \mathcal{P}_1(\mathbb{R}^2) = 3 , \quad \dim \mathcal{P}_2(\mathbb{R}^2) = 6 , \quad \dim \mathcal{P}_3(\mathbb{R}^2) = 10, \dots ,$$

und allgemein

$$\dim \mathcal{P}_r(\mathbb{R}^2) = \frac{1}{2}(r+1)(r+2) . \tag{4.2}$$

Induktiv sieht man nämlich, daß beim Übergang von $\mathcal{P}_{r-1}$ zu $\mathcal{P}_r$ genau $r+1$ Monome der Ordnung r hinzukommen, nämlich x^r, $x^{r-1}y$, $x^{r-2}y^2, \dots, xy^{r-1}$, y^r; also erhält man für die Dimension von $\mathcal{P}_r$,

$$\dim \mathcal{P}_r = \sum_{\nu=1}^{r+1} \nu = \frac{1}{2}(r+1)(r+2) .$$

Der Teilraum der Polynome r–ten Grades im Raum aller Polynome ist offenbar durch die Eigenschaft charakterisiert

$$p \in \mathcal{P}_r(\mathbb{R}^n) \Longleftrightarrow D^s p = 0 \quad \forall |s| = r+1 . \tag{4.3}$$

Eine weitere Klasse $\mathcal{Q}_r(\mathbb{R}^n)$ von Polynomen in n reellen Veränderlichen erhält man durch Linearkombinationen aller Monome vom Grade

$$|s|_\infty := \max_{k=1,\dots,n} s_k \le r .$$

Diese Polynomräume sind also erklärt durch

$$\mathcal{Q}_r := \left\{ p \,\middle|\, p(x) = \sum_{|s|_\infty \le r} \alpha_s x^s , \qquad x \in \mathbb{R}^n \right\} . \tag{4.4}$$

Für $n = 2$ und $x = x_1$, $y = x_2$ ist zum Beispiel

$$\begin{aligned} \mathcal{Q}_1(\mathbb{R}^2) &= [1, x, y, xy] , \\ \mathcal{Q}_2(\mathbb{R}^2) &= [1, x, y, x^2, xy, y^2, x^2y, xy^2, x^2y^2], \dots \end{aligned}$$

mit den Dimensionen

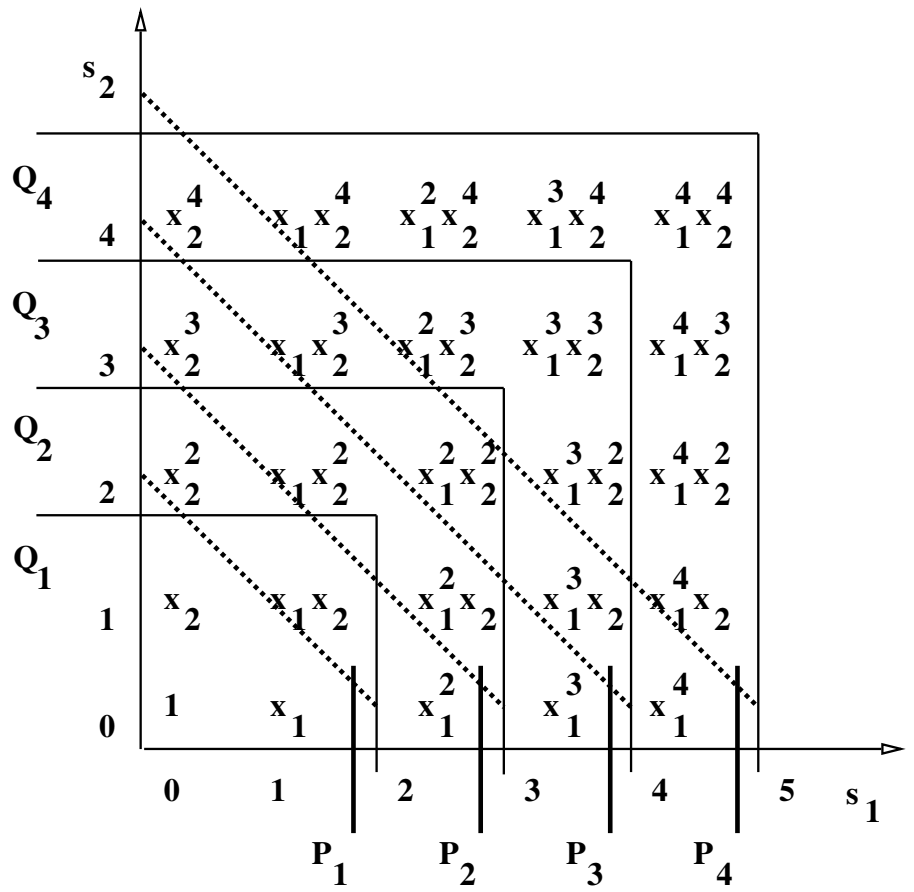


Abbildung 4.1: Polynomräume $\mathcal{P}_r(\mathbb{R}^2)$ und $\mathcal{Q}_r(\mathbb{R}^2)$

$$\dim \mathcal{Q}_1(\mathbb{R}^2) = 4 \ , \quad \dim \mathcal{Q}_2(\mathbb{R}^2) = 9 \ , \quad \dim \mathcal{Q}_3(\mathbb{R}^2) = 16, \ldots$$

und allgemein

$$\dim \mathcal{Q}_r(\mathbb{R}^2) = (r+1)^2 \ , \tag{4.5}$$

wie man durch vollständige Induktion über r sieht.

Die Polynome oder Funktionen in den Räumen $\mathcal{Q}_r(\mathbb{R}^2)$ nennt man *bilinear* für $r = 1$, *biquadratisch* für $r = 2$, *bikubisch* für $r = 3$ und so weiter. Die Abbildung 4.1 veranschaulicht diese Räume, die erzeugenden Monome und die zugehörigen Mengen von Multiindizes.

Schließlich kann man etwas allgemeiner zu jedem Multiindex $(r_1, \ldots, r_n)$ die zugehörige Indexmenge

$$0 \le s \le (r_1, \ldots, r_n) \Longleftrightarrow 0 \le s_j \le r_j \ , \qquad j = 1, \ldots, n \ ,$$

und damit den zugehörigen Polynomraum

$$\mathcal{Q}_{r_1,\ldots,r_n}(\mathbb{R}^n) = \left\{ p \,\middle|\, p(x) = \sum_{0 \le s \le (r_1,\ldots,r_n)} \alpha_s x^s \ , \qquad x \in \mathbb{R}^n \right\} \ , \tag{4.6}$$

betrachten. Zum Beispiel ist dann

$$\begin{aligned}
\mathcal{Q}_{r,0}(\mathbb{R}^2) &= [1, x, \ldots, x^r] \ , \\
\mathcal{Q}_{r,1}(\mathbb{R}^2) &= [1, y, x, xy, x^2, \ldots, x^r, x^r y] \ , \\
\mathcal{Q}_{r,2}(\mathbb{R}^2) &= [1, y, y^2, x, xy, xy^2, \ldots, x^r, x^r y, x^r y^2] \ , \ldots
\end{aligned}$$

mit den Dimensionen

$$\dim \mathcal{Q}_{r_1,\ldots,r_n}(\mathbb{R}^n) = (r_1 + 1) \cdots (r_n + 1) \ . \tag{4.7}$$

Bei der Herleitung von Fehlerabschätzungen benötigt man die Tatsache, daß sich Polynome aus diesen Räumen durch die Eigenschaft

$$p \in \mathcal{Q}_{r_1,\ldots,r_n} \Longleftrightarrow D_1^{r_1+1} p(x) = 0, \ldots, D_n^{r_n+1} p(x) = 0 \ , \qquad x \in \mathbb{R}^n \ , \tag{4.8}$$

charakterisieren lassen.

4.2 Lagrange–Interpolation in $\mathcal{P}_1(\mathbb{R}^n)$

Wir beginnen unsere Untersuchungen mit der Lagrangeschen Interpolationsaufgabe im Raum der Polynome ersten Grades in n reellen Veränderlichen

$$\mathcal{P}_1(\mathbb{R}^n) = [1, x_1, \ldots, x_n] \ .$$

Dieses ist ein $(n+1)$–dimensionaler Raum, so daß $n+1$ Knotenfunktionale für die eindeutige Lösbarkeit der Interpolationsaufgabe notwendig sind. Dementsprechend betrachtet man die Aufgabe [1)]

$$u \in \mathcal{P}_1(\mathbb{R}^n) \ ; \quad u(a_k) = y_k \ , \qquad k = 0, \ldots, n \ , \tag{4.9}$$

für beliebige rechte Seiten $(y_0, \ldots, y_n) \in \mathbb{K}^{n+1}$ und mit vorgegebenen Punkten $a_0, \ldots, a_n \in \mathbb{R}^n$. Die Knotenfunktionale q_k sind die δ–Funktionale δ_k zu den Stützstellen $a_0, \ldots, a_n$, $a_j = (a_{1j}, \ldots, a_{nj})^\top$, $j = 0, \ldots, n$,

$$\delta_k(v) = v(a_k) \ , \qquad v \in \mathcal{P}(\mathbb{R}^n) \ , \ k = 0, \ldots, n \ . \tag{4.10}$$

Die Funktionen

$$u_0(x) = 1 \ , \quad u_j(x) = x_j \ , \qquad j = 0, \ldots, n \ , \ x = (x_1, \ldots, x_n) \in \mathbb{R}^n \ , \tag{4.11}$$

bilden eine Basis von $\mathcal{P}_1(\mathbb{R}^n)$. Wie wir in Abschnitt 2.2 gezeigt haben, ist die Interpolationsaufgabe genau dann eindeutig lösbar für jede rechte Seite, wenn die zugehörige Gramsche Matrix regulär ist, das heißt

$$\det(u_j(a_k)) = \begin{vmatrix} 1 & 1 & \ldots & 1 \\ a_{10} & a_{11} & \ldots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n0} & a_{n1} & \ldots & a_{nn} \end{vmatrix} \neq 0 \tag{4.12}$$

ist. Diese Bedingung ist äquivalent dazu, daß das homogene Gleichungssystem

$$\sum_{k=0}^{n} \alpha_k u_j(a_k) = 0 \ , \qquad j = 0, \ldots, n \ ,$$

[1)] Wir schreiben gelegentlich auch $\underline{a_k}, \underline{x}$ — insbesondere in den Abbildungen — um zu verdeutlichen, daß Vektoren im $\mathbb{R}^n$ vorliegen.

nur die triviale Lösung besitzt. Diese Forderung hat offenbar die Form

$$\sum_{k=0}^{n} \alpha_k = 0 \, , \; \sum_{k=0}^{n} \alpha_k a_k = 0 \Longrightarrow \alpha_0 = \cdots = \alpha_n = 0 \, . \tag{4.13}$$

Durch diese Relation ist die *affine Unabhängigkeit* der Punkte $a_0, \ldots, a_n$ in $\mathbb{R}^n$ definiert. Setzt man

$$b_k := a_k - a_0 \, , \qquad k = 1, \ldots, n \, ,$$

für die *Differenzvektoren*, dann wird (4.13) äquivalent zu

$$\sum_{k=1}^{n} \alpha_k b_k = 0 \Longrightarrow \alpha_1 = \cdots = \alpha_n = 0 \, , \tag{4.14}$$

also zur linearen Unabhängigkeit der Differenzvektoren $b_1, \ldots, b_n$ in $\mathbb{R}^n$. Das Ergebnis dieser Überlegungen fassen wir im folgenden Satz zusammen:

Satz 4.1 *Die Lagrangesche Interpolationsaufgabe in $\mathcal{P}_1(\mathbb{R}^n)$ ist genau dann eindeutig lösbar, wenn die Stützstellen $a_0, \ldots, a_n$ affin unabhängig oder die Differenzvektoren $b_1, \ldots, b_n$ linear unabhängig sind.*

Die Interpolationsaufgabe in $\mathcal{P}_1(\mathbb{R}^2)$ ist genau dann eindeutig lösbar, wenn die drei Punkte a_0, a_1, a_2 in $\mathbb{R}^2$ die Eckpunkte eines nichtausgearteten Dreiecks bilden, also die drei Eckpunkte nicht auf einer Geraden liegen. Allgemeiner sind $n+1$ Punkte $a_0, \ldots, a_n$ in $\mathbb{R}^n$ genau dann affin unabhängig, wenn sie ein nichtausgeartetes n–Simplex aufspannen. Dieses n–Simplex ist die konvexe Hülle von affin unabhängigen $a_0, \ldots, a_n$ und diese Punkte sind die Eckpunkte des Simplex.

Wir setzen im folgenden die eindeutige Lösbarkeit der Interpolationsaufgabe (4.9) voraus und wollen jetzt die zu den Knotenfunktionalen $\delta_0, \ldots, \delta_n$ biorthonormale Basis von Polynomen $p_0, \ldots, p_n \in \mathcal{P}_1(\mathbb{R}^n)$ bestimmen. Die Differenzvektoren $b_1 = a_1 - a_0, \ldots, b_n = a_n - a_0$ sind linear unabhängig und bilden somit eine Basis des Zahlenraums $\mathbb{R}^n$. Für jeden Punkt $x \in \mathbb{R}^n$ gibt es daher eindeutig bestimmte Koeffizienten $z_1, \ldots, z_n \in \mathbb{R}$ derart, daß die Darstellung

$$x = a_0 + \sum_{k=1}^{n} z_k b_k \Longleftrightarrow x_j = a_{j0} + \sum_{k=1}^{n} b_{jk} z_k \, , \qquad j = 1, \ldots, n \, , \tag{4.15}$$

gilt. Offenbar sind $z_1, \ldots, z_n$ die Koordinaten von x bezüglich des durch $b_1, \ldots, b_n$ definierten *schiefwinkligen Koordinatensystems* mit dem Ursprung a_0 (s. Abb. 4.2).

Ersetzt man hier b_k durch $a_k - a_0$ und definiert

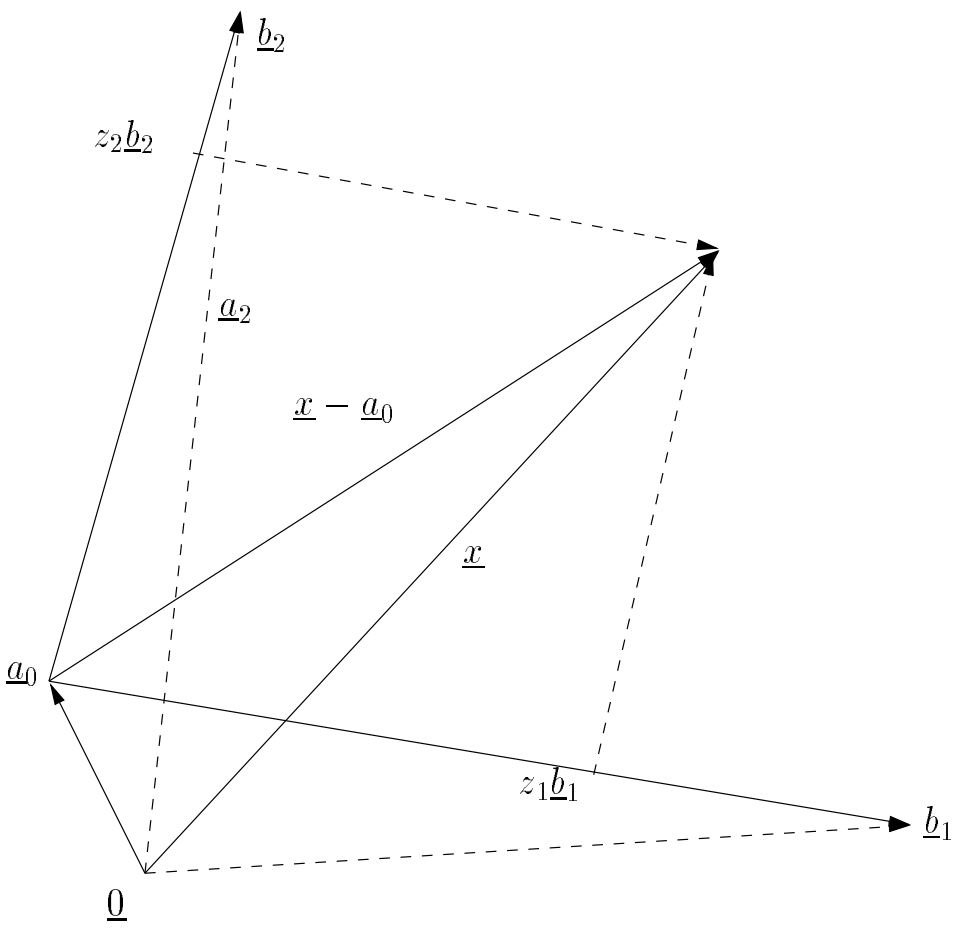


Abbildung 4.2: Schiefwinkliges Koordinatensystem in $\mathbb{R}^2$

$$z_0 := 1 - \sum_{k=1}^{n} z_k \ ,$$

dann wird (4.15) äquivalent zu

$$x = \sum_{k=0}^{n} z_k a_k \ , \quad \sum_{k=0}^{n} z_k = 1 \ . \tag{4.16}$$

Auf Grund der affinen Unabhängigkeit der Punkte $a_0, \ldots, a_n$ sind $z_0, \ldots, z_n$ eindeutig durch x bestimmt. Man bezeichnet $z_0, \ldots, z_n$ als *baryzentrische Koordinaten* von x.

Die $n \times n$ Matrix B sei definiert durch

$$B := (b_1, \ldots, b_n) = (b_{jk})_{j,k=1,\ldots,n}. \tag{4.17}$$

Nach Voraussetzung über die Vektoren $b_1, \ldots, b_n$ ist B regulär. Durch die Beziehung (4.15) wird also eine umkehrbar eindeutige affine Abbildung A des $\mathbb{R}^n$ auf sich definiert,

$$x = Az \Longleftrightarrow x = a_0 + Bz \ , \qquad z = (z_1, \ldots, z_n) \in \mathbb{R}^n \ . \tag{4.18}$$

Die zugehörige inverse Transformation A^{-1} erhält man in der Form

$$z = A^{-1}x \Longleftrightarrow z = B^{-1}(x - a_0) \ , \qquad x = (x_1, \ldots, x_n) \in \mathbb{R}^n \ . \tag{4.19}$$

Aus der Darstellung (4.16) folgt sofort, daß die Stützstellen a_k die baryzentrischen Koordinaten $\delta_{0k}, \ldots, \delta_{nk}$ haben. Unter der affinen Transformation A gehen also die Stützstellen $a_1, \ldots, a_n$ aus den Einheitsvektoren $e_1 = (1, 0, \ldots, 0), \ldots, e_n = (0, \ldots, 0, 1)$ und der Vektor a_0 aus dem Nullvektor $e_0 = 0$ hervor.

$$a_k = Ae_k \ , \ e_k = (\delta_{1k}, \ldots, \delta_{nk}) \in \mathbb{R}^n \ , \qquad k = 0, \ldots, n \ . \tag{4.20}$$

Schließlich sei

$$B^{-1} = \left(b_{jk}^{(-1)}\right)_{j,k=1,\ldots,n} \ , \ -B^{-1}a_0 =: c = (c_j)_{j=1,\ldots,n} \ . \tag{4.21}$$

Dann folgt aus (4.19) die Koordinatendarstellung

$$z = A^{-1}x \Longleftrightarrow z_j = c_j + \sum_{k=1}^{n} b_{jk}^{(-1)} x_k \ , \qquad j = 1, \ldots, n \ . \tag{4.22}$$

Setzt man noch

$$c_0 := 1 - \sum_{j=1}^{n} c_j \ , \ b_{0k}^{(-1)} := -\sum_{j=1}^{n} b_{jk}^{(-1)} \ , \qquad k = 1, \ldots, n \ , \tag{4.23}$$

dann haben die baryzentrischen Koordinaten z_j der Punkte $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ die Darstellung

$$z_j = c_j + \sum_{k=1}^{n} b_{jk}^{(-1)} x_k \ , \qquad j = 0, \ldots, n \ . \tag{4.24}$$

Für $j = 1, \ldots, n$ steht dies schon in (4.22); für $j = 0$ ist nach (4.16)

$$\begin{aligned} z_0 &= 1 - \sum_{j=1}^{n} z_j = 1 - \sum_{j=1}^{n} \left[c_j + \sum_{k=1}^{n} b_{jk}^{(-1)} x_k \right] \\ &= 1 - \sum_{j=1}^{n} c_j + \sum_{k=1}^{n} \left(-\sum_{j=1}^{n} b_{jk}^{(-1)} \right) x_k \\ &= c_0 + \sum_{k=1}^{n} b_{0k}^{(-1)} x_k \ . \end{aligned}$$

Damit gilt der

Satz 4.2 *Die zu $\{\delta_0, \ldots, \delta_n\}$ biorthonormale Basis $\{p_0, \ldots, p_n\}$ in $\mathcal{P}_1(\mathbb{R}^n)$ lautet*

$$p_j(x) = z_j \ , \qquad j = 0, \ldots, n \ , \tag{4.25}$$

wobei $z_0, \ldots, z_n$ die eindeutig bestimmten baryzentrischen Koordinaten von $x \in \mathbb{R}^n$ sind.

Beweis: Aus der Darstellung (4.24) sieht man unmittelbar, daß $p_0, \ldots, p_r$ Polynome ersten Grades in $x_1, \ldots, x_n$ sind. In (4.20) haben wir bereits festgestellt, daß die Stützstellen a_k die baryzentrischen Koordinaten $\delta_{0k}, \ldots, \delta_{nk}$ haben für $k = 0, \ldots, n$. Also ist

$$p_j(a_k) = \delta_{jk} \ , \qquad j, k = 0, \ldots, n \ .$$

□

Im Fall $n = 2$ wird also jede Funktion aus $\mathcal{P}_1(K)$ eindeutig durch ihre Werte an den drei (affin unabhängigen) Dreieckspunkten a_0, a_1, a_2 bestimmt. Hat man speziell die Eckpunkte

$$a_0 = (0,0)\ ,\ \ a_1 = (1,0)\ ,\ \ a_2 = (0,1)\ ,$$

dann ergeben sich die Basisfunktionen gerade als

$$p_0 = 1 - x - y\ ,\ \ p_1 = x\ ,\ \ p_2 = y\ ;$$

in diesem Fall sind die baryzentrischen Koordinaten z_1, z_2 gerade die kartesischen Koordinaten.

Später werden wir *Triangulierungen* $\Delta_h = \{K_\iota\}$ von Gebieten G betrachten, d. h. eine Aufteilung von G in Dreiecke mit den Eigenschaften [2)]

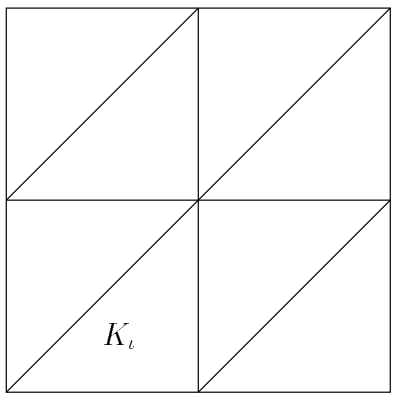


Abbildung 4.3: Einfache gleichförmige Triangulierung

$$\bigcup_\iota K_\iota = G\ ,\ \ \overset{\circ}{K}_\iota \cap \overset{\circ}{K}_\nu = \emptyset\ ,\qquad \iota \neq \nu\ .$$

Der Parameter h kennzeichnet dabei z. B. den maximalen Durchmesser der Dreiecke. Im Raum ($G \subset \mathbb{R}^2$)

$$V_h = \{v \in C^0(\overline{G}) |\, v|K \in \mathcal{P}_1(K)\ \forall K \in \Delta_h\}$$

der stetigen, stückweise linearen Funktionen wird dann jede Funktion eindeutig durch ihre Werte an den Eckpunkten der Dreiecke bestimmt. Die Stetigkeit der Funktionen aus V_h ist offenbar schon dadurch gesichert, daß die Funktionen an den Gitterpunkten der Triangulation stetig sind. Stimmen nämlich zwei Funktionen aus V_h an den zwei Eckpunkten einer Dreiecksseite überein, dann müssen sie wegen ihrer Linearität auf der ganzen Seite übereinstimmen.

Jedes Polynom ersten Grades in n Veränderlichen hat also die kanonische Darstellung

[2)] $\overset{\circ}{K}_\iota$ bezeichnet das *Innere* von K_ι

$$u(x) = \sum_{k=0}^{n} u(a_k) p_k(x) , \qquad x \in \mathbb{R}^n , \; u \in \mathcal{P}_1(\mathbb{R}^n) . \tag{4.26}$$

Die zugehörige Darstellung von Linearformen q auf $\mathcal{P}_1(\mathbb{R}^n)$ lautet

$$q(u) = \sum_{k=0}^{n} q(p_k) u(a_k) , \qquad u \in \mathcal{P}_1(\mathbb{R}^n) .$$

Zum Beispiel erhält man hiermit die *Quadraturformel*

$$\int_G u(x)\, dx = |G| \sum_{k=0}^{n} \alpha_k u(a_k) , \qquad u \in \mathcal{P}_1(\mathbb{R}^n) , \tag{4.27}$$

mit den Koeffizienten

$$\alpha_k = \frac{1}{|G|} \int_G p_k(x)\, dx , \qquad k = 0, \ldots, n , \tag{4.28}$$

wobei $|G|$ das *Volumen* des Gebiets G bezeichnet.

Im Spezialfall, daß G die konvexe Hülle K der Punkte $a_0, \ldots, a_n$, also das von diesen Punkten aufgespannte Simplex des $\mathbb{R}^n$ darstellt, untersuchen wir die Quadraturformel (4.27), (4.28) etwas genauer. Dieses Simplex kann man in der Form

$$K = \operatorname{conv}\{a_0, \ldots, a_n\} = \left\{ x \in \mathbb{R}^n | x = a_0 + \sum_{k=1}^{n} z_k b_k, z_k \geq 0 , \; \sum_{k=1}^{n} z_k \leq 1 \right\}$$

darstellen. K ist offenbar das Bild des *Einheitssimplex*

$$E = \left\{ z \in \mathbb{R}^n | z_k \geq 0 , \; \sum_{k=1}^{n} z_k \leq 1 \right\}$$

unter der in (4.18) definierten affinen Abbildung A (s. Abb. 4.4). Die *Funktionalmatrix* dieser Abbildung ist

$$A' = \frac{\partial A}{\partial z} = B , \tag{4.29}$$

also konstant, unabhängig von z. Nach bekannten Regeln über Variablensubstitutionen von Integrationsveränderlichen wird dann in (4.28)

$$\alpha_j = \frac{1}{|K|} \int_K p_j(x)\, dx = \frac{|\det B|}{|K|} \int_E z_j\, dz , \qquad j = 0, \ldots, n . \tag{4.30}$$

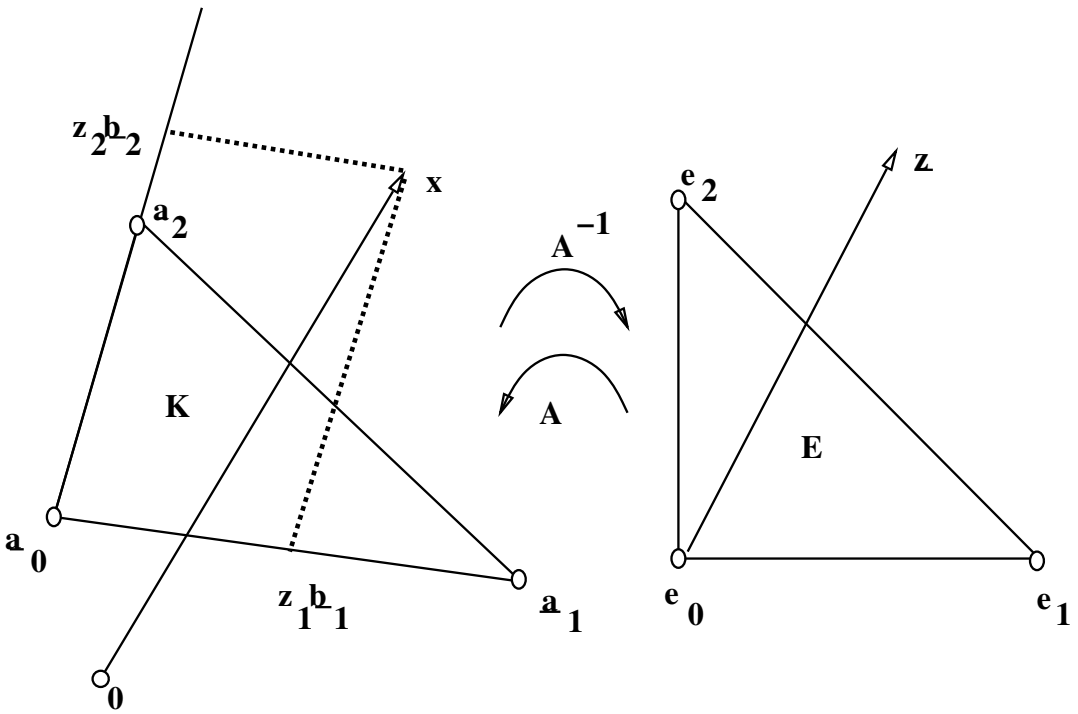


Abbildung 4.4: Lagrange-Interpolation in $\mathcal{P}_1(\mathbb{R}^2)$

Es gelten die Beziehungen

$$\begin{aligned} |E| &= \int_E 1\,dz = \frac{1}{n!}\,, \quad \int_E z_j\,dz = \frac{1}{(n+1)!}\,, \\ &\quad \int_E z_j^2\,dz = \frac{2}{(n+2)!}\,, \quad \int_E z_j z_k\,dz = \frac{1}{(n+2)!}\,, \quad j \neq k\,, \end{aligned} \tag{4.31}$$

für $j,k = 0,\ldots,n$ (s. [Str71, S. 22]). Bekanntlich ist $|\det B|$ das Volumen des Parallelepipeds $\{x : x = a_0 + \sum_1^n t_k b_k\,,\ 0 \le t_k \le 1\,,\ k = 1,\ldots,n\}$, das von den Differenzenvektoren $b_1 = a_1 - a_0, \ldots, b_n = a_n - a_0$ aufgespannt wird und

$$|K| = \int_K 1\,dx = |\det B| \int_E 1\,dz = |\det B| \cdot |E| = \frac{1}{n!}|\det B| \tag{4.32}$$

das Volumen des Simplex K. Damit haben wir schließlich die *Quadraturformel*

$$\int_K u(x)\,dx = \frac{|K|}{n+1} \sum_{j=0}^{n} u(a_j)\,, \qquad u \in \mathcal{P}_1(\mathbb{R}^n)\,, \tag{4.33}$$

d. h. es gilt $\alpha_k = \frac{1}{n+1}$, $k = 0,\ldots,n$, in (4.27) im Fall $G = K$. Benutzt man nämlich (4.31) und (4.32) in (4.30), so erhält man für die Koeffizienten

$$\alpha_j = \frac{|\det B|}{|K|} \int_E z_j\,dz = \frac{n!}{(n+1)!} = \frac{1}{n+1}\,, \qquad j = 0,\ldots,n\,.$$

Die *1–Punkt–Formel*

$$\int_K u(x)\,dx \sim |K|\,u(a_S)$$

mit dem *Schwerpunkt* (oder *Baryzentrum*)

$$a_S = \frac{1}{n+1} \sum_{j=0}^{n} a_j$$

als einziger Stützstelle ist ebenfalls exakt für Polynome aus $\mathcal{P}_1(\mathbb{R}^n)$,

$$\int_K u(x)\,dx = |K|\,u(a_S), \quad u \in \mathcal{P}_1(\mathbb{R}^n). \tag{4.34}$$

Wählt man nämlich $q_S(u) = u(a_S)$ als Linearform, dann erhält man über die kanonische Darstellung

$$q_S(u) = \sum_{k=0}^{n} q_S(p_k) u(a_k) = \frac{1}{n+1} \sum_{k=0}^{n} u(a_k), \quad u \in \mathcal{P}_1(\mathbb{R}^n),$$

da die baryzentrischen Koordinaten des Schwerpunktes gerade alle den Wert $\frac{1}{n+1}$ haben ($q_S(p_k) = z_k(a_S) = \frac{1}{n+1}$, $k = 0,\ldots,n$).

4.3 Lagrange–Interpolation in $\mathcal{P}_r(\mathbb{R}^n)$

Bei einer speziellen Wahl der Stützstellen kann man sehr einfach die eindeutige Lösbarkeit der Lagrangeschen Interpolationsaufgabe für Polynome r-ten Grades in n reellen Veränderlichen zeigen. Dies geschieht durch explizite Angabe der zugehörigen biorthonormalen Basis in $\mathcal{P}_r(\mathbb{R}^n)$.

Gegeben seien $n+1$ affin unabhängige Punkte $\underline{a}_0, \ldots, \underline{a}_n$ in $\mathbb{R}^n$. Zur Verdeutlichung werden wir in den folgenden Abschnitten für Vektoren aus $\mathbb{R}^n$ immer $\underline{x}$ anstelle von x schreiben. Mit Hilfe der zugehörigen linear unabhängigen Differenzenvektoren $\underline{b}_k = \underline{a}_k - \underline{a}_0$ erhält man das System von *Stützstellen*

$$\underline{a}_t = \underline{a}_0 + \frac{t_1}{r}\,\underline{b}_1 + \cdots + \frac{t_n}{r}\,\underline{b}_n \tag{4.35}$$

für jeden Multiindex $t = (t_1, \ldots, t_n)$ mit $|t| \le r$. Hiermit ist die Interpolationsaufgabe:

$$u \in \mathcal{P}_r(\mathbb{R}^n) \;:\; u(\underline{a}_t) = y_t\,,\ |t| \le r\,, \tag{4.36}$$

gestellt für beliebige rechte Seiten $(y_t) \in \mathbb{K}^d$, wobei d die Anzahl der Bedingungen in (4.36), also die Anzahl der Multiindizes t in $|t| \le r$ und damit

$$d = \sum_{|t| \le r} 1 = \dim \mathcal{P}_r(\mathbb{R}^n)$$

ist. Im Fall $n = 2$ hat man

$$d = \sum_{t_1 + t_2 \le r} 1 = \frac{1}{2}(r+1)(r+2)\,.$$

Die zugehörigen Knotenfunktionale werden mit δ_t bezeichnet,

$$\delta_t(v) = v(\underline{a}_t)\,, \qquad v \in \mathcal{P}(\mathbb{R}^n)\,,\ |t| \le r\,. \tag{4.37}$$

Die in Abschnitt 4.2 definierte, zugehörige affine Abbildung A hat die Eigenschaft

$$\underline{a}_t = A\left(\frac{1}{r}\,t\right) \Longleftrightarrow \frac{1}{r}\,t = A^{-1}\underline{a}_t\,, \quad |t| \le r\,. \tag{4.38}$$

Zu jedem Multiindex $s = (s_1, \ldots, s_n)$ mit der Ordnung $|s| = |s|_1 \le r$ setzen wir

$$s_0 := r - |s| \Longleftrightarrow s_0 + s_1 + \cdots + s_n = r\,.$$

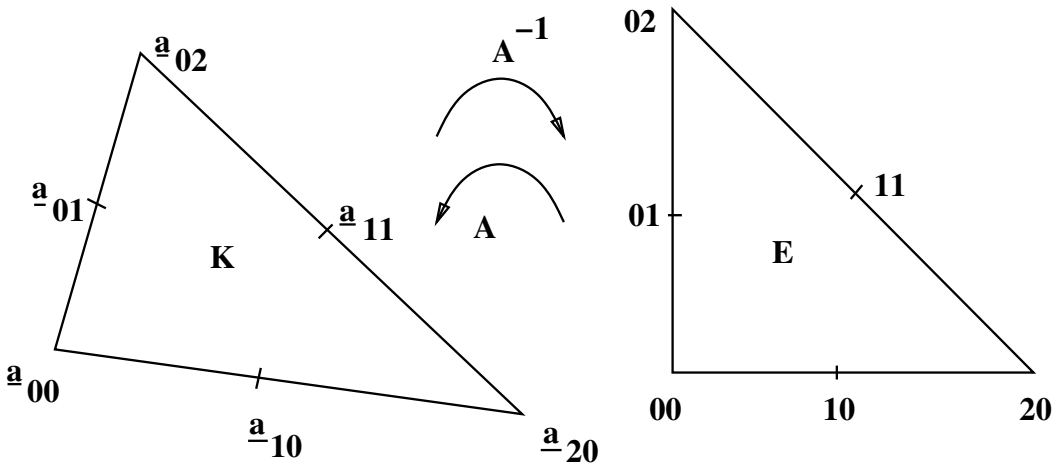


Abbildung 4.5: Stützstellen für Lagrange–Interpolation in $\mathcal{P}_2(\mathbb{R}^2)$

Zu jedem dieser Multiindizes wird dann durch

$$L_s(\underline{z}) = \prod_{i=0}^{n} \prod_{\ell_i=0}^{s_i-1} \frac{rz_i - \ell_i}{s_i - \ell_i} \tag{4.39}$$

mit

$$z_0 = 1 - (z_1 + \cdots + z_n)\,, \quad \underline{z} = (z_1, \ldots, z_n) \in \mathbb{R}^n\,,$$

ein Polynom in den n Veränderlichen $z_1, \ldots, z_n$ erklärt. Definitionsgemäß ist ein Produkt gleich 1, wenn die zugehörige Menge der Faktoren leer ist. Offensichtlich ist $L_s(\underline{z})$ ein Produkt von $s_0 + s_1 + \ldots + s_n = r$ Faktoren. Jeder Faktor ist ein Polynom ersten Grades. Also ist L_s ein Polynom r-ten Grades in $z_1, \ldots, z_n$. Im einfachsten Fall $r = 1$ ist $L_s(\underline{z}) = z_k (= p_k(\underline{x}))$, wobei k den Index kennzeichnet, für den die k-te Komponente von $s = (\delta_{1k}, \ldots, \delta_{nk})$ gerade Eins ist.

Lemma 4.3 *Die Polynome L_s haben die charakteristische Eigenschaft*

$$L_s\left(\frac{1}{r}\,t\right) = \delta_{st} \tag{4.40}$$

für alle Multiindizes s, t mit $|s| \le r$, $|t| \le r$.

Beweis: Aus der Darstellung (4.39) liest man sofort ab, daß $L_s\left(\frac{1}{r}\,s\right) = 1$ ist — da dann $rz_i = s_i\ \forall i$. Sei jetzt t ein von s verschiedener Multiindex der Ordnung $|t| \le r$ und sei $t_0 = r - |t|$. Dann ist

$$\begin{aligned}\sum_{i=0}^{n}(s_i - t_i) &= \underbrace{r - (s_1 + \cdots + s_n)}_{s_0} + \underbrace{s_1 + \cdots + s_n}_{s} \\ &\quad - [\underbrace{r - (t_1 + \cdots + t_n)}_{t_0} + \underbrace{t_1 + \cdots + t_n}_{t}] \\ &= r - r = 0\ .\end{aligned}$$

Da $s \neq t$ sein soll, muß es also einen Index $i \in \{0, \ldots, n\}$ geben derart, daß $s_i - t_i > 0$ wird. Hiermit ist

$$\prod_{\ell_i=0}^{s_i-1} \frac{t_i - \ell_i}{s_i - \ell_i} = \frac{t_i}{s_i}\,\frac{t_i - 1}{s_i - 1} \cdots \frac{t_i - t_i}{s_i - t_i} \cdots \frac{t_i - (s_i - 1)}{1} = 0$$

und folglich $L_s\left(\frac{1}{r}\,t\right) = 0$. □

Nach diesen Vorbereitungen können wir die eindeutige Lösbarkeit der Interpolationsaufgabe (4.36) zeigen, indem wir die zugehörige biorthonormale Basis angeben. Aus der Darstellung (4.38) und Lemma 4.3 erhält man unmittelbar, daß

$$p_s(\underline{x}) := L_s\left(A^{-1}\underline{x}\right)\ , \qquad \underline{x} \in \mathbb{R}^n\ , \tag{4.41}$$

das zu $\{\delta_t \,|\, |t| \le r\}$ biorthonormale System von Polynomen aus $\mathcal{P}_r(\mathbb{R}^n)$ ist. Man hat nämlich wegen (4.38) und (4.40)

$$p_s(\underline{a}_t) = L_s(A^{-1}\underline{a}_t) = L_s\left(\frac{1}{r}\,t\right) = \delta_{st}\ .$$

Da dieses System aus genau $d = \dim \mathcal{P}_r(\mathbb{R}^n)$ Polynomen besteht, hat man hiermit eine biorthonormale Basis von $\mathcal{P}_r(\mathbb{R}^n)$ gewonnen. Nach Satz 3.1 ist folglich die Interpolationsaufgabe (4.36) stets eindeutig lösbar.

Die kanonische Darstellung von Polynomen $u \in \mathcal{P}_r(\mathbb{R}^n)$ wird jetzt zu

$$u(\underline{x}) = \sum_{|t| \le r} u(\underline{a}_t)\,p_t(\underline{x})\ , \qquad \underline{x} \in \mathbb{R}^n\ . \tag{4.42}$$

Analog zu Abschnitt 4.2 gewinnt man die zugehörige *Quadraturformel* für das Integral über das Simplex $K = \text{ conv } \{a_0, \ldots, a_n\}$ in der Form

$$\int_K u(\underline{x})\,d\underline{x} = |K| \sum_{|t|\le r} \alpha_t u(\underline{a}_t)\ , \qquad u \in \mathcal{P}_r(\mathbb{R}^n)\ , \tag{4.43}$$

mit den Koeffizienten

$$\alpha_t = \frac{1}{|K|}\int_K p_t(\underline{x})\,d\underline{x} = \frac{|\det B|}{|K|}\int_E L_t(\underline{z})\,d\underline{z} = \frac{1}{|E|}\int_E L_t(\underline{z})\,d\underline{z}\ , \quad |t| \le r\ ,$$

wobei wir von der Darstellung (4.32) Gebrauch gemacht haben.

Beispiel 4.4 Ein erstes Beispiel für diese Überlegungen haben wir bereits ausführlich in Abschnitt 4.2 mit der Interpolation in $\mathcal{P}_1(\mathbb{R}^n)$ behandelt. Für $r = 1$ liest man aus der Darstellung (4.39) das Ergebnis ab:

$$L_s(\underline{z}) = z_k\ ,\ s = (\delta_{1k}, \ldots, \delta_{nk})\ ,\ k = 0, \ldots, n\ .$$

In diesem Falle liefert (4.41) gerade die Basis (4.25) für $\mathcal{P}_1(\mathbb{R}^n)$. □

Beispiel 4.5 Interpolation in $\mathcal{P}_2(\mathbb{R}^2)$ Der Raum der Polynome zweiten Grades in 2 Veränderlichen hat die Dimension $\dim \mathcal{P}_2(\mathbb{R}^2) = 6$, wie wir in Abschnitt 4.1 festgestellt haben. Dementsprechend wird an 6 Stützstellen interpoliert (vgl. Abb. 4.5).

Zur Bestimmung der Basispolynome $L_{s_1 s_2}$ wird zuerst die Menge aller Tripel (s_0, s_1, s_2) mit $s_0 = 2 - (s_1 + s_2)$ für Multiindizes $s = (s_1, s_2)$ mit $|s| = s_1 + s_2 \le 2$ angegeben. Diese Menge besteht aus den Elementen

$$\begin{array}{lll} (2,0,0)\ , & (0,2,0)\ , & (0,0,2)\ , \\ (1,1,0)\ , & (1,0,1)\ , & (0,1,1)\ . \end{array}$$

Die zugehörigen *Lagrange Polynome* lauten

$$\begin{array}{lll} L_{00}(\underline{z}) = z_0(2z_0-1), & L_{20}(\underline{z}) = z_1(2z_1-1), & L_{02}(\underline{z}) = z_2(2z_2-1), \\ L_{10}(\underline{z}) = 4z_0z_1, & L_{01}(\underline{z}) = 4z_0z_2, & L_{11}(\underline{z}) = 4z_1z_2\ , \end{array} \tag{4.44}$$

für $\underline{z} = (z_1, z_2) \in \mathbb{R}^2$ und $z_0 = 1 - (z_1 + z_2)$. Mit den baryzentrischen Koordinaten $p_j(\underline{x}) = z_j\ ,\ j = 0, 1, 2$, hat jede Funktion $v \in \mathcal{P}_2(\mathbb{R}^2)$ daher die Darstellung

$$\begin{aligned} v(\underline{x}) \;=\; & \sum_{j=0}^{2} v(\underline{a}_j)p_j(\underline{x})(2p_j(\underline{x})-1) + \\ & + v(\underline{a}_{11})4p_1(\underline{x})p_2(\underline{x}) + v(\underline{a}_{10})4p_1(\underline{x})p_0(\underline{x}) + \\ & + v(\underline{a}_{01})4p_2(\underline{x})p_0(\underline{x})\ . \end{aligned}$$

Mit Hilfe des Formelsatzes (4.31) kann man leicht die Koeffizienten der Quadraturformel berechnen. Es ist (wegen $|E| = \frac{1}{2!}$)

$$\alpha_{00} = \alpha_{20} = \alpha_{02} = 2! \int_E z_k(2z_k - 1) d\underline{z} = 2! \left(\frac{4}{4!} - \frac{1}{3!} \right) = 0 \ ,$$
$$\alpha_{10} = \alpha_{01} = \alpha_{11} = 2!4 \int_E z_j z_k d\underline{z} = \frac{2!4}{4!} = \frac{1}{3} \quad (j \neq k) \ .$$

Die *Quadraturformel* (4.43) wird damit zu

$$\int_K u(\underline{x})\, d\underline{x} = \frac{|K|}{3} \{u(\underline{a}_{10}) + u(\underline{a}_{01}) + u(\underline{a}_{11})\} \tag{4.45}$$

für alle Polynome $u \in \mathcal{P}_2(\mathbb{R}^2)$. □

4.4 Lagrange–Interpolation in $\mathcal{Q}_{r_1,\ldots,r_n}$

Dieser Abschnitt zeigt die eindeutige Lösbarkeit der Lagrangeschen Interpolationsaufgabe für Polynome aus den Räumen $\mathcal{Q}_{r_1,\ldots,r_n}$. Die zugehörigen Stützstellen für diese Klasse von Interpolationsaufgaben werden durch das kartesische Produkt geeigneter Stützstellenmengen auf den n Koordinatenachsen gewonnen. Die einfachsten Beispiele erhält man mit der Interpolation durch bilineare und biquadratische Polynome. Ein weiteres einfaches Beispiel bildet die Interpolation in $\mathcal{Q}_{1,r}$. Darüber hinaus interessieren wir uns für biorthonormale Basissysteme und Anwendungen der Interpolation auf Quadraturformeln und Differenzenquotienten.

Der Polynomraum $\mathcal{Q}_{r_1,\ldots,r_n}$ hat die Dimension $d = (r_1 + 1) \cdots (r_n + 1)$. Für jedes j seien dementsprechend $r_j + 1$ paarweise verschiedene Punkte

$$c_{j0}, \ldots, c_{jr_j} \in \mathbb{R} \ , \qquad j = 1, \ldots, n \ , \tag{4.46}$$

gegeben. Die Lagrange–Polynome r_j-ten Grades in einer Veränderlichen zu diesen Mengen von Stützstellen bezeichnen wir mit $L_{j\nu}$. Damit setzen wir

$$p_t(\underline{x}) := L_{1t_1}(x_1) \cdots L_{nt_n}(x_n) \ , \qquad \underline{x} \in \mathbb{R}^n \ , \tag{4.47}$$

für jeden Multiindex $t = (t_1, \ldots, t_n)$ mit $0 \leq t \leq (r_1, \ldots, r_n)$. Offenbar ist jedes Polynom p_t ein Element von $\mathcal{Q}_{r_1,\ldots,r_n}$. Zu jedem durch (4.46) definierten Punkt

$$\underline{c}_s = (c_{1s_1}, \ldots, c_{ns_n}) \in \mathbb{R}^n \ , \quad 0 \leq s = (s_1, \ldots, s_n) \leq (r_1, \ldots, r_n) \tag{4.48}$$

ist dann

$$p_t\,(\underline{c}_s) = L_{1t_1}\,(c_{1s_1}) \cdots L_{nt_n}\,(c_{ns_n}) = \delta_{s_1 t_1} \cdots \delta_{s_n t_n} =: \delta_{st} \ . \tag{4.49}$$

Folglich ist dieses System von Polynomen $p_t \in \mathcal{Q}_{r_1,\ldots,r_n}$ biorthonormal zum System der δ–Funktionale zu den Punkten $\underline{c}_s$, welche durch

$$\delta_s(v) = v(\underline{c}_s) \;, \qquad v \in \mathcal{P}(\mathbb{R}^n) \;, \qquad 0 \le s \le r \;, \tag{4.50}$$

definiert sind. Das System der p_t ist linear unabhängig und besteht aus $d = (r_1 + 1) \cdots (r_n + 1)$ Polynomen, so daß dieses System eine Basis von $\mathcal{Q}_{r_1,\ldots,r_n}$ bildet.

Auf Grund von Satz 3.1 ist daher die Interpolationsaufgabe

$$u \in \mathcal{Q}_{r_1,\ldots,r_n} \;: \quad u(\underline{c}_s) = y_s \;, \quad 0 \le s \le (r_1, \ldots, r_n) \;, \tag{4.51}$$

für jedes System von rechten Seiten $(y_s) \in \mathbb{K}^d$ eindeutig lösbar. Die Lösung u hat die kanonische Darstellung

$$u(\underline{x}) = \sum_{0 \le s \le (r_1,\ldots,r_n)} u(\underline{c}_s) p_s(\underline{x}) \;, \qquad \underline{x} \in \mathbb{R}^n \;, \tag{4.52}$$

und jede Linearform f auf $\mathcal{Q}_{r_1,\ldots,r_n}$ hat die zugehörige Darstellung

$$f(u) = \sum_{0 \le s \le (r_1,\ldots,r_n)} \beta_s u(\underline{c}_s) \;, \qquad u \in \mathcal{Q}_{r_1,\ldots,r_n} \;, \tag{4.53}$$

mit den Koeffizienten $\beta_s = f(p_s)$. Hiermit gewinnt man zum Beispiel für das bestimmte Integral

$$q(u) = f(u) = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} u(x_1, \ldots, x_n)\, dx_1 \cdots dx_n$$

durch (4.53) eine *Quadraturformel* mit den Koeffizienten

$$\beta_s = \beta_{1s_1} \cdots \beta_{ns_n} \;, \quad \beta_{js_j} = \int_{a_j}^{b_j} L_{js_j}(x_j)\, dx_j \;, \qquad j = 1, \ldots, n \;, \tag{4.54}$$

für jeden Multiindex $s = (s_1, \ldots, s_n)$.

Beispiel 4.6 Interpolation in $\mathcal{Q}_1(\mathbb{R}^2)$ In diesem Falle betrachtet man die Interpolationsaufgabe (4.51) für $r_1 = r_2 = 1$ mit dem Raum

$$\mathcal{Q}_1(\mathbb{R}^2) = [1, x, y, xy]$$

der bilinearen Funktionen auf $\mathbb{R}^2$. Weiter sind gemäß (4.46) vier Punkte $\underline{c}_{00}$, $\underline{c}_{10}$, $\underline{c}_{01}$, $\underline{c}_{11}$ in $\mathbb{R}^2$ gegeben.

Zur Abkürzung führen wir die neuen Variablen

$$z_j = \frac{1}{h_j}(x_j - c_{j0}) \ , \quad h_j = c_{j1} - c_{j0} \ , \qquad j = 1,2 \ ,$$

ein. Dann haben die Punkte $\underline{c}_{00}, \ldots, \underline{c}_{11}$ die Koordinaten $(z_1, z_2) = (0,0), \ldots, (1,1)$. Weiter ist

$$L_{j0}(x_j) = 1 - z_j \ , \quad L_{j1}(x_j) = z_j \ , \qquad j = 1,2 \ ,$$

und es gilt

$$\begin{aligned} L_{j0}(c_{j0}) &= 1 - z_j|_{z_j=0} = 1 \ , \\ L_{j0}(c_{j1}) &= 1 - z_j|_{z_j=1} = 0 \ , \\ L_{j1}(c_{j0}) &= 0 \, , \ L_{j1}(c_{j1}) = 1 \ ; \end{aligned}$$

die vier Basisfunktionen lauten

$$\begin{aligned} p_{00}(\underline{x}) &= (1-z_1)(1-z_2) \ , \quad p_{10}(\underline{x}) = z_1(1-z_2) \ , \\ p_{01}(\underline{x}) &= (1-z_1)z_2 \ , \qquad\quad p_{11}(\underline{x}) = z_1 z_2 \ . \end{aligned} \tag{4.55}$$

Die zugehörige *Quadraturformel* (4.53) über $[a,b] = [a_1,b_1] \times [a_2,b_2]$ für

$$a_j = c_{j0} \, , \quad b_j = c_{j1} \ , \qquad j = 1,2 \ ,$$

lautet dann

$$\int_{a_1}^{b_1} \int_{a_2}^{b_2} u(x_1,x_2)\, dx_1 dx_2 = \frac{h_1 h_2}{4} \left(u(\underline{c}_{00}) + u(\underline{c}_{10}) + u(\underline{c}_{01}) + u(\underline{c}_{11}) \right) \ , \quad u \in \mathcal{Q}_1(\mathbb{R}^2) \ , \tag{4.56}$$

denn

$$\begin{aligned} \beta_{00} = \beta_{10}\beta_{20} &= \int_{a_1}^{b_1} L_{10}\, dx_1 \int_{a_2}^{b_2} L_{20}\, dx_2 = \frac{h_1}{2}\,\frac{h_2}{2} \ , \\ \beta_{10} = \beta_{11}\beta_{20} &= \int_{a_1}^{b_1} L_{11}\, dx_1 \int_{a_2}^{b_2} L_{20}\, dx_2 = \frac{h_1}{2}\,\frac{h_2}{2} \ \text{usw.} \end{aligned}$$

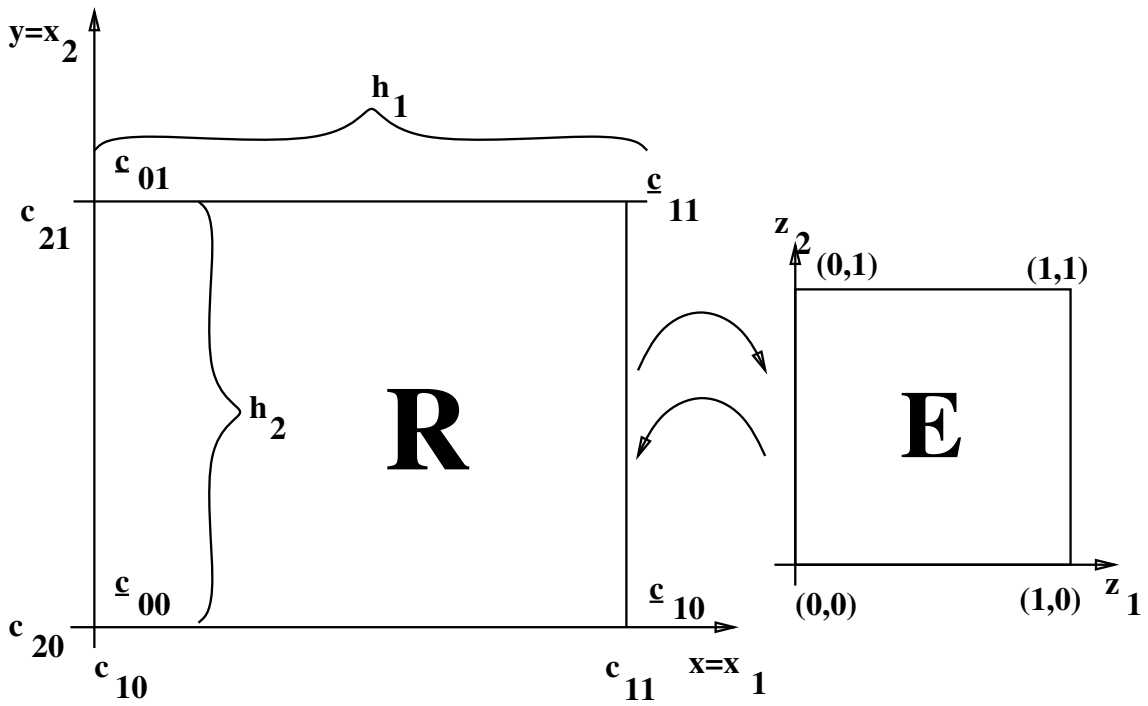


Abbildung 4.6: Stützstellen der Interpolation in $\mathcal{Q}_1(\mathbb{R}^2)$

Differenzenquotienten in $\mathcal{Q}_1(\mathbb{R}^2)$:

(kanonische Darstellung) $\quad \dfrac{\partial u}{\partial x_i}(\underline{c}) = \displaystyle\sum_{0\le s_1,s_2\le 1} \frac{\partial p_s}{\partial x_i}(\underline{c}) u(\underline{c}_s)$

wobei

$$\frac{\partial p_{00}}{\partial x_1}(\underline{c}) = \frac{\partial p_{00}}{\partial z_1}(\underline{c}) \underbrace{\frac{dz_1}{dx_1}}_{1/h_1} = -(1-z_2)\frac{1}{h_1} = -\frac{1}{h_1 h_2}(c_{21}-c_2)$$

$\underline{c} = (c_1, c_2)$, da $1 - z_j = \frac{1}{h_j}(c_{j1} - x_j)$, $j = 1, 2$;

analog:

$$\begin{aligned}
\frac{\partial p_{00}}{\partial x_2}(\underline{c}) &= \frac{-1}{h_1 h_2}(c_{11}-c_1)\,, \\
\frac{\partial p_{10}}{\partial x_1}(\underline{c}) &= \frac{1}{h_1 h_2}(c_{21}-c_2)\,, \quad \frac{\partial p_{10}}{\partial x_2}(\underline{c}) = -\frac{1}{h_1 h_2}(c_1-c_{10})\,, \\
\frac{\partial p_{01}}{\partial x_1}(\underline{c}) &= \frac{-1}{h_1 h_2}(c_2-c_{20})\,, \quad \frac{\partial p_{01}}{\partial x_2}(\underline{c}) = \frac{1}{h_1 h_2}(c_{11}-c_1)\,,
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial p_{11}}{\partial x_1}(\underline{c}) &= \frac{1}{h_1 h_2}(c_2 - c_{20}) \,, \quad \frac{\partial p_{11}}{\partial x_2}(\underline{c}) = \frac{1}{h_1 h_2}(c_1 - c_{10}) \,, \\
\Longrightarrow \frac{\partial u}{\partial x_1}(\underline{c}) &= \frac{1}{h_1 h_2}\Big\{(c_{21} - c_2)[-u(\underline{c}_{00}) + u(\underline{c}_{10})] + \\
&\qquad + (c_2 - c_{20})[-u(\underline{c}_{01}) + u(\underline{c}_{11})]\Big\} \\
\text{und } \frac{\partial u}{\partial x_2}(\underline{c}) &= \frac{1}{h_1 h_2}\Big\{(c_{11} - c_1)[-u(\underline{c}_{00}) + u(\underline{c}_{01})] + \\
&\qquad + (c_1 - c_{10})[-u(\underline{c}_{10}) + u(\underline{c}_{11})]\Big\} \,.
\end{aligned}
$$

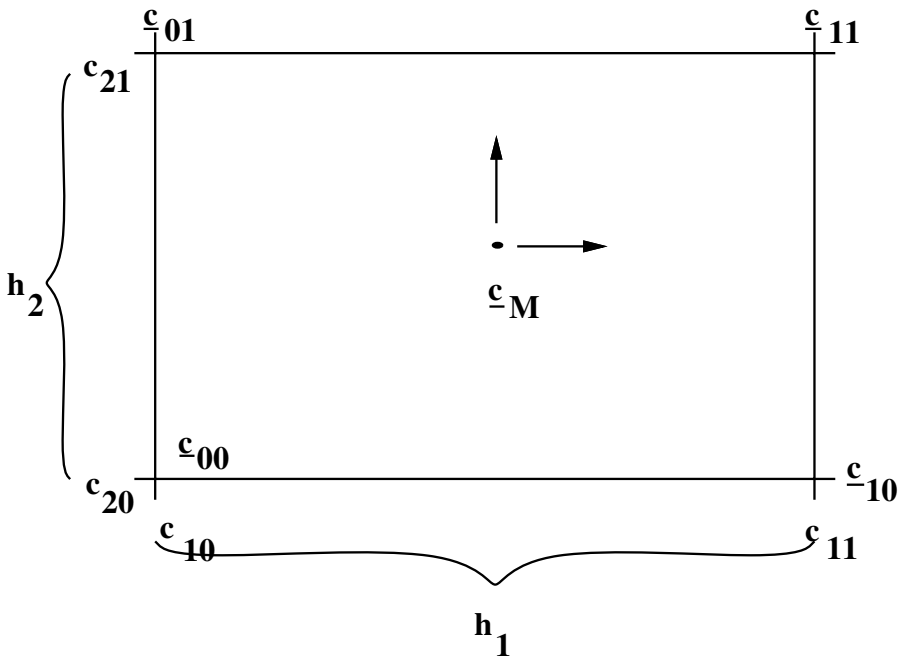


Abbildung 4.7: Differenzenquotienten in $\mathcal{Q}_1(\mathbb{R}^2)$

Speziell für $\underline{c} = \underline{c}_M = \left(\frac{c_{10} + c_{11}}{2} \,,\ \frac{c_{20} + c_{21}}{2}\right)$:

$$
\begin{aligned}
\frac{\partial u}{\partial x_1}(\underline{c}_M) &= \frac{1}{h_1 h_2}\left\{\frac{h_2}{2}[u(\underline{c}_{10}) - u(\underline{c}_{00})] + \frac{h_2}{2}[u(\underline{c}_{11}) - u(\underline{c}_{01})]\right\} \\
&= \frac{1}{2}\left\{\frac{u(\underline{c}_{10}) - u(\underline{c}_{00})}{h_1} + \frac{u(\underline{c}_{11}) - u(\underline{c}_{01})}{h_1}\right\} \,, \\
\frac{\partial u}{\partial x_2}(\underline{c}_M) &= \frac{1}{2}\left\{\frac{u(\underline{c}_{01}) - u(\underline{c}_{00})}{h_2} + \frac{u(\underline{c}_{11}) - u(\underline{c}_{10})}{h_2}\right\} \,, \quad u \in \mathcal{Q}_1(\mathbb{R}^2) \,;
\end{aligned}
$$

speziell für $\underline{c} = \underline{c}_{00}$ (analog für $\underline{c}_{01}, \underline{c}_{11}, \underline{c}_{10}$) :
Es ist

$$\begin{aligned}
\frac{\partial p_{00}}{\partial x_1}(\underline{c}_{00}) &= -\frac{1}{h_1 h_2}\overbrace{(c_{21} - c_{20})}^{h_2} = -\frac{1}{h_1} \\
\frac{\partial p_{10}}{\partial x_1}(\underline{c}_{00}) &= \frac{1}{h_1 h_2} h_2 = \frac{1}{h_1}\,, \quad \frac{\partial p_{01}}{\partial x_1}(\underline{c}_{00}) = -\frac{1}{h_1 h_2} 0 = 0 \\
\frac{\partial p_{11}}{\partial x_1}(\underline{c}_{00}) &= 0 \\
\Longrightarrow \frac{\partial u}{\partial x_1}(\underline{c}_{00}) &= \frac{1}{h_1}(u(\underline{c}_{10}) - u(\underline{c}_{00}))\ ;
\end{aligned}$$

analog: $\dfrac{\partial u}{\partial x_2}(\underline{c}_{00}) = \dfrac{1}{h_2}(u(\underline{c}_{01}) - u(\underline{c}_{00}))$.

Bemerkung: Offenbar ist $\dfrac{\partial^2 u}{\partial x_i^2} = 0\,, \quad i = 1, 2\,,\ u \in \mathcal{Q}_1(\mathbb{R}^2)$. □

Beispiel 4.7 Interpolation in $\mathcal{Q}_{1,r}(\mathbb{R}^2)$ Dieser Polynomraum ist definiert durch

$$\mathcal{Q}_{1,r}(\mathbb{R}^2) = [1, x, y, xy, \ldots, y^r, xy^r]$$

und hat die Dimension $2(r+1)$. Dementsprechend sind Abszissen

$$c_{10}, c_{11};\ c_{20}, \ldots, c_{2r} \in \mathbb{R}$$

gegeben und die $2(r+1)$ zugehörigen Punkte $\underline{c}_s = (c_{1s_1}, c_{2s_2})$ für $s_1 = 0, 1$;
$s_2 = 0, \ldots, r$.
Die zugehörigen Lagrange–Polynome L_{1s_1} lauten

$$L_{10}(x) = \frac{1}{h_1}(c_{11} - x)\,, \quad L_{11}(x) = \frac{1}{h_1}(x - c_{10})\,, \qquad x \in \mathbb{R}\,,$$

wobei $h_1 = c_{11} - c_{10}$. Also hat die zu den δ–Funktionalen an den Stellen $\underline{c}_s$ biorthonormale Basis von Polynomen $p_t \in \mathcal{Q}_{1,r}$ die Form

$$p_{0\ell}(\underline{x}) = \frac{1}{h_1}(c_{11} - x) L_{2\ell}(y)\,, \quad p_{1\ell}(\underline{x}) = \frac{1}{h_1}(x - c_{10}) L_{2\ell}(y)$$

für alle $\underline{x} = (x, y) \in \mathbb{R}^2$ und $\ell = 0, \ldots, r$, mit $L_{2\ell}(y) = \prod\limits_{\substack{j=0 \\ j \neq \ell}}^{r} \frac{y - c_{2j}}{c_{2\ell} - c_{2j}}$. □

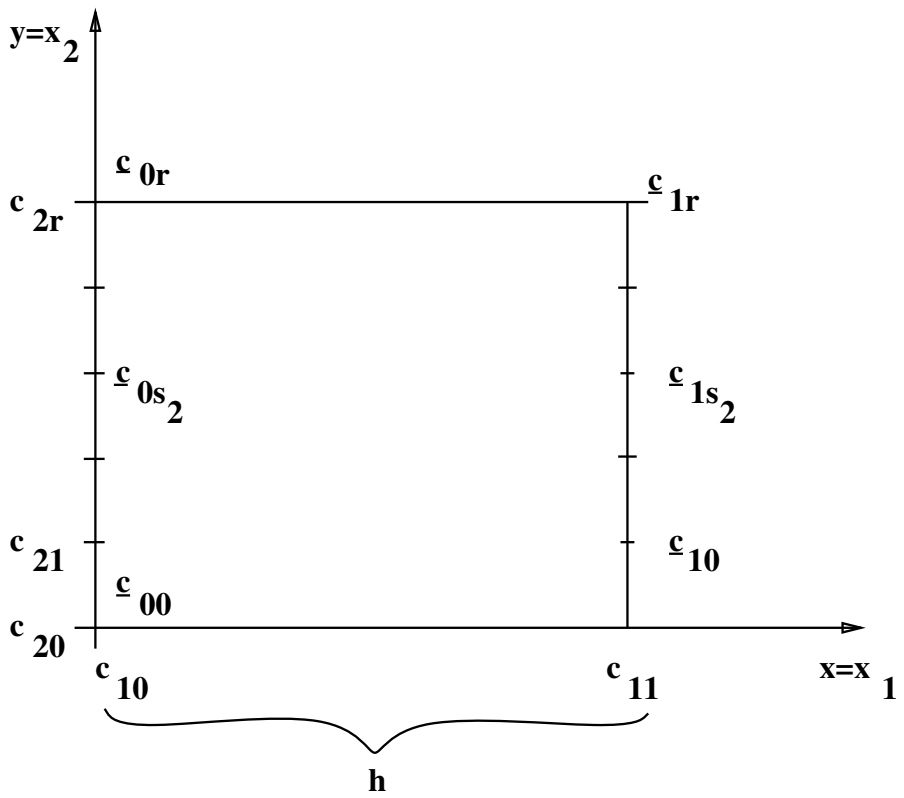


Abbildung 4.8: Stützstellen der Interpolation in $\mathcal{Q}_{1,r}(\mathbb{R}^2)$

Beispiel 4.8 Differenzenquotienten in $\mathcal{Q}_{22}(\mathbb{R}^2)(\dim \mathcal{Q}_{22} = 9)$

(speziell sei $c_{j0} = a_j\,,\ c_{j2} = b_j\,,\ c_{j1} = \frac{a_j+b_j}{2}\,,\ j = 1,2,$ mit $b_j > a_j$).

Es ist

$$
\begin{aligned}
L_{j0}(x_j) &= \frac{x_j - c_{j1}}{c_{j0} - c_{j1}} \cdot \frac{x_j - c_{j2}}{c_{j0} - c_{j2}} \\
L_{j1}(x_j) &= \frac{x_j - c_{j0}}{c_{j1} - c_{j0}} \cdot \frac{x_j - c_{j2}}{c_{j1} - c_{j2}} \\
L_{j2}(x_j) &= \frac{x_j - c_{j0}}{c_{j2} - c_{j0}} \cdot \frac{x_j - c_{j1}}{c_{j2} - c_{j1}}
\end{aligned}
$$

also $L_{j0}(c_{j1}) = 0\,,\ L_{j1}(c_{j1}) = 1\,,\ L_{j2}(c_{j1}) = 0\,,\ j = 1,2.$

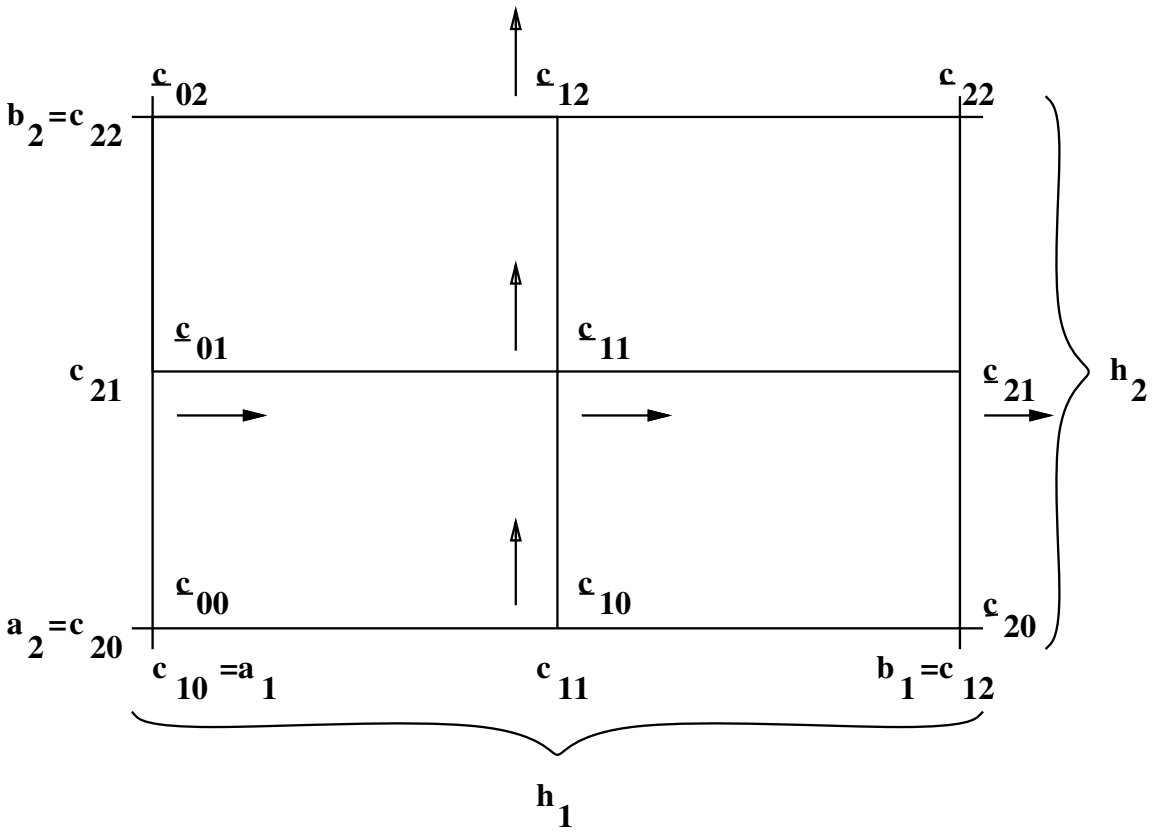


Abbildung 4.9: Differenzenquotienten in $\mathcal{Q}_{22}(\mathbb{R}^2)$

$$\Longrightarrow \qquad \begin{aligned} L'_{j0}(x_j) &= \frac{2}{h_j^2}(x_j - c_{j2} + x_j - c_{j1}) , \quad L''_{j0} = \frac{4}{h_j^2} , \\ L'_{j1}(x_j) &= \frac{4}{h_j^2}(x_j - c_{j0} + x_j - c_{j2}) , \quad L''_{j1} = \frac{8}{h_j^2} , \\ L'_{j2}(x_j) &= \frac{2}{h_j^2}(x_j - c_{j0} + x_j - c_{j1}) , \quad L''_{j2} = \frac{4}{h_j^2} , \quad j = 1,2 . \end{aligned}$$

Man hat also bei c_{j1}:

$$\begin{aligned} L'_{j0}(c_{j1}) &= -\frac{1}{h_j} , \\ L'_{j1}(c_{j1}) &= -\frac{4}{h_j^2}\left\{\frac{h_j}{2} + -\frac{h_j}{2}\right\} = 0 , \\ L'_{j2}(c_{j1}) &= \frac{1}{h_j} , \qquad j = 1,2 . \end{aligned}$$

Zu berechnen ist $\frac{\partial u}{\partial x_j}(\underline{c}_{11})$ und $\frac{\partial^2 u}{\partial x_j^2}(\underline{c}_{11})$. Nach der kanonischen Darstellung hat man

$$\begin{aligned} \frac{\partial u}{\partial x_j}(\underline{c}_{11}) &= \sum_{0 \le s_1, s_2 \le 2} \frac{\partial p_s}{\partial x_j}(\underline{c}_{11}) u(\underline{c}_s) , \\ \frac{\partial^2 u}{\partial x_j^2}(\underline{c}_{11}) &= \sum_{0 \le s_1, s_2 \le 2} \frac{\partial^2 p_s}{\partial x_j^2}(\underline{c}_{11}) u(\underline{c}_s) , \qquad u \in \mathcal{Q}_2(\mathbb{R}^2) , \end{aligned}$$

wobei
$$\begin{aligned} &p_{00} = L_{10} \cdot L_{20} , \quad p_{10} = L_{11} \cdot L_{20} , \quad p_{20} = L_{12} \cdot L_{20} , \\ &p_{01} = L_{10} \cdot L_{21} , \quad p_{11} = L_{11} \cdot L_{21} , \quad p_{21} = L_{12} \cdot L_{21} , \\ &p_{02} = L_{10} \cdot L_{22} , \quad p_{12} = L_{11} \cdot L_{22} , \quad p_{22} = L_{12} \cdot L_{22} . \end{aligned}$$

Bei $\underline{x} = \underline{c}_{11}$ erhält man

$$\begin{aligned} &\frac{\partial p_{00}}{\partial x_1} = L'_{10} \cdot \underbrace{L_{20}}_{=0} = 0 , \quad \frac{\partial p_{00}}{\partial x_2} = \underbrace{L_{10}}_{=0} \cdot L'_{20} = 0 , \\ &\frac{\partial p_{10}}{\partial x_1} = \underbrace{L'_{11}}_{=0} \cdot L_{20} = 0 , \quad \frac{\partial p_{10}}{\partial x_2} = \underbrace{L_{11}}_{1} \cdot L'_{20} = -\frac{1}{h_2} , \\ &\frac{\partial p_{20}}{\partial x_1} = L'_{12} \cdot \underbrace{L_{20}}_{=0} = 0 , \quad \frac{\partial p_{20}}{\partial x_2} = \underbrace{L_{12}}_{=0} \cdot L'_{20} = 0 , \end{aligned}$$

$$
\begin{aligned}
&\frac{\partial p_{01}}{\partial x_1} = L'_{10}\cdot\underbrace{L_{21}}_{1} = -\frac{1}{h_1}\ , \quad \frac{\partial p_{01}}{\partial x_2} = L_{10}\cdot\underbrace{L'_{21}}_{=0} = 0\ ,\\
&\frac{\partial p_{11}}{\partial x_1} = \underbrace{L'_{11}}_{=0}\cdot L_{21} = 0\ , \qquad \frac{\partial p_{11}}{\partial x_2} = L_{11}\cdot\underbrace{L'_{21}}_{=0} = 0\ ,\\
&\frac{\partial p_{21}}{\partial x_1} = L'_{12}\cdot\underbrace{L_{21}}_{=1} = \frac{1}{h_1}\ , \qquad \frac{\partial p_{21}}{\partial x_2} = L_{12}\cdot\underbrace{L'_{21}}_{=0} = 0\ ,\\
&\frac{\partial p_{02}}{\partial x_1} = L'_{10}\cdot\underbrace{L_{22}}_{=0} = 0\ , \qquad \frac{\partial p_{02}}{\partial x_2} = \underbrace{L_{10}}_{=0}\cdot L'_{22} = 0\ ,\\
&\frac{\partial p_{12}}{\partial x_1} = L'_{11}\cdot\underbrace{L_{22}}_{=0} = 0\ , \qquad \frac{\partial p_{12}}{\partial x_2} = \underbrace{L_{11}}_{=1}\cdot L'_{22} = \frac{1}{h_2}\ ,\\
&\frac{\partial p_{22}}{\partial x_1} = L'_{12}\cdot\underbrace{L_{22}}_{=0} = 0\ , \qquad \frac{\partial p_{22}}{\partial x_2} = \underbrace{L_{12}}_{=0}\cdot L'_{22} = 0\ ,
\end{aligned}
$$

also $$\frac{\partial u}{\partial x_1}(\underline{c}_{11}) = \frac{1}{h_1}(u(\underline{c}_{21}) - u(\underline{c}_{01}))$$

und $$\frac{\partial u}{\partial x_2}(\underline{c}_{11}) = \frac{1}{h_2}(u(\underline{c}_{12}) - u(\underline{c}_{10}))\ .$$

Für die *2. Ableitungen* berechnet man:

$$\frac{\partial^2 p_{00}}{\partial x_1^2} = L''_{10}\cdot\underbrace{L_{20}}_{=0} = 0\ , \quad \frac{\partial^2 p_{00}}{\partial x_2^2} = \underbrace{L_{10}}_{=0}\cdot L''_{20} = 0\ ;$$

analog:

$$
\begin{aligned}
&\frac{\partial^2 p_{10}}{\partial x_1^2} = \frac{\partial^2 p_{20}}{\partial x_1^2} = \frac{\partial^2 p_{02}}{\partial x_1^2} = \frac{\partial^2 p_{12}}{\partial x_1^2} = \frac{\partial^2 p_{22}}{\partial x_1^2} = 0\ ,\\
&\frac{\partial^2 p_{01}}{\partial x_1^2} = L''_{10}\cdot\underbrace{L_{21}}_{=1} = \frac{4}{h_1^2}\left(= \frac{\partial^2 p_{21}}{\partial x_1^2} = L''_{12}\cdot L_{21}\right)\ , \ \frac{\partial^2 p_{11}}{\partial x_1^2} = L''_{11}\cdot L_{21} = -\frac{8}{h_1^2}\ ,\\
&\frac{\partial^2 p_{20}}{\partial x_2^2} = \frac{\partial^2 p_{21}}{\partial x_2^2} = \frac{\partial^2 p_{02}}{\partial x_2^2} = \frac{\partial^2 p_{22}}{\partial x_2^2} = 0\ ,\\
&\frac{\partial^2 p_{10}}{\partial x_2^2} = L_{11}\cdot L''_{20} = \frac{4}{h_2^2}\left(= \frac{\partial^2 p_{12}}{\partial x_2^2} = L_{11}\cdot L''_{22}\right)\ ,\\
&\frac{\partial^2 p_{11}}{\partial x_2^2} = L_{11}\cdot L''_{21} = -\frac{8}{h_j^2}\ .
\end{aligned}
$$

$\Longrightarrow$

$$\frac{\partial^2 u}{\partial x_1^2}(\underline{c}_{11}) = \frac{4}{h_1^2}\left\{u(\underline{c}_{01}) - 2u(\underline{c}_{11}) + u(\underline{c}_{21})\right\}$$

und

$$\frac{\partial^2 u}{\partial x_2^2}(\underline{c}_{11}) = \frac{4}{h_2^2}\Big\{u(\underline{c}_{10}) - 2u(\underline{c}_{11}) + u(\underline{c}_{12})\Big\} .$$

□

4.5 Hermite–Interpolation

In diesem Abschnitt werden allgemeine Hermitesche Interpolationsaufgaben betrachtet. Wie in den vorherigen Abschnitten entspricht in konkreten Beispielen wieder einer vorgegebenen Geometrie — jetzt in Verbindung mit den durch die Hermite–Interpolation gegebenen Knotenfunktionalen – ein geeigneter endlichdimensionaler Teilraum von $\mathcal{P}(\mathbb{R}^n)$. Ähnlich wie bei der Lagrange–Interpolation gehen wir von der gegebenen Geometrie (Simplex oder Parallelepiped) mittels der schon angegebenen affinen Abbildungen zu einem Einheitselement (Einheitssimplex oder n–dimensionales Quadrat) über, lösen die Interpolationsaufgabe dort und transformieren die erhaltene Lösung zurück, so daß man die Lösung der gegebenen Aufgabe erhält. Wir beweisen in diesem Abschnitt Existenz- und Eindeutigkeitsaussagen für die allgemeine Hermite–Interpolation und diskutieren wichtige Beispiele.

Gegeben sei eine abgeschlossene Teilmenge e von $\mathbb{R}^n$ (z. B. Simplex oder Parallelepiped) und ein „Einheitselement“ (Einheitssimplex oder n–dimensionales Quadrat) $E \subset \mathbb{R}^n$ derart, daß mit einer umkehrbaren affinen Abbildung

$$A\underline{z} = \underline{a}_0 + B\underline{z} , \qquad \underline{z} \in \mathbb{R}^n ,$$

gilt $E = A^{-1}(e)$. Damit ist auch die lineare Abbildung B regulär und A^{-1} läßt sich ebenfalls als affine Abbildung schreiben,

$$(\underline{x} =)A\underline{z} = \underline{a}_0 + B\underline{z} \Longleftrightarrow (\underline{y} =)B\underline{z} = A\underline{z} - \underline{a}_0$$

und

$$\underline{z} = A^{-1}\underline{x} = B^{-1}(\underline{x} - \underline{a}_0) = B^{-1}\underline{x} + \underline{c} , \quad \text{wobei } \underline{c} = -B^{-1}\underline{a}_0 .$$

Im Fall eines Simplex bedeutet die Invertierbarkeit von A bzw. B, daß die Ecken affin unabhängige Punkte darstellen; im Fall von n–dimensionalen Rechtecken liefert die Vorgehensweise in Abschnitt 4.4 solche Abbildungen A und B.

Gegeben seien weiter ein endlichdimensionaler Vektorraum M_E von Polynomen mit $\dim M_E = N$ und Punkte $\underline{z}_1, \ldots, \underline{z}_N \in E$ sowie Differentialoperatoren D^i von der

Ordnung $\alpha_i \geq 0$. (Dies können partielle Ableitungen aber auch Richtungsableitungen sein.)

Mit diesen Bezeichnungen besteht die allgemeine *Hermitesche Interpolationsaufgabe auf E* darin, zu gegebenen $\eta_i \in \mathbb{K}$, $i = 1, \ldots, N$, ein $U \in M_E$ zu finden, das

$$D^i U(\underline{z}_i) = \eta_i \ , \qquad i = 1, \ldots, N \ , \tag{4.57}$$

erfüllt.

Die zugehörigen *Knotenfunktionale* sind offenbar durch

$$q_i(V) = D^i V(\underline{z}_i) \ , \qquad i = 1, \ldots, N \ , \tag{4.58}$$

gegeben. Die M_E–Unisolvenz dieser Funktionale wird durch die folgende Eigenschaft gesichert, was im darauf folgenden Satz bewiesen wird:

Eigenschaft 4.9 *Das einzige Polynom $P \in M_E$ mit*

$$D^i P(\underline{z}_i) = 0 \ , \qquad i = 1, \ldots, N \ ,$$

ist das Nullpolynom.

Der folgende Satz zeigt nun die eindeutige Lösbarkeit der Hermiteschen Interpolationsaufgabe (4.57) und gibt ein biorthonormales System zu $\{q_i\}$ an.

Satz 4.10 *Die allgemeine Hermitesche Interpolationsaufgabe auf E ist d. u. n. d. für jede rechte Seite $\eta_1, \ldots, \eta_N$ eindeutig lösbar, wenn Eigenschaft 4.9 gilt. In diesem Fall hat die Lösung U die Darstellung*

$$U = \sum_{i=1}^{N} D^i U(\underline{z}_i) P_i \ , \tag{4.59}$$

wobei die $P_j \in M_E$ definiert sind als Lösungen von

$$D^i P_j(\underline{z}_i) = \delta_{ij} \ , \qquad i, j = 1, \ldots, N \ . \tag{4.60}$$

Beweis: Für jede, beliebige Basis $\{\psi_1, \ldots, \psi_N\}$ in M_E ist $\det(q_i(\psi_j)) \neq 0$, denn das homogene Gleichungssystem

$$\sum_{j=1}^{N} \alpha_j q_i(\psi_j) \left(= \sum \alpha_j D^i \psi_j(\underline{z}_i) \right) = 0 \ , \qquad i = 1, \ldots, N \ ,$$

besitzt nur die triviale Lösung. Andernfalls gäbe es eine nichttriviale Lösung $(\alpha_1, \ldots, \alpha_n) \neq 0$ und

$$P = \sum \alpha_j \psi_j (\not\equiv 0)$$

erfüllte

$$D^i P(\underline{z}_i) = \sum_{j=1}^{N} \alpha_j D^i \psi_j(\underline{z}_i) = 0$$

für alle i im Widerspruch zur Eigenschaft 4.9. Nach dem allgemeinen Satz 3.3 ist damit die M_E–Unisolvenz gezeigt. Umgekehrt folgt aus der M_E–Unisolvenz offenbar sofort Eigenschaft (4.57).

Das biorthonormale System erhält man offenbar durch die Polynome, die sich als Linearkombinationen

$$P_k = \sum \beta_j^{(k)} \psi_j \ , \qquad k = 1, \ldots, N \ ,$$

der ψ_j , $j = 1, \ldots, N$, ergeben, wobei die $\beta^{(k)} = (\beta_1^{(k)}, \ldots, \beta_N^{(k)})$ Lösungen von

$$\sum_{j=1}^{N} \beta_j^{(k)} D^i \psi_j(\underline{z}_i) = \delta_{ik} \ , \qquad i = 1, \ldots, N \ ,$$

sind. Dann ist nämlich

$$D^i P_k(\underline{z}_i) = \sum_{j=1}^{N} \beta_j^{(k)} D^i \psi_j = \delta_{ik} \ , \qquad i, k = 1, \ldots, N \ .$$

□

Wir betrachten nun das eigentliche Problem, nämlich die *Hermitesche Interpolationsaufgabe auf e.* Wir bezeichnen mit

$$\underline{x}_i = A\underline{z}_i \ , \qquad i = 1, \ldots, N \ ,$$

die Punkte, die aus den Stützstellen $\underline{z}_i$ unter der Abbildung A hervorgehen. Entsprechend definieren wir einen Polynomraum durch

$$M_e := \left\{ p \,\middle|\, \exists P \in M_E : p(\underline{x}) = P(A^{-1}\underline{x}) \right\} \tag{4.61}$$

und erhalten Basisfunktionen in M_e durch

$$p_i(\underline{x}) = P_i(A^{-1}\underline{x}) \, , i = 1, \ldots, N \, . \tag{4.62}$$

Die *allgemeine Hermitesche Interpolationsaufgabe auf* e lautet: Zu gegebenen $y_1, \ldots,$ $y_N \in \mathbb{K}$, bestimme $u \in M_e$ mit

$$D^i u(\underline{x}_i) = y_i \, , \qquad i = 1, \ldots, N \, . \tag{4.63}$$

Mit Hilfe von Satz 4.10 erhält man leicht die folgenden Aussagen zur Existenz und Eindeutigkeit der Lösung sowie deren explizite Darstellung.

Satz 4.11 *Eigenschaft 4.9 ist äquivalent dazu, daß die allgemeine Hermitesche Interpolationsaufgabe (4.63) auf* e *für jede rechte Seite eindeutig lösbar ist. Die Lösung hat dann die Darstellung*

$$u(\underline{x}) = \sum_{i=1}^{N} y_i p_i(\underline{x}) \, , \tag{4.64}$$

wobei die p_i , $i = 1, \ldots, N$, *in (4.62) definiert sind.*

Beweis: Wenn Eigenschaft 4.9 gilt, dann ist (4.57) eindeutig lösbar mit der Lösung (4.59). Die in (4.64) angegebene Funktion erfüllt dann offenbar

$$D^i u(\underline{x}_i) = \sum_{j=1}^{N} y_j \overbrace{D^i P_j \overbrace{(A^{-1}\underline{x}_i)}^{\underline{z}_i}}^{\delta_{ij}} = y_i \, , \qquad i = 1, \ldots, N \, ,$$

löst also (4.63). Die Eindeutigkeit ist klar, da mit jeder Lösung $v \in M_e$ der homogenen Interpolationsaufgabe die Funktion $V(\underline{z}) = v(A\underline{z})$ in M_E liegt und $D^i V(\underline{z}_i) = D^i v(A\underline{z}_i) = D^i v(\underline{x}_i) = 0$, $i = 1, \ldots N$, erfüllt, also $V \equiv 0$ sein muß; damit ist auch $v(\underline{x}) = V(A^{-1}\underline{x}) = 0 \ \forall \underline{x}$. Die Umkehrung ist wieder trivial. □

Abschließend bemerken wir noch zu der allgemeinen Interpolationsaufgabe, daß für jede integrierbare Funktion v (über e) und die zugehörige transformierte Funktion

$$V(\underline{z}) = v(A\underline{z}) \, , \qquad \underline{z} \in E \, ,$$

mit Hilfe von Variablensubstitution gilt

$$\int_e v(\underline{x})\,d\underline{x} = \int_{A(E)} \underbrace{v(A\underline{z})}_{V(\underline{z})}\,d\underline{x} = |\det B| \int_E V(\underline{z})\,d\underline{z} \tag{4.65}$$

Beispiel 4.12 Die Lagrange–Interpolationen in $\mathcal{P}_r(\mathbb{R}^n)$ und $\mathcal{Q}_{r_1,\ldots,r_n}(\mathbb{R}^n)$ fallen offenbar auch in diese Aufgabenstellung (mit $D^i v(\underline{z}_i) = v(\underline{z}_i)$). Im Einheitssimplex bzw. dem n–dimensionalen Quadrat mit paarweise verschiedenen Punkten auf den Kanten ist die Eigenschaft 4.9 trivialerweise erfüllt. Für $\mathcal{P}_1(\mathbb{R}^n)$ besteht die eigentliche Voraussetzung darin, daß die umkehrbare affine Abbildung A existiert, was nach Abschnitt 4.2 äquivalent zur affinen Unabhängigkeit der Vektoren $\underline{a}_0, \underline{a}_1, \ldots, \underline{a}_n$ ist. □

Beispiel 4.13 „Dreieck von Morley“

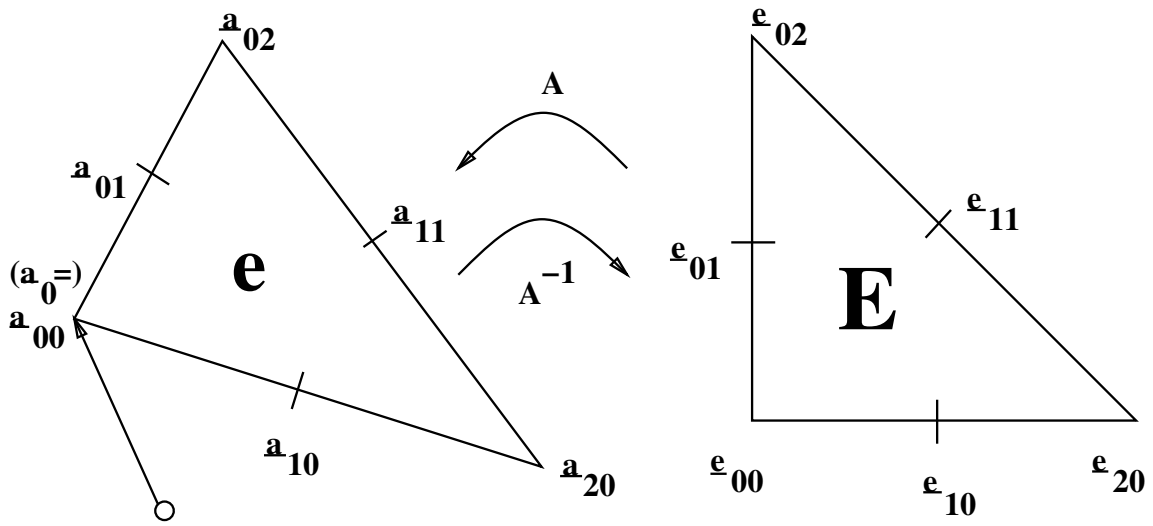


Abbildung 4.10: Stützstellen für Dreieck von Morley

Die Bezeichnungen seien wie in Abschnitt 4.3 bei $\mathcal{P}_2(\mathbb{R}^2)$ gewählt, d. h.

$$\begin{aligned} e \;=\;& \text{Dreieck mit Ecken } \underline{a}_{00}, \underline{a}_{20}, \underline{a}_{02} \\ & \text{und Seitenmittelpunkten } \underline{a}_{10}, \underline{a}_{11}, \underline{a}_{01}\ . \end{aligned}$$

Der zugehörige endlich–dimensionale Polynom–Raum ist gegeben durch

$$M_e = \mathcal{P}_2(\mathbb{R}^2) = \left[1, x_1, x_2, x_1x_2, x_1^2, x_2^2\right]\ .$$

Mit den Normalenableitungen in Richtung der (inneren) Normalen an den Seitenmittelpunkten lautet die *Interpolationsaufgabe in e:*

$$
\begin{aligned}
u \in \mathcal{P}_2(\mathbb{R}^2) \quad : \quad u(\underline{a}_s) &= y_s \ , \quad s = (0,0),(2,0),(0,2) \ , \\
\frac{\partial u}{\partial \underline{n}_s}(\underline{a}_s) &= y'_s \ , \quad s = (1,0),(1,1),(0,1) \ .
\end{aligned}
$$

Die *zugehörige Interpolationsaufgabe in E* ist dann gegeben durch:

$$
\begin{aligned}
U \in \mathcal{P}_2(\mathbb{R}^2) \quad : \quad U(\underline{e}_s) &= \eta_s \ , \quad s = (0,0),(2,0),(0,2) \ , \\
\frac{\partial U}{\partial \underline{n}_s}(\underline{e}_s) &= \eta'_s \ , \quad s = (1,0),(1,1),(0,1) \ .
\end{aligned}
$$

Die Normalenableitungen sind hierbei einfach wie folgt gegeben,

$$
\frac{\partial U}{\partial \underline{n}_s} = \underline{n}_s \nabla U \ , \quad \text{mit} \quad \underline{n}_{10} = (0,1)^\top \ , \ \underline{n}_{11} = -2^{-1/2}(1,1)^\top \ , \ \underline{n}_{01} = (1,0)^\top \ .
$$

Wir überprüfen nun die Eigenschaft 4.9. Sei $P \in \mathcal{P}_2(\mathbb{R}^2)$ ein Polynom, das die homogene Aufgabe löst. Wir schreiben

$$
P(z_1, z_2) = \alpha_{00} + \alpha_{10} z_1 + \alpha_{01} z_2 + \alpha_{11} z_1 z_2 + \alpha_{20} z_1^2 + \alpha_{02} z_2^2
$$

und bestimmen die Koeffizienten. Für die Eckpunkte erhält man

$$
\begin{aligned}
(0,0) \quad &: \quad \alpha_{00} = 0 \ , \\
(2,0) \quad &: \quad \alpha_{00} + \alpha_{10} + \alpha_{20} = 0 \ , \\
(0,2) \quad &: \quad \alpha_{00} + \alpha_{01} + \alpha_{02} = 0 \ .
\end{aligned}
$$

Man hat

$$
\begin{aligned}
\frac{\partial P}{\partial z_1} &= \alpha_{10} + \alpha_{11} z_2 + 2\alpha_{20} z_1 \\
\frac{\partial P}{\partial z_2} &= \alpha_{01} + \alpha_{11} z_1 + 2\alpha_{02} z_2
\end{aligned}
$$

und erhält für die Seitenmittelpunkte

$$
\begin{aligned}
(1,0) : 0 &= \frac{\partial P}{\partial \underline{n}_{10}} \left(\frac{1}{2}, 0 \right) = \frac{\partial P}{\partial z_2} \left(\frac{1}{2}, 0 \right) = \alpha_{01} + \frac{1}{2} \alpha_{11} \\
(0,1) : 0 &= \frac{\partial P}{\partial \underline{n}_{01}} \left(0, \frac{1}{2} \right) = \frac{\partial P}{\partial z_1} \left(0, \frac{1}{2} \right) = \alpha_{10} + \frac{1}{2} \alpha_{11} \\
(1,1) : 0 &= \frac{\partial P}{\partial \underline{n}_{11}} \left(\frac{1}{2} \ , \ \frac{1}{2} \right) = \frac{-1}{\sqrt{2}} \left(\frac{\partial P}{\partial z_1} + \frac{\partial P}{\partial z_2} \right) \left(\frac{1}{2} \ , \ \frac{1}{2} \right) \\
&= -\frac{1}{\sqrt{2}} \left[\left(\alpha_{10} + \frac{1}{2} \alpha_{11} + \alpha_{20} \right) + \left(\alpha_{01} + \frac{1}{2} \alpha_{11} + \alpha_{02} \right) \right] \ .
\end{aligned}
$$

Insgesamt ist folgendes Gleichungssystem zu lösen:

$$
\begin{array}{rcrcrcrcrcl}
\alpha_{00} & & & & & & & & & = & 0 \\
\alpha_{00} & + & \alpha_{10} & + & & & & \alpha_{20} & & = & 0 \\
\alpha_{00} & + & & \alpha_{01} & + & & & & \alpha_{02} & = & 0 \\
 & & & \alpha_{01} & + & \tfrac{1}{2}\alpha_{11} & & & & = & 0 \\
 & & \alpha_{10} & + & & \tfrac{1}{2}\alpha_{11} & & & & = & 0 \\
 & & \alpha_{10} & + & \alpha_{01} & + & \alpha_{11} & + & \alpha_{20} + \alpha_{02} & = & 0\,.
\end{array}
$$

Man muß also zeigen, daß die Matrix $\mathcal{A}$, definiert durch

$$
\mathcal{A} = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 1 & \frac{1}{2} & 0 & 0 \\
0 & 1 & 0 & \frac{1}{2} & 0 & 0 \\
0 & 1 & 1 & 1 & 1 & 1
\end{pmatrix}, \quad
\mathcal{A}^{-1} = \frac{1}{2}\begin{pmatrix}
2 & 0 & 0 & 0 & 0 & 0 \\
-2 & 1 & 1 & 0 & 2 & -1 \\
-2 & 1 & 1 & 2 & 0 & -1 \\
4 & -2 & -2 & 0 & 0 & 2 \\
0 & 1 & -1 & 0 & -2 & 1 \\
0 & -1 & 1 & -2 & 0 & 1
\end{pmatrix},
$$

regulär ist (z. B. mit Hilfe des Gaußschen Eliminationsverfahrens); deren Inverse ist gleich mitangegeben. Die Basisfunktionen haben also die folgende Gestalt:

$$
\begin{aligned}
P_{00}(z_1, z_2) &= 1 - z_1 - z_2 + 2z_1z_2\,, \\
P_{20}(z_1, z_2) &= \frac{1}{2}\left(z_1 + z_2 - 2z_1z_2 + z_1^2 - z_2^2\right)\,, \\
P_{02}(z_1, z_2) &= \frac{1}{2}\left(z_1 + z_2 - 2z_1z_2 - z_1^2 + z_2^2\right)\,, \\
P_{10}(z_1, z_2) &= z_2 - z_2^2\,, \quad P_{01}(z_1, z_2) \;=\; z_1 - z_1^2\,, \\
P_{11}(z_1, z_2) &= \frac{1}{\sqrt{2}}\left(z_1 + z_2 - 2z_1z_2 - z_1^2 - z_2^2\right)\,.
\end{aligned}
$$

Für die Lösung der zugehörigen Interpolationsaufgabe erhält man nach Satz 4.11

$$
u(\underline{x}) = \sum_{s\in\{(0,0),(2,0),(0,2)\}} y_s P_s(A^{-1}\underline{x}) + \sum_{s\in\{(1,0),(1,1),(0,1)\}} y'_s P_s(A^{-1}\underline{x});
$$

z. B. erhält man darin für $p_{00}(\underline{x}) = P_{00}(A^{-1}\underline{x})$:

$$
\begin{aligned}
p_{00}(\underline{x}) &= 1 - (b_{11}^{(-1)}x_1 + b_{12}^{(-1)}x_2 + c_1) - (b_{21}^{(-1)}x_1 + b_{22}^{(-1)}x_2 + c_2) \\
&\quad + 2(b_{11}^{(-1)}x_1 + b_{12}^{(-1)}x_2 + c_1)(b_{21}^{(-1)}x_1 + b_{22}^{(-1)}x_2 + c_2)
\end{aligned}
$$

□

Beispiel 4.14 „Rechteck von Adini“

Wir untersuchen wieder nur die Interpolationsaufgabe im Einheitsquadrat:

$$
\begin{aligned}
U \in M_E \ : \quad U(\underline{e}_s) &= \eta_s \ , \ 0 \le s \le (1,1) \ , \\
\frac{\partial U}{\partial z_1}(\underline{e}_s) &= \eta_s^1 \ , \ 0 \le s \le (1,1) \ , \\
\frac{\partial U}{\partial z_2}(\underline{e}_s) &= \eta_s^2 \ , \ 0 \le s \le (1,1) \ ,
\end{aligned}
$$

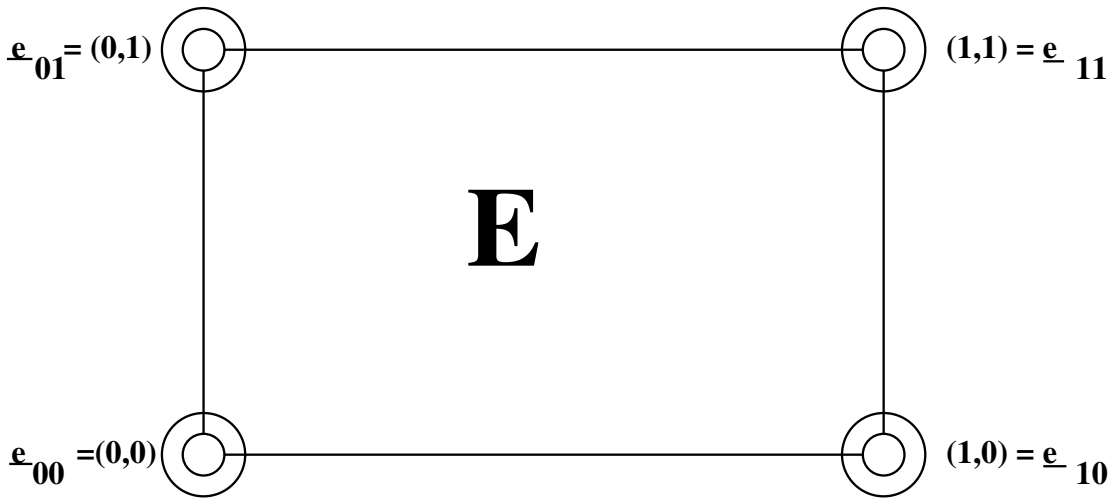


Abbildung 4.11: Stützstellen für Rechteck von Adini

wobei der endlichdimensionale Teilraum M_E von $\mathcal{P}(\mathbb{R}^2)$ gegeben ist durch

$$M_E = [1, z_1, z_2, z_1 z_2, z_1^2, z_2^2, z_1^2 z_2, z_1 z_2^2, z_1^3, z_2^3, z_1^3 z_2, z_1 z_2^3] \ .$$

Hier gilt also $\mathcal{P}_3 \subset M_E \subset \mathcal{Q}_3$; $\dim \mathcal{P}_3 = 10$, $\dim \mathcal{Q}_3 = 16$, $\dim M_E = 12$. Man hat 12 Gleichungen für U vorgeschrieben. Jede Funktion aus M_E läßt sich schreiben als

$$
\begin{aligned}
V(\underline{z}) \ = \ & \alpha_{00} + \alpha_{10} z_1 + \alpha_{20} z_1^2 + \alpha_{30} z_1^3 + \alpha_{01} z_2 + \alpha_{02} z_2^2 + \alpha_{03} z_2^3 + \\
& + \alpha_{11} z_1 z_2 + \alpha_{21} z_1^2 z_2 + \alpha_{12} z_1 z_2^2 + \alpha_{31} z_1^3 z_2 + \alpha_{13} z_1 z_2^3 \ .
\end{aligned}
$$

Zur Überprüfung der Eigenschaft 4.9 hat man also eine 12×12 Matrix zu invertieren. Man übersieht sofort, daß aufgrund der Eindeutigkeitsaussagen für Hermitesche Interpolationsaufgaben im 1–dimensionalen (vgl. Abschnitt 3.3, Beispiel 3.6) für die Lösung $P \in M_E$ der homogenen Aufgabe gilt

$$P(z_1, 0) = 0 \ , \quad 0 \le z_1 \le 1 \ , \ \text{und} \ \ P(0, z_2) = 0 \ , \ 0 \le z_2 \le 1 \ .$$

Z. B. erfüllt $u(z) = P(z,0) : \frac{d^s u}{dz^s}(\zeta_j) = 0$ ($\zeta_0 = 0$, $\zeta_1 = 1$, $s = 0, 1$, $r_0 = r_1 = 2$, $r = 4$); diese Aufgabe ist eindeutig lösbar in $\mathcal{P}_{r-1}(\mathbb{R}^1) = \mathcal{P}_3(\mathbb{R}^1)$. Also ist $\alpha_{00} = \alpha_{10} = \alpha_{20} = \alpha_{30} = \alpha_{01} = \alpha_{02} = \alpha_{03} = 0$. Es bleibt dann noch ein 5×5 Gleichungssystem zu lösen, um auch

$$\alpha_{11} = \alpha_{21} = \alpha_{12} = \alpha_{31} = \alpha_{13} = 0$$

zu zeigen. Man erhält die Gleichungen

$$\begin{aligned}
0 &= \frac{\partial P}{\partial z_1}(0,1) &&= \alpha_{11} z_2 + 2\alpha_{21} z_1 z_2 + \alpha_{12} z_2^2 + 3\alpha_{31} z_1^2 z_2 + \\
&&&\quad + \alpha_{13} z_2^3|_{(0,1)} = \alpha_{11} + \alpha_{12} + \alpha_{13} \ , \\
0 &= \frac{\partial P}{\partial z_2}(1,0) &&= \alpha_{11} z_1 + \alpha_{21} z_1^2 + 2\alpha_{12} z_1 z_2 + \alpha_{31} z_1^3 + \\
&&&\quad + 3\alpha_{13} z_1 z_2^2|_{(1,0)} = \alpha_{11} + \alpha_{21} + \alpha_{31} \ , \\
0 &= \frac{\partial P}{\partial z_1}(1,1) &&= \alpha_{11} + 2\alpha_{21} + \alpha_{12} + 3\alpha_{31} + \alpha_{13} \ , \\
0 &= \frac{\partial P}{\partial z_2}(1,1) &&= \alpha_{11} + \alpha_{21} + 2\alpha_{12} + \alpha_{31} + 3\alpha_{13} \ , \\
0 &= P(1,1) &&= \alpha_{11} + \alpha_{21} + \alpha_{12} + \alpha_{31} + \alpha_{13} \ .
\end{aligned}$$

Die eindeutige Lösbarkeit durch die triviale Lösung ist äquivalent zur Invertierbarkeit von

$$\mathcal{A} = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 2 & 1 & 3 & 1 \\ 1 & 1 & 2 & 1 & 3 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}, \quad \mathcal{A}^{-1} = \begin{pmatrix} 1 & 1 & 0 & 0 & -0.5 \\ -14 & 32 & -13 & 4 & -4.5 \\ 3 & -8 & 3 & -1 & 1.5 \\ 13 & -32 & 13 & -4 & 5 \\ -3 & 7 & -3 & 1 & -1 \end{pmatrix} .$$

Die zugehörige Inverse ist wieder gleich angegeben; damit ist Eigenschaft 4.9 erfüllt.

Die biorthonormale Basis erhält man entweder durch Invertierung der folgenden 12×12 Matrix:

	α_{00}	α_{10}	α_{20}	α_{30}	α_{01}	α_{02}	α_{03}	α_{11}	α_{21}	α_{12}	α_{31}	α_{13}
$P(0,0):$	1	0	0	0	0	0	0	0	0	0	0	0
$P(1,0):$	1	1	1	1	0	0	0	0	0	0	0	0
$P(0,1):$	1	0	0	0	1	1	1	0	0	0	0	0
$P_{z_1}(0,0):$	0	1	0	0	0	0	0	0	0	0	0	0
$P_{z_2}(0,0):$	0	0	0	0	1	0	0	0	0	0	0	0
$P_{z_1}(1,0):$	0	1	2	3	0	0	0	0	0	0	0	0
$P_{z_2}(0,1):$	0	0	0	0	1	2	3	0	0	0	0	0
$P_{z_1}(0,1):$	0	1	0	0	0	0	0	1	0	1	0	1
$P_{z_2}(1,0):$	0	0	0	0	1	0	0	1	1	0	1	0
$P_{z_1}(1,1):$	0	1	2	3	0	0	0	1	2	1	3	1
$P_{z_2}(1,1):$	0	0	0	0	1	2	3	1	1	2	1	3
$P(1,1):$	1	1	1	1	1	1	1	1	1	1	1	1

Die vierte bis elfte Zeile der obigen Matrix ergibt sich aus

$$\begin{aligned}\frac{\partial P}{\partial z_1} &= \alpha_{10} + 2\alpha_{20}z_1 + 3\alpha_{30}z_1^2 + \\ &\quad + \alpha_{11}z_2 + 2\alpha_{21}z_1z_2 + \alpha_{12}z_2^2 + 3\alpha_{31}z_1^2z_2 + \alpha_{13}z_2^3 \ , \\ \frac{\partial P}{\partial z_2} &= \alpha_{01} + 2\alpha_{02}z_2 + 3\alpha_{03}z_2^3 + \\ &\quad + \alpha_{11}z_1 + \alpha_{21}z_1^2 + 2\alpha_1 z_1z_2 + \alpha_{31}z_1^3 + 3\alpha_{13}z_1z_2^2 \ .\end{aligned}$$

Die biorthonormale Basis kann man sich auch durch Überlegungen verschaffen, die die Darstellung von Lösungen eindimensionaler Hermitescher Interpolationsaufgaben 2. Ordnung benutzt. Die Lösung von

$$\frac{d^s p}{dt^s}(t_j) = \lambda_{js} \ , \qquad j = 0,1 \ , \quad s = 0,1 \ ,$$

mit $t_0 = 0$, $t_1 = 1$ läßt sich in kanonischer Form durch die biorthonormale Basis darstellen (s. (3.56)):

$$\begin{array}{llllll}
p_{00}(t) &= (1+2t)(t-1)^2 & p'_{00}(t) &= 6t(t-1) & p_{00}(0) &= 1 \\
p_{10}(t) &= (1+2(1-t))t^2 & p'_{10}(t) &= 6t(t-1) & p_{10}(1) &= 1 \\
p_{01}(t) &= t(t-1)^2 & p'_{01}(t) &= (t-1)(3t-1) & p'_{01}(0) &= 1 \\
p_{11}(t) &= (t-1)t^2 & p'_{11}(t) &= t(3t-2) & p'_{11}(1) &= 1 \ .
\end{array}$$

Z. B. erhält man für $P_1 \in M_E$, das durch

$$P_1(\underline{e}_{00}) = 1\ ,\ P_1(\underline{e}_s) = 0 \text{ sonst}, \quad \frac{\partial P_1}{\partial z_i}(\underline{e}_s) = 0\ ,\ i = 1,2, \quad \forall s$$

definiert ist (vgl. Abb. 4.12), die Koeffizienten $\alpha_{00}, \alpha_{10}, \ldots, \alpha_{03}$ durch

$$\begin{aligned} P_1(z_1,0) &= p_{00}(z_1) = (1+2z_1)(1-z_1)^2 = 1 - 3z_1^2 + 2z_1^3 \\ P_1(0,z_2) &= p_{00}(z_2) = 1 - 3z_2^2 + 2z_2^3\ , \end{aligned}$$

also

$$\alpha_{00} = 1\ ,\ \alpha_{10} = \alpha_{01} = 0\ ,\ \alpha_{20} = \alpha_{02} = -3\ ,\ \alpha_{30} = \alpha_{03} = 2\ .$$

(Übrigens ist für alle anderen Basisfunktionen der Koeffizient $\alpha_{00} = 0$.) Für die anderen Koeffizienten von P_1 hat man die Gleichungen

$$\begin{aligned} 0 &= P_1(1,1) &&= \overbrace{1}^{\alpha_{00}} + \alpha_{11} + \alpha_{21} + \alpha_{12} + \alpha_{31} + \alpha_{13}\ , \\ 0 &= \frac{\partial P_1}{\partial z_1}(0,1) &&= \overbrace{0}^{\alpha_{10}} + \alpha_{11} + \alpha_{12} + \alpha_{13}\ , \\ 0 &= \frac{\partial P_1}{\partial z_2}(1,0) &&= \overbrace{0}^{\alpha_{01}} + \alpha_{11} + \alpha_{21} + \alpha_{31}\ , \end{aligned}$$

$$\begin{aligned} 0 &= \frac{\partial P_1}{\partial z_1}(1,1) &&= \overbrace{0 - 6 + 6}^{0} + \alpha_{11} + 2\alpha_{21} + \alpha_{12} + 3\alpha_{31} + \alpha_{13}\ , \\ 0 &= \frac{\partial P_1}{\partial z_2}(1,1) &&= 0 + \alpha_{11} + \alpha_{21} + 2\alpha_{12} + \alpha_{31} + 3\alpha_{13}\ . \end{aligned}$$

Man muß also wieder die obige 5×5 Matrix invertieren (bzw. mit dem Gaußschen Eliminationsverfahren auf Dreiecksgestalt bringen), wobei als rechte Seite jetzt $(0,0,0,0,-1)$ zu nehmen ist. Ähnlich kann man die anderen 11 Basisfunktionen ermitteln.

□

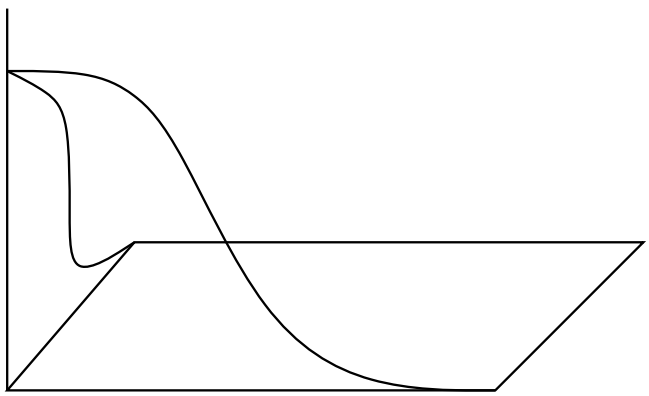


Abbildung 4.12: Basisfunktion P_1 (Rechteck von Adini)

Kapitel 5

Interpolationsabschätzungen

Ziel dieses Kapitels ist die Herleitung von Interpolationsabschätzungen in einer und mehreren örtlichen Veränderlichen. Insbesondere interessiert die Abhängigkeit des Interpolationsfehlers von Potenzen des Diskretisierungsparameters, der z. B. die maximale Gitterweite beschreibt. Die Abschätzungen für die Hermiteschen Interpolationsaufgaben ergeben sich als Anwendung von Approximationsfehlerabschätzungen für die verallgemeinerte Interpolationsaufgabe. Als weitere Anwendung können Approximationen von Linearformen untersucht werden, z. B. Differenzenapproximationen oder Quadraturformeln.

5.1 Interpolatorische Approximation und Restglieder

In diesem Abschnitt untersuchen wir den Fehler bei der Approximation mit Hilfe der Lösung der verallgemeinerten Interpolationsaufgabe in endlichdimensionalen Räumen. Konkret werden in den folgenden Abschnitten die hermiteschen interpolatorischen Approximationen von hinreichend glatten Funktionen betrachtet. Durch einfache, funktionalanalytische Begriffsbildungen gelingt es, grundlegende Fehlerabschätzungen sowie hinreichende Bedingungen dafür herzuleiten.

Die *interpolatorische Approximation* in einem normierten Raum E läßt sich wie folgt erklären. Sei ein endlichdimensionaler Teilraum $M \subset E$ und ein System von stetigen Linearformen $q_1, \ldots, q_l$ auf E als Knotenfunktionale gegeben. Im folgenden wird stets vorausgesetzt, daß die zugehörige Interpolationsaufgabe (3.16) für jede rechte Seite eindeutig lösbar, das System $\{q_1, \ldots, q_l\}$ also M–unisolvent ist. Für jedes Element u aus E existiert dann ein eindeutig bestimmtes Element $p \in M$ mit der Eigenschaft

$$q_j(p) = q_j(u) \,, \qquad j = 1, \ldots, l \,. \tag{5.1}$$

Die *interpolatorische Approximation* $p \in M$ und das gegebene Element u, aufgefaßt als lineare Funktionen auf E', haben die gleichen Knotenwerte an den Knotenpunkten $q_1, \ldots, q_l$. Dementsprechend hat der zugehörige Fehler $r := u - p$ *Nullstellen* an den Knotenpunkten, das heißt,

$$q_j(r) = 0 \ , \qquad j = 1, \ldots, l \ . \tag{5.2}$$

Der Fehler $r = u - p$ liegt folglich im *Orthogonalraum* $L^\perp := \{v \in E : q(v) = 0 \ \forall q \in L\} \subset E$ des Teilraums $L = [q_1, \ldots, q_l] \subset E'$. Jedes Element u aus E gestattet jetzt die Zerlegung

$$u = p + r, \ , \qquad p \in M \ , \ r \in L^\perp \ . \tag{5.3}$$

Die Komponenten p, r in dieser Darstellung sind eindeutig durch u bestimmt. Ist nämlich

$$u = p + r = p' + r' \ , \qquad p, p' \in M \ , \ r, r' \in L^\perp \ ,$$

so folgt auf Grund der Eindeutigkeitsvoraussetzung

$$p - p' = r' - r \in M \cap L^\perp = \{0\} \ , \tag{5.4}$$

also $p = p'$ und $r = r'$. Aus den obigen Ausführungen ergibt sich, daß die interpolatorische Approximation *exakt* ist für Elemente u aus dem Teilraum M, denn in diesem Falle ist offenbar $p = u$.

Die Darstellung (5.3) zeigt, daß der Vektorraum E die *direkte Summe* von M und $L^\perp$ ist,

$$E = M \oplus L^\perp \ . \tag{5.5}$$

In Satz 3.1 haben wir bewiesen, daß der Teilraum M eine zu $\{q_1, \ldots, q_l\}$ biorthonormale Basis $\{p_1, \ldots, p_l\}$ besitzt. Hiermit werden zwei lineare Abbildungen P, R von E in die Teilräume M und $L^\perp$ definiert durch die Vorschrift

$$Pu = \sum_{k=1}^{l} q_k(u) p_k \ , \quad Ru = u - \sum_{k=1}^{l} q_k(u) p_k \ , \qquad u \in E \ . \tag{5.6}$$

Offensichtlich ist

$$u = Pu + Ru \ , \quad Pu \in M \ , \qquad Ru \in L^\perp \ , \ u \in E \ . \tag{5.7}$$

Aus der Eindeutigkeit der Zerlegung (5.3) folgt dann $p = Pu\ , r = Ru$. Die interpolatorische Approximation läßt sich also durch den linearen *Interpolationsoperator* P und der zugehörige Fehlerdurch den linearen *Restgliedoperator* R aus (5.6) darstellen. Die Basissysteme $\{p_1, \ldots, p_l\}$, $\{q_1, \ldots, q_l\}$ sind biorthonormal. Hieraus erhält man sofort, daß P und $R = I - P$ *Projektionsoperatoren* sind, das heißt

$$P^2 = P\ , \quad R^2 = R\ , \quad \text{und es gilt} \quad PR = RP = 0\ . \tag{5.8}$$

Beweis von (5.8): $P^2u = \sum^l q_k(Pu)p_k = \sum q_k(u)p_k$, da $q_j(Pu) = \sum q_k(u)q_j(p_k)$ $= q_j(u)$, und

$$PRu = Pu - P^2u = 0\ , \quad RPu = Pu - P^2u = 0\ .$$

□

Wir bezeichnen mit $N(P)$, $N(R)$ die *Nullräume* und mit $P(E)$, $R(E)$ die *Bildräume* der Operatoren P, R. Wie man leicht nachprüft, gilt dann

$$P(E) = N(R) = M\ , \quad N(P) = R(E) = L^\perp\ . \tag{5.9}$$

Beweis von (5.9): Wegen (5.7) gilt für $v \in N(P)$, daß $v = Pv + Rv = Rv$, d. h. $N(P) \subset R(E)$. Für $v \in R(E)$ gibt es ein $w \in E$ mit $v = Rw$; wegen (5.8) folgt $Pv = PRw = 0$, d. h. $v \in N(P)$ und $N(P) = R(E)$. Weiter ist $R(E) = L^\perp$, denn $R(E) \subset L^\perp$, und $v \in L^\perp$ erfüllt nach (5.6) $Rv = v$, d. h. $L^\perp \subset R(E)$.

Offenbar ist $P(E) = M$ und $v \in P(E)$ erfüllt $Rv = RPu = 0$ für ein $u \in E$, d. h. $P(E) \subset N(R)$. Umgekehrt ist $v = Pv$ für $v \in N(R)$, d. h. $v \in P(E)$. □

Wir wollen jetzt Fehlerabschätzungen für die verallgemeinerte Interpolation von Elementen u aus E herleiten. Die Abbildungen P, R sind stetige lineare Operatoren in E, wie man aus der Darstellung (5.6) unmittelbar sieht. Ein wichtiges Hilfsmittel beim Studium der Fehlerabschätzungen ist der *kürzeste Abstand*

$$|u, M| := \inf_{w \in M} \|u - w\| \tag{5.10}$$

von Elementen u zum Teilraum M in E. Wie man leicht zeigt, wird durch $|., M|$ eine Halbnorm auf E definiert. Mit Hilfe dieser Halbnorm wird bekanntlich die Norm des *Quotientenraums* E/M erklärt. Die Bedeutung der Halbnorm (5.10) für Fehlerabschätzungen des Restglieds $Ru = u - Pu$ ergibt sich aus folgendem

Lemma 5.1 *Es gilt die zweiseitige Restgliedabschätzung*

$$|u, M| \leq \|Ru\| \leq \|R\| \, |u, M| \, , \qquad u \in E \, , \tag{5.11}$$

mit der natürlichen Norm $\|R\| := \sup_{0 \neq v \in E} \frac{\|Rv\|}{\|v\|}$ *des Operators* R *in* E.

Beweis: Zunächst ist

$$\|Ru\| = \|u - Pu\| \geq \inf_{w \in M} \|u - w\| = |u, M| \, .$$

Jedes Element $w \in M$ liegt im Nullraum von R. Daher gilt die Abschätzung

$$\|Ru\| = \|R(u - w)\| \leq \|R\| \, \|u - w\|$$

für alle $u \in E$ und alle $w \in M$. Die linke Seite hängt nicht von w ab. Folglich wird auch

$$\|Ru\| \leq \|R\| \inf_{w \in M} \|u - w\| = \|R\| \, |u, M| \, .$$

□

Es sei hier bemerkt, daß $\|R\| < \infty$, d. h. R ist ein beschränkter Operator. Dies folgt aus der Abschätzung

$$\|Ru\| \leq \|u - Pu\| \leq \|u\| + \|Pu\|$$

und der Tatsache, daß P wegen der Stetigkeit der q_k , $k = 1, \ldots, l$, selbst beschränkt ist.

In den Anwendungen interssiert man sich etwas allgemeiner für Abschätzungen des Fehlers $Ru = u - Pu$ der verallgemeinerten Interpolation in dafür geeigneten Halbnormen $|.|_F$ auf E. Wir setzen voraus, daß die Halbnorm $|.|_F$ stetig ist auf E, also $|u|_F \leq \beta \|u\|_E$ ist für alle $u \in E$ mit einer geeigneten Konstanten β. Mit Hilfe der obigen Ungleichung (5.11) läßt sich der Fehler dann stets abschätzen in der Form

$$|u - Pu|_F = |Ru|_F \leq \rho_1 |u, M| \, , \qquad u \in E \, , \tag{5.12}$$

mit der Konstanten $\rho_1 = \beta \|R\|$. In Anwendungen auf konkrete Beispiele arbeitet man im allgemeinen allerdings nicht mit der Halbnorm $|., M|$, sondern stattdessen mit einer passenden, *äquivalenten Halbnorm* $|.|_E$. Eine solche Halbnorm genügt einer Ungleichung der Gestalt

$$\gamma_0 |u|_E \leq |u, M| \leq \gamma_1 |u|_E \, , \qquad u \in E \, , \tag{5.13}$$

mit positiven Konstanten γ_0, γ_1. Der endlichdimensionale Teilraum M von E ist abgeschlossen. Daher folgt aus (5.13) sofort die Bedingung

$$|u|_E = 0 \Longleftrightarrow u \in M \ . \tag{5.14}$$

Beweis von (5.14): „$\Longleftarrow$“: $u \in M \Longrightarrow |u, M| = 0 \overset{(5.13)}{\Longrightarrow} |u|_E = 0$.

„$\Longrightarrow$“: $|u, M| = 0 \Longrightarrow \exists v_n \in M : \|u - v_n\| \longrightarrow 0 \Longrightarrow u \in \overline{M} = M$.

□

Beziehung (5.14) bedeutet, daß die Halbnorm $|.|_E$ den Nullraum M haben muß. Mit $Pu \in M$ ist dann notwendig $|Pu|_E = 0$ und folglich

$$|Ru|_E = |u - Pu|_E = |u|_E \ , \qquad u \in E \ . \tag{5.15}$$

Gemäß (5.9) hat jedes Element $v \in L^\perp$ die Darstellung $v = Ru$. Aus den Ungleichungen (5.11), (5.13) und mit Hilfe der Beziehung $v = Rv$ für $v \in L^\perp$ gewinnt man leicht die zweiseitige Abschätzung

$$\frac{1}{\gamma_1 \|R\|} \|v\| \leq |v|_E \leq \frac{1}{\gamma_0} \|v\| \ , \qquad v \in L^\perp \ . \tag{5.16}$$

Beweis von (5.16): Für beliebiges $u \in E$ hat man

$$\gamma_0 |u|_E \overset{(5.13)}{\leq} |u, M| \overset{(5.11)}{\leq} \|Ru\| \overset{(5.11)}{\leq} \|R\| \, |u, M| \overset{(5.13)}{\leq} \gamma_1 \|R\| \, |u|_E \ .$$

Für $u = v \in L^\perp$ ist $v = Rv$, womit (5.16) gezeigt ist. □

Aufgrund von (5.16) sind also die beiden Normen $|.|_E$ und $\|.\|$ auf dem Teilraum $L^\perp$ äquivalent.

Das wichtigste Ergebnis der obigen Überlegungen fassen wir im folgenden Satz zusammen.

Satz 5.2 *Sei $|.|_F$ eine beliebige stetige Halbnorm auf E und sei $|.|_E$ eine beliebige Halbnorm, die äquivalent zur Halbnorm $|., M|$ auf E ist. Dann genügt die interpolatorische Approximation der Fehlerabschätzung*

$$|u - Pu|_F = |Ru|_F \leq \rho |u|_E \ , \qquad u \in E \ , \tag{5.17}$$

mit einer positiven Konstanten ρ. Die kleinste Konstante $\rho = \rho_0$ hat die Darstellung

$$\rho_0 := \sup_{0 \neq u \in E} \frac{|Ru|_F}{|u|_E} = \sup_{0 \neq v \in L^\perp} \frac{|v|_F}{|v|_E} \,. \tag{5.18}$$

Beweis: (i) Nach (5.12) und (5.13) ist $|Ru|_F \leq \rho_1 |u, M| \leq \rho_1 \gamma_1 |u|_E$, wobei $\rho_1 = \beta \|R\|$.

(ii) ρ_0 ist kleinste Schranke, denn andernfalls $\exists \rho' < \rho_0$ mit $\frac{|u-Pu|_F}{|u|_E} \leq \rho' \;\; \forall 0 \neq u \in E$. Dies steht im Widerspruch zur Definition von ρ_0 als Supremum. Weiter ist

$$|u - Pu|_F / |u|_E = |Ru|_F / |u|_E \overset{(5.15)}{=} |Ru|_F / |Ru|_E \leq \sup_{0 \neq v \in L^\perp} \frac{|v|_F}{|v|_E} \quad \forall 0 \neq u \in E \,,$$

und

$$\frac{|v|_F}{|v|_E} = \frac{|Rv|_F}{|v|_E} \leq \sup_{0 \neq u \in E} |Ru|_F / |u|_E \quad \forall 0 \neq v \in L^\perp \,,$$

was die zweite Gleichheit in (5.18) beweist. □

Der obige Satz läßt sich übrigens unmittelbar auf *interpolatorische Approximationen q stetiger Linearformen f* auf E anwenden, die durch

$$q(u) := f(Pu) \,, \qquad u \in E \,, \tag{5.19}$$

definiert sind. Als Anwendung von Satz 5.3 erhält man das folgende Ergebnis.

Satz 5.3 *Sei $|.|_E$ eine beliebige zu $|., M|$ äquivalente Halbnorm auf E. Für eine stetige Linearform f auf E wird durch*

$$|u|_F := |f(u)| \,, \qquad u \in E \,, \tag{5.20}$$

eine stetige Halbnorm auf E erklärt. Für die durch (5.19) definierte interpolatorische Approximation q von f gilt die Fehlerabschätzung

$$|f(u) - q(u)| \leq \rho |u|_E \,, \qquad u \in E \,, \tag{5.21}$$

mit der besten Konstanten

$$\rho_0 = \sup_{0 \neq v \in L^\perp} \frac{|f(v)|}{|v|_E}$$

Beweis: Die Dreiecksungleichung für $|.|_F$ folgt aus der Linearität von f,

$$|u+v|_F = |f(u)+f(v)| \le |f(u)| + |f(v)| \; ;$$

die Stetigkeit der Halbnorm folgt aus der von f. Fehlerabschätzung (5.21) ist eine Anwendung von (5.17),

$$|f(u)-q(u)| = |f(u-Pu)| = |u-Pu|_F \le \rho |u|_E \; , \qquad u \in E \; .$$

Die optimale Fehlerkonstante in (5.21) ergibt sich aus (5.18). □

Die Approximation q von f ist offensichtlich *exakt* auf Elementen u aus dem Teilraum M, das heißt

$$q(u) = f(u) \Longleftrightarrow s(u) = 0 \; , \qquad u \in M \; ,$$

wobei $s := f - q$ den *Fehler* dieser Approximation bezeichnet.

Für Funktionale $f \in E'$, die auf M verschwinden, wird in (5.21) $q(u) = f(Pu) = 0$, und man erhält das folgende Ergebnis, was im Rahmen konkreter, stückweise polynomialer Approximationen in Sobolewschen Funktionenräumen auch als *„Lemma von Bramble–Hilbert“* bezeichnet wird.

Satz 5.4 *Sei $|.|_E$ eine zu $|.,M|$ äquivalente Halbnorm auf E. Dann gilt die Äquivalenz*

$$f \in E' \, , \; f|M = 0 \Longleftrightarrow |f(u)| \le \rho |u|_E \; , \qquad u \in E \; .$$

Beweis: „ $\Longrightarrow$ “ : Diese Richtung folgt sofort aus (5.21) wegen $q(u) = 0$.

„ $\Longleftarrow$ “ Umgekehrt ist $|u|_E = 0$, $u \in M$, da $|.|_E$ wegen der Äquivalenz mit $|.,M|$ den Nullraum M besitzt. Nach Voraussetzung hat man daher

$$|f(u)| \le \rho |u|_E = 0 \; , \qquad u \in M \; .$$

Die Beschränktheit von f folgt aus der Äquivalenz von $|.|_E$ zu $|.,M|$,

$$|u|_E \le \frac{1}{\gamma_0} |u,M| \le \frac{1}{\gamma_0} \|u\| \; , \qquad u \in E \; .$$

□

Zum Abschluß dieser Überlegungen sollen jetzt noch zwei Sätze bewiesen werden, die man in Anwendungen auf Funktionenräume zum Nachweis der grundlegenden Äquivalenzbedingung (5.13) der Halbnormen $|.|_E$ und $|., M|$ benötigt.

Sei $|.|_E$ eine beliebige Halbnorm auf E mit dem Nullraum M. Dann wird durch

$$|||u||| := \sum_{k=1}^{l} |q_k(u)| + |u|_E \,, \qquad u \in E \,, \tag{5.22}$$

eine Norm auf E definiert. Auf Grund der M–Unisolvenz von $\{q_1, \ldots, q_l\}$ ist nämlich

$$|||u||| = 0 \Longleftrightarrow u \in M \cap L^{\perp} = \{0\} \,,$$

d. h. die Definitheitsbedingung für die Norm $|||.|||$ ist gesichert. Hiermit kann man das folgende erste Kriterium aussprechen.

Satz 5.5 *Eine Halbnorm $|.|_E$ mit dem Nullraum M ist dann und nur dann äquivalent zur Halbnorm $|., M|$, wenn die zugehörige Norm $|||.|||$ äquivalent zur Norm $\|.\|$ von E ist.*

Beweis: (i) Wenn die Norm (5.22) äquivalent zur Norm von E ist, dann existieren positive Konstanten α_0, α_1 derart, daß die Ungleichungen

$$\alpha_0 \|u\| \le \sum_{k=1}^{l} |q_k(u)| + |u|_E \le \alpha_1 \|u\| \,, \qquad u \in E \,,$$

gelten. Setzt man jetzt $Ru = u - Pu$ anstelle von u, so wird $|Ru|_E = |u|_E, q_k(Ru) = 0$ und folglich

$$\alpha_0 \|Ru\| \le |u|_E \le \alpha_1 \|Ru\| \,, \qquad u \in E \,.$$

Hieraus folgt mit Hilfe von Lemma 5.1, daß $|.|_E$ äquivalent zur Halbnorm $|., M|$ auf E ist,

$$|u|_E \le \alpha_1 \|Ru\| \overset{(5.11)}{\le} \alpha_1 \|R\| \, |u, M| \overset{(5.11)}{\le} \alpha_1 \|R\| \, \|Ru\| \le \alpha_1 \|R\| \frac{1}{\alpha_0} |u|_E \,.$$

(ii) Sei jetzt umgekehrt die Bedingung (5.13) erfüllt. Die Linearformen q_k sind stetig auf E und es ist $|u, M| \le \|u\|_E$ (da $0 \in M$), so daß

$$|||u||| \le \sum_{k=1}^{l} \|q_k\| \, \|u\|_E + \frac{1}{\gamma_0} \|u\|_E = \alpha_1 \|u\|_E \,, \qquad u \in E \,,$$

wird mit der Konstanten $\alpha_1 = \sum_{k=1}^{l} \|q_k\| + 1/\gamma_0$. Weiter folgt mit (5.6), (5.7), (5.11) die Abschätzung

$$\|u\| \leq \|Pu\| + \|Ru\| \leq \sum_{k=1}^{l} |q_k(u)| \, \|p_k\| + \|R\| \gamma_1 |u|_E \, ,$$

also

$$\alpha_0 \|u\| \leq |||u||| \, , \qquad u \in E \, ,$$

mit der Konstanten

$$\alpha_0 = \{\max(\|p_k\|, \, \gamma_1 \|R\|)\}^{-1} \, .$$

□

Das folgende Äquivalenzkriterium für Halbnormen werden wir in Abschnitt 5.3 für mehrdimensionale interpolatorische Approximationen anwenden.

Eine stetige Halbnorm $|.|_E$ nennen wir *koerzitiv* auf dem normierten Raum E, wenn die beiden folgenden Bedingungen gelten: (i) E ist Teilraum eines normierten Raumes D und die natürliche Einbettung von E in D ist vollstetig[1)].

(ii) Es gibt eine positive Zahl η derart, daß die *Koerzitivitätsungleichung*

$$\|u\|_E \leq \eta(|u|_E + \|u\|_D) \, , \qquad u \in E \, , \tag{5.23}$$

gilt.

Da ein linearer kompakter Operator notwendigerweise beschränkt ist, reicht es in (i) aus, die Kompaktheit der Einbettungen zu fordern. Das zweite Kriterium lautet nun:

Satz 5.6 *Sei E ein Banachscher Raum und die stetige Halbnorm $|.|_E$ koerzitiv auf E mit Nullraum M. Dann ist die Norm $|||.|||$ in (5.22) äquivalent zur Norm von E und die Halbnorm $|.|_E$ äquivalent zu $|., M|$.*

Beweis: Die Halbnorm $|.|_E$ und die Linearformen q_k sind stetig auf E nach Voraussetzung, so daß $|||u||| \leq \alpha_1 \|u\|_E$ für $u \in E$ ist,

[1)] vollstetig = kompakt und stetig; Einbettung $J : E \to D$ *kompakt*, d. h. $\forall u_n \in E : \|u_n\|_E \leq c \, , \; n \in \mathbb{N} \exists \mathbb{N}' \subset \mathbb{N} , \; v \in D : \|u_n - v\|_D \to 0 \, (n \to \infty)$; Einbettung $J : E \to D$ *stetig*, d. h. $\|u\|_D \leq \gamma \|u\|_E$

$$|||u||| \leq \sum_k \|q_k\| \, \|u\|_E + |u|_E \leq \sum_k \|q_k\| \, \|u\|_E + \beta \|u\|_E = \alpha_1 \|u\|_E \, .$$

Die entsprechende Abschätzung nach unten wird indirekt bewiesen. Nehmen wir also an, daß eine Folge $u_t \in E \, , \; t \in \mathbb{N}$, existiert mit den Eigenschaften

$$\|u_t\|_E = 1, |||u_t||| = \sum_{k=1}^{l} |q_k(u_t)| + |u_t|_E \longrightarrow 0$$

für $t \to \infty$. Nach Voraussetzung ist diese Folge kompakt in D, so daß eine in D konvergente Teilfolge $u_t \, , \; t \in \mathbb{N}' \subset \mathbb{N}$, existiert. Da $|u_t|_E \to 0$ strebt, erhält man aus der Koerzitivitätsungleichung

$$\|u_s - u_t\|_E \leq \eta \, (|u_s|_E + |u_t|_E + \|u_s - u_t\|_D) \longrightarrow 0$$

für $s, t \to \infty \, , \; s, t \in \mathbb{N}'$. Also ist diese Teilfolge Cauchy–konvergent in E. Nach Voraussetzung ist E vollständig, daher konvergiert diese Teilfolge gegen einen Limes $u \in E$. Dieser Limes hat die Eigenschaft

$$|u|_E = \lim_{t \in \mathbb{N}'} |u_t|_E = 0 \; , \quad q_k(u) = \lim_{t \in \mathbb{N}} q_k(u_t) = 0 \; .$$

Die erste Beziehung gilt wegen

$$| \, |u|_E - |u_t|_E| \leq |u - u_t|_E \leq \beta \|u - u_t\|_E \longrightarrow 0 \quad (t \to \infty \, , \; t \in \mathbb{N}') \; ;$$

die zweite folgt aus der Stetigkeit der q_k. Somit ist $u \in M \cap L^{\perp} = \{0\}$, das heißt, $u = 0$ im Widerspruch zu

$$\|u\|_E = \lim_{t \in \mathbb{N}'} \|u_t\|_E = 1 \; .$$

Folglich ist die Norm $|||.|||$ äquivalent zur Norm $\|.\|$ von E. Die Äquivalenz der Halbnormen $|.|_E$ und $|., M|$ schließlich ergibt sich hiermit aus Satz 5.5. □

5.2 Approximation durch Polynome in einer Veränderlichen

Ziel dieses Abschnitts ist die Herleitung von Interpolationsabschätzungen für die hermitesche interpolatorische Approximation in einer Ortsdimension durch Anwendung der Ergebnisse des letzten Abschnitts. Wesentlich ist hierfür, die Äquivalenz der

Halbnormen $|.|_{r,p}$ auf $C^r[a,b]$ mit dem kürzesten Abstand $|.,\mathcal{P}_{r-1}|_{r,p}$ zu zeigen, sowie $\mathcal{P}_{r-1}[a,b]$ als Nullraum von $|.|_{r,p}$ zu identifizieren. Der Einfachheit halber betrachten wir hier anstelle der Sobolewschen Räume $H^{r,p}(a,b)$ den Teilraum $C^r[a,b]$ mit den Sobolewschen Normen. Da $C^r[a,b]$ dicht in $H^{r,p}(a,b)$ ist, lassen sich die zentralen Ungleichungen und Normäquivalenzen stetig auf $H^{r,p}(a,b)$ fortsetzen.

Zur Wiederholung sei noch einmal die Aufgabenstellung bzw. die Definition der *hermiteschen interpolatorischen Approximation* genannt,

$$p \in \mathcal{P}_{r-1} : \frac{d^s p}{dx^s}(x_j) = \frac{d^s u}{dx^s}(x_j) \;, \quad s = 0, \ldots, r_j - 1 \;, \;\; j = 0, \ldots, m \;, \tag{5.24}$$

für $u \in C^r[a,b]$, wobei $x_0, \ldots, x_m \in \mathbb{K}$ paarweise verschieden und $r = r_0 + \cdots + r_m$.

Wir betrachten zunächst ein beliebiges beschränktes Intervall $[a,b] \subset \mathbb{R}$. Es bezeichnet $C^r[a,b]$ den Vektorraum der r–mal stetig differenzierbaren Funktionen auf $[a,b]$. Für jedes $s = 0, \ldots, r$ werden durch

$$\begin{aligned} |u|_{s,\infty} &= \max_{x \in [a,b]} \left| \frac{d^s u}{dx^s}(x) \right| \\ |u|_{s,p} &= \left(\int_a^b \left| \frac{d^s u}{dx^s}(x) \right|^p dx \right)^{1/p} \;, \qquad 1 \le p < \infty \;, \end{aligned} \tag{5.25}$$

Halbnormen auf dem Raum $C^r[a,b]$ erklärt. Hiermit gewinnt man zugehörige Normen durch

$$\begin{aligned} \|u\|_{r,\infty} &= \max_{s=0,\ldots,r} |u|_{s,\infty} \\ \|u\|_{r,p} &= \left(\sum_{s=0}^{r} |u|_{s,p}^p \right)^{1/p} \;, \qquad 1 \le p < \infty \;, \end{aligned} \tag{5.26}$$

für $u \in C^r[a,b]$. Die Konkretisierung der allgemeinen Situation des letzten Abschnitts wird sich hier wie folgt ergeben,

$$E = C^r[a,b] \;, \;\; \|.\|_E = \|.\|_{r,p} \;, \;\; |.|_E = |.|_{r,p} \;, \;\; |.|_F = |.|_{s,q} \;.$$

Die hermitesche Interpolationsaufgabe verwendet Diracsche δ–Funktionalen und ihre Ableitungen als Knotenfunktionalen. Sei

$$\delta_c^s(u) = \frac{d^s u}{dx^s}(c) \;, \qquad u \in C^r[a,b] \;, \tag{5.27}$$

für $s = 0, \ldots, r$ und beliebige Stellen $c \in [a,b]$. Der folgende Satz zeigt, daß diese Linearformen für $s \le r-1$ in jeder Norm $\|.\|_{r,p}$ auf $C^r[a,b]$ stetig sind. Hier und im folgenden ist $1/p = 0$ für $p = \infty$.

Satz 5.7 *Es gelten die Ungleichungen*

$$|u(c)| \le \varepsilon |u|_{1,p} + \varepsilon^{-p'/p} |u|_{0,p} , \qquad u \in C^1[a,b] , \tag{5.28}$$

für jede Stelle $c \in [a,b]$, jedes p in $1 \le p \le \infty$, $1/p + 1/p' = 1$ und jedes ε in $0 < \varepsilon \le \left(\frac{b-a}{2}\right)^{1/p'}$. Insbesondere sind für jedes $s = 0, \ldots, r-1$ die Linearformen δ_c^s beschränkt durch

$$|\delta_c^s(u)| = \left| \frac{d^s u}{dx^s}(c) \right| \le \alpha \|u\|_{r,p} , \qquad u \in C^r[a,b] , \tag{5.29}$$

mit einer geeigneten Konstanten α, die von p und $(b-a)$ abhängt.

Beweis: Der Punkt c liegt in der linken Hälfte oder in der rechten Hälfte des Intervalls $[a,b]$. Sei h beliebig gewählt in $0 < h \le (b-a)/2$ im ersten Fall $\left(c \in \left[a, \frac{a+b}{2}\right]\right)$ oder in $-(b-a)/2 \le h < 0$ im zweiten Fall $\left(c \in \left(\frac{a+b}{2}, b\right]\right)$. Offenbar ist

$$\begin{aligned} u(c) &= -\int_c^{c+h} \frac{d}{dx} \left\{ \frac{c+h-x}{h} u(x) \right\} dx = \\ &= \frac{1}{h} \int_c^{c+h} u(x)\, dx - \int_c^{c+h} \frac{c+h-x}{h} u'(x)\, dx . \end{aligned}$$

Für $p = \infty$ erhält man damit die Ungleichung

$$|u(c)| \le \frac{1}{2}|h|\, |u|_{1,\infty} + |u|_{0,\infty} ,$$

denn

$$\left| \int_c^{c+h} \frac{c+h-x}{h} u' dx \right| \le |u|_{1,\infty} \Bigg| \underbrace{\int_c^{c+h} \frac{c+h-x}{h}\, dx}_{-\frac{1}{2|h|}(c+h-x)^2 \Big|_c^{c+h} = \frac{1}{2}|h|} \Bigg| .$$

Für $1 \le p < \infty$ läßt sich das erste Integral mit Hilfe der Hölderschen Ungleichung abschätzen durch (wegen $\frac{1}{p'} - 1 = -\frac{1}{p}$)

$$\left| \frac{1}{h} \int_c^{c+h} u(x)\, dx \right| \le |h|^{-\frac{1}{p}} |u|_{0,p}$$

und das zweite Integral durch (wegen $c+h-x \le h\,,\ x \in [c, c+h]$, und $\frac{1}{p'} = 1 - \frac{1}{p}$)

$$\left| \int_c^{c+h} \frac{c+h-x}{h} u'(x)\, dx \right| \le |h|^{1-\frac{1}{p}} |u|_{1,p}\ .$$

Also folgt

$$|u(c)| \le |h|^{1-\frac{1}{p}} |u|_{1,p} + |h|^{-\frac{1}{p}} |u|_{0,p}\ .$$

Dies ist die oben behauptete Ungleichung (5.28) mit $\varepsilon = |h|^{1/p'}$. Wendet man diese Ungleichung auf $\frac{d^s u}{dx^s}$ anstelle von u an und setzt $\varepsilon = \left(\frac{b-a}{2}\right)^{1/p'}$, so folgt die zweite Ungleichung (5.29) mit der Konstanten

$$\alpha = \left(\frac{b-a}{2}\right)^{1/p'} + \left(\frac{b-a}{2}\right)^{-1/p}\ ,$$

denn man erhält

$$\begin{aligned} |u^{(s)}(c)| &\le \varepsilon |u|_{s+1,p} + \varepsilon^{-p'/p} |u|_{s,p} \\ &\le \underbrace{(\varepsilon + \varepsilon^{-p'/p})}_{\le \alpha} \max(|u|_{s+1,p}, |u|_{s,p}) \ \le\ \alpha \left(\sum_{s=0}^{r} |u|_{s,p}^p \right)^{1/p} \end{aligned}$$

im Fall $1 \le p < \infty$, und, mit $\alpha = 1 + \frac{b-a}{2}$ im Fall $p = \infty$,

$$|u^{(s)}(c)| \le \alpha \|u\|_{r,\infty}\ .$$

□

Die *Stetigkeit der Halbnormen* $|.|_{s,q}$ auf $C^r[a,b]$ erhält man sofort aus der obigen Ungleichung (5.29),

$$|u|_{s,q} = \left(\int_a^b |u^{(s)}(x)|^q\, dx \right)^{1/q} \le (b-a)^{1/q} \alpha \|u\|_{r,p} \tag{5.30}$$

für $s = 0, \ldots, r-1$ und jedes q in $1 \le q \le \infty$. Um eine entsprechende Ungleichung für $s = r$ herzuleiten, erhält man aus der Hölderschen Ungleichung zunächst

$$\begin{aligned} \int_a^b |u(x)|^q\, dx &\le \left(\int_a^b 1\, dx \right)^{1-\frac{q}{p}} \left(\int_a^b |u(x)|^{q\frac{p}{q}}\, dx \right)^{q/p} \\ &= (b-a)^{\frac{p-q}{p}} \left(\int_a^b |u(x)|^p\, dx \right)^{q/p} \end{aligned}$$

für $1 \le q \le p \le \infty$. Einsetzen von $u^{(r)}$ anstelle von u liefert

$$|u|_{r,q} \le (b-a)^{\frac{p-q}{pq}} |u|_{r,p} \le (b-a)^{\frac{p-q}{pq}} \|u\|_{r,p} , \qquad 1 \le q \le p . \tag{5.31}$$

Als letzte Vorbereitung vor dem Nachweis von Äquivalenzen von Halbnormen wird der *Nullraum* von $|.|_{r,p}$ bestimmt. Alle Funktionen $u \in C^r[a,b]$ und ihre Ableitungen $u^{(j)} = \frac{d^j u}{dx^j}$ gestatten die Taylorentwicklung

$$u^{(j)}(x) = \sum_{k=0}^{m} \frac{1}{k!}(x-c)^k u^{(j+k)}(c) + R_m(x; u^{(j)}) , \qquad u \in C^r[a,b] , \tag{5.32}$$

an einer beliebigen Stelle $c \in [a,b]$ mit den Restgliedern

$$R_m\left(x; u^{(j)}\right) = \frac{1}{m!} \int_c^x u^{(r)}(y)(x-y)^m \, dy \tag{5.33}$$

für $m = r-1-j$, $j = 0, \ldots, r-1$. Dies sieht man mit Hilfe der bekannten Restglieddarstellung (vgl. z. B. [SH82], § 3)

$$R_m(x; u^{(j)}) = \frac{1}{m!}(x-c)^{m+1} \int_0^1 u^{(j+m+1)}(c + t(x-c))(1-t)^m \, dt$$

und der Variablensubstitution $y = c + t(x-c)$, $r = j+m+1$ sowie $(x-c) \times (1-t) = x-y$.

Die Höldersche Ungleichung ergibt die Restgliedabschätzung

$$|R_m(x; u^{(j)})| \le \frac{(b-a)^m}{m!} (b-a)^{1/p'} |u|_{r,p} , \qquad x \in [a,b] ; \tag{5.34}$$

daraus folgt weiter, daß

$$\begin{aligned} \|R_m(.; u^{(j)})\|_{0,p} &\le \frac{(b-a)^{m+1-\frac{1}{p}}}{m!} |u|_{r,p} \left(\int_a^b 1 \, dx \right)^{1/p} \\ &\le \frac{1}{m!}(b-a)^{m+1} |u|_{r,p} , \quad u \in C^r[a,b],\ 1 \le p \le \infty . \end{aligned} \tag{5.35}$$

Von zentraler Bedeutung für die Herleitung von Fehlerabschätzungen ist die Tatsache, daß der Nullraum der Halbnormen $|.|_{r,p}$ auf $C^r[a,b]$ aus den Polynomen $(r-1)$–ten Grades auf $[a,b]$ besteht,

$$u \in C^r[a,b] , \ |u|_{r,p} = 0 \Longleftrightarrow u \in \mathcal{P}_{r-1}[a,b] , \tag{5.36}$$

wie man aus (5.32), (5.34) für $j = 0$ (und $m = r - 1$) unmittelbar abliest.

Wir wollen jetzt zeigen, daß diese Halbnormen äquivalent sind zur Halbnorm $|., \mathcal{P}_{r-1}|_{r,p}$. Dieses Ergebnis wird sich aus Satz 5.6 ergeben. Zu diesem Zweck führen wir die Norm

$$|||u|||_{r,p} = \sum_{k=0}^{r-1} \left| \frac{d^k u}{dx^k}(c) \right| + |u|_{r,p} , \qquad u \in C^r[a,b] , \tag{5.37}$$

für einen beliebigen Punkt $c \in [a,b]$ ein.

Satz 5.8 *Die Norm (5.37) ist äquivalent zur Norm $\|.\|_{r,p}$ auf $C^r[a,b]$. Daher ist $|.|_{r,p}$ äquivalent zur Halbnorm $|., \mathcal{P}_{r-1}|_{r,p}$, das heißt, mit zwei positiven Konstanten β_0, β, gelten die Ungleichungen*

$$\beta_0 |u|_{r,p} \le |u, \mathcal{P}_{r-1}|_{r,p} \le \beta_1 |u|_{r,p} , \qquad u \in C^r[a,b] . \tag{5.38}$$

Beweis: Mit Hilfe der Ungleichungen (5.29) läßt sich die Norm in (5.37) nach oben abschätzen durch

$$|||u|||_{r,p} \le (\alpha r + 1) \|u\|_{r,p} , \qquad u \in C^r[a,b] .$$

Die entsprechende Abschätzung nach unten gewinnt man mit Hilfe der Taylorformel (5.32). Daraus und aus (5.33) folgt nämlich (wegen $m + 1 = r - j \le r \ \ \forall j$)

$$\left| u^{(j)}(x) \right| \le \beta \left(\sum_{k=0}^{r-1} \left| u^{(k)}(c) \right| + |u|_{r,p} \right) = \beta |||u|||_{r,p}$$

für $j = 0, \ldots, r-1$ mit der Konstanten

$$\beta = \max_{k=0,\ldots,r} \frac{(b-a)^k}{k!} \max(1, (b-a)^{1/p'}) .$$

Für $p = \infty$ ist somit (mit $\tilde{\beta} = \max(1, \beta)$)

$$|u|_{s,\infty} \le \tilde{\beta} |||u|||_{r,\infty} , \quad \|u\|_{r,\infty} \le \tilde{\beta} |||u|||_{r,\infty} ,$$

für $s = 0, \ldots, r-1$. Für $1 \le p < \infty$ folgt unmittelbar

$$|u|_{s,p} \le \beta (b-a)^{1/p} |||u|||_{r,p} , \qquad s = 0, \ldots, r-1 ,$$

also

$$\|u\|_{r,p} \le (r+1)^{1/p} \max_{s=0,\dots,r} |u|_{s,p} \le (r+1)^{1/p}(1+\beta(b-a)^{1/p})|||u|||_{r,p} \ .$$

Hiermit ist die behauptete Äquivalenz der Normen $\|.\|_{r,p}$ und $|||.|||_{r,p}$ für $C^r[a,b]$ bewiesen. Durch Anwendung von Satz 5.6 folgt daraus die Ungleichung (5.38). □

Wir kehren zurück zur hermiteschen interpolatorischen Approximation, deren Lösung $p = Pu$ sich (kanonisch) darstellen läßt als

$$Pu = \sum_{j=0}^{m} \sum_{s=0}^{r_j-1} \frac{d^s u}{dx^s}(x_j) w_{js} \ , \qquad u \in C^r[a,b] \ , \tag{5.39}$$

wobei w_{js} die zu den Knotenfunktionalen

$$q_{js}(u) = \delta_j^s(u) = \frac{d^s u}{dx^s}(x_j) \ , \qquad u \in C^r[a,b] \ , \tag{5.40}$$

biorthonormalen Polynome in $\mathcal{P}_{r-1}$ sind und P den im Sinne von Abschnitt 5.1 definierten Projektionsoperator bezeichnet.

Der Fehler $Ru = u - Pu$ dieser Approximationen läßt sich mit Hilfe von Satz 5.2 abschätzen. Dazu setzen wir

$$E = C^r[a,b] \ , \quad M = \mathcal{P}_{r-1}[a,b] \ ,$$
$$\|u\|_E = \|u\|_{r,p} \ , \quad |u|_E = |u|_{r,p} \ , \quad |u|_F = |u|_{s,q} \ ,$$

für s, p, q wie in den Ungleichungen (5.30), (5.31) beschrieben, d. h. $1 \le p,\ q \le \infty$ für $s = 0, \dots, r-1$ und $1 \le q \le p \le \infty$ für $s = r$. Die Fehlerabschätzung (5.17) nimmt dann die folgende Gestalt an,

$$|u - Pu|_{s,q} = |Ru|_{s,q} \le \rho |u|_{r,p} \ , \qquad u \in C^r[a,b] \ , \tag{5.41}$$

mit einer positiven Konstanten ρ. Die kleinste Konstante $\rho = \rho_0$ hat die Darstellung

$$\rho_0 = \sup_{0 \ne v \in L^\perp} \frac{|v|_{s,q}}{|v|_{r,p}} \tag{5.42}$$

mit einem Teilraum $L^\perp$ aller Funktionen aus $C^r[a,b]$, die an Knotenpunkten Null sind, das heißt,

$$L^\perp = \left\{ v \in C^r[a,b] \Big| \frac{d^s v}{dx^s}(x_j) = 0\,, \quad s = 0, \ldots, r_j - 1\,, \quad j = 0, \ldots, m \right\}\,. \tag{5.43}$$

Die obige Abschätzung zeigt, daß der Fehler bzw. dessen Ableitungen durch die r–te Ableitung von u abgeschätzt wird. Daher ist die Approximation exakt für Polynome aus $\mathcal{P}_{r-1}$. Die typische Abhängigkeit der Fehlerschranke von der *Intervallänge* $h = b - a$ läßt sich schließlich mit Hilfe der Darstellung (5.42) der besten Konstanten herleiten. Dabei verwendet man die *Technik der Variablentransformation* auf das Einheitsintervall $[0,1]$. Das Intervall $[a,b]$ ist das Bild des Einheitsintervalls unter der *affinen Transformation*

$$x = Az \Longleftrightarrow x = a + hz\,, \qquad 0 \le z \le 1\,. \tag{5.44}$$

Die Gitterpunkte $x_0, \ldots, x_m \in [a,b]$ sind die Bilder der Punkte

$$z_j = \frac{x_j - a}{b - a} \in [0,1]\,, \qquad j = 0, \ldots, m\,. \tag{5.45}$$

Der folgende Satz zeigt die Abhängigkeit der besten Konstanten ρ_0 in (5.42) von der Intervallänge $h = b - a$. Die Konstante σ in (5.46), (5.47) hängt nur von $r, s, p, q, z_0, \ldots, z_m$ ab, ist also konstant für alle Intervalle $[a,b]$ und Gitterpunkte $x_0, \ldots, x_m$, die unter der Transformation A das gleiche Urbild haben.

Satz 5.9 *Die beste Konstante $\rho = \rho_0$ in (5.41), (5.42) hat die Darstellung*

$$\rho_0 = \sigma h^{r-s+\frac{1}{q}-\frac{1}{p}} \tag{5.46}$$

mit der Konstanten

$$\sigma := \sup_{0 \neq y \in \hat{L}^\perp} \frac{|y|_{s,q}}{|y|_{r,p}} \tag{5.47}$$

und dem Teilraum

$$\hat{L}^\perp = \left\{ y \in C^r[0,1] \Big| \frac{d^s y}{dz^s}(z_j) = 0\,, \quad s = 0, \ldots, r_j - 1\,, \quad j = 0, \ldots, m \right\}\,.$$

Beweis: Jede Funktion $v \in C^r[a,b]$ läßt sich mit Hilfe der affinen Transformation A in (5.44) durch eine Funktion $y \in C^r[0,1]$ in der Form

$$v(x) = v(a + hz) = y(z)\,, \quad x = a + hz\,, \quad 0 \le z \le 1\,,$$

schreiben. Nach der Kettenregel ist dabei $y'(z) = hv'(x)$. Die höheren Ableitungen von y haben die Darstellung

$$\frac{d^j y}{dz^j}(z) = h^j \frac{d^j v}{dx^j}(x) \ , \qquad j = 0, \ldots, r \ .$$

Eine Funktion v liegt genau dann im Teilraum $L^\perp$, wenn die zugehörige Funktion $y = v \circ A$ zum Teilraum $\hat{L}^\perp$ von $C^r[0,1]$ gehört. Die Halbnormen $|v|_{s,q}$ gehen bei dieser Transformation über in

$$\begin{aligned} |v|_{s,\infty} &= \max_{x\in[a,b]} \left|\frac{d^s v}{dx^s}(x)\right| \ = \ h^{-s} \max_{z\in[0,1]} \left|\frac{d^s y}{dz^s}(z)\right| \ = \ h^{-s}|y|_{s,\infty} \ , \\ |v|_{s,q} &= \left(\int_a^b \left|\frac{d^s v}{dx^s}(x)\right|^q \ dx\right)^{1/q} \ = \ h^{-s+\frac{1}{q}} \left(\int_0^1 \left|\frac{d^s y}{dz^s}(z)\right|^q \ dz\right)^{1/q} = \\ &= h^{-s+\frac{1}{q}}|y|_{s,q} \ . \end{aligned}$$

Daher ist

$$\frac{|v|_{s,q}}{|v|_{r,p}} = \frac{h^{-s+\frac{1}{q}}|y|_{s,q}}{h^{-r+\frac{1}{p}}|y|_{r,p}} \le \sigma h^{r-s+\frac{1}{q}-\frac{1}{p}} \ , \quad 0 \ne v \in L^\perp \ .$$

Die Behauptungen des Satzes folgen somit unmittelbar aus (5.41), (5.42). □

Betrachten wir speziell die *Taylor–Interpolation* , d. h. $j = 0$, $s = 0, r = m+1$, dann folgt aus (5.41), (5.46), (5.47) für $p = q$, daß

$$\|u - Pu\|_{0,p} = \|R_m(.,u)\|_{0,p} \le \sigma h^{m+1}|u|_{m+1,p} \ ,$$

wobei

$$\sigma = \sup_{0\ne y\in\hat{L}^\perp} \frac{\|y\|_{0,p}}{|y|_{m+1,p}} \le \frac{1}{m!} \ . \tag{5.48}$$

Beweis von (5.48): Es ist

$$\hat{L}^\perp = \left\{ y \in C^{m+1}[0,1] \Big| \frac{d^s y}{dz^s}(z_0) = 0 \ , \quad s = 0, \ldots, m \right\} \ ;$$

$$\overset{\text{(Taylor)}}{\Longrightarrow} y(z) = \sum_{k=0}^{m} \frac{1}{k!}(z-z_0)^k \underbrace{y^{(k)}(z_0)}_{=0,\, y\in\hat{L}^\perp} + \frac{1}{m!}\int_{z_0}^{z} y^{(m+1)}(\zeta)\underbrace{(z-\zeta)^m}_{\le 1}\, d\zeta$$

$$\overset{\text{(Hölder)}}{\Longrightarrow} |y(z)| \le \frac{1}{m!}\left(\int_0^1 \left|y^{(m+1)}(\zeta)\right|^p d\zeta\right)^{1/p}$$

$$\Longrightarrow \quad \|y\|_{0,p} \le \|y\|_{0,\infty} \;\le\; \frac{1}{m!}|y|_{m+1,p}, \;\; y \in \hat{L}^\perp \; .$$

□

Wir haben so Abschätzung (5.35) für $r = m+1$ wieder erhalten.

Wir wollen nun das Ergebnis von Satz 5.9 auf Räume stückweise polynomialer Funktionen anwenden, die, wie in den Kapiteln 1 und 2 ausgeführt, die zugrundeliegenden endlichdimensionalen Räume für die Ritzschen Projektionsmethoden darstellen. Hierzu sei eine Unterteilung $\Delta : a = x_0 < \cdots < x_n = b$ mit Gitterpunkten in $[a,b]$ gegeben, $I_j = [x_{j-1},\, x_j]\,,\; h_j = x_j - x_{j-1}\,,\; j = 1,\ldots,n$, und die stückweise polynomialen Räume sind durch

$$\begin{aligned}
\mathcal{P}_r(\Delta) &= \left\{u | \forall j \exists u^j \in \mathcal{P}_r(\mathbb{K}) : u|I_j \;=\; u^j|I_j\right\}\,, \\
\mathcal{P}_r^m(\Delta) &= \{u \in \mathcal{P}_r(\Delta) | g_{js}(u) \;=\; 0\,, \quad j = 1,\ldots,n-1\,, \\
&\qquad\qquad\qquad\qquad\qquad\qquad s = 0,\ldots,m-1\}
\end{aligned}$$

definiert, wobei

$$g_{js}(u) = \frac{1}{2}\left(\frac{d^s u}{dx^s}(x_{j+0}) - \frac{d^s u}{dx^s}(x_{j-0})\right) \quad (s \le r)$$

die *halben Sprunghöhen* der möglicherweise unstetigen Ableitungen an den Gitterpunkten bezeichnet. Es sei noch $h = \max_{1\le j\le n} h_j$ die *maximale Gitterweite.*

Beispiel 5.10 Lineare Splineapproximationen

Die Aufgabenstellung besteht darin, für $u \in C^2[a,b]$ ein $v \in \mathcal{P}_1^1(\Delta)$ zu bestimmen, so daß $v(x_j) = u(x_j)\,,\; j = 0,\ldots,n;\;\; v = Pu$ ist die stetige, stückweise lineare interpolatorische Approximation von u. Für jedes Teilintervall I_j liegt die Hermitesche Interpolationsaufgabe (5.24) mit $m = 1\,,\; r_0 = r_1 = 1\,, r = 2\,,\; M = \mathcal{P}_1(I_j)$ vor, und Satz 5.9 liefert die Abschätzung

$$|u - Pu|_{s,q;\,I_j} \le \sigma h_j^{2-s+\frac{1}{q}-\frac{1}{p}} |u|_{2,p;\,I_j} \tag{5.49}$$

für $u \in C^2(I_j)\,,\; s = 0,1\,,\; 1 \le p,q \le \infty$. Für die Halbnormen auf den einzelnen Teilintervallen verwenden wir die Bezeichnungen

$$|v|_{m,p;I_j} := \left(\int_{I_j} |v^{(m)}(x)|^p \, dx \right)^{1/p} ; \tag{5.50}$$

wir lassen das Teilintervall weg, wenn ganz $[a,b]$ gemeint ist.

Aus (5.49) ergeben sich nun sofort die folgenden *Interpolationsabschätzungen* für lineare Splineapproximationen.

Satz 5.11 *Für $u \in C^2[a,b]$, $Pu \in \mathcal{P}_1^1(\Delta)$ gelten die Abschätzungen*

$$|u - Pu|_{s,q} \le \sigma h^{2-s+\frac{1}{q}-\frac{1}{p}} |u|_{2,p} , \quad s = 0,1 , \ 1 \le p \le q \le \infty , \tag{5.51}$$

wobei noch $\sigma \le 1$ im Fall $s = 0$ ist.

Beweis: Es ist $[a,b] = \bigcup_{j=1}^n I_j$, und aus (5.49) folgt

$$\begin{aligned} |u - Pu|_{s,q}^q &= \sum_{j=1}^n |u - Pu|_{s,q;I_j}^q \\ &\le \sigma^q h^{q(2-s+\frac{1}{q}-\frac{1}{p})} \sum_{j=1}^n |u|_{2,p;I_j}^q . \end{aligned}$$

Die Höldersche Ungleichung für Summen mit $\alpha_j = |u|_{2,p,I_j}^q$, $t = p/q$, $\frac{1}{t'} = 1 - \frac{1}{t} = \frac{p-q}{p}$ liefert

$$\begin{aligned} \sum_{j=1}^n |u|_{2,p;I_j}^q &\le \left(\sum_j \alpha_j^t \right)^{1/t} \left(\sum_j 1 \right)^{1/t'} \\ &\le |u|_{2,p}^q \, n^{\frac{p-q}{p}} , \qquad 1 \le p,q < \infty . \end{aligned}$$

Im Fall $p \le q < \infty$ ist offenbar $n^{\frac{p-q}{p}} \le 1$, was (5.51) für diesen Fall beweist. Ist $q = \infty$ und $1 \le p \le \infty$, dann folgt aus (5.49)

$$\begin{aligned} |u - Pu|_{s,\infty} &= \max_{1 \le j \le n} |u - Pu|_{s,\infty;I_j} \\ &\le \sigma \max_j \left\{ h_j^{2-s-\frac{1}{p}} |u|_{2,p;I_j} \right\} \\ &\le \sigma h^{2-s-\frac{1}{p}} |u|_{2,p} , \quad s = 0,1 . \end{aligned}$$

Zum Nachweis von $\sigma \le 1$ im Fall $s = 0$ benutzen wir die Darstellung der Lösung des 2–Punkt–Randwertproblems

$$y'' = f \text{ in } [0,1] , \quad y(0) = y(1) = 0 ,$$

mit Hilfe der Greenschen Funktion

$$G(z,t) = \begin{cases} t(z-1)\ , 0 \le t \le z \le 1\ , \\ z(t-1)\ , 0 \le z \le t \le 1\ , \end{cases}$$

d. h. $y(z) = \int_0^1 G(z,t) f(t) dt$ (vgl. [Wal76, § 26], oder Anhang A.6, Beispiel A.23). Der in der Darstellung von σ (vgl. (5.47)) vorkommende Orthogonalraum hat hier die spezielle Gestalt

$$\hat{L}^\perp = \left\{ y \in C^2[0,1] \,|\, y(0) = y(1) = 0 \right\}\ .$$

Daher folgt, mit $f = y''$ und wegen $G(z,t) \le 1$, daß

$$|y(z)| \le |f|_{0,1} \le |f|_{0,p} = |y|_{2,p} \quad \forall z \in [0,1]\ ,\ 1 \le p \le \infty\ ,$$

und

$$|y|_{0,q} \le |y|_{0,\infty} \le |y|_{2,p}\ , \qquad 1 \le p, q \le \infty\ ,\ y \in \hat{L}^\perp\ .$$

Dies beweist im Falle $s = 0$, daß

$$\sigma = \sup_{0 \ne y \in \hat{L}^\perp} \frac{|y|_{0,q}}{|y|_{2,p}} \le 1\ .$$

□

Wir bemerken noch, daß man für $s = 0$ und verschiedene Kombinationen von p, q noch kleinere Schranken für σ ausrechnen kann. (vgl. Übungsaufgabe 25). □

Beispiel 5.12 Lagrange–Interpolation in $\mathcal{P}_2^1(\Delta)$ (vgl. Abschnitt 3.4.1)

Die Aufgabenstellung besteht darin, zu $u \in C^3[a,b]$ ein $v(= Pu) \in \mathcal{P}_2^1(\Delta)$ zu bestimmen mit

$$v(x_j) = u(x_j)\ , \qquad j = 0, \tfrac{1}{2}, 1, \ldots, n - \tfrac{1}{2}, n\ .$$

Es liegen also $2n + 1$ Interpolationspunkte vor, wobei die Intervallmittelpunkte mit $x_{j+1/2} = \frac{1}{2}(x_j + x_{j+1})$, $j = 0, \ldots, n-1$, bezeichnet werden. Die zugehörige (Hermite–)Interpolation in jedem Teilintervall I_j in der Form (5.24) liegt vor, wenn man setzt $m = 2$, $r_0 = r_1 = r_2 = 1$, $r = 3$, $M = \mathcal{P}_2(I_j)$. Satz 5.9 liefert

$$|u - Pu|_{s,q;I_j} \le \sigma h_j^{3-s+\frac{1}{p}-\frac{1}{q}} |u|_{3,p;I_j} \ , \quad s = 0,1,2 \ , \ 1 \le p,q \le \infty \ .$$

Wie in Beispiel 5.10 folgen daraus die *Interpolationsabschätzungen*

$$|u - Pu|_{s,q} \le \sigma h^{3-s+\frac{1}{p}-\frac{1}{q}} |u|_{3,p} \ , \qquad s = 0,1,2 \ , \tag{5.52}$$
$$1 \le p \le q \le \infty \ , \ u \in C^3[a,b] \ ,$$

wobei
$$\hat{L}^\perp = \left\{ y \in C^3[0,1] \middle| y(0) = y(1/2) = y(1) = 0 \right\}$$
$$\sigma = \sup_{0 \ne y \in \hat{L}^\perp} \frac{|y|_{s,q}}{|y|_{3,p}} \ .$$

□

Beispiel 5.13 Quadratische Splineapproximation (vgl. Abschnitt 3.4.2)

Zu $u \in C^3[a,b]$ sucht man eine quadratische Splinefunktion $u = Pu \in \mathcal{P}_2^2(\Delta)$ mit

$$v'(x_0) = u'(x_0) \ , \quad v(x_j) = u(x_j) \ , \quad j = 0, \ldots, n \ .$$

Die Lösung läßt sich auf jedem Teilintervall I_j explizit angeben in der Form

$$v(x) = y_j z + y_{j-1}(1-z) + z(1-z)h_j \left(y'_{j-1} - \frac{1}{h_j}(y_j - y_{j-1}) \right) \ , \tag{5.53}$$
$$x \in I_j \ ,$$

wobei $y_\nu = u(x_\nu)$, $y'_0 = u'(x_0)$, $y'_\nu = \frac{2}{h_\nu}(y_\nu - y_{\nu-1}) - y'_{\nu-1}$, und $z = \frac{1}{h_j}(x - x_{j-1})$. Die Einschränkungen $v = v^j = v|I_j$ aus $\mathcal{P}_2(I_j)$ lösen die Hermitesche Interpolationsaufgabe

$$v(x_{j-1}) = u(x_{j-1}) \ , \ v'(x_{j-1}) = y'_{j-1} \ , \ v(x_j) = u(x_j)$$

auf jedem Teilintervall I_j , $j = 1, \ldots, n$.

Wir betrachten also zunächst die *Hermitesche Interpolationsaufgabe* auf I_j mit $m = 1$, $r_0 = 2$, $r_1 = 1$, $r = 3$, $M = \mathcal{P}_2(I_j)$, d. h.

$$\hat{v} = \hat{v}^j \in \mathcal{P}_2(I_j) : \hat{v}(x_\nu) = u(x_\nu), \ \nu = j-1, j, \ \hat{v}'(x_{j-1}) = u'(x_{j-1}) \ .$$

Dafür hat man die Abschätzungen

$$|u-\hat{v}|_{s,q;I_j} \le \sigma h_j^{3-s+\frac{1}{q}-\frac{1}{p}} |u|_{3,p,I_j} \ , \qquad s=0,1,2, \ \ 1 \le p,q \le \infty, \ \ u \in C^3(I_j) \ .$$

Bevor wir die zugehörigen Interpolationsabschätzungen angeben, klären wir die Frage, ob y'_ν in (5.53) mit $u'(x_\nu)$ übereinstimmt. Die *Differenzen* $\hat{e}|I_j = v|I_j - \hat{v}^j$ erfüllen die Aufgaben

$$\hat{e}(x_{j-1}) = \hat{e}(x_j) = 0 \ , \ \hat{e}'(x_{j-1}) = y'_{j-1} - u'(x_{j-1}) \ , \quad j = 1, \dots, n \ .$$

Behauptung: $\hat{e} \equiv 0$ in I_j , für alle $j = 1, \dots, n$.

Beweis: Nach der obigen Darstellung (5.53) der Lösung dieser Interpolationsaufgabe muß gelten (mit $y_{j-1} = y_j = 0 \ , \ \hat{y}'_0 = 0$)

$$\hat{e}(x) = z(1-z)h_j \hat{y}'_{j-1} \quad \text{mit} \quad \hat{y}'_\nu = y'_\nu - u'(x_\nu) \ ,$$

wobei $\hat{y}'_0 = 0 \ , \ \hat{y}'_\nu + \hat{y}'_{\nu-1} = 0 \ (\Leftrightarrow \hat{y}'_\nu = -\hat{y}'_{\nu-1}) \ , \ \nu = 1, \dots, n$. Daher muß $\hat{y}'_\nu = 0 \ \ \forall \nu$, also $\hat{e}(x) = 0 \ \ \forall x \in I_j \ \ \forall j$ sein. □

Also ist $v = \hat{v}$, und man hat die *Interpolationsabschätzungen*

$$|u-v|_{s,q} \le \sigma h^{3-s+\frac{1}{q}-\frac{1}{p}} |u|_{3,p} \ , \quad s=0,1,2 \ , \ 1 \le p \le q \le \infty \ , \quad u \in C^3[a,b] \ . \tag{5.54}$$

Als weitere Anwendung der Interpolationsabschätzungen (5.41), (5.46) betrachten wir Fehlerabschätzungen für *interpolatorische Quadraturformeln* und *Differenzenquotienten*. Sei zuerst

$$f(u) = \int_a^b u(x) g(x) \, dx \ , \qquad u \in C^r[a,b] \ ,$$

mit einer *Gewichtsfunktion* $g(x) \ge 0$. Mit Hilfe der Hermite–Interpolation erhält man eine zugehörige *Quadraturformel* der Gestalt

$$q(u) = f(Pu) = \sum_{j=0}^{m} \sum_{s=0}^{r_j - 1} \beta_{js} \frac{d^s u}{dx^s}(x_j)$$

mit den Koeffizienten

$$\beta_{js} = \int_a^b w_{js}(x) g(x)\, dx\ , \qquad s = 0, \ldots, r_j - 1\ ,\ j = 0, \ldots, m\ .$$

Die Gewichtsfunktion g soll über (a, b) integrierbar sein. Damit gilt stets

$$|f(u)| \le \int_a^b g(x)\, dx \cdot \max_{x\in[a,b]} |u(x)| = \|g\|_{0,1}\, \|u\|_{0,\infty}\ . \tag{5.55}$$

Ist g beschränkt auf $[a, b]$, so hat man auch

$$|f(u)| \le \sup_{x\in[a,b]} g(x) \int_a^b |u(x)|\, dx = \|g\|_{0,\infty} \|u\|_{0,1}\ . \tag{5.56}$$

Mit Hilfe von (5.30), (5.31) ist also f stetig in den Normen $\|.\|_{r,p}$ für alle $r \ge 1$, $1 \le p \le \infty$. Der Fehler der Quadraturformel genügt dann der Fehlerabschätzung

$$|f(u) - q(u)| \le \sigma h^{r+\frac{1}{q}} \|g\|_{0,q'}\ |u|_{r,\infty}\ , \qquad u \in C^r[a, b]\ , \tag{5.57}$$

wobei wir $s = 0$, $p = \infty$, $q' = q/(q-1)$ gesetzt haben; es ist $q = \infty$ oder $q = 1$, je nachdem ob (5.55) oder (5.56) gilt. Im Falle von (5.55) hat man nämlich für die Konstante in (5.21) die Abschätzung

$$\frac{|f(v)|}{|v|_{r,\infty}} \le \|g\|_{0,1} \frac{\|v\|_{0,\infty}}{|v|_{r,\infty}} \le \sigma \|g\|_{0,1} h^r$$

(vgl. (5.41) und (5.46) mit $p = q = \infty$, $s = 0$), wobei

$$\sigma = \sup_{0\ne v\in \hat{L}^\perp} \frac{\|y\|_{0,\infty}}{|y|_{r,\infty}}\ ;$$

analog folgt die Abschätzung im Fall (5.56) (mit $p = \infty$, $q = 1$, $s = 0$).

Wird die *Quadraturformel* speziell *durch Lagrange–Interpolation* gewonnen, dann ist $r_j = 1$, $r = m + 1$ und mit $s = 0$, $q = 1$, $p = \infty$ erhält man

$$\left| \int_a^b u(x) g(x)\, dx - \sum_{j=0}^m \beta_{j0} u(x_j) \right| \le \sigma h^{m+2} \|g\|_{0,\infty} \max_{x\in[a,b]} \left| \frac{d^{m+1} u}{dx^{m+1}}(x) \right| \tag{5.58}$$

für alle $u \in C^{m+1}[a, b]$. In diesem Fall sind $w_{j0} = L_j$ die Lagrange–Polynome.

Für die *Gaußschen Quadraturformeln* (s. Abschnitt 3.3, Beispiel 3.6) zum Intervall $[a, b]$ ist $r_j = 2$, $r = 2(m + 1)$, $g(x) = 1$. Die zugehörige Fehlerabschätzung lautet daher

$$\left| \int_a^b u(x)\,dx - \sum_{j=0}^{m} \beta_{j0} u(x_j) \right| \le \sigma h^{2m+3} \max_{x\in[a,b]} \left| \frac{d^{2m+2}u}{dx^{2m+2}}(x) \right| \tag{5.59}$$

für alle $u \in C^{2m+2}[a,b]$.

Die *summierte Sehnentrapezformel* erhält man als interpolatorische Approximation durch lineare Splinefunktionen für eine äquidistante Unterteilung $x_k = a + kh\,,\ k = 0,\ldots,n$, des Intervalls $[a,b]$,

$$q(u) = f(Pu) = \sum_{k=0}^{n} \beta_k u(x_k)\,, \qquad v = Pu \in \mathcal{P}_1^1(\Delta)\,,$$

wobei $\beta_k = h/2\,,\ k = 0,n$, und $\beta_k = h\,,\ k = 1,\ldots,n-1$. In jedem Teilintervall approximiert man $f(u|I_j)$ nämlich durch

$$\begin{aligned} f(v^j) &= \left(\int_{I_j} p_{j-1}\,dx \right) u(x_{j-1}) + \left(\int_{I_j} p_j\,dx \right) u(x_j) \\ &= \frac{h}{2}(u(x_{j-1}) + u(x_j))\,, \qquad j = 1,\ldots,n\,, \end{aligned}$$

wobei $p_k\,,\ k = j-1, j$, die Dachfunktion bei x_k bezeichnet; Summation über j liefert die bekannte summierte Sehnentrapezformel. Die Interpolationsabschätzungen liefern für jedes I_j

$$|f(u|I_j) - f(v^j)| \le \sigma h^3 |u|_{2,\infty;I_j}\,, \qquad u \in C^2(I_j)\,,$$

und insgesamt (durch Summation und wegen $nh = b-a$)

$$|f(u) - q(u)| \le \sigma h^3 \sum_{j=1}^{n} |u|_{2,\infty;I_j} \le \sigma(b-a)h^2 |u|_{2,\infty}\,. \tag{5.60}$$

Mit Hilfe einer Abschätzung für die Greensche Funktion macht man sich noch klar, daß in (5.60) $\sigma \le 1/12$ gilt (vgl. Bem. zu Satz 5.11).

Differenzenquotienten zur Approximation einer Funktion u an einer Stelle $c \in [a,b]$ erhält man durch die entsprechende Ableitung der Interpolation $p = Pu$. Sei also

$$f(u) = \frac{d^k u}{dx^k}(c)\,, \qquad u \in C^r[a,b]\,,$$

für ein $k = 0,\ldots,r$. Die zugehörige *Differenzenapproximation* lautet dann

$$q(u) = f(Pu) = \sum_{j=0}^{m} \sum_{s=0}^{r_j - 1} \gamma_{js}^{(k)} \frac{d^s u}{dx^s}(x_j)$$

mit den Koeffizienten

$$\gamma_{js}^{(k)} = \frac{d^k w_{js}}{dx^k}(c) \,, \qquad s = 0, \ldots, r_j - 1, \;, \; j = 0, \ldots, m \,.$$

Die Linearform f genügt der Abschätzung

$$|f(u)| \leq |u|_{k,\infty} \,, \qquad u \in C^r[a,b] \,. \tag{5.61}$$

Nach (5.30) ist daher f stetig in der Norm $\|.\|_{r,\infty}$ für $k = 0, \ldots, r$. Die Interpolationsabschätzung (5.41), (5.46) mit $s = k$, $p = q = \infty$ in Verbindung mit (5.61) ergibt daher

$$\begin{aligned} |f(u) - q(u)| = |f(u - Pu)| \leq |u - Pu|_{k,\infty} \leq \sigma h^{r-k} |u|_{r,\infty} \,, \\ k = 0, \ldots, r \,, \;\; u \in C^r[a,b] \,. \end{aligned} \tag{5.62}$$

Für Lagrange–Interpolationen ist $r_j = 1$, $r = m + 1$ also

$$\left| \frac{d^k u}{dx^k}(c) - \sum_{j=0}^{m} \gamma_{j0}^{(k)} u(x_j) \right| \leq \sigma h^{m+1-k} \max_{x \in [a,b]} \left| \frac{d^{m+1} u}{dx^{m+1}}(x) \right| \,. \tag{5.63}$$

□

5.3 Approximation durch Polynome in mehreren Veränderlichen

Wir gehen wie in Kapitel 4 von folgender Situation aus:

Jedes Finite Element e, i. e. Dreieck oder Rechteck, wird über die affine Abbildung A bzw. A^{-1} umkehrbar eindeutig auf ein Referenzelement E abgebildet. Dabei ist

$$E = A^{-1} e = B^{-1} e + c \,, \qquad B \in \mathbb{R}^{n \times n} \,, \;\; c \in \mathbb{R}^n \,.$$

Wir bezeichnen mit $M_E = \mathcal{P}(E)$ den Vektorraum aller Polynome über E mit $\dim M_E = N$. Gegeben sei ein System von Punkten $z_i \in E$, $i = 1, \ldots, N$, die auch mehrfach vorkommen können, und ein System $\{D^i | i = 1, \ldots, N\}$ von partiellen Ableitungen $D^i = D^{\alpha_i}$, $\alpha_i =$ Multiindex, mit der folgenden

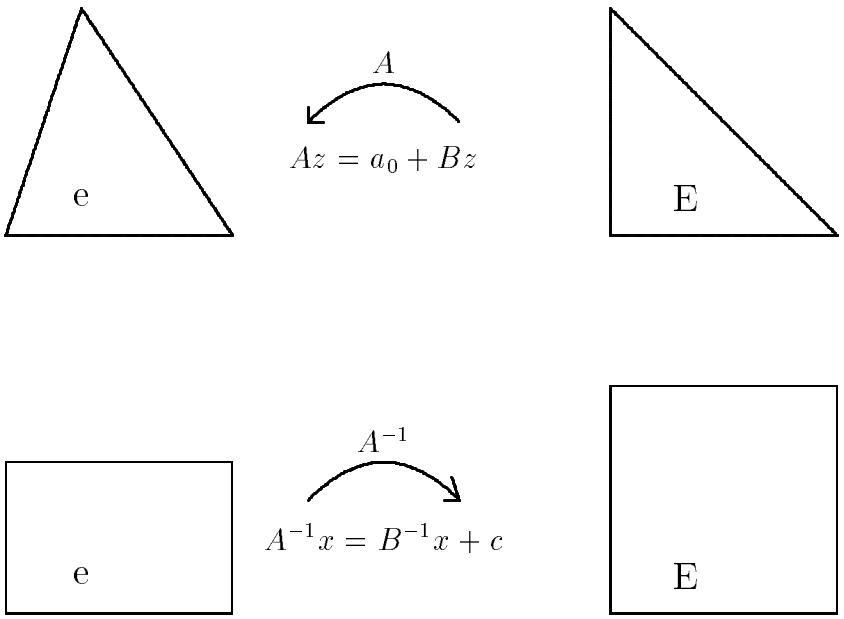


Abbildung 5.1: Affine Transformation finiter Elemente

Eigenschaft 5.14 *Für ein* $r \in \mathbb{N}$ *und ein* p *in* $1 \leq p \leq \infty$ *ist*

$$\mathcal{P}_{r-1}(E) \subset M_E \,, \tag{5.64}$$

und

$$|\alpha_i| < r - \frac{n}{p} \,, \qquad i = 1, \ldots, N \,. \tag{5.65}$$

Dabei bezeichnet $\mathcal{P}_m(E)$ den Vektorraum aller Polynome über E der Ordnung kleiner oder gleich m,

$$\mathcal{P}_m(E) = \left\{ P(z) = \sum_{0 \leq |\alpha| \leq k} a_\alpha z^\alpha \,, \qquad 0 \leq k \leq m \,, \;\; z \in E \right\} \,.$$

In (5.64) wird gerade der maximale Grad $m = r - 1$ von Polynomen festgelegt, die noch in M_E liegen. Bei den in früheren Abschnitten behandelten Beispielen ergibt sich r wie folgt.

$E =$ Einheitsdreieck:

Lagrange–Interpolation in $\mathcal{P}_1(E)$*:* $M_E = \mathcal{P}_1(E)\,,\; N = 3\,,\; r = 2\,;$
Lagrange–Interpolation in $\mathcal{P}_2(E)$*:* $M_E = \mathcal{P}_2(E)\,,\; N = 6\,,\; r = 3\,;$
Dreieck von Morley: $M_E = \mathcal{P}_2(E)\,,\; N = 6\,,\; r = 3\,;$

E = Einheitsquadrat:

Lagrange–Interpolation in $\mathcal{Q}_1(E)$: $M_E = \mathcal{Q}_1(E)\,,\ N = 4\,,$
$\mathcal{P}_1(E) \subset M_E\,,\ r = 2\,;$

Lagrange–Interpolation in $\mathcal{Q}_{22}(E)$:

$$M_E = \mathcal{Q}_{22}(\mathbb{R}) = [1, z_1, z_2, z_1z_2, z_1^2, z_1^2z_2, z_1z_2^2, z_2^2, z_1^2z_2^2]\,,$$
$$N = 9\,,\ \mathcal{P}_2(E) \subset M_E\,,\ r = 3\,;$$

Rechteck von Adini:

$$M_E = [1, z_1, z_2, \ldots, z_1z_2^2, z_1^3, z_2^3, z_1^3z_2, z_1z_2^3]\,,$$
$$N = 12\,,\ \mathcal{P}_3(E) \subset M_E\,,\ r = 4\,.$$

Die Schranke $r - \frac{n}{p}$ in (5.65) für die Ordnungen der Ableitungen D^i garantiert nach der *Sobolevschen Ungleichung* (vgl. z. B. [Rei95], 2.5, [Ada75] V, [OR76], 3.6) für jede Funktion $v \in H^{r,p}(E)$, daß

$$\|D^i v\|_{0,\infty,E} \le C\|v\|_{r,p,E}\,, \qquad i = 1, \ldots, N\,. \tag{5.66}$$

Genauer besagt (5.66), daß jedes $v \in H^{r,p}(E)$ fast überall gleich einer Funktion $\tilde{v} \in C^r(E)$ ist, für die die Werte $D^i\tilde{v}(z)$ für jedes $z \in E$ durch die rechte Seite in (5.66) abschätzbar sind. In diesem Fall sind also die Funktionale $\delta_i(v) = D^i v(z_i)$ stetige Linearformen auf $H^{r,p}(E)$. Die folgende Tabelle gibt für den Fall $p = 2$ die Ordnungen $m = |\alpha_i|$ in Abhängigkeit von r und der Dimension n wieder, für die die Linearformen noch stetig sind, d. h. für die (5.65) gilt:

$n \backslash^r$	1	2	3	4	5
1	0	1	2	3	4
2	/	0	1	2	3
3	/	0	1	2	3

Neben Eigenschaft 5.14 fordern wir noch von E und $\{z_i\}$, daß folgendes gilt:

Eigenschaft 5.15 *Das einzige Polynom* $P \in M_E$ *mit* $D^iP(z_i) = 0\,,\ i = 1, \ldots, N,$ *ist das Nullpolynom* $P = 0$.

Wie wir in Abschnitt 4.5 ausgeführt haben, sichert diese Eigenschaft die eindeutige Lösbarkeit der *hermiteschen Interpolationsaufgabe* auf E für jede rechte Seite. Letztere betrachten wir hier in der Form

$$D^i V(z_i) = D^i u(z_i)\,, \qquad i = 1, \ldots, N \tag{5.67}$$

für $u \in H^{r,p}(E)$ mit Lösung $V = P_E u \in \mathcal{P}_{r-1}(E)$. Die Lösung läßt sich darstellen als

$$V = \sum_{i=1}^{n} D^i u(z_i) \hat{p}_i$$

mit den durch

$$D^i \hat{p}_j(z_i) = \delta_{ij}\,, \qquad i, j = 1, \ldots, N\,, \tag{5.68}$$

eindeutig bestimmten kanonischen Basisfunktionen $\hat{p}_j$ (vgl. Satz 4.10).

Zur Abschätzung des Interpolationsfehlers $u - P_E u$ werden wir die Normäquivalenz von $\|.\|_{r,p}$ mit

$$|||u|||_{r,p} = |u|_{r,p} + \sum_{|s|<r} \left| \int_G D^s u(x)\, dx \right|\,, \qquad u \in H^{r,p}(G)\,, \tag{5.69}$$

mit Hilfe von Satz 5.6 zeigen. Für die folgenden Überlegungen können wir ein beliebiges beschränktes, offenes, zusammenhängendes Gebiet G des $\mathbb{R}^n$ betrachten, das einer sogenannten Kegelbedingung genügt — damit die Sobolewsche Ungleichung (vgl. (5.66)) gilt. Insbesondere kann $G = E$ oder $G = e$ sein.

Nach Satz 5.5 ist $|||.|||_{r,p}$ äquivalent $\|.\|_{r,p}$ dann und nur dann, wenn $|.|_{r,p}$ äquivalent $|., \mathcal{P}_{r-1}|_{r,p}$ ist, vorausgesetzt $\mathcal{P}_{r-1} = \mathcal{P}_{r-1}(G)$ ist der Nullraum von $|.|_{r,p}$ und

$$q_s(u) = \int_G D^s u(x)\, dx\,, \qquad |s| < r\,, \tag{5.70}$$

sind stetige Linearformen auf $H^{r,p}(G)$.

Die genannten zwei Voraussetzungen werden in den folgenden Sätzen nachgewiesen.

Satz 5.16 *Es gelten die genannten Voraussetzungen an G. Dann ist jede Funktion $v \in H^{r,p}(G)$ mit $D^s v = 0$, für alle $|s| = r$, f. ü. gleich einem Polynom aus $\mathcal{P}_{r-1}(G)$, und der Nullraum von $|.|_{r,p}$ ist gleich $\mathcal{P}_{r-1}(G)$.*

Beweis: Sei $v \in H^{r,p}(G)$ mit $|v|_{r,p} = 0$. Dann ist offenbar $D^s v = 0$ für alle $|s| = r$ im L^p–Sinne. Weiter folgt auch $D^s u = 0$ für alle $|s| \geq r$, denn die Definitionsgleichung der verallgemeinerten Ableitungen liefert

$$\begin{aligned} \int_G (D^s v)\varphi\, dx &= (-1)^{|s|} \int_G v D^s \varphi\, dx \\ &= (-1)^{|s_1|}(-1)^r \int_G v D^{s_0} D^{s_1} \varphi\, dx = 0\,, \qquad \varphi \in C_0^\infty(G)\,, \end{aligned}$$

wenn $|s| > r$, $s = s_0 + s_1$ mit $|s_0| = r$. Damit ist $v \in H^{t,p}(G)$ für alle $t \geq r$, und der Sobolewsche Einbettungssatz liefert ein $\tilde{v} \in C^\infty(G)$ mit $\tilde{v} = v$ f. ü. in G und $D^s\tilde{v} = 0$, $|s| \geq r$. Taylorentwicklung um einen beliebigen Punkt $c \in G$ zeigt, daß $\tilde{v} \in \mathcal{P}_{r-1}(B_\rho(c))$ in einer Kugel $B_\rho(c)$ mit Radius $\rho > 0$ um c. Da G als zusammenhängend vorausgesetzt wurde, ist $\tilde{v} \in \mathcal{P}_{r-1}(G)$. Umgekehrt gilt trivialerweise für $v \in \mathcal{P}_{r-1}(G)$, daß $|v|_{r,p} = 0$ ist. □

Satz 5.17 *Die in (5.70) definierten Funktionale q_s sind stetig auf $H^{r,p}(G)$ für $|s| \leq r-1$, und es existiert eine Basis $\{p_t\}$ von $\mathcal{P}_{r-1}(G)$ mit* $\det(q_s(p_t)) \neq 0$.

Beweis: Bezeichnen wir mit $|G|$ das Maß (oder Volumen) von G, dann liefert die Höldersche Ungleichung

$$|q_s(u)| \leq \left(\int_G 1\,dx\right)^{1/p'} \left(\int_G |D^s u|^p\,dx\right)^{1/p} \leq |G|^{1/p'}|u|_{k,p}$$

für $k = |s|$, $1 \leq p < \infty$; diese Abschätzung gilt auch für $p = \infty$. Die Stetigkeit von $q_s(u)$ folgt aus

$$|q_s(u)| \leq |G|^{1/p'}\|u\|_{r,p}\,, \quad u \in H^{r,p}(G)\,,\ 1 \leq p \leq \infty\,,\ |s| < r\,;$$

dies gilt übrigens auch für $|s| = r$.

Zum Beweis der zweiten Aussage wählen wir

$$p_t(x) := \frac{1}{t!}x^t = \frac{1}{t_1!\cdots t_n!}x_1^{t_1}\cdots x_n^{t_n}\,,\ \ |t| \leq r-1\,.$$

Dann ist $D^s p_t = 0$, falls $s_i > t_i$ für ein $i \in \{1,\ldots,n\}$; weiter ist offenbar $D^s p_s = 1$. Damit ist

$$q_s(p_s) = \int_G 1\,dx = |G|$$

und

$$q_s(p_t) = 0\,, \quad \text{falls} \quad s_i > t_i \quad \text{für ein } i \in \{1,\ldots,n\}\,.$$

Die Gramsche Matrix $(q_s(p_t))_{s,t}$ hat die Gestalt

$$\begin{pmatrix} |G| & 0 & & 0 \\ * & \cdot & & \\ & \cdot & & \\ & & \cdot & 0 \\ * & & * & |G| \end{pmatrix} .$$

Für festes t und $|s| \leq r-1$ ergibt sich nämlich gerade eine Zeile dieser Matrix, für deren Elemente rechts der Diagonalen immer der Fall $t_i < s_i$, für ein i, auftritt; in der Diagonalen erscheint immer $|G|$. Zum Beispiel erhält man für $n = 2\,,\ r = 3\,,\ |s| = s_1 + s_2 \leq 2$,

$$\begin{pmatrix} q_{00}(p_{00}) & q_{10}(p_{00}) & q_{01}(p_{00}) & q_{20}(p_{00}) & q_{11}(p_{00}) & q_{02}(p_{00}) \\ q_{00}(p_{10}) & q_{10}(p_{10}) & & & & q_{02}(p_{10}) \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ q_{00}(p_{02}) & & & & & q_{02}(p_{02}) \end{pmatrix}$$

Für die Determinante der unteren Dreiecksmatrix erhält man dann

$$\det(q_s(p_t)) = \prod_{|s|\leq r-1} |G| \neq 0 \ ,$$

was Satz 5.17 beweist. □

Als Folgerung von Satz 5.17 ergibt sich das folgende Ergebnis.

Satz 5.18 *Zu jedem $v \in H^{r,p}(G)$ existiert ein eindeutig bestimmtes Polynom $p \in \mathcal{P}_{r-1}(G)$ mit*

$$\int_G D^s(v+p)\,dx = 0 \quad \forall\, 0 \leq |s| \leq r-1 \ . \tag{5.71}$$

Beweis: Die Gleichung (5.71) ist äquivalent zur verallgemeinerten Interpolationsaufgabe:

Gegeben $v \in H^{r,p}(G)$, gesucht $p \in \mathcal{P}_{r-1}(G)$ mit

$$q_s(p) = -q_s(v) \ , \qquad 0 \leq |s| \leq r-1 \ .$$

Wegen der Aussage über die Determinante in Satz 5.17 und Satz 3.3 ist diese Aufgabe immer eindeutig lösbar. □

Als letzte Vorbereitung für die Herleitung von Interpolationsabschätzungen müssen wir noch die Äquivalenz der Normen $|||.|||_{r,p}$ und $\|.\|_{r,p}$ beweisen. Wollen wir dazu Satz 5.6 benutzen, dann müssen wir eine Koerzitivitätsungleichung und das Vorliegen von kompakten Einbettungen beweisen.

Die Kompaktheit der Einbettungen $J : H^{r,p}(G) \to H^{j,p}(G)$ für $j = 0, \ldots, r-1$, liefert der folgende *Satz von Rellich,* den wir ohne Beweis wiedergeben.

Satz 5.19 (Satz von Rellich) *Zu jeder beschränkten Folge in $H^{r,p}(G)$ existiert eine Teilfolge, die in $H^{j,p}(G)$ für jedes $j = 0, \ldots, r-1$, konvergiert.*

Mit Hilfe dieses Ergebnisses — und Satz 5.6 — können wir die Äquivalenz der Halbnormen beweisen.

Satz 5.20 *Unter den genannten Voraussetzungen an G gilt die „verallgemeinerte Poincarésche Ungleichung“*

$$c\|u\|_{r,p} \le |u|_{r,p} + \sum_{|s|\le r-1} \left| \int_G D^s u(x)\, dx \right| , \qquad u \in H^{r,p}(G) \tag{5.72}$$

mit einer Konstanten $c = c(n, r, p, G) > 0$, und die Halbnormen $|.|_{r,p}$, $|., \mathcal{P}_{r-1}|_{r,p}$ sind äquivalent, d. h.

$$\beta_0 |u|_{r,p} \le \inf_{w \in \mathcal{P}_{r-1}} \|u - w\|_{r,p} \le \beta_1 |u|_{r,p} , \qquad u \in H^{r,p}(G) . \tag{5.73}$$

Beweis: Zunächst stellen wir fest, daß auf der rechten Seite von (5.72) gerade die Norm $|||u|||_{r,p}$ auftaucht. Offenbar hat man $\|u\|_{r,p}^p = \|u\|_{r-1,p}^p + |u|_{r,p}^p$ und deshalb auch

$$\|u\|_{r,p} \le |u|_{r,p} + \|u\|_{r-1,p} , \qquad u \in H^{r,p}(G) .$$

Dies kann man als Koerzitivitätsungleichung (mit $\eta = 1$) für die Räume $D = H^{r-1,p}(G)$, $E = H^{r,p}(G)$ und die (Halb–)Normen $|.|_E = |.|_{r,p}$, $\|.\|_D = \|.\|_{r-1,p}$ verstehen (vgl. (5.23)). Zusammen mit dem Satz von Rellich läßt sich Satz 5.6 anwenden, der die Äquivalenz von $\|.\|_{r,p}$ mit $|||.|||_{r,p}$ sowie $|.|_{r,p}$ mit $|., \mathcal{P}_{r-1}|_{r,p}$ sichert. □

Als Folgerung der verallgemeinerten Poincaréschen Ungleichung sieht man sofort, daß

$$\|v\|_{r,p} \le C|v|_{r,p} \tag{5.74}$$

für alle Funktionen $v \in H^{r,p}(G)$ mit der Eigenschaft

$$\int_G D^s v \, dx = 0 \,, \qquad 0 \le |s| \le r-1 \,.$$

Nach diesen Vorbereitungen können wir Interpolationsabschätzungen beweisen. Die konkrete Situation stellt sich jetzt wie folgt dar,

$$E = H^{r,p}(G) \,, \;\; M = \mathcal{P}_{r-1}(G) \,, \;\; L = [q_s \,|\, |s| \le r-1]$$
$$\|.\|_E = \|.\|_{r,p} \,, \;\; |.|_E = |.|_{r,p} \,, \qquad |.|_F = |.|_{j,p} \,, \quad j = 0, \ldots, r \,.$$

Als *Beispiele* für q_s haben wir folgendes kennengelernt

— Lagrange–Interpolation: $q_s(u) = u(a_s) = \delta_s(u)$
(hierbei: $r > n/p$, damit $\delta_s, |s| \le r-1$, stetig auf $H^{r,p}(G)$)

— Hermite–Interpolation: $q_i(u) = D^{\alpha_i} u(z_i) \,, \quad i = 1, \ldots, N$ (hierbei: $|\alpha_i| < r - \frac{n}{p}$)

— q_s aus (5.70), d. h. $q_s(u) = \int_G D^s u \, dx \,, \quad |s| \le r-1$

Die Wahl der q_s aus (5.70) in diesem Abschnitt ist dadurch motiviert, daß die mit q_s definierten Normen $|||.|||_{r,p}$ (vgl. (5.69)) als äquivalent mit $\|.\|_{r,p}$ nachgewiesen werden konnten, wobei als Hilfsmittel die Koerzivität von $|.|_{r,p}$ auf $H^{r,p}(G)$ benutzt wurde.

Für die Lösung $P_E u$ der hermiteschen Interpolationsaufgabe auf dem Referenzelement E erhält man aus der allgemeinen Interpolationsabschätzung in Satz 5.2

$$|u - P_E u|_{j,p;\,E} \le \rho_0 |u|_{r,p;\,E} \,, \qquad 0 \le j \le r \,, \;\; 1 \le p \le \infty \,, \tag{5.75}$$

wobei die (beste) Konstante ρ_0 sich ergibt als

$$\rho_0 = \sup_{0 \ne v \in L^\perp} \frac{|v|_{j,p;\,E}}{|v|_{r,p;\,E}} \,,$$
$$L^\perp = \Big\{ v \in H^{r,p}(E) | D^i v(z_i) \;=\; 0 \,, \;\; i = 1, \ldots, N \Big\} \,.$$

In der durch die Eigenschaften 5.14 und 5.15 charakterisierten Situation hat man nämlich

$$P_E v = v \quad \text{für alle } v \in \mathcal{P}_{r-1}(E),$$

so daß P_E auch eine Projektion von $H^{r,p}$ auf $\mathcal{P}_{r-1}(E)$ ist und damit die Sätze aus Abschnitt 5.1 anwendbar sind.

Als nächstes wenden wir uns der eigentlichen Aufgabenstellung, nämlich der *hermiteschen Interpolationsaufgabe* auf einem einzelnen finiten Element e zu, d. h. gesucht ist ein $v \in M_e$, für $u \in H^{r,p}(e)$, mit

$$D^i v(x_i) = D^i u(x_i) \,, \qquad i = 1, \ldots, N \,. \tag{5.76}$$

Die Punkte $x_i \in e$ sind mit $z_i \in E$ durch die Beziehung

$$z_i = A^{-1} x_i = B^{-1} x_i + c \,, \qquad i = 1, \ldots, N \,,$$

verbunden. Die Lösung von (5.76) ergibt sich als (vgl. Satz 4.11)

$$v(x) = \sum_{i=1}^{N} D^i u(x_i) p_i \,, \tag{5.77}$$

wobei $p_i(x) = \hat{p}_i(A^{-1}x)$, $i = 1, \ldots, N$, mit den Basisfunktionen $\hat{p}_i$ der Hermiteschen Interpolationsaufgabe auf E.

Zur Abschätzung der besten Konstante ρ_0 in (5.75) mit Hilfe von Potenzen des Durchmessers von e benötigen wir Transformationsformeln für die Ableitungen.

Transformationsformeln für Ableitungen:

$$\overbrace{Az}^{x} = a_0 + Bz \,, \quad z \in \mathbb{R}^n$$

$$\text{Fktl.matrix}: \; B = \left(\frac{\partial x_j}{\partial z_k} \right)$$

$$v(x) = v(Az) \;=\; (\overbrace{v \circ A}^{y})(z) \,, \quad x = Az$$

Kettenregel: $\quad y(z) = v(\overbrace{a_0 + Bz}^{x})$

$$\Longrightarrow \frac{\partial y}{\partial z_k}(z) \;= \sum_{j=1}^{n} \frac{\partial v}{\partial x_j}(x) \frac{dx_j}{dz_k} \;=\; \sum_{j=1}^{n} b_{jk} \frac{\partial v}{\partial x_j}(x) \,,$$

$$k = 1, \ldots, n \,, \;\; x \in \mathbb{R}^n$$

$$\Longleftrightarrow \nabla_z y(z) = B^\top \nabla_x v(x)$$

höhere Abl.: $\nabla_z^m y(z) \;=\; (B^\top \nabla_x)^m v(x) \in \mathbb{R}^{n^m}$

d. h. $$\frac{\partial^m y}{\partial z_{k_1} \cdots \partial z_{k_m}}(z) \;=$$
$$= \sum_{j_1,\ldots j_m=1}^{n} b_{j_1,k_1} \cdots b_{j_m,k_m} \frac{\partial^m v}{\partial x_{j_1} \cdots \partial x_{j_m}}(x)\,, \quad k_1,\ldots,k_m = 1,\ldots,n$$

z. B. $$n=2: \nabla^2 v(x) = \left(\frac{\partial^2 u}{\partial x_1^2}\,,\; \frac{\partial^2 u}{\partial x_1 \partial x_2}\,,\; \frac{\partial^2 u}{\partial x_2\, \partial x_1}\,,\; \frac{\partial^2 u}{\partial x_2^2} \right)\,.$$

Inverse Bez.: $\nabla_x^m v(x) = (C\nabla_z)^m y(z)$

mit $C = (B^\top)^{-1} = (B^{-1})^\top\,,\;\; |C| = |B^{-1}|$

wobei

$$|C| = \left(\sum_{j,k=1}^{n} |c_{jk}|^2 \right)^{1/2}$$

die *Quadratsummennorm* einer Matrix $C = (c_{jk})$ bezeichnet.

Behauptung:

$$|B|^{-m} |\nabla_z^m y(z)| \le |\nabla_x^m v(x)| \le |B^{-1}|^m |\nabla_z^m y(z)| \tag{5.78}$$

Beweis:

$$|\nabla_x^m v(x)|^2 = \sum_{k_1,\ldots,k_m=1}^{n} \left| \frac{\partial^m v(x)}{\partial x_{k_1} \cdots \partial x_{k_m}} \right|^2$$

$$\overset{\text{s. o.}}{=} \sum_{k_1,\ldots,k_m} \left\{ \sum_{j_1,\ldots,j_m=1}^{n} |c_{k_1,j_1}| \cdots |c_{k_m,j_m}| \left| \frac{\partial^m y(z)}{\partial z_{j_1} \cdots \partial z_{j_m}} \right| \right\}^2$$

$$\le \sum_{k_1,\ldots,k_m} \left\{ \sum_{j_1,\ldots,j_m} |c_{k_1,j_1}|^2 \cdots |c_{k_m,j_m}|^2 \cdot \sum_{j_1,\ldots,j_m} \left| \frac{\partial^m y(z)}{\partial z_{j_1} \cdots \partial z_{j_m}} \right|^2 \right\}$$

$$= \sum_{j_1,\ldots,j_m=1} \left| \frac{\partial^m y(z)}{\partial z_{j_1} \cdots \partial z_{j_m}} \right|^2 \sum_{k_1,j_1=1}^{n} |c_{j_1 k_1}|^2 \cdots \sum_{k_m,j_m=1}^{n} |c_{j_m k_m}|^2$$

$$= |C|^{2m} \left| \nabla_y^m y(z) \right|^2 \,.$$

Wegen $|C| = |(B^\top)^{-1}| = |B^{-1}|$ folgt die 2. Ungleichung.

Analog zeigt man $|\nabla_z^m y(z)|^2 \le |B|^{2m} |\nabla_x^m v(x)|^2$. □

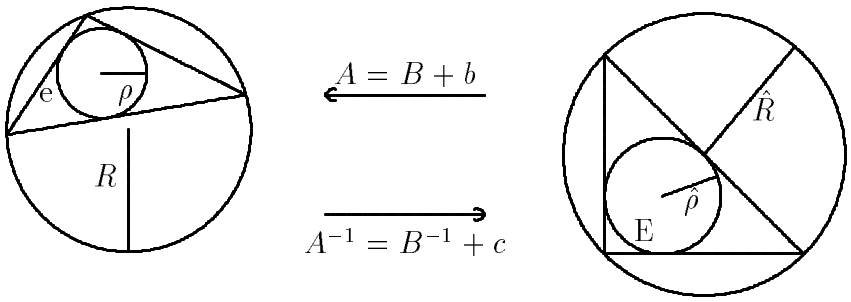


Abbildung 5.2: In– und Umkreis finiter Elemente

Bevor wir zur Herleitung von Interpolationsabschätzungen im folgenden Satz 5.21 kommen, wollen wir noch die Quadratsummennormen von B und B^{-1} abschätzen.

Wir betrachten wieder die Situation mit einem finiten Element e (i. e. Dreieck oder Viereck bzw. entsprechende n–dimensionale Körper), dem Referenzelement E (i. e. Einheitsdreieck oder Quadrat bzw. n–dimensionales Simplex oder Quadrat) und der umkehrbaren affinen Abbildung $A = B+b : E \to e$. Es bezeichne ρ bzw. R den Radius des Inkreises bzw. Umkreises (bzw. der entspr. n–dimensionalen Kugeln) von e; $\hat\rho$ und $\hat R$ seien die entsprechenden Radien für E. Weiter seien h bzw. $\hat h$ die maximalen Durchmesser von e bzw. $\hat E$. Für $n = 2$ und das Einheitsdreieck hat man bekanntlich $\frac{1}{2}\hat h = \hat R = 2\hat\rho = \frac{1}{2}\sqrt{2}$; für beliebiges e gilt immer $h \le 2R$.

Für $\xi, \eta \in e$ mit $|\xi - \eta| = \rho$, $x = \xi - \eta$, sind $\hat\xi = B^{-1}\xi + c$, $\hat\eta = B^{-1}\eta + c$ aus E und für $B = (b_{ij})$, $B^{-1} = \left(b_{ij}^{(-1)}\right)$ gilt

$$|B^{-1}x| = |B^{-1}\xi + c - B^{-1}\eta - c| = |\hat\xi - \hat\eta| \le 2\hat R \ .$$

Da im endlichdimensionalen Zahlenraum alle Normen äquivalent sind, gibt es eine Konstante $c = c(n)$ mit

$$\max_{i,j} \left|b_{ij}^{(-1)}\right| \le c \sup_{0\neq x} \frac{|B^{-1}x|}{|x|} \le c\frac{1}{\rho} \sup_{|x|=\rho} |B^{-1}x| \le 2c\frac{\hat R}{\rho} \ . \tag{5.79}$$

Analog folgert man, daß

$$\max_{ij} |b_{ij}| \le 2c\frac{h}{\hat\rho}$$

und deshalb

$$|B| \le Ch \ , \quad |B^{-1}| \le C\rho^{-1} \ , \tag{5.80}$$

wobei C nur von der Dimension n und dem Referenzelement E abhängt.
Außerdem benötigen wir noch die folgende *Determinantenabschätzung*

$$|B^{-1}|^{-n} \le |\det B| \le |B|^n \ , \tag{5.81}$$

die aus dem Determinantenentwicklungssatz in Verbindung mit $|\det B| = |\det B^{-1}|^{-1}$ folgt. Man erhält schließlich mit Konstanten $c_0, c_1, C > 0$ die Abschätzungen

$$c_0\rho^n \le |\det B| \le c_1 h^n \quad \text{und} \quad |B|\,|B^{-1}| \le C\frac{h}{\rho} \ . \tag{5.82}$$

Hierbei ist $\kappa(B) = |B|\,|B^{-1}|$ die *Konditionszahl* der Matrix B.

Satz 5.21 *Die beste Konstante in (5.75) läßt sich abschätzen durch*

$$\rho_0 \le \sigma\gamma\frac{h^r}{\rho^j} \ , \qquad j = 0, \ldots, r-1 \ , \tag{5.83}$$

wobei

$$\begin{aligned}
&\gamma = \gamma(n) \quad \textit{und} \\
&\sigma = \sup_{0 \ne y \in \hat{L}^\perp} \frac{|y|_{j,p;E}}{|y|_{r,p;E}} \ , \\
&\hat{L}^\perp = \left\{ v \in H^{r,p}(E) \,\middle|\, D^i v(z_i) \ = \ 0 \ , \quad i = 1, \ldots, N \right\} \ .
\end{aligned}$$

Beweis: (nur für $p = 2$) Durch Variablensubstitution erhält man

$$\int_{A(E)} f(x)\,dx = |\det B| \int_E f(Az)\,dz \ .$$

Zusammen mit der Ungleichung (5.78) folgt $(y(z) = v(A^{-1}z))$

$$\begin{aligned}
\int_e \left|\nabla_x^j v(x)\right|^2 dx &\le |B^{-1}|^{2j} \int_e \left|\left(\nabla_z^j y\right)(A^{-1}x)\right|^2 dx \\
&= |B^{-1}|^{2j} |\det B| \int_E \left|\nabla_z^j y(z)\right|^2 dz \ , \qquad j = 0, \ldots, r-1 \ ,
\end{aligned}$$

und

$$\int_e |\nabla_x^r v(x)|^2 \, dx \geq |B|^{-2r} \int_e \left| (\nabla_z^r y)\,(A^{-1}x) \right|^2 \, dx$$
$$= |B|^{-2r} |\det B| \int_E |\nabla_z^r y(z)|^2 \, dz \,, \qquad v \in C_0^\infty(\mathbb{R}^n) \,.$$

Durch ein Stetigkeitsargument gelten diese Abschätzungen auch für $v \in H^{r,2}(e)$. Für die Halbnormen $|.|_{j,2;e}$ hat man die Darstellung

$$|v|_{j,2;e} = \left(\sum_{|\alpha|=j} \int_e |D^\alpha v(x)|^2 \, dx \right)^{1/2} = \left(\int_e |\nabla_x^j v(x)|^2 \, dx \right)^{1/2} \,,$$

so daß für $v \in L^\perp$ auch $y \in \hat{L}^\perp$ und

$$\frac{|v|_{j,2;e}}{|v|_{r,2;e}} \leq |B|^r |B^{-1}|^j \frac{|\det B|}{|\det B|} \frac{|y|_{j,2;E}}{|y|_{r,2;E}} \,, \qquad 0 \neq v \in L^\perp \,.$$

Mit Hilfe von (5.80) folgt (5.83). □

Im Beweis des letzten Satzes beobachtet man, daß eigentlich keine Einschränkung für j und r benötigt wurde, also auch $j \geq r$ zugelassen ist. Dies liefert das folgende Ergebnis.

Satz 5.22 *(Inverse Abschätzung) Für alle $v \in \mathcal{P}_{r-1}(e)$ gilt*

$$|v|_{k,p;e} \leq c \frac{h^s}{\rho^k} |v|_{s,p;e} \,, \qquad 0 \leq s < k \leq r-1 \,, \;\; 1 \leq p \leq \infty \,, \tag{5.84}$$

mit einer Konstanten $c = c(n, r, p, E)$.

Beweis: Auf dem endlichdimensinalen Raum $\mathcal{P}_{r-1}(E)$ sind alle Normen äquivalent, so daß mit einer Konstanten $c_1 = c_1(n, r, p, E)$ gilt

$$\|V\|_{k,p;E} \leq c_1 \|V\|_{s,p;E} \quad \forall 0 \leq s, k \leq r-1, \; 1 \leq p \leq \infty, \; V \in \mathcal{P}_{r-1}(E).$$

Damit gilt auch für die Halbnormen eine analoge Beziehung

$$|V|_{k,p;E} \leq c_2 |V|_{s,p;E} \quad \forall 0 \leq s, k \leq r-1, \; 1 \leq p \leq \infty, \; V \in \mathcal{P}_{r-1}(E).$$

Nach Satz 5.20 wissen wir nämlich, daß $|.|_{s,p;E}$ äquivalent ist zum kürzesten Abstand $|., \mathcal{P}_{s-1}(E)|_{s,p}$, der wiederum eine Norm auf dem Quotientenraum $\mathcal{P}_{r-1}(E)/\mathcal{P}_{s-1}(E)$ erklärt (setze $\{0\}$ anstelle von $\mathcal{P}_{s-1}(E)$ im Fall $s = 0$; vgl. auch [Cia78, Bew. von Thm. 3.2.6]). Der Quotientenraum ist ebenfalls endlichdimensional und daher sind alle Normen darauf äquivalent.

Benutzt man die obige Abschätzung für die Halbnormen und wählt im Beweis von Satz 5.21 $j = k,\ s = r$ (und $p = 2$), so folgt

$$\frac{|v|_{k,p;e}}{|v|_{s,p;e}} \le |B|^s |B^{-1}|^k \frac{|y|_{k,p,E}}{|y|_{s,p;E}} \le c_3 \frac{h^s}{\rho^k}, \qquad 0 \le s\,,\ k \le r-1,\ v \in \mathcal{P}_{r-1}(e)$$

mit einer Konstanten $c_3 = c_3(n, r, p, E)$ und $y(z) = v(Az)$. □

Wir wenden die Ergebnisse nun auf zwei einfache Beispiele an, wobei wir eine Zerlegung $\Delta = \{G_\iota\}_{\iota \in I}$ eines Gebiets $G \subset \mathbb{R}^n$ in n–Simplexe betrachten,

$$\overline{G} = \bigcup_{\iota \in I} G_\iota\ .$$

I sei eine endliche Indexmenge. Die „Triangulierung“ erfülle noch

$$\mathring{G}_\iota \cap \mathring{G}_\mu = \emptyset\ , \quad \iota \neq \mu\ .$$

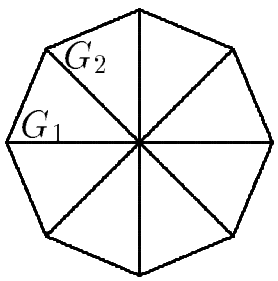


Abbildung 5.3: Quasi–uniforme Triangulierung

Jedes finite Element G_ι wird durch eine affine Abbildung A_ι bzw. A_ι^{-1} auf das Referenzelement E abgebildet.

$$G_\iota = \operatorname{conv}\left(a_0^{(\iota)}, \ldots, a_n^{(\iota)}\right) \qquad E = \operatorname{conv}(e_0, \ldots, e_n)$$

Die maximalen Durchmesser von G_ι seien h_ι und $h = \max_\iota h_\iota$, $h_{\min} = \min_\iota h_\iota$. Man nennt eine *Triangulierung regulär*, wenn für das zugehörige System von affinen Abbildungen $A_\iota = B_\iota + b_\iota$ gilt

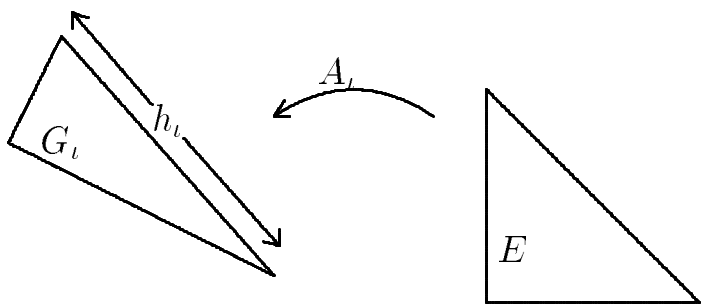


Abbildung 5.4: Transformation auf Einheitselement

$$1 \le |B_\iota|\,|B_\iota^{-1}| \le \kappa \quad \forall \iota\,. \tag{5.85}$$

D. h. bei einer regulären Triangulierung sind die zugehörigen Konditionszahlen der Matrizen B_ι gleichmäßig beschränkt. Weiter heißt eine Triangulierung *quasi–uniform*, wenn mit zwei Konstanten β_0, $\beta_1 > 0$ gilt

$$\beta_0 h \le h_\iota \le \beta_1 h \quad \forall \iota\,. \tag{5.86}$$

Wir geben noch zwei hinreichende Bedingungen für die Regularität (5.85) an. Wegen (5.82) ist die Bedingung

$$\frac{h_\iota}{\rho_\iota} \le c \quad \forall \iota \tag{5.87}$$

offenbar hinreichend für die Regularität (5.85). Eine Verschärfung stellt die folgende Bedingung dar,

$$h_\iota \le ch \quad \text{und} \quad \rho_\iota \ge c'h \quad \forall \iota\,. \tag{5.88}$$

Gilt (5.87) — oder (5.88) —, dann kann man in (5.83) und (5.84) offenbar ρ_ι durch h_ι ersetzen.

Im Falle von Dreiecken $G_\iota = \operatorname{conv}\left(a_0^{(\iota)}, a_1^{(\iota)}, a_2^{(\iota)}\right) \subset \mathbb{R}^2$ kann man die Konditionszahl und das Verhältnis von Umkreis– zu Inkreisradius durch die Winkel im Dreieck abschätzen bzw. darstellen. Man hat (vgl. [SF73, Chapt. 3.1])

$$\frac{h_\iota}{\rho_\iota} = \left\{4 \sin\left(\frac{\alpha_\iota}{2}\right) \sin\left(\frac{\beta_\iota}{2}\right) \sin\left(\frac{\gamma_\iota}{2}\right)\right\}^{-1} \quad \text{und} \quad |B_\iota|\,|B_\iota^{-1}| \le \frac{2}{(\sin\gamma_\iota)^2}\,,$$

wobei γ_ι den kleinsten Winkel des Dreiecks bezeichnet. Daraus sieht man sofort, daß die *„Winkelbedingung“*

$$\gamma_\iota \geq \gamma_0 > 0 \quad \forall \iota \tag{5.89}$$

hinreichend für (5.87) — und damit für die Regularität (5.85) — ist. Die Winkelbedingung besagt, daß die Winkel der Dreiecke einer Triangulierung nicht zu spitz werden dürfen. Wir bemerken hier noch, daß wir später die Bedingungen (5.85), (5.87), (5.88) oder (5.89) gleichmäßig für alle Zerlegungen eines Gebietes G fordern werden — also nicht nur für *eine* Zerlegung, für die diese Bedingungen ja trivial wären (weil nur jeweils endlich viele finite Elemente vorkommen).

Die inverse Abschätzung (5.84) auf jedes einzelne Dreieck angewandt liefert im Fall $s = 0\,,\ k = 1$ unter der Regularitätsbedingung (5.87) die Abschätzungen

$$\begin{aligned} \int_G |\nabla v|^2\,dx = |v|^2_{1,2;G} \ &= \ \sum_\iota |v|^2_{1,2;G_\iota} \\ &\leq c \sum_\iota h_\iota^{-2} |v|^2_{0,2;G_\iota} \ \leq \ c\, h_{\min}^{-2} |v|^2_{0,2;G} \end{aligned} \tag{5.90}$$

für Funktionen $v \in M$, deren Einschränkungen $v|G_\iota$ Polynome mindestens vom Grad eins sind. Für eine quasi–uniforme Triangulierung folgt daraus

$$|v|_{1,2;G} \leq c\, h^{-1} |v|_{0,2;G}\,, \qquad v \in M\ .$$

Beispiel 5.23 Lagrange–Interpolation in

$$M = \left\{ v \in C(\overline{G}) \,\middle|\, v|G_\iota \in \mathcal{P}_1(G_\iota)\ \forall \iota \right\}\,, \qquad G \subset \mathbb{R}^2\,.$$

Auf jedem finiten Element G_ι ist ein Polynom

$$(P_\iota u)(x) = p(x) = \gamma_{10}^{(\iota)} x_1 + \gamma_{01}^{(\iota)} x_2 + \gamma_{00}^{(\iota)}\,, \qquad x \in G_\iota\,,$$

mit $p\left(a_i^{(\iota)}\right) = u\left(a_i^{(\iota)}\right)\,,\ i = 0,1,2$, zu bestimmen. Wir setzen Bedingung (5.87) — oder (5.88) — voraus. Der Fehler genügt nach Satz 5.21 der Abschätzung

$$|u - P_\iota u|_{j,2;G_\iota} \leq C h_\iota^{2-j} |u|_{2,2;G_\iota}\,, \qquad j = 0,1\,,\ u \in H^{2,2}(G)\,.$$

Über dem gesamten Gebiet folgt daraus $(I_h u|G_\iota := P_\iota u|G_\iota)$

$$|u - I_h u|_{j,2;G} \leq C h^{2-j} |u|_{2,2;G}\,, \qquad j = 0,1\,,\ u \in H^{2,2}(G)\,,$$

wie man durch Summation über ι sofort sieht. □

Beispiel 5.24 Lagrange–Interpolation in

$$M = \left\{ v \in C(\overline{G}) \middle| \; v|G_\iota \in \mathcal{P}_2(G_\iota) \;\; \forall \iota \right\} , \qquad G_\iota \subset G \subset \mathbb{R}^2 .$$

Gesucht ist also für jedes G_ι ein Polynom

$$(P_\iota u)(x) = p(x) = \gamma_{20}^{(\iota)} x_1^2 + \gamma_{11}^{(\iota)} x_1 x_2 + \gamma_{0,2}^{(\iota)} x_2^2 + \gamma_{10}^{(\iota)} x_1 + \gamma_{01}^{(\iota)} x_2 + \gamma_{00}^{(\iota)} ,$$

das an den Seitenmittelpunkten und den Eckpunkten mit einer Funktion u übereinstimmt.

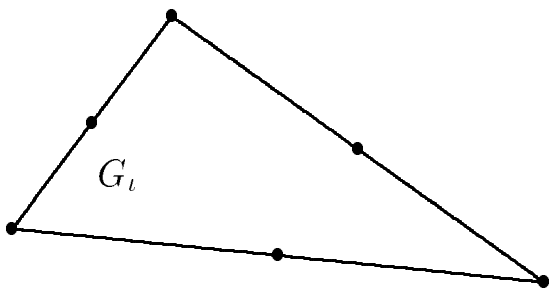


Abbildung 5.5: Knotenpunkte der Lagrange—Interpolation in $\mathcal{P}_2(G_\iota)$

Die Interpolationsabschätzung für jedes einzelne G_ι liefert ($r = 3$)

$$|u - P_\iota u|_{j,p;G_\iota} \leq C h_\iota^{3-j} |u|_{3,p;G_\iota} , \qquad u \in H^{3,p}(G) .$$

Insgesamt folgt daraus unter der Voraussetzung (5.87)

$$|u - I_h u|_{j,p;G} \leq C h^{3-j} |u|_{3,p;G} , \qquad u \in H^{3,p}(G) , \;\; j = 0, 1, 2 .$$

□

Wir wenden nun zum Schluß dieses Kapitels die Interpolationsabschätzungen auf die Finite–Elemente–Lösungen elliptischer Randwertprobleme an. Wesentliches Hilfsmittel ist hierfür Satz 2.6, der im wesentlichen über die Quasioptimalität der FE–Lösung sicherstellt, daß letztere sich asymptotisch so verhält wie der Interpolationsfehler.

Wir betrachten die schwache Lösung u eines elliptischen Randwertproblems (vgl. Abschnitt 2.1) und eine zugehörige Ritzapproximation u_h in

$$M = M_h = \left\{ v \in C(\overline{G}) \,\middle|\, v|G_\iota \in \mathcal{P}_s(G_\iota) \;\; \forall \iota \right\} \cap H_0^{1,2}(G) \; ,$$

wobei $s = 1$ oder $s = 2$ ist (vgl. Beispiel 5.23 und 5.24 in diesem Abschnitt). Für eine hinreichend glatte Lösung $u \in H^{s+1}(G)$ liefert die Quasioptimalität bzgl. der zur Energienorm äquivalenten Norm $\|.\|_{1,2}$, daß (vgl. Satz 2.6)

$$\|u - u_h\|_{1,2} \leq C \|u - I_h u\|_{1,2} \leq C h^s |u|_{s+1,2} \; , \qquad s = 1, 2 \; ,$$

wobei $I_h u|G_\iota = P_\iota u|G_\iota$ die Lagrange–Interpolation auf jedem finiten Element G_ι bezeichnet. Es sei G konvex und ∂G aus der Klasse C^2, dann ist mindestens $u \in H^{2,2}(G)$ (vgl. Satz 2.5), und man erhält

$$\|u - u_h\|_{1,2} = \mathrm{O}(h^s)$$

für $s = 1$ und $s = 2$. In jedem Fall erhält man optimale Konvergenzordnung bzgl. $\|\cdot\|_1$ in Abhängigkeit der Glattheit der Lösung.

Optimale Konvergenzordnung bzgl. der L^2–Norm liefert der Aubin–Nitsche–Trick (vgl. (2.9) in Satz 2.6). Hierzu benötigt man wieder mindestens $H^{2,2}$–Regularität der schwachen Lösung und erhält wegen

$$\Delta_h^m = \mathrm{O}(h^{m-1}) \; , \qquad m = 2, 3$$

die Abschätzung $(s + 1 = m)$

$$\|u - u_h\|_{0,2} \leq C \, h^{s+1} |u|_{s+1,2} \; , \qquad m = 2, 3 \; .$$

5.4 Beste Approximation von Linearformen

Für interpolatorische Approximationen stetiger Linearformen wurde schon in Abschnitt 5.1 eine zentrale Fehlerabschätzung (vgl. (5.21)) hergeleitet und als Konsequenz eine allgemeine Formulierung des Bramble–Hilbert–Lemmas (vgl. Satz 5.5) gegeben. Am Ende von Abschnitt 5.2 wurden diese Ergebnisse in einer Ortsdimension auf die Approximation von Integralen bzw. Differentialquotienten durch Quadraturformeln bzw. Differentialquotienten angewandt. In diesem Abschnitt untersuchen wir beste Approximationen stetiger Linearformen auf Hilbertschen Räumen, die man wieder als interpolatorische Approximationen auffassen kann. Zentrales Ergebnis dieses Abschnitts ist die Fehlerabschätzung (5.103), die die sog. Hyperkreis–Fehlerabschätzung von Golumb–Weinberger[2)], S. 134, darstellt. Diese Abschätzung

[2)] Golumb, M., Weinberger, H. F., Optimal approximation and error bounds. Proc Symp. "On Numerical Approximation" (Madison, 1958), 117–190. Univ. of Wisconsin Press, Madison, 1959.

steht in engem Zusammenhang zu a–priori–Fehlerabschätzungen in schwächeren Normen, die wir in (2.9) hergeleitet haben und mit dem Namen „Aubin–Nitsche–Trick“ verbunden sind.

Sei E' der Raum der stetigen Linearformen auf einem Hilbert–Raum E. Bekanntlich ist E' ebenfalls ein Hilbertscher Raum. Zu jedem Paar $\ell_1, \ell_2 \in E'$ existieren nach dem Satz von Fréchet–Riesz eindeutig bestimmte Elemente $u_1, u_2 \in E$ mit der Eigenschaft

$$(\varphi, u_1) = \ell_1(\varphi) \,, \quad (\varphi, u_2) = \ell_2(\varphi) \,, \qquad \varphi \in E \,. \tag{5.91}$$

Hiermit ist das Skalarprodukt in E' erklärt durch

$$(\ell_1, \ell_2) = \overline{\ell_1(u_2)} = \ell_2(u_1) = (u_1, u_2) \,. \tag{5.92}$$

Aus diesen Beziehungen folgt unmittelbar die Darstellung

$$\|\ell_1\| = \|u_1\| = \sup_{0 \neq \varphi \in E} \frac{|\ell_1(\varphi)|}{\|\varphi\|} \tag{5.93}$$

für die Normen stetiger Linearformen auf E. Durch (5.91) wird eine umkehrbar eindeutige semilineare isometrische Abbildung von E auf E' erklärt, die wir als *kanonischen Isomorphismus* bezeichnen.

Sei jetzt ein System linear unabhängiger Linearformen $q_1, \ldots, q_m$ aus E' gegeben. Diese Linearformen spannen den endlichdimensionalen Teilraum $L = [q_1, \ldots, q_m]$ von E' auf. Wir fragen nach *besten Approximationen* von Linearformen $f \in E'$ durch Elemente $q \in L$ im Sinne von

$$q \in L : \ \|f - q\| = \min_{\ell \in L} \|f - \ell\| = |f, L| \,. \tag{5.94}$$

Dieses ist die duale Form der in Abschnitt 5.1 betrachteten Aufgabe. Die dort gewonnenen Ergebnisse können wir auf $E', L, f, q_1, \ldots, q_m$ anstelle von $E, M, u, w_1, \ldots, w_m$ anwenden. Die eindeutig bestimmte Lösung q der Aufgabe (5.94) ist die orthogonale Projektion von f auf L, die durch die *Orthogonalitätsrelation*

$$(f - q, q_j) = 0 \,, \quad j = 1, \ldots, m \Longleftrightarrow f - q \perp L \tag{5.95}$$

charakterisiert ist. Die Linearformen $q_1, \ldots, q_m$ lassen sich durch Elemente $w_1, \ldots, w_m$ darstellen in der Form

$$q_j(\varphi) = (\varphi, w_j) \,, \quad \varphi \in E \,, \ \ j = 1, \ldots, m \,. \tag{5.96}$$

Sei $M = [w_1, \ldots, w_m]$ der durch diese Elemente aufgespannte Teilraum von E. Hiermit werden die Teilräume M und L kanonisch isomorph, das heißt,

$$\ell \in L \Longleftrightarrow \exists w \in M : \ell(\varphi) = (\varphi, w) \ , \quad \varphi \in E \ . \tag{5.97}$$

Die Beziehung (5.95) ist jetzt gleichbedeutend mit

$$q \in L : \ (f - q)(w_j) = 0 \ , \quad j = 1, \ldots, m \ . \tag{5.98}$$

Das heißt, q ist *interpolatorische Approximation* von f an den *Knotenpunkten* $w_1, \ldots, w_m$. Die Lösung dieser Aufgabe ist wiederum eindeutig bestimmt, weil das System $\{w_1, \ldots, w_m\}$ L–unisolvent ist. Auf Grund von (5.98) ist q die eindeutig bestimmte Lösung der Variationsgleichung

$$q \in L : \ q(w) = f(w) \ , \quad w \in M \ . \tag{5.99}$$

Die beste Approximation $q \in L$ ist somit *exakt* gleich f auf dem Teilraum M. Aus den obigen Überlegungen folgen noch die wichtigen Beziehungen

$$M^\perp = E' \ominus L \ , \quad L = E' \ominus M^\perp \ . \tag{5.100}$$

Anwendung von Satz 5.2 auf die kürzesten Abstände $|f, L| = \|f - g\|$ in E' ergibt die folgende Darstellung und zweiseitige Abschätzung für die beste Konstante in (5.17),

$$\frac{|f(r)|}{\|r\|_E} \leq \max_{0 \neq r \in L^\perp} \frac{|f(r)|}{\|r\|_E} = |f, L| \leq \|f - \ell\| \tag{5.101}$$

für alle $0 \neq r \in L^\perp$ und alle $\ell \in L$. Zusammen mit den obigen Ergebnissen gewinnt man noch eine interessante Darstellung und Fehlerabschätzung für die besten Approximationen q stetiger Linearformen f auf E.

Satz 5.25 *Die beste Approximation q von f hat die Darstellung*

$$q(u) = f(p) = q(p) \ , \quad p = Pu \ , \ u \in E \ , \tag{5.102}$$

mit der orthogonalen Projektion P von E auf M. Es gilt die zugehörige Fehlerabschätzung

$$|f(u) - q(p)| \leq |f, L| \ |u, M| \ . \tag{5.103}$$

Beweis: Sei $q_1(u) := f(Pu)$ für $u \in E$. Dann wird

$$(f - q_1)(w) = f(w - Pw) = 0 \ , \quad w \in M \ .$$

Mit Hilfe von (5.97) sieht man, daß $q_1 \in L$ ist. Also ist q_1 Lösung von (5.98), (5.99) und muß daher die eindeutig bestimmte beste Approximation q sein. Weiter ist $u = p + r$ mit $r \in L^\perp$ und $q \in L$, so daß $q(r) = 0$ und $q(u) = q(p)$ wird. Aus den Eigenschaften (5.102) und den Eigenschaften von p bzw. q als beste Approximationen folgt unmittelbar die Beziehung

$$\begin{aligned} |f(u) - q(p)| &= |(f - q)(u) + q(u) - q(p)| \\ &= |(f - q)(u - p)| \leq |f, L|\,|u, M| \ , \end{aligned}$$

also die obige Ungleichung (5.103). □

Kapitel 6

Numerische Aspekte der FEM

In diesem Kapitel wird die Frage diskutiert, wie genau man die Elemente der Steifigkeitsmatrix $K = (a(w_k, w_j))_{j,k}$ und des Lastvektors $((f, w_j))_j$ berechnen muß bzw. kann, um die optimale Konvergenzordnung der Methode der Finiten Elemente zu erhalten. Es werden verschiedene Fälle diskutiert, wobei der Fall der näherungsweisen Berechnung mit Hilfe von Quadraturformeln für die Praxis der wichtigste ist. Einleitend werden noch gewisse Begriffe geklärt und einige aus früheren Kapiteln bekannte Sachverhalte wiederholt.

6.1 Bezeichnungen und Wiederholung

Wir betrachten hier die schwache Lösung eines elliptischen Randwertproblems in zwei Ortsdimensionen mit homogenen Dirichletschen Randbedingungen,

Gesucht $u \in H_0^{1,2}(G)$ *mit*

$$a(u,v) := \sum_{i,k=1}^{2} \int_G a^{ik} D_k u D_i v \, dx = (f,v) \ , \quad v \in H_0^{1,2}(G) \ . \tag{V}$$

Das Gebiet $G \subset \mathbb{R}^2$ sei beschränkt, konvex und polygonal. ∂G bezeichne den Rand von G.

Weiter sei eine Triangulierung $\Delta_h = \{G_\iota\}$ von $\overline{G}$ gegeben mit den Eigenschaften

$$\overline{G} = \bigcup_\iota G_\iota \ ; \quad \overset{\circ}{G}_\iota \cap \overset{\circ}{G}_\nu = \emptyset \ , \quad \nu \neq \iota \ . \tag{6.1}$$

Der Diskretisierungsparameter h bezeichne den maximalen Durchmesser der Dreiecke bzw. Vierecke G_ι oder deren maximalen Umkreisradius.

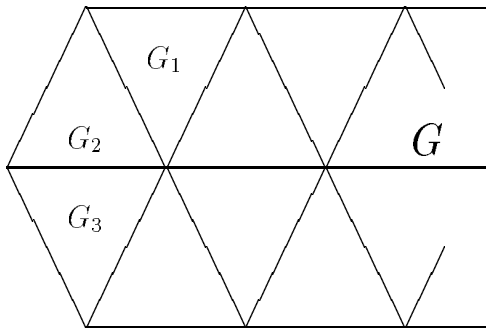


Abbildung 6.1: Quasi–uniforme Triangulierung

S_h^m bezeichne Finite–Elemente–Räume von stückweise polynomialen Funktionenräumen M_ι auf jedem einzelnen G_ι derart, daß $\mathcal{P}_m(G_\iota) \subset M_\iota \quad \forall \iota \in I$,

$$S_h^m = \left\{ v \in H_0^{1,2}(G) \,\middle|\, v|G_\iota \in M_\iota \ , \ \mathcal{P}_m(G_\iota) \subset M_\iota \quad \forall \iota \in I \right\} \ ;$$

I bezeichne eine endliche Indexmenge, dessen Kardinalität von h abhängt. Beispiele haben wir den vorherigen Kapiteln angegeben (vgl. insbesondere Abschnitt 5.3). Wir benötigen mindestens $m \geq 1$ um sicherzustellen, daß die Diracfunktionale stetig auf $H^{2,2}(G)$ sind (vgl. (5.65) mit $r = m+1$). Zu jedem Raum M_ι von Polynomen auf G_ι sei noch ein M_ι–unisolventes System $\left\{q_j^{(\iota)}\right\}_j$ gegeben, zum *Beispiel* die Diracfunktionale an gewissen ausgezeichneten (Interpolations–)Punkten $z_j^{(\iota)} \in G_\iota$. Weiter bezeichne $L_\iota = \left[q_j^{(\iota)}\right]_j$.

In Anlehnung an [Cia78, 2.3] nennen wir jetzt das Tripel $(G_\iota, L_\iota, M_\iota)$ ein *finites Element*. Vorher hatten wir kurz einfach G_ι als finites Element bezeichnet; die Kennzeichnung als Tripel macht aber deutlich, daß zur geometrischen Form auch immer ein geeigneter Polynomraum und ein geeignetes System von Funktionalen gehört.

Wir wollen jetzt eine Situation zugrunde legen, in der alle finiten Elemente einer Familie von Triangulierungen aus einem Referenzelement $(E, \hat{L}, \hat{M})$ hervorgehen. Die zugehörigen invertierbaren affinen Abbildungen bezeichnen wir wieder mit

$$x = A_\iota z = B_\iota z + b_\iota \ , \quad z = A_\iota^{-1} x = B_\iota^{-1} x + c_\iota \ , \quad \iota \in I \ , \tag{6.2}$$

wobei $x \in G_\iota$, $z \in E$. Zwischen Funktionen auf G_ι bzw. E besteht die Beziehung

$$v(x) = \hat{v}(z) \ , \quad z = A_\iota^{-1} x \ , \quad x \in G_\iota \ . \tag{6.3}$$

Eine Familie $\{G_\iota, L_\iota, M_\iota\}_{\iota,h}$ von finiten Elementen heißt *affin äquivalent* zu einem *Referenzelement* $\hat{G} = (E, \hat{L}, \hat{M})$, wenn für alle $\iota \in I$ und alle betrachteten h gilt

(i) $G_\iota = A_\iota E$,

(ii) $L_\iota^\perp = \left\{ v \,\middle|\, v(x) = \hat{v}\,(A_\iota^{-1}x) \ , \ \hat{v} \in \hat{L}^\perp \right\}$,

(iii) $M_\iota = \left\{ p = \hat{p}\,(A_\iota^{-1}\cdot) \ , \ \hat{p} \in \hat{M} \right\}$.

$\hat{M}$ bezeichne einen Polynomraum auf E mit $\mathcal{P}_m(E) \subset \hat{M}$ und $\hat{L}$ ein $\hat{M}$–unisolventes System von Funktionalen. Man nennt $\{(G_\iota, L_\iota, M_\iota)\}_{\iota,h}$ eine *reguläre* Familie finiter Elemente, wenn

(i) $\{(G_\iota, L_\iota, M_\iota)\}_{\iota,h}$ affin äquivalent zu einem $\hat{G} = (E, \hat{L}, \hat{M})$,

(ii) $\{A_\iota\}$ regulär, d. h. $1 \le |B_\iota|\,|B_\iota^{-1}| \le \kappa \quad \forall \iota, h$,

(iii) $h \longrightarrow 0$.

Zum Abschluß dieser vorbereitenden Bemerkungen wiederholen wir noch die Transformationsformel für Integrale,

$$\int_{G_\iota} v(x)\,dx = |\det B_\iota| \int_E \hat{v}(z)\,dz \ , \tag{6.4}$$

wobei gilt (vgl. (5.82))

$$C_0 h^2 \le |B_\iota^{-1}|^{-2} \le |\det B_\iota| \le |B_\iota|^2 \le C_1 h^2 \ , \tag{6.5}$$

vorausgesetzt die Regularitätsbedingung

$$h_\iota \le c\,h \quad \text{und} \quad \rho_\iota \ge c' h \quad \forall \iota, h \tag{6.6}$$

ist erfüllt (h_ι = Durchmesser von G_ι, ρ_ι = Inkreisradius von G_ι). Dann gilt noch (vgl. (5.78))

$$c_0\,|\nabla_z \hat{v}(z)| \le h\,|\nabla_x v(x)| \le c_1\,|\nabla_z \hat{v}(z)| \ , \quad x \in G_\iota,\ z = A_\iota^{-1}x \ . \tag{6.7}$$

Im folgenden setzen wir die Regularitätsbedingung (6.6) für die reguläre Familie finiter Elemente voraus. Es bezeichne $\{w_1, \ldots, w_N\}$ eine Basis von S_h^m und $\dim S_h^m = N (= N_h)$.

Die FEM liefert nun Näherungen $u_h = \sum_{k=1}^N \beta_k w_k$ zu u als (eindeutige) Lösungen der Probleme
Gesucht $u_h \in S_h^m$ *mit*

$$a(u_h, v_h) = (f, v_h) , \quad \forall v_h \in S_h^m . \tag{V_h}$$

Mit den Größen

$$\begin{aligned} u &= (\beta_1, \ldots, \beta_N) \in \mathbb{R}^N , \\ K &= (a(w_i, w_k))_{i,k=1}^N \in \mathbb{R}^{N \times N} , \\ F &= ((f, w_i))_{i=1,\ldots,N} \in \mathbb{R}^N \end{aligned}$$

ist (V_h) äquivalent zu dem linearen Gleichungssystem

$$Ku = F . \tag{6.8}$$

Die Matrix K (= Gesamtsteifigkeitsmatrix) ist offenbar symmetrisch und positiv definit.

Wenn, wie im obigen Beispiel, die Basisfunktion w_i in den Dreiecken, die den Interpolationspunkt z_i nicht enthalten, verschwindet („lokale" Basis), sind nur wenige Elemente $a(w_i, w_k)$ der Matrix K von Null verschieden.

Sie hat also eine Bandstruktur, wobei die Anzahl der nicht nur mit Nullen besetzten Nebendiagonalen gleich der maximalen Anzahl von Dreiecken ist, die in einem Punkt z_i zusammentreffen.

6.2 Exakte Bestimmung der Matrix K

Wir betrachten im folgenden Triangulierungen bestehend aus Dreiecken und den Spezialfall

$$a^{\nu\mu} = \delta_{\nu\mu} \ : \ a(u, v) = \int_G \nabla u \cdot \nabla v \, dx .$$

Die Elemente $K_{ij} = a(w_i, w_j)$ der Matrix K werden als Summe der Bestandteile berechnet, welche von den einzelnen Dreiecken $G_\iota \in \Delta_h$ geliefert werden („Assemblierung"):

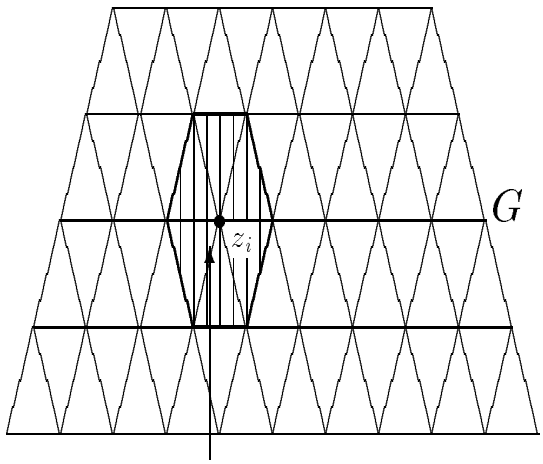


Abbildung 6.2: Träger von w_i

$$K_{ij} = \sum_{G_\iota} a_{G_\iota}(w_i, w_j) \ , \quad a_{G_\iota}(w_i, w_j) = \int_{G_\iota} \nabla w_i \cdot \nabla w_j \, dx \ . \tag{6.9}$$

Offenbar ist $a_{G_\iota}(w_i, w_j) = 0$, wenn einer oder beide der zugehörigen Knotenpunkte z_i, z_j nicht zu G_ι gehören.

Seien nun $\zeta_i \notin \partial G \, , \ i = 1, \ldots N_\iota$, die zu einem Dreieck G_ι gehörenden Knotenpunkte. Die zugehörigen Basisfunktionen $w_i \in S_h^m$ haben dann auf G_ι Darstellungen

$$w_i\Big|_{G_\iota}(x) = \sum_{0 \le |\alpha| \le m} a_i^\alpha x^\alpha \in \mathcal{P}_m(G_\iota) \ .$$

Die Koeffizienten a_i^α erhält man dabei als Lösungen der N_ι linearen Gleichungssysteme

$$w_i(\zeta_j) = \sum_{0 \le |\alpha| \le m} a_i^\alpha \zeta_j^\alpha = \delta_{ij} \ , \quad j = 1, \ldots, N_\iota \ , \tag{6.10}$$

mit bekannten Koeffizienten ζ_j^α, die sich aus den Koordinaten der Dreieckpunkte ergeben.

Die Bestimmung von $a_{G_\iota}(w_i, w_j)$ ist dann reduziert auf die Berechnung der Integrale

$$\int_{G_\iota} x^\beta \, dx \ , \quad 0 \le |\beta| \le 2m - 2 \ .$$

Hierfür existieren exakte Formeln, z. B. bei Dreiecken mit Eckpunkten (x_i, y_i), (x_j, y_j), (x_m, y_m), die

$$\frac{x_i + x_j + x_m}{3} = \frac{y_i + y_j + y_m}{3} = 0$$

erfüllen, hat man

$$\int x\,dx\,dy = \int y\,dx\,dy = 0$$

$$\int dx\,dy = \frac{1}{2}\begin{vmatrix} 1 & x_i & y_i \\ 1 & x_j & y_j \\ 1 & x_m & y_m \end{vmatrix} =: \Delta = \text{ Dreiecksfläche}$$

$$\int x^2\,dx\,dy = \frac{\Delta}{12}(x_i^2 + x_j^2 + x_m^2)$$

$$\int y^2\,dx\,dy = \frac{\Delta}{12}(y_i^2 + y_j^2 + y_m^2)$$

$$\int xy\,dx\,dy = \frac{\Delta}{12}(x_i y_i + x_j y_j + x_m y_m)$$

6.3 Näherungsweise Bestimmung von K und F

Bei dem oben betrachteten einfachen Fall stückweise linearer finiter Elemente konnte die Matrix K exakt bestimmt werden. Dies ist im allgemeinen jedoch u. a. aus folgenden Gründen nicht mehr möglich:

- Selbst im Fall $a^{ij} \equiv \delta_{ij}$ sind zur Bestimmung der Elemente K_{ij} von K Integrale über Polynome vom Grad $2m-2$ zu berechnen. Die dazu existierenden exakten Formeln sind für große m sehr aufwendig. Andererseits werden wir sehen, daß zur Sicherung der Konvergenz des gesamten Verfahrens auch schon näherungsweise Berechnung mit Quadraturformeln geringerer Genauigkeit genügt.
- Bei Vorliegen nichtkonstanter Koeffizienten a^{ij} und rechter Seite f können die Elemente K_{ij} und F_i i. a. nicht mehr exakt bestimmt werden.
- Numerische Integration wird herangezogen, wenn sog. „isoparametrische“ Elemente verwendet werden (z. B. Splines über Dreiecken mit krummem Rand). Auf diese Variante der FEM werden wir weiter unten noch näher eingehen.

Bei lediglich näherungsweiser Bestimmung der Matrix

$$K = (K_{ij})_{i,j=1}^{N} \quad \text{und des Vektors} \quad F = (F_i\,,\ i = 1, \ldots, N)$$

werden an Stelle von

$$Ku = F \tag{6.11}$$

die gestörten linearen Gleichungssysteme

$$\tilde{K}\tilde{u} = \tilde{F} \tag{6.12}$$

mit den Matrizen $\tilde{K} = \left(\tilde{K}_{ij}\right)_{i,j=1}^{N}$ und den Vektoren $\tilde{F} = \left(\tilde{F}_i\right)_{i=1}^{N}$ gelöst. Dies führt dann zu fehlerhaften Ritzapproximationen

$$\tilde{u}_h = \sum_{i=1}^{N} \tilde{\beta}^i w_i \in S_h^m\ . \tag{6.13}$$

Wir wollen den Fehler $u_h - \tilde{u}_h$ abschätzen.

Dazu werden zunächst für $v_h = \sum_{i=1}^{N} \xi^i w_i\ ,\quad w_h = \sum_{i=1}^{N} \eta^i w_i$ die gestörten Bilinearformen und Linearformen erklärt,

$$\tilde{a}(v_h, w_h) = \sum_{i,j=1}^{N} \tilde{K}_{ij}\xi^i\eta^j\ , \quad \tilde{\ell}(v_h) = \sum_{i=1}^{N} \tilde{F}_i\xi^i\ .$$

Das lineare Gleichungssystem (6.12) ist dann äquivalent zu der Aufgabe

Gesucht $\tilde{u}_h \in S_h^m$ *mit*

$$\tilde{a}(\tilde{u}_h, v_h) = \tilde{\ell}(v_h)\ , \quad \forall v_h \in S_h^m\ . \tag{$\tilde{V}_h$}$$

Lemma 6.1 *Für die gestörten Bilinearformen* $\tilde{a}$ *und Linearformen* $\tilde{\ell}$ *gelte gleichmäßig für* $0 < h \le h_0$

$$\|v_h\|_{1,2}^2 \le \gamma\tilde{a}(v_h)\ ,$$
$$|(a - \tilde{a})(u_h, v_h)| + |(f, v_h) - \tilde{\ell}(v_h)| \le c\,h^\tau \|v_h\|_{1,2}, \quad v_h \in S_h^m.$$

Dann besitzen die Probleme $(\tilde{V}_h)$ *eindeutige Lösungen* $\tilde{u}_h \in S_h^m$, *und der Fehler zu den Ritzapproximationen* $u_h \in S_h^m$ *läßt sich abschätzen durch*

$$\|u_h - \tilde{u}_h\|_{1,2} \leq c\gamma h^\tau \ . \tag{6.14}$$

Beweis: Die eindeutige Lösbarkeit der Probleme $(\tilde{V}_h)$ liefert das Lemma von Lax–Milgram. Ferner gilt die folgende Identität für $v_h := u_h - \tilde{u}_h$,

$$\begin{aligned}\tilde{a}(v_h) &= \tilde{a}(u_h, v_h) - \tilde{a}(\tilde{u}_h, v_h) \\ &= (\tilde{a} - a)(u_h, v_h) + a(u_h, v_h) - \tilde{a}(\tilde{u}_h, v_h) \\ &= (\tilde{a} - a)(u_h, v_h) + (f, v_h) - \tilde{\ell}(v_h) \ .\end{aligned}$$

Mit Hilfe der Voraussetzungen folgt daraus direkt die Behauptung. □

Bei der näherungsweisen Bestimmung von K und F kann auf verschiedene Art vorgegangen werden. Dabei nehmen wir im folgenden die Koeffizienten a^{ij} und die Funktion f als hinreichend glatt an, etwa:

$$a^{ij} \ , \ f \in H^{s,\infty}(G) \ , \quad i,j = 1,2 \ .$$

Weiter gehen wir im folgenden von einer regulären Familie finiter Elemente bestehend aus Dreiecken aus. Wir können deshalb auch annehmen, daß der Polynomraum für das Referenzelement gerade $\hat{M} = \mathcal{P}_m(E)$ ist; die Verallgemeinerung auf den allgemeinen Fall $\hat{M} \supset \mathcal{P}_m(E)$ bringt keine grundsätzlichen Schwierigkeiten.

6.3.1 Interpolation mit exakter Polynomintegration

Auf jedem Dreieck $G_\iota \in \Delta_h$ werden für ein $r > 0$, $r \in \mathbb{N}$, polynomiale Interpolierende (bzgl. geeigneter Stützstellen)

$$\tilde{a}^{ij} \ , \ \ \tilde{f} \in \mathcal{P}_{r-1}(G_\iota) \ ; \quad i,j = 1,2 \ , \tag{6.15}$$

zu a^{ij} und f gebildet. Aufgrund unserer Sätze aus Kapitel 5 kann dies so geschehen, daß unabhängig von h gilt

$$\left\|a^{ij} - \tilde{a}^{ij}\right\|_{0,\infty} + \|f - \tilde{f}\|_{0,\infty} = \mathrm{O}(h^r) \ . \tag{6.16}$$

Die damit gebildeten gestörten Elemente

$$\begin{aligned}\tilde{K}_{ij} &= \int_G \sum_{\nu,\mu=1}^{2} \tilde{a}^{\nu\mu} D_\mu w_i D_\nu w_j \, dx \\ \tilde{F}_i &= \int_G \tilde{f} w_i \, dx \ , \quad i,j = 1,\ldots,N \ ,\end{aligned}$$

werden dann exakt integriert. Hierzu sind allerdings Formeln für Integrale der folgenden Form erforderlich,

$$\int_{G_\iota} x^\beta \, dx \ , \quad 0 \le |\beta| \le 2m + r - 3 \ .$$

Wir überprüfen nun die Voraussetzungen von Lemma 6.1. Mit Hilfe von (6.16) folgt für $v_h, w_h \in S_h^m$

$$\begin{aligned} |(a - \tilde{a})(v_h, w_h)| &= \left| \int_G \sum_{\nu,\mu=1}^{2} (a^{\nu\mu} - \tilde{a}^{\nu\mu}) \, D_\mu v_h D_\nu w_h \, dx \right| \\ &\le c \, h^r \|v_h\|_{1,2} \|w_h\|_{1,2} \end{aligned}$$

sowie

$$\left| (f, v_h) - \tilde{\ell}(v_h) \right| = \left| \int_G (f - \tilde{f}) v_h \, dx \right| \le c \, h^r \|v_h\|_{1,2} \ .$$

Wir setzen $v_h = w_h$ und erhalten

$$\begin{aligned} \|v_h\|_{1,2}^2 \le c \, a(v_h) \ &\le \ c \, |(a - \tilde{a})(v_h)| + c \, \tilde{a}(v_h) \\ &\le c \, h^r \|v_h\|_{1,2}^2 + c \, \tilde{a}(v_h) \end{aligned}$$

und daher für $0 < h \le h_0$ hinreichend klein

$$\|v_h\|_{1,2}^2 \le c \, \tilde{a}(v_h) \ .$$

Lemma 6.1 liefert somit

$$\|u_h - \tilde{u}_h\|_{1,2} \le c \, h^r \ ,$$

und der Gesamtfehler des Verfahrens ist

$$\|u - \tilde{u}_h\|_{1,2} = \mathrm{O}(h^m + h^r) \ . \tag{6.17}$$

Zur Erlangung der optimalen Konvergenzordnung muß also $r \ge m$ gewählt werden. Der Fehler durch die Interpolierenden $\tilde{a}^{ij}, \tilde{f}$ ist gegenüber dem Verfahrensfehler vernachlässigbar, wenn $r > m$ gewählt wird.

6.3.2 Integration mit Quadraturformeln

Ähnlich wie die eindimensionalen Quadraturformeln von Newton–Cotes und Gauß lassen sich auch für mehrdimensionale Bereiche, wie hier z. B. Dreiecke, Integrationsformeln angeben

$$f_{G_\iota}(g) = \int_{G_\iota} g\,dx \sim q_{G_\iota}(g) = \sum_{i=1}^{L} \omega_i g(\xi_i)$$

mit geeigneten Stützstellen $\xi_i \in G_\iota$ und Gewichten ω_i.

Hinsichtlich der Konstruktion derartiger Formeln sei auf die Kapitel 3 und 4 sowie auf die Literatur verwiesen (siehe z. B. [Cia78, §4.1],[ZT89, Chapt. 8] etc.).

Eine Quadraturformel $q_{G_\iota}(.)$ auf G_ι heißt *von der Ordnung* r, wenn sie für Polynome bis zum Grad $r-1$ exakt ist:

$$q_{G_\iota}(p) = \int_{G_\iota} p\,dx\ , \quad p \in \mathcal{P}_{r-1}(G_\iota)\ . \tag{6.18}$$

Der *Genauigkeitsgrad* von q_{G_ι} ist dann gerade durch $r-1$ gegeben.

Äquivalent zu (6.18) ist, daß q_{G_ι} interpolatorische Approximation von f_{G_ι} als Linearformen auf $\mathcal{P}_{r-1}(G_\iota)$ ist,

$$q_{G_\iota}(g) = f_{G_\iota}(P_\iota g)$$

für geeignetes g, wobei $p = P_\iota g \in \mathcal{P}_{r-1}(G_\iota)$ mit Hilfe der Diracfunktionale erklärt ist,

$$\delta_i(p) = p(\xi_i) = g(\xi_i) = \delta_i(g)\ , \quad i = 1, \ldots, L\ .$$

Das folgende Ergebnis ist eine Konsequenz der Interpolationsabschätzungen und sichert eine Abschätzung für den Quadraturfehler vorausgesetzt $r \geq 3$.

Lemma 6.2 *Für eine Quadraturformel der Ordnung $r \geq 3$ und $g \in H^{r,1}(G_\iota)$ gilt*

$$|f_{G_\iota}(g) - q_{G_\iota}(g)| \leq C h^r |g|_{r,1;G_\iota} \quad \forall \iota \tag{6.19}$$

mit einer Konstanten C unabhängig von ι und h.

Beweis: Mit $E = H^{r,1}(G_\iota)$, $M = \mathcal{P}_{r-1}(G_\iota)$, $|.|_E = |.|_{r,1}$, $\|.\|_E = \|.\|_{r,1}$ ist $\delta_\mu(g) = g(\xi_\mu)$ ein stetiges Funktional auf E, falls (Sobolevsche Ungleichung) $0 = |s| < r - \frac{n}{p} = r-2$, d. h. es muß $r \geq 3$ sein. Weiter ist f_{G_ι} eine stetige Linearform bzgl. der L^1–Norm,

$$|f_{G_\iota}(g)| \leq |g|_{0,1}\ , \quad g \in L^1(G_\iota)\ ,$$

so daß aus den Interpolationsabschätzungen von Abschnitt 5.3 folgt

$$\begin{aligned} |f_{G_\iota}(g) - q_{G_\iota}(g)| = |f_{G_\iota}(g - P_\iota g)| &\leq |g - P_\iota g|_{0,1} \\ &\leq \sigma\gamma h^r |g|_{r,1} , \quad g \in H^{r,1}(G_\iota) \quad \forall \iota . \end{aligned}$$

□

Beispiele für Quadraturformeln auf Dreiecken von der Ordnung 2 bis 6 findet man in Table 8.2 von [ZT89, p. 176].

Sei $r \geq m$, dann gilt $\mathcal{P}_m(E) \subset \mathcal{P}_r(E)$ mit dem zum Referenzelement (i. e. Einheitsdreieck) gehörigen Polynomraum $\hat{M} = \mathcal{P}_m(E)$, und q_E eine (lineare) Quadraturformel auf E,

$$q_E(g) = \sum_{i=1}^{L} \hat{\omega}_i g(\hat{\xi}_i) \tag{6.20}$$

mit Stützstellen $\hat{\xi}_i \in E$ und positiven Gewichten $\hat{\omega}_i > 0$, $i = 1, \ldots, L$. Für die affine Familie finiter Elemente erhält man dann Quadraturformeln

$$q_{G_\iota}(g) = \sum_{i=1}^{L} \omega_i g(\xi_i)$$

mit den Stützstellen $\xi_i = A_\iota \hat{\xi}_i$ und Gewichten $\omega_i = |\det B_\iota| \hat{\omega}_i$, $i = 1, \ldots, L$.

Von der Quadraturformel q_E bzw. den Stützstellen $\hat{\xi}_i$ fordern wir, daß mit einem $k' \in \mathbb{N}$ eine der folgenden Eigenschaften gilt — später werden wir $k' \geq m$ verlangen:

Eigenschaft 6.3 *Für beliebiges $\hat{p} \in \mathcal{P}_{k'-1}(E)$ folgt aus $\hat{p}(\hat{\xi}_i) = 0$, $i = 1, \ldots, L$, daß $\hat{p} \equiv 0$.*

Eigenschaft 6.4 *Die Quadraturformel q_E hat die Ordnung $r = 2k' - 1$.*

Eigenschaft 6.3 ist eng verwandt mit entsprechenden Eigenschaften aus den Abschnitten 4.5 bzw. 5.3 und sichert die eindeutige Lösbarkeit der Lagrange–Interpolation an den Interpolationspunkten als Stützstellen. Äquivalent zu Eigenschaft 6.3 ist, daß eine $\mathcal{P}_{k'-1}(E)$–unisolvente Teilmenge von $\{\delta_\mu\}_{\mu=1}^{L}$, der Dirac–Funktionale $\delta_\mu(g) = g(\hat{\xi}_\mu)$, $\mu \in \{1, \ldots, L\}$, existiert. Notwendig für Eigenschaft 6.3 ist offenbar, daß die Anzahl L der Stützstellen größer oder gleich $\dim \mathcal{P}_{k'-1}$ ist, d. h.

$$L \geq N := \frac{1}{2} k'(k'+1) .$$

Eigenschaft 6.4 stellt eine Beziehung zum Genauigkeitsgrad her.

Gewisse Quadraturformeln erfüllen beide Eigenschaften, z. B. die Dreipunktformel

$$\int_{G_\iota} g(x)\,dx = \frac{|G_\iota|}{3}(g(\underline{a}^\iota_{10}) + g(\underline{a}^\iota_{01}) + g(\underline{a}^\iota_{11}))$$

(vgl. (4.45) in Beispiel 4.5) mit den Seitenmittelpunkten eines Dreiecks als Stützstellen. Diese ist exakt für Polynome aus $\mathcal{P}_2(\mathbb{R}^2)$, d. h. hier ist $k' = 2$ in Eigenschaft 6.4. An drei nicht ausgearteten Punkten ist die Lagrange–Interpolationsaufgabe in $\mathcal{P}_1(\mathbb{R}^2)$ immer eindeutig lösbar, so daß auch $k' = 2$ in Eigenschaft 6.3 für dieses Beispiel.

Die 1–Punkt–Formel mit dem Schwerpunkt als einziger Stützstelle (s. (4.35)) erfüllt Eigenschaft 6.3 und 6.4 mit $k' = 1$; allerdings ist hierfür die Ordnung sogar $r = 2$.

Die beiden genannten Eigenschaften sind hinreichend für die folgende Eigenschaft, die wesentlich im Beweis des folgenden Lemmas benutzt wird.

Eigenschaft 6.5 *Die Abbildung*

$$\hat{p} \longmapsto \left(\sum_{i=1}^{L} \hat{\omega}_i \sum_{j=1}^{2} D_j\hat{p}(\hat{\xi}_i)^2\right)^{1/2} \tag{6.21}$$

definiert eine Norm auf dem Quotientenraum $\hat{M}/\mathcal{P}_0(E)$.

Lemma 6.6 *Unter der Eigenschaft 6.3 oder 6.4 für die Quadraturformel q_E folgt für $k' \geq m$ Eigenschaft 6.5. Weiter gibt es eine von h unabhängige Konstante c, so daß*

$$q_{G_\iota}\left(|\nabla_x v|^2\right) \geq c\int_{G_\iota} |\nabla_x v|^2\,dx \quad \forall v \in S_h^m \quad \forall \iota \in I\ . \tag{6.22}$$

Beweis: (i) Gilt Eigenschaft 6.3 und beachtet man die Positivität der Gewichte $\hat{\omega}_i$, dann folgt die Beziehung

$$\hat{p} \in \hat{M} \text{ und } \sum_{i=1}^{L} \hat{\omega}_i \sum_{\nu=1}^{2} D_\nu\hat{p}(\hat{\xi}_i)^2 = 0 \Longrightarrow D_\nu\hat{p}(\hat{\xi}_i) = 0\ ,$$
$$1 \leq \nu \leq 2\,,\ 1 \leq i \leq L\ .$$

Jedes $D_\nu\hat{p}$ liegt in $\mathcal{P}_{k'-1}(E)$ — da $\hat{p} \in \hat{M} = \mathcal{P}_m(E) \subset \mathcal{P}_{k'}(E)$ — und verschwindet bei $\hat{\xi}_i\,,\ i = 1,\ldots,L$, so daß nach Eigenschaft 6.3 $D_\nu\hat{p} \equiv 0\,,\ \nu = 1,2$. Damit erfüllt die durch (6.21) definierte Funktion die Definitheitsbedingung auf $\hat{M}/\mathcal{P}_0(E)$; die anderen Eigenschaften einer Norm sind trivialerweise erfüllt.

Falls Eigenschaft 6.4 erfüllt ist, gilt

$$\sum_{i=1}^{L} \hat{\omega}_i \sum_{\nu=1}^{2} D_\nu \hat{p}(\hat{\xi}_i)^2 = \int_E \sum_{\nu=1}^{2} (D_\nu \hat{p})^2 \, dx \left(= |\hat{p}|^2_{1,2,E} \right) \ ,$$

da $\sum_{\nu=1}^{2} D_\nu \hat{p}^2$ ein Polynom aus $\mathcal{P}_{2k'-2}(E)$ ist (für $\hat{p} \in \hat{M} = \mathcal{P}_m(E) \subset \mathcal{P}_{k'}(E)$) und links die (exakte) Quadraturformel für das (rechte) Integral steht. Da $|\hat{p}|_{1,2,E}$ offenbar eine Norm auf $\hat{M}/\mathcal{P}_0(E)$ darstellt, ist damit auch durch (6.21) eine Norm erklärt.

(ii) Wegen Eigenschaft 6.5, da $\hat{M}/\mathcal{P}_0(E)$ endlichdimensional ist, und weil auch $|\hat{p}|_{1,2,E}$ eine Norm auf $\hat{M}/\mathcal{P}_0(E)$ darstellt, gibt es eine Konstante $\hat{C}$ mit

$$\hat{C} \, |\hat{p}|^2_{1,2,E} \leq \sum_{i=1}^{L} \hat{\omega}_i \sum_{\nu=1}^{2} D_\nu \hat{p}(\hat{\xi}_i)^2 \ , \quad \hat{p} \in \hat{M} \ .$$

Aus (6.5) und (6.7) folgt für die zugehörigen Quadraturformeln über jedem finiten Element G_ι

$$\begin{aligned} q_{G_\iota}\left(|\nabla_x p|^2 \right) &= \sum_{i=1}^{L} \omega_i \, |\nabla_x p(\xi_i)|^2 \ = \ |\det B_\iota| \sum_{i=1}^{L} \hat{\omega}_i \, |\nabla_x p(\xi_i)|^2 \\ &\geq C |\det B_\iota| h^{-2} \sum_{i=1}^{L} \hat{\omega}_i \left| \nabla_z \hat{p}(\hat{\xi}_i) \right|^2 \\ &\geq c_0' q_E(|\nabla_z \hat{p}|^2) \ , \quad \hat{p} \in \hat{M} \ (p = \hat{p}(A^{-1}\cdot)) \ . \end{aligned}$$

Zusammen mit der vorigen Ungleichung erhält man

$$\begin{aligned} q_{G_\iota}\left(|\nabla_x p|^2 \right) &\geq c_0' q_E(|\nabla_z \hat{p}|^2) \ \geq \ c_0' \hat{C} |\hat{p}|^2_{1,2;E} \\ &\geq C \int_{G_\iota} |\nabla_x p|^2 \, dx \ , \quad p \in S_h^m \ , \end{aligned}$$

da noch (wegen (6.5) und (6.7))

$$\int_{G_\iota} |\nabla_x p|^2 \, dx \leq c_1 h^{-2} \int_{G_\iota} |\nabla_z \hat{p}|^2 \, dx = \underbrace{c_1 |\det B_\iota| h^{-2}}_{\mathrm{O}(1)} \int_E |\nabla_z \hat{p}|^2 \, dz \ .$$

□

Die gestörten Matrixelemente $\tilde{K}_{ij}$ und Vektorkomponenten $\tilde{F}_i$ sind nun definiert durch

$$\begin{aligned} \tilde{K}_{ij} &= \sum_\iota q_{G_\iota} \left(\sum_{\nu,\mu=1}^{2} a^{\nu\mu} D_\mu w_i D_\nu w_j \right) \\ \tilde{F}_i &= \sum_\iota q_{G_\iota}(f w_i) \ , \quad i,j = 1, \ldots, N \ . \end{aligned}$$

Das folgende Lemma sichert die Voraussetzungen von Lemma 6.1 für die approximierende Bilinearform $\tilde{a}$ bzw. Linearform $\tilde{\ell}$ und erlaubt somit die Abschätzung des Gesamtfehlers $u - \tilde{u}_h$.

Lemma 6.7 *Sei r die Ordnung der Quadraturformel (6.20), Eigenschaft 6.3 oder 6.4 sei mit $k' \geq m$ erfüllt, m der Polynomgrad der finiten Elemente ($\hat{M} = \mathcal{P}_m(E)$), und es gelte $r \geq m$, $r \geq 3$. Weiter sei $a^{\nu\mu} \in H^{r,\infty}(G)$. Dann gelten für die Finite–Element–Approximationen $u_h \in S_h^m$ der Lösung u des Variationsproblems (V) die folgenden Abschätzungen*

$$|(a - \tilde{a})(u_h, v_h)| \leq C h^{r-m+1} \|v_h\|_{1,2} \ , \quad v_h \in S_h^m \tag{6.23}$$

mit nur von u abhängigen Konstanten c, C, sowie

$$|(f, v_h)_0 - \tilde{\ell}(v_h)| \leq C h^{r-m+1} \|v_h\|_{1,2} \ , \quad v_h \in S_h^m \tag{6.24}$$

$$\tilde{a}(v_h) \geq c \|v_h\|_{1,2}^2 \ , \quad v_h \in S_h^m \ . \tag{6.25}$$

Beweis: Für $v_h \in S_h^m$ gilt (vgl. Lemma 6.2)

$$\begin{aligned} |(a - \tilde{a})(u_h, v_h)| &= \left| \sum_{\iota} \sum_{\nu,\mu=1}^{2} \left\{ \int_{G_\iota} a^{\nu\mu} D_\mu u_h D_\nu v_h \, dx \right. \right. \\ &\qquad \left. \left. - q_{G_\iota} \left(a^{\nu\mu} D_\mu u_h D_\nu v_h \right) \right\} \right| \\ &\leq c \, h^r \sum_{\iota} \left| \sum_{\nu,\mu=1}^{2} a^{\nu\mu} D_\mu u_h D_\nu v_h \right|_{r,1;G_\iota} \end{aligned}$$

und bei Beachtung von $\|a^{\nu\mu}\|_{r,\infty} \leq c$ erhält man für die Norm $\|.\|_{r+1,2}$ aus der Schwarzschen Ungleichung

$$|(a - \tilde{a})(u_h, v_h)| \leq c \, h^r \sum_{\iota} \|u_h\|_{r+1,2;G_\iota} \|v_h\|_{r+1,2;G_\iota} \ .$$

Nun ist $w_h \in S_h^m$ auf jedem $G_\iota \in \Delta_h$ aus $\mathcal{P}_{k'}(G_\iota)$, und folglich gilt ($r \geq m$)

$$\|w_h\|_{r+1,2;G_\iota} = \|w_h\|_{m,2;G_\iota} \ .$$

Ferner ergibt die inverse Abschätzung (vgl. Satz 5.22)

$$\|w_h\|_{m,2;G_\iota} \leq c \, h^{1-m} \|w_h\|_{1,2;G_\iota} \ .$$

Damit erhalten wir zunächst

$$\|v_h\|_{r+1,2;G_\iota} \leq c \, h^{1-m} \|v_h\|_{1,2;G_\iota}$$

und

$$|(a-\tilde{a})(u_h,v_h)| \le c\,h^{r-m+1}\|v_h\|_{1,2}\left(\sum_\iota \|u_h\|^2_{m,2;G_\iota}\right)^{1/2} .$$

Mit einer geeigneten Interpolierenden $I_h u \in S_h^m$ zu u (siehe Kapitel 5) ist weiter

$$\begin{aligned}\|u_h\|_{m,2;G_\iota} &\le \|u_h - I_h u\|_{m,2;G_\iota} + \|I_h u\|_{m,2;G_\iota}\\ &\le c\,h^{1-m}\|u_h - I_h u\|_{1,2;G_\iota} + \|I_h u - u\|_{m,2;G_\iota} + \|u\|_{m,2;G_\iota}\\ &\le c\,h^{1-m}\{\|u_h - u\|_{1,2;G_\iota} + \|u - I_h u\|_{1,2;G_\iota}\} + c\|u\|_{m+1,2;G_\iota}\end{aligned}$$

$$\Longrightarrow \qquad \left(\sum_{G_\iota}\|u_h\|^2_{m,2;G_\iota}\right)^{1/2} \le c\,h^{1-m}\|u-u_h\|_{1,2} + c\|u\|_{m+1,2} .$$

Wir setzen dies ein und erhalten mit Hilfe von $\|u-u_h\|_{1,2} = \mathrm{O}(h^m)$ die Abschätzung ($c = c(u)$!)

$$|(a-\tilde{a})(u_h,v_h)| \le c\,h^{r-m+1}\|v_h\|_{1,2} .$$

Analog wird bewiesen

$$\left|(f,v_h) - \tilde{\ell}(v_h)\right| \le c\,h^{r-m+1}\|v_h\|_{1,2} .$$

Aus (6.22) folgt schließlich noch für $v_h \in S_h^m$ wegen $v_h|_{G_\iota} \in \mathcal{P}_m(G_\iota)$, wegen der Elliptizität von $a(.,.)$ und wegen $\omega_i > 0$, daß

$$\begin{aligned}\tilde{a}(v_h) &= \sum_\iota q_{G_\iota}\left(\sum_{\nu,\mu=1}^{2} a^{\nu\mu}D_\mu v_h D_\nu v_h\right) \;\ge\; c\sum_\iota q_{G_\iota}\left(|\nabla_x v_h|^2\right)\\ &\ge c\sum_\iota \int_{G_\iota}|\nabla_x v_h|^2\,dx \;=\; c\|\nabla_x v_h\|^2_{0,2} \;\ge c\|v_h\|^2_{1,2} .\end{aligned}$$

Die letzte Abschätzung benutzt die Poincaré–Friedrichs–Ungleichung für Funktionen aus $S_h^m \subset H_0^{1,2}(G)$. □

Zu der in Lemma 6.7 geforderten Beziehung $r \ge m$ sei noch bemerkt, daß diese im Falle des Vorliegens von Eigenschaft 6.4, d. h. $r = 2k'-1$, schon aus $k' \ge m$ folgt, falls $m \ge 1$ ist. Fordert man Eigenschaft 6.3, dann kann man $k' = m$ wählen.

Lemma 6.1 liefert schließlich die Abschätzung (mit $\tau = r-m+1$)

$$\|u_h - \tilde{u}_h\|_{1,2} = \mathrm{O}(h^{r-m+1}) ,$$

und der Gesamtfehler des Verfahrens ist von der Ordnung

$$\|u - \tilde{u}_h\|_{1,2} = \mathrm{O}(h^m + h^{r-m+1}) \,. \tag{6.26}$$

Konvergenz liegt also vor für $r \geq m$, und die optimale Konvergenzordnung wird erreicht für $r \geq 2m-1$; der Quadraturfehler ist vernachlässigbar, falls $r > 2m-1$ ist.

Kapitel 7

FEM auf nicht polygonalen Gebieten; Isoparametrische Finite Elemente

Bei der bisherigen Untersuchung der FEM haben wir stets vorausgesetzt, daß das zugrunde liegende Gebiet G konvex ist und (im zweidimensionalen Fall) einen Polygonrand ∂G besitzt. Letzeres war erforderlich, um G triangulieren zu können, und die Konvexität sicherte gemäß Satz 2.5 eine gewisse Mindestregularität der schwachen Lösung der Randwertaufgabe (D) : $u \in H_0^{1,2}(G) \cap H^{2,2}(G)$.

Nun sollen einige (von vielen möglichen) Modifikationen der FEM skizziert werden, welche auch die Behandlung von Gebieten mit nicht notwendig konvexem und polygonalem Rand ermöglichen.

7.1 Fall: $G \subset \mathbb{R}^2$ konvex und ∂G aus der Klasse C^2.

Auf dem Rand $\partial G = g(s), s \in [a,b]$, werden Punkte mit Bogenlängenabstand d, $0 < ch \leq d \leq c'h$ (für ein h aus $0 < h \leq 1$) markiert und durch Strecken verbunden. Die dadurch entstehenden Polygonzüge ∂G_h beranden ganz in $\overline{G}$ enthaltene Polygonbereiche G_h (wegen der Konvexität von G): $G_h \subset \overline{G}$ (vgl. Abb. 7.1).

Für diese lassen sich dann wieder Triangulationen Δ_h der üblichen Art konstruieren. Anhand der Abb. 7.2 macht man sich leicht durch Taylorentwicklung von $g(s)$ um s_0 klar, daß gleichmäßig in h gilt

$$\max_{x \in \partial G_h} \operatorname{dist}(x, \partial G) \leq c\, h^2 \ . \tag{7.1}$$

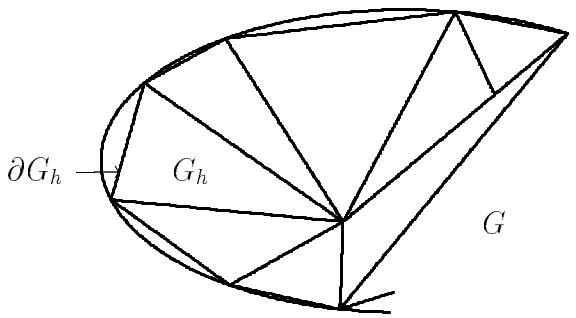


Abbildung 7.1: Approximation von $\partial G \in C^2$ durch Streckenzüge

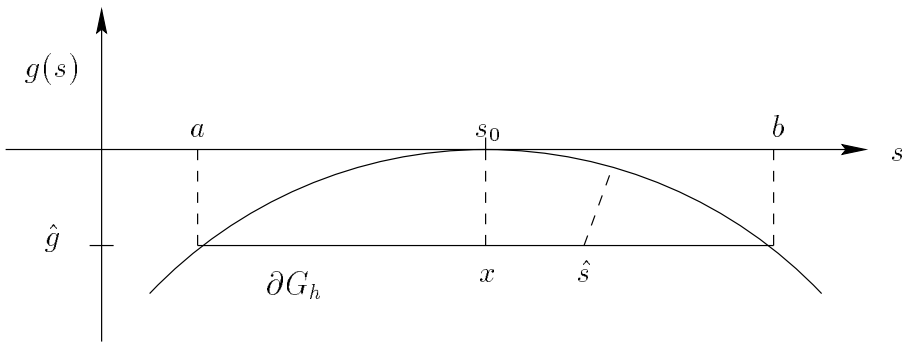


Abbildung 7.2: Randabstand $\mathrm{O}(h^2)$

Beweis von (7.1): Wir betrachten ein Polygonrandstück von ∂G_h, $a \le s \le b$, und suchen zu einem beliebigen Punkt $(\hat{s}, \hat{g}) \in \partial G_h$ den Punkt $(s^*, g(s^*)) \in \partial G$ mit minimalem Abstand zu $(\hat{s}, \hat{g})$. D. h.

$$\varphi(s) = (\hat{s} - s)^2 + (\hat{g} - g(s))^2$$

soll minimiert werden; notwendig dafür ist

$$\begin{aligned} &\varphi'(s^*) = 0 \Longleftrightarrow (\hat{s} - s) + (\hat{g} - g(s))g'(s)|_{s=s^*} = 0 \\ \Longleftrightarrow \quad &\hat{s} - s^* = -g'(s^*)(\hat{g} - g(s^*)) \ . \end{aligned}$$

Letzteres ist gerade die Normalengleichung bei $(s^*, g(s^*))$; die Lösung s^* erhält man iterativ z. B. mit der Methode der sukzessiven Approximation oder dem Newton–Verfahren.

Für den minimalen Abstand erhält man

$$\varphi(s^*) = \left(1 + g'(s^*)^2\right)(\hat{g} - g(s^*))^2 \ .$$

Beachtet man, daß $g(b) = \hat{g}$, und entwickelt um s_0 mit $g'(s_0) = 0$, dann erhält man (s. Abb. 7.2)

$$\begin{aligned} \hat{g} - g(s_0) &= (b - s_0)\underbrace{g'(s_0)}_{=0} - \frac{1}{2}(b - s_0)^2 g''(\tilde{s}_0) \ , \\ g(s_0) - g(s^*) &= (s_0 - s^*)g'(s_0) + \frac{1}{2}(s_0 - s^*)^2 g''(\overline{s}_0) \ , \end{aligned}$$

mit $\tilde{s}_0, \overline{s}_0 \in (a, b)$. Wegen $|b - s_0| \le c'h$, $|s_0 - s^*| \le c'h$ folgt

$$|\hat{g} - g(s^*)| \le |\hat{g} - g(s_0)| + |g(s_0) - g(s^*)| \le (c'h)^2 \max_{a \le s \le b} |g''(s)| \ ,$$

und für den kleinsten Abstand erhält man die Abschätzung

$$\inf_{(s, g(s)) \in \partial G} \left((\hat{s} - s)^2 + (\hat{g} - g(s))^2\right)^{1/2} = \varphi(s^*)^{1/2} \le C\, h^2 \quad \forall (\hat{s}, \hat{g}) \in \partial G_h$$

mit einer nur von c', g', g'' abhängigen Konstanten C. □

Für die FEM mit Teilräumen $S_h^1 \subset H_0^{1,2}(G)$ haben wir nach Satz 2.6

$$\|u - u_h\|_{1,2} \le c\Delta_h^2 \|u\|_{2,2} \tag{7.2}$$

mit

$$\Delta_h^2 = \sup \left\{ \inf_{v_h \in S_h} \frac{\|v - v_h\|_{1,2}}{\|v\|_{2,2}} , \quad v \in H_0^{1,2}(G) \cap H^{2,2}(G) \right\} .$$

Die bzgl. der Triangulationen Δ_h stückweise linearen Interpolierenden $I_h v$ einer Funktion $v \in H_0^{1,2}(G) \cap H^{2,2}(G) \subset C(\overline{G})$ sind aufgrund der Konvexität von G offenbar gleich Null auf ∂G_h. Durch

$$\hat{I}_h v(x) \ := \begin{cases} I_h v(x) \ , \ x \in G_h \\ 0 \quad\quad , x \in G - G_h \end{cases}$$

wird somit eine Funktion $\hat{I}_h v \in H_0^{1,2}(G)$ erklärt.
Für diese gilt gemäß Satz 5.21:

$$\begin{aligned} \|v - \hat{I}_h v\|_{1,2} &= \|v - I_h v\|_{1,2;G_h} + \|v\|_{1,2;G-G_h} \\ &\leq c\, h \|v\|_{2,2} + \|v\|_{1,2;G-G_h} . \end{aligned} \tag{7.3}$$

Wir müssen also das Integral über dem Streifen $G - G_h$ abschätzen. Dazu dient:

Lemma 7.1 *Für $v \in H^{1,1}(G)$ gilt*

$$\int_{G-G_h} |v| \, dx \leq c \max_{x \in \partial G_h} \operatorname{dist}(x, \partial G) \, \|v\|_{1,1} . \tag{7.4}$$

Beweis: Sei $v \in C^{1,1}(G) \subset H^{1,1}(G)$. Da der Rand ∂G aus der Klasse C^2 ist, besitzt das Gebiet G die sog. „Segmenteigenschaft“, d. h.: Es existiert eine endliche Familie $\{O_i\}_{i=1}^L$ offener Mengen und zugehörige normierte Vektoren e_i, so daß mit einem festen $T > 0$ gilt:

i. $\partial G \subset \bigcup_{i=1}^{L} O_i$

ii. $\{\overline{G} \cap O_i + t e_i\} \subset G \quad 0 < t \leq T.$

Für $0 < t \leq T$ definieren wir die Mengen

$$Q_i^t := \bigcup_{0 \leq s \leq t} \{\partial G \cap O_i + s e_i\} .$$

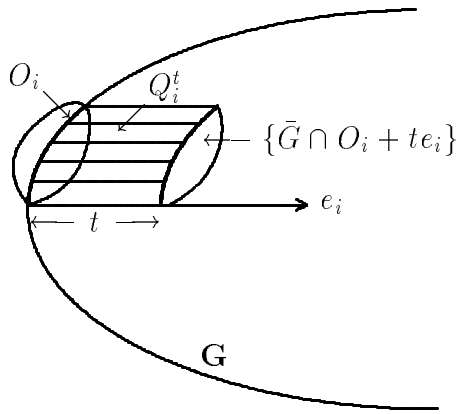


Abbildung 7.3: Segmenteigenschaft

Wegen $d_h := \max_{x \in \partial G_h} \operatorname{dist}(x, \partial G) \longrightarrow 0$ $(h \to 0)$ ist für hinreichend kleines $0 < h \leq h_0$ sicher $d_h \leq T$ und

$$G - G_h \subset \bigcup_{i=1}^{L} Q_i^{\kappa d_h} \tag{7.5}$$

mit einer von h unabhängigen Konstanten $\kappa > 0$.
Für $x \in \partial G \cap O_i$ wird die Funktion erklärt

$$f(t) := v(x + te_i)\,, \quad f \in C^1[0,T]\,.$$

Dann ist mit $s, t \in [0,T]$

$$f(t) - f(s) = \int_s^t f'(r)dr$$

bzw.

$$|f(t)| \leq |f(s)| + \int_0^T |f'(r)|\,dr\,.$$

Integration nach s über $[0,T]$ und anschließend nach t über $[0, \kappa d_h]$ ergibt

$$T|f(t)| \leq \int_0^T |f(s)|\,ds + T\int_0^T |f'(r)|dr$$

und

$$\int_0^{\kappa d_h} |f(t)|\,dt \le d_h \frac{\kappa}{T} \int_0^T |f(s)|\,ds + \kappa d_h \int_0^T |f'(r)|\,dr \ .$$

Wir integrieren nun noch über die Flächenschar $\{\partial G \cap O_i + te_i\,,\ 0 \le t \le T\}$ und erhalten (beachte $|f'(r)| \le |\nabla v(x + re_i)|$)

$$\int_{Q_i^{\kappa d_h}} |v|\,dx \le d_h \frac{\kappa}{T} \int_{Q_i^T} |v|\,dx + \kappa d_h \int_{Q_i^T} |\nabla v|\,dx \ .$$

Summation über alle $Q_i^{\kappa d}$ ergibt dann wegen (7.5)

$$\int_{G-G_h} |v|\,dx \le c(\kappa, T, L) d_h \int_G \{|v| + |\nabla v|\}\,dx \ .$$

Durch das übliche Stetigkeitsargument übertragt sich diese Abschätzung auf ganz $H^{1,1}(G)$. □

Wegen (7.1) erhalten wir mit (7.4)

$$\begin{aligned} \int_{G-G_h} |v|^2\,dx &\le c\,h^2 \int_G \{|v|^2 + |\nabla(|v|^2)|\}\,dx \\ &\le c\,h^2 \int_G \{|v|^2 + 2|v|\,|\nabla v|\}\,dx \\ &\le c\,h^2 \int_G \{|v|^2 + |\nabla v|^2\}\,dx \end{aligned}$$

und analog

$$\int_{G-G_h} |\nabla v|^2\,dx \le c\,h^2 \int_G \{|\nabla v|^2 + |\nabla^2 v|^2\}\,dx \ .$$

Also ist

$$\|v\|_{1,2;G-G_h} \le c\,h \|v\|_{2,2} \ , \tag{7.6}$$

und wir haben bei Kombination mit (7.3) und (7.2) die gewünschte Fehlerordnung

$$\|u - u_h\|_{1,2} \le c\,h \|u\|_{2,2} \ . \tag{7.7}$$

Es sei bemerkt, daß bei Verwendung Finiter Elemente der Ordnung $m \ge 2$ diese Methode nicht die bestmögliche Fehlerordnung liefert, denn wir kommen in (7.1) nicht über $\mathrm{O}(h^2)$ hinaus, (selbst wenn der Rand ∂G noch glatter ist).

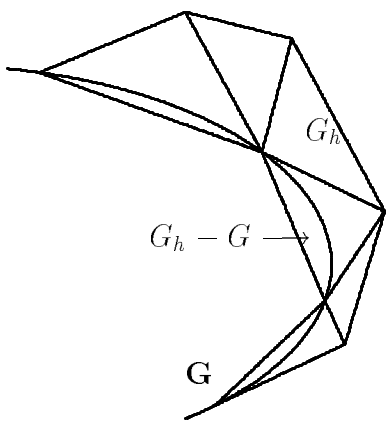


Abbildung 7.4: Fall $G_h \not\subset \overline{G}$.

7.2 Fall: $G \subset \mathbb{R}^2$ nicht konvex und ∂G aus der Klasse C^2.

Wir gehen genau wie im konvexen Fall von Abschnitt 7.1 vor. Diesmal sind die Polygonbereiche G_h jedoch nicht in $\overline{G}$ enthalten:

Offenbar ist wieder

$$\max_{x \in \partial G_h} \operatorname{dist}(x, \partial G) \leq c\, h^2 \ .$$

Der auf übliche Weise auf G_h erklärte Finite–Elemente–Raum $S_h^1 \subset H_0^{1,2}(G_h)$ (stückweise lineare Elemente) ist dann auch nicht in $H_0^{1,2}(G)$ enthalten, $S_h^1 \not\subset H_0^{1,2}(G)$, da die $v_h \in S_h^1$ auf ∂G nicht notwendig verschwinden. Wir definieren daher mit geeigneten Fortsetzungen $\tilde{a}^{\nu\mu}$ und $\tilde{f}$ der Funktionen $a^{\nu,\mu}, f$ auf eine alle G_h , $0 < h \leq 1$, und G umfassende Kugel $B \subset \mathbb{R}^2$ die Bilinearformen

$$a_h(v, w) := \int_{G_h} \sum_{\nu,\mu=1}^{2} \tilde{a}^{\nu\mu} D_\mu v D_\nu w \, dx$$

und die Funktionale

$$\ell_h(v) := \int_{G_h} \tilde{f} v \, dx \ , \quad v, w \in H^{1,2}(G_h) \ .$$

Es sei $\tilde{a}^{\nu\mu}$ so, daß $a_h(v,w)$ gleichmäßig (bzgl. h) beschränkt und elliptisch auf $H_0^{1,2}(G_h)$ ist.

Die Näherungsprobleme sind dann:

Gesucht $u_h \in S_h^1$ *mit*

$$a_h(u_h, v_h) = \ell_h(v_h) \ , \quad \forall v_h \in S_h^1 \ . \tag{V_h}$$

Das übliche Projektionsargument zur Gewinnung von Fehlerabschätzungen ist hier wegen $S_h^1 \not\subset H_0^{1,2}(G)$ nicht anwendbar. Wir können jedoch mit Interpolierenden $I_h\tilde{u} \in S_h^1$ einer geeigneten Fortsetzung $\tilde{u}$ von u auf G_h bzw. B abschätzen (setze $v_h := u_h - I_h\tilde{u}$) :

$$\begin{aligned} c\|v_h\|^2_{1,2;G_h} &\le a_h(v_h) \ = \ a_h(u_h, v_h) - a_h(I_h\tilde{u}, v_h) \\ &= \ell_h(v_h) + a_h(\tilde{u} - I_h\tilde{u}, v_h) - a_h(\tilde{u}, v_h) \\ &\le c\|\tilde{u} - I_h\tilde{u}\|_{1,2;G_h}\|v_h\|_{1,2;G_h} + \\ &\quad + |a_h(\tilde{u}, v_h) - \ell_h(v_h)| \ . \end{aligned}$$

Dies ergibt bei Ausnutzung der Interpolationsabschätzungen für $\tilde{u}$ auf G_h

$$\begin{aligned} \|\tilde{u} - u_h\|_{1,2;G_h} &\le \|\tilde{u} - I_h\tilde{u}\|_{1,2;G_h} + \|v_h\|_{1,2;G_h} \\ &\le c\left\{ h\|\tilde{u}\|_{2,2;G_h} + \sup_{v_h \in S_h^1} \frac{|a_h(\tilde{u}, v_h) - \ell_h(v_h)|}{\|v_h\|_{1,2;G_h}} \right\} . \end{aligned} \tag{7.8}$$

Weiter ist mit dem Operator

$$\tilde{A} := -\sum_{\nu,\mu=1}^{2} D_\nu \left\{ \tilde{a}^{\nu\mu} D_\mu \right\} \quad \text{auf} \quad G_h \ \text{bzw.} \ B$$

bei partieller Integration für beliebiges $v_h \in S_h^1$ (beachte $\tilde{A}\tilde{u} - \tilde{f} = Au - f = 0$ auf G):

$$\begin{aligned} |a_h(\tilde{u}, v_h) - \ell_h(v_h)| &= \left| \int_{G_h} \left\{ \sum_{\nu,\mu=1}^{2} \tilde{a}^{\nu\mu} D_\mu\tilde{u} D_\nu v_h - \tilde{f} v_h \right\} dx \right| \\ &= \left| \int_{G_h} (\tilde{A}\tilde{u} - \tilde{f}) v_h \, dx \right| \ = \ \left| \int_{G_h - G} (\tilde{A}\tilde{u} - f) v_h \, dx \right| \\ &\le \|\tilde{A}\tilde{u} - \tilde{f}\|_{0,2;B} \|v_h\|_{0,2;G-G_h} \ . \end{aligned}$$

Analog zu Lemma 7.1 (vgl. auch (7.6)) läßt sich nun auch in diesem Fall zeigen, daß

$$\|v_h\|_{2;G_h - G} \le c\, h \|v_h\|_{1,2;G_h} \ .$$

Somit folgt

$$|a_h(\tilde{u}, v_h) - \ell_h(v_h)| \leq c\, h \|v_h\|_{1,2;G_h} , \quad v_h \in S_h^1 ,$$

und über (7.8) die gewünschte Fehlerordnung

$$\|\tilde{u} - u_h\|_{1,2;G_h} = \mathrm{O}(h) . \tag{7.9}$$

Durch ein ähnliches Vorgehen läßt sich auch noch die optimale Fehlerordnung für die L^2–Norm sichern

$$\|\tilde{u} - u_h\|_{0,2;G_h} = \mathrm{O}(h^2) , \tag{7.10}$$

was hier jedoch nicht ausgeführt werden soll.

7.3 Fall: $G \subset \mathbb{R}^2$ nicht notwendig konvex und ∂G aus der Klasse C^{m+1}, $m \geq 1$.

Die einfachen Randapproximationen in 7.1 und 7.2 waren lediglich von der Ordnung $\mathrm{O}(h)$ und damit für Finite Elemente höherer Ordnung $m \geq 2$ unbrauchbar. In diesem Falle bildet man Triangulationen Δ_h von $\overline{G}$, wobei die Dreiecke $e(= G_\iota) \in \Delta_h$ entlang des Randes ∂G als gekrümmt zugelassen sind. Auf diese Weise wird ganz $\overline{G}$ ausgeschöpft:

Wir setzen voraus, daß sich jedes Dreieck e vermittels einer umkehrbar eindeutigen (diesmal nicht notwendig linearen!) Transformation

$$\phi : \mathbb{R}^2 \longrightarrow \mathbb{R}^2 , \quad \phi \in C^m ,$$

auf das Einheitsdreieck $E \subset \mathbb{R}^2$ abbilden läßt:

$$E = \phi(e) .$$

Genügen die Dreiecke wieder der Bedingung

$$\mathrm{diam}(e) \leq c\, h , \quad \text{Inkreisradius } \rho(e) \geq c' h ,$$

($0 < c' < c$ unabhängig von h), so gilt für die Funktionalmatrizen ϕ' der Transformationen ϕ gleichmäßig für $e \in \Delta_h$ und $0 < h \leq 1$:

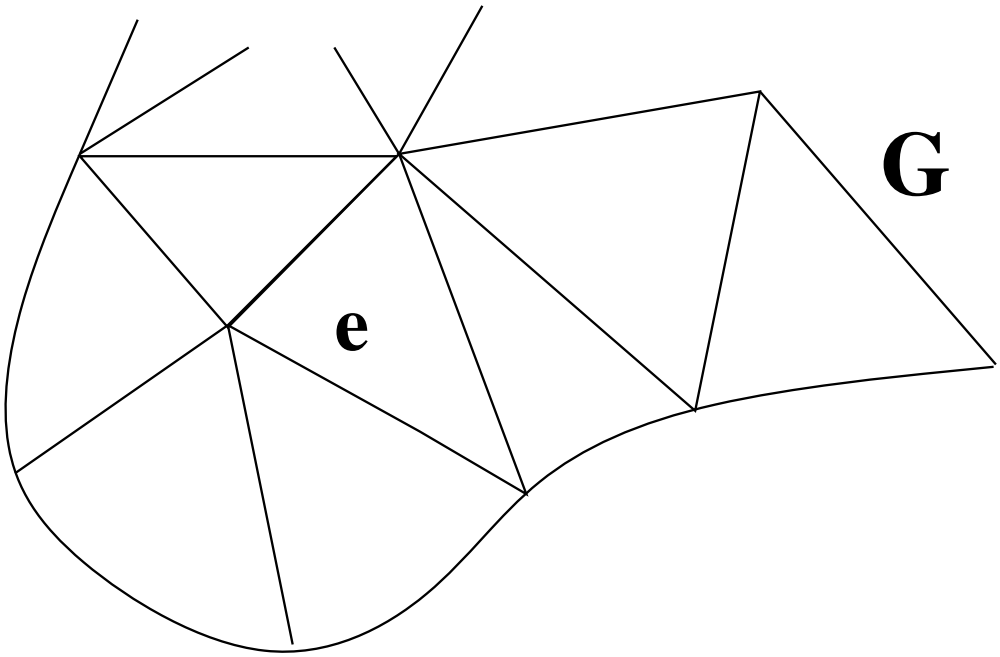


Abbildung 7.5: Triangulierung mit krummlinigen Seiten

$$c\,h^2 \le |\det \phi'| \le c' h^2 \ , \tag{7.11}$$

wobei (vgl. (5.78), (5.80), (5.81))

$$\begin{aligned} |D_z^\alpha \phi(z)| &= \mathrm{O}(h^{|\alpha|}) \ , \quad 0 \le |\alpha| \le m \\ |D_x^\alpha \phi^{-1}(x)| &= \mathrm{O}(h^{-|\alpha|}) \ , \ 0 \le |\alpha| \le m \ . \end{aligned} \tag{7.12}$$

Zur Konstruktion solcher Abbildungen ϕ siehe insbes. [Cia78, 4.3]. Wesentlich dabei ist das Vorliegen expliziter (lokaler) Parameterdarstellungen des Randes ∂G, was in der Praxis meistens der Fall ist (z. B.: ∂G Kreisbogen oder anderer Kegelschnitt).

Zu jedem Polynom $\hat{q} \in \mathcal{P}_m(E)$ wird dann durch

$$q(x) = \hat{q}(z) \ , \quad z = \phi(x) \ , \quad x \in e \ ,$$

eine Funktion $q \in C^m(e)$ (i. a. kein Polynom!) definiert. Die Finite–Elemente–Räume $S_h^m \subset H_0^{1,2}(G)$ erklären wir durch

$$S_h^m := \left\{ v \in C(\overline{G}) |\ \hat{v}\Big|_E(z) = v\Big|_e(\phi^{-1}(z)) \in \mathcal{P}_m(E) \ , \quad e \in \Delta_h \ , v\Big|_{\partial G} = 0 \right\} .$$

Die Elemente $v \in S_h^m$ werden in der Literatur „isoparametrische" Finite Elemente genannt. Durch Transformation von e auf E und Beachtung der Beziehungen (7.11), (7.12) gewinnt man dann analog zu dem Vorgehen in Abschnitt 5.3 für die Interpolierende $I_h u \in S_h^m$ zu u die Abschätzung:

$$\|u - I_h u\|_{k,p;e} \leq c\, h^{m+1-k} \|u\|_{m+1,p;e} \;, \quad 0 \leq k \leq m \,,\; 1 \leq p \leq \infty \;. \tag{7.13}$$

Für die approximierenden Probleme mit isoparametrischen Finiten Elementen,

Gesucht $u_h \in S_h^m$ *mit*

$$a(u_h, v_h) = (f, v_h) \;, \quad \forall v_h \in S_h^m \;, \tag{V_h}$$

erhalten wir damit ebenfalls die Fehlerabschätzungen

$$\|u - u_h\|_{k,2} \leq c\, h^{m+1-k} \|u\|_{m+1,2} \;, \quad k = 0, 1 \;. \tag{7.14}$$

Zur Aufstellung der Gesamtsteifigkeitsmatrix $K = (a(w_i, w_k))_{i,k=1}^N$ für isoparametrische Finite Elemente ist i. a. numerische Integration erforderlich, denn die Basisfunktionen $w_i \in S_h^m$ sind auf den krummberandeten Dreiecken keine Polynome mehr.

Kapitel 8

Finite–Element–Verfahren für zeitabhängige Probleme

In diesem letzten Abschnitt soll die Idee zur Herleitung von Finite–Element–Verfahren zur näherungsweisen Lösung von parabolischen und hyperbolischen Differentialgleichungen dargestellt werden. Außerdem werden erste Stabilitätsabschätzungen für die sequentiell in Zeitrichtung fortschreitenden Näherungslösungen angegeben. Es zeigt sich, daß sich der Gesamtfehler in einen Zeit- und einen Ortsanteil aufspalten läßt; für letzteren benötigt man wieder die in den vorherigen Kapiteln hergeleiteten Interpolationsabschätzungen.

8.1 Parabolische Probleme

Wir betrachten als Modellbeispiel für parabolische Probleme die (im Ort eindimensionale) Wärmeleitungsgleichung

$$u_t(x,t) = a u_{xx}(x,t) + F(x,t), \quad x \in (0,1),\ t \in (0,T] \tag{8.1}$$

mit homogenen Dirichletschen Randbedingungen

$$u(0,t) = u(1,t) = 0, \quad t \in (0,T], \tag{8.2}$$

und Anfangsbedingung

$$u(x,0) = g(x), \quad x \in (0,1) \tag{8.3}$$

Hierbei sei a der Einfachheit halber eine positive Konstante, $a > 0$.

Eine Lösung u von (8.1) erfüllt (nach partieller Integration) offenbar auch die Gleichungen

$$(u_t, \varphi) + a(u', \varphi') = (F, \varphi), \quad \varphi \in V,\ t \in (0, T], \tag{8.4}$$

$$(u, \varphi)_{t=0} = (g, \varphi) \quad \forall \varphi \in V, \tag{8.5}$$

wobei $(.,.)$ das L^2–Skalarprodukt und $V = H_0^1(0,1)$ bezeichnet. Man nennt Gleichung (8.4) in Anlehnung an entsprechende Begriffsbildungen bei elliptischen Gleichungen auch die „variationelle Formulierung“ der Wärmeleitungsgleichung. Der Strich bezeichnet die verallgemeinerte Ableitung nach x. Eine Lösung von (8.4), (8.5) heißt auch *schwache Lösung* der Wärmeleitungsgleichung . Zur Approximation durch ein Finite–Elemente–Verfahren zerlegt man wie beim Ritz–Galerkin–Verfahren in Abschnitt 1.2 und 2.2 das Intervall $[0,1]$ in nicht notwendig äquidistante Teilintervalle $I_j = (x_{j-1}, x_j),\ j = 1, \ldots, J$,

$$\Delta:\ 0 = x_0 < x_1 < \ldots < x_J = 1.$$

Der Einfachheit halber suchen wir eine Näherungslösung wieder im Raum $S_h^1 = \mathcal{P}_1^1(\Delta) \cap H_0^1(0,1)$ der stetigen, stückweise linearen Funktionen. Beim *halbdiskreten Galerkin–Verfahren* wird eine stetig differenzierbare Funktion $v:\ [0,T] \to S_h^1$ gesucht, die

$$(v_t, \varphi) + a(v', \varphi') = (F, \varphi), \quad \varphi \in S_h^1,\ t \in (0, T], \tag{8.6}$$

erfüllt. Für $v(0)$ wählt man die L_2–Projektion von g in S_h^1,

$$(v, \varphi)_{t=0} = (g, \varphi), \quad \varphi \in S_h^1, \tag{8.7}$$

oder eine andere Näherung von g. Mit den Dachfunktionen $\varphi_j,\ j = 1, \ldots, J-1$, als Basis von S_h^1 läßt sich v als

$$v(x,t) = \sum_{i=1}^{J-1} c_i(t)\varphi_i(x), \quad 0 \le x \le 1,\ 0 \le t \le T,$$

mit stetig differenzierbaren Koeffizientenfunktionen c_j schreiben. Setzt man diese Darstellung von v in (8.6), (8.7) ein, so ergibt sich $c(t) = (c_1(t), \ldots, c_{J-1}(t))$ als

Lösung des folgenden Anfangswertproblems für ein System gewöhnlicher Differentialgleichungen:

$$\sum_{i=1}^{J-1} c_i'(t)(\varphi_i, \varphi_j) + a \sum_{i=1}^{J-1} c_i(t)(\varphi_i', \varphi_j') = (F(t), \varphi_j) \ , \qquad (8.8)$$
$$0 < t \le T, \ j = 1, \ldots, J-1 \ ,$$

$$\sum_{i=1}^{J-1} c_i(0)(\varphi_i, \varphi_j) = (g, \varphi_j), \quad j = 1, \ldots, J-1 \ . \qquad (8.9)$$

Da man zunächst nur in x-Richtung diskretisiert hat und t noch kontinuierlich durchläuft, nennt man die durch (8.9), (8.9) beschriebene Vorgehensweise auch die *Linienmethode*.

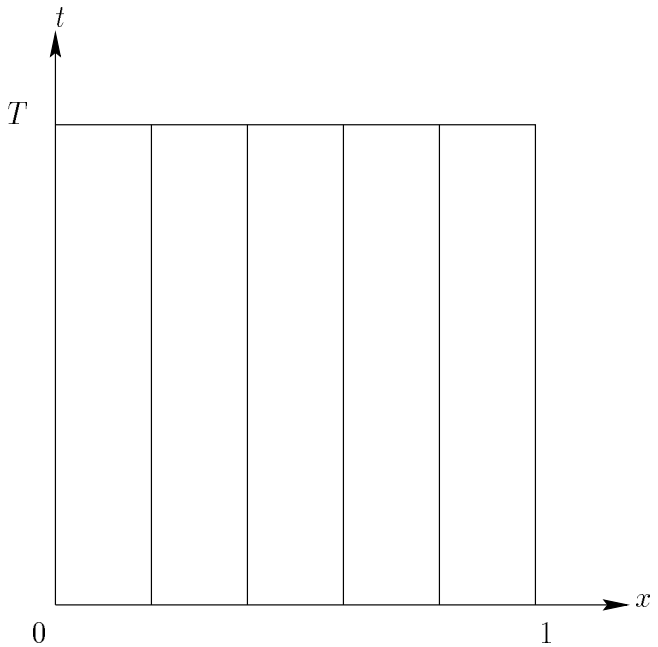


Abbildung 8.1: Linienmethode

Bei Diskretisierung auch in t–Richtung durch eine äquidistante Schrittweite $k = T/N$, $t_n = nk$, $n = 1, \ldots, N$, ist eine geeignete Approximation der Ableitung nach t zu wählen. Das *Crank–Nicolson–Galerkin–Verfahren* (Abk.: CNG–Verfahren) wählt entsprechend dem Crank–Nicolson–Verfahren (vgl. [MM81, I.6]) das arithmetische Mittel des vorwärts- und rückwärtsgenommenen Differenzenquotienten erster Ordnung. Gesucht sind dabei also Funktionen $u_n \in S_h^1$, $n = 0, \ldots, N$, die

$$\left(\frac{u_{n+1} - u_n}{k}, \varphi\right) + a\left(\frac{1}{2}(u_{n+1}' + u_n'), \varphi'\right) \qquad (8.10)$$

$$= \left(\frac{1}{2}(F_{n+1} + F_n), \varphi\right), \quad \varphi \in S_h^1,\ n = 0, \ldots, N-1,$$

$$(u_0, \varphi) = (g, \varphi), \quad \varphi \in S_h^1, \tag{8.11}$$

erfüllen. Hierbei bezeichnet $F_n = F(., t_n)$. Mit Hilfe der Basisfunktionen $\varphi_i,\ i = 1, \ldots, J-1$, läßt sich jedes u_n als

$$u_n(x) = \sum_{i=1}^{J-1} u_n(x_i)\varphi_i(x), \quad 0 \le x \le 1,$$

darstellen. Das Gleichungssystem (8.11), (8.11) läßt sich daher auch in der Form

$$\begin{aligned} &\frac{1}{k}\sum_{i=1}^{J-1}(v_{i,n+1} - v_{i,n})(\varphi_i, \varphi_j) + a\sum_{i=1}^{J-1} v_{i,n+\frac{1}{2}}(\varphi_i', \varphi_j') \\ &= (F_{n+\frac{1}{2}}, \varphi_j), \quad j = 1, \ldots, J-1,\ n = 0, 1, \ldots, N-1, \end{aligned} \tag{8.12}$$

$$\sum_{i=1}^{J-1} v_{i,0}(\varphi_i, \varphi_j) = (g, \varphi_j), \quad j = 1, \ldots, J-1, \tag{8.13}$$

schreiben, wobei $v_{i,\nu} = u_\nu(x_i),\ \nu = n, n+\frac{1}{2}$, und $u_{n+\frac{1}{2}} = (u_n + u_{n+1})/2$. Die Koeffizienten dieses Gleichungssystems sind bereits in Abschnitt 1.2 berechnet worden. Wir beschränken uns im folgenden auf äquidistante Schrittweiten $h_i = h,\ i = 1, \ldots, J$.

$$(\varphi_i, \varphi_j) = \begin{cases} h/6 & , i = j-1, j+1, \\ 2h/3 & , i = j, \\ 0 & \text{sonst}, \end{cases} \qquad (\varphi_i', \varphi_j') = \begin{cases} -1/h & , i = j-1, j+1, \\ 2/h & , i = j, \\ 0 & \text{sonst}. \end{cases}$$

Nach Multiplikation mit k/h hat (8.12), (8.13) dann die Form

$$\begin{aligned} &\left(\frac{2}{3} + ar\right) v_{j,n+1} + \left(\frac{1}{6} - \frac{1}{2}ar\right)(v_{j+1,n+1} + v_{j-1,n+1}) \\ &= \left(\frac{2}{3} - ar\right) v_{j,n} + \left(\frac{1}{6} + \frac{1}{2}ar\right)(v_{j+1,n} + v_{j-1,n}) + \frac{k}{h}(F_{n+\frac{1}{2}}, \varphi_j), \\ &\quad j = 1, \ldots, J-1,\ n = 0, 1, \ldots, N-1, \end{aligned} \tag{8.14}$$

$$\frac{1}{6}(v_{j-1,0} + 4v_{j,0} + v_{j+1,0}) = \frac{1}{h}(g, \varphi_j), \quad j = 1, \ldots, J-1. \tag{8.15}$$

Hier ist $r = \frac{k}{h^2}$ das für parabolische Probleme typische Schrittweitenverhältnis.

Die Skalarprodukte auf den rechten Seiten kann man wie beim Ritzschen Verfahren durch Quadraturformeln, z.B. die Simpson–Formel, auswerten. Für das Gleichungssystem ist das starke Zeilensummenkriterium erfüllt, und es läßt sich iterativ oder mit Hilfe des Gaußschen Eliminationsverfahrens für Tridiagonalmatrizen lösen.

Die Lösbarkeit des Crank–Nicolson–Galerkin–Verfahrens (8.11), (8.11) läßt sich auch allgemein für beliebige Basisfunktionen $w_1, \ldots, w_m$ eines endlichdimensionalen Teilraums $M \subset V$ zeigen. Mit der Bilinearform

$$b(u,v) := (u,v) + \frac{1}{2}ak(u',v'), \quad u,v \in V,$$

und

$$\ell_n(v) := (u_n,v) - \frac{1}{2}ak(u_n',v') + k(F_{n+\frac{1}{2}},v), \quad v \in V,$$

ist das Crank–Nicolson–Galerkin–Verfahren äquivalent zu

$$b(u_{n+1},\varphi) = \ell_n(\varphi), \quad \varphi \in M, \tag{8.16}$$

oder, mit $u_n = \sum_{i=1}^m c_i^{(n)} w_i$,

$$\sum_{i=1}^m c_i^{(n+1)} b(w_i,w_j) = \ell_n(w_j), \quad j = 1,\ldots,m. \tag{8.17}$$

Die symmetrische, beschränkte Bilinearform $b(.,.)$ ist auch elliptisch auf V bzgl. $\|.\|_1$, denn

$$|b(u,u)| = |u|_0^2 + \frac{1}{2}ak|u'|_0^2 \geq \min\left(1, \frac{ak}{2}\right)\|u\|_1^2.$$

Außerdem ist ℓ_n ein beschränktes Funktional auf V bzgl. $\|.\|_1$, so daß nach dem Lemma von Lax–Milgram eine eindeutige Lösung $u_{n+1} \in M$ existiert. Aus der Elliptizität von $b(.,.)$ folgt die positive Definitheit der Matrix $(b(w_i,w_j))$, was auch die eindeutige Lösbarkeit von (8.16) sichert.

Satz 8.1 *Die Lösungen des CNG–Verfahrens (8.11), (8.11) erfüllen die folgende Stabilitätsabschätzung*

$$\max_{0\leq n\leq N} |u_n|_1 \leq \gamma \left\{ |u_0|_1 + \left(k \sum_{\nu=0}^{N-1} |F_{\nu+\frac{1}{2}}|_0^2 \right)^{\frac{1}{2}} \right\} \tag{8.18}$$

wobei $\gamma = \max(1, \frac{1}{2a})$, $F_{\nu+\frac{1}{2}} = \frac{1}{2}(F_\nu + F_{\nu+1})$.

Beweis: (s. [Rei85, Thm. 12.14]) □

Wählt man U_n als Ritz–Galerkin–Approximation von $U(.,t_n)$, d. h. $(aU_n', v') = \left(a\frac{\partial}{\partial x}u(.,t_n), v'\right)$, $v \in V_h$, und setzt den Fehler $U_n - u_n$ in (8.18) ein, dann lassen sich die auf der rechten Seite auftretenden Terme analysieren und abschätzen, und man kann für den Gesamtfehler zeigen, daß (s. [Rei85, Thm. 13.11])

$$\max_{0\le n\le N} |u(.,t_n) - u_n|_1 = \mathrm{O}(k^2 + h) \tag{8.19}$$

Der Anteil $\mathrm{O}(h)$ stammt von dem Interpolationsfehler in der Ortsvariablen. Nimmt man noch — wie üblich bei parabolischen Problemen — an, daß k/h^2 konstant bleibt, dann überwiegt der Ortsanteil in (8.19). Um eine Ausgewogenheit zwischen Orts- und Zeitanteil des Fehlers zu erreichen, müßten im Ort Polynome höheren Grades $S_h^\mu = \mathcal{P}_\mu(\Delta)$ mit $\mu > 1$ gewählt werden — optimal wäre $\mu = 4$.

8.2 Hyperbolische Probleme

Die Wellengleichung (vgl. Anhang B.1) ist das Modellbeispiel für hyperbolische Differentialgleichungen . Zu deren numerischer Lösung gibt es wieder zahlreiche Möglichkeiten der Approximation durch Differenzenverfahren (vgl. z. B. [MM81, Rei85]) sowie durch Finite–Element–Verfahren. Die Konstruktion von letzteren werden wir hier an einem Beispiel darstellen.

Das *Anfangswertproblem* (oder *Cauchy–Problem*) *für die Wellengleichung* lautet:

$$u_{tt} - c^2 u_{xx} = 0, \quad -\infty < x < \infty,\ t > 0,\ (c > 0) \tag{8.20}$$

$$\begin{cases} u(x,0) = f(x), \\ u_t(x,0) = g(x), \end{cases} \quad -\infty < x < \infty. \tag{8.21}$$

Die Lösung von (8.20), (8.21) ist bekanntlich

$$u(x,t) = \frac{1}{2}(f(x+ct) + f(x-ct)) + \frac{1}{2c}\int_{x-ct}^{x+ct} g(s)\,ds$$

für einmal bzw. zweimal stetig differenzierbares f bzw. g. Man sieht leicht, daß (8.20), (8.21) äquivalent zu dem folgenden System ist:

$$\begin{cases} u_t - cv_x = 0, \\ v_t - cu_x = 0, \end{cases} \quad -\infty < x < \infty,\ t > 0, \tag{8.22}$$

$$\begin{cases} u(x,0) = f(x), \\ v(x,0) = \dfrac{1}{c}\displaystyle\int_0^x g(s)\,ds + C, \end{cases} \quad -\infty < x < \infty, \tag{8.23}$$

wobei C eine beliebige Konstante bezeichnet. Mit $\underline{u} = (u, v)$ und der Matrix $A = c\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ läßt sich (8.22) als

$$\frac{\partial}{\partial t}\underline{u} = A\frac{\partial}{\partial x}\underline{u}, \quad -\infty < x < \infty,\ t > 0, \tag{8.24}$$

schreiben. Durch Addition und Subtraktion der zwei Gleichungen in (8.22) erhält man noch das folgende äquivalente Gleichungssystem für $\underline{w} = (w_1, w_2),\ w_1 = u + v,\ w_2 = u - v$:

$$\frac{\partial}{\partial t}\underline{w} = D\frac{\partial}{\partial x}\underline{w}, \quad -\infty < x < \infty,\ t > 0 \tag{8.25}$$

wobei $D = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. Man bezeichnet das System (8.25) auch als *charakteristische Form der Wellengleichung* und $u \pm v$ als *charakteristische Variablen*. Beim Übergang von (8.24) nach (8.25) hat man nichts anderes gemacht, als die Matrix A mit Hilfe einer orthogonalen Matrix (bestehend aus den Eigenvektoren von A) diagonalisiert. Es ist hier nämlich $D = PAP^{-1}$ mit

$$P^{-1} = P = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Das System (8.24) oder (8.25) ist der Ausgangspunkt zur Herleitung von zahlreichen Differenzenapproximationen.

Am Beispiel der *verallgemeinerten Wellengleichung*

$$u_{tt} - (au_x)_x = s(x,t), \quad x \in (0,1),\ t \in (0,T], \tag{8.26}$$

mit Anfangsbedingungen

$$u(x,0) = f(x), \quad u_t(x,0) = g(x), \quad x \in (0,1), \tag{8.27}$$

und homogenen Dirichletschen Randbedingungen

$$u(0,t) = u(1,t) = 0, \quad t \in (0,T], \tag{8.28}$$

wollen wir ein Finite–Element–Verfahren konstruieren. Hierbei sei die Funktion $a = a(x)$ nur von der Ortsvariablen abhängig und genüge $0 < \alpha_0 \le a(x) \le \alpha_1,\ x \in (0,1)$. Wir nehmen an, daß die Lösung von (8.26)–(8.28) existiert und bzgl. t und x (mindestens) zweimal stetig differenzierbar ist.

Nach partieller Integration sieht man, daß jede Lösung von (8.26) und (8.28) die folgende *variationelle Formulierung* der verallgemeinerten Wellengleichung erfüllt:

$$(u_{tt}, \varphi) + (au', \varphi') = (s, \varphi), \quad \varphi \in V. \tag{8.29}$$

Die Bezeichnungen seien wie beim Crank–Nicolson–Galerkin–Verfahren gewählt. Beim zugehörigen *halb–diskreten Galerkin–Verfahren* wird eine zweimal (bzgl. t) stetig differenzierbare Funktion $v : [0,T] \to V_h = \mathcal{P}_1^1(\Delta) \cap H_0^1(0,1)$ gesucht, die

$$(v_{tt}, \varphi) + (av', \varphi') = (s, \varphi), \quad \varphi \in V_h,\ t \in (0,T], \tag{8.30}$$

erfüllt. Als zeitdiskrete Finite–Elemente–Approximation sucht man entsprechend Dupont[1)] Funktionen $u_0, u_1, \ldots, u_J \in V_h$, so daß

$$(D_t^2 u_n, \varphi) + \frac{1}{2}(a(u'_{n+\frac{1}{2}} + u'_{n-\frac{1}{2}}), \varphi') = \frac{1}{2}(s_{n+\frac{1}{2}} + s_{n-\frac{1}{2}}, \varphi), \quad \varphi \in V_h. \tag{8.31}$$

Hierbei bezeichnet D_t^2 den zentralen Differenzenquotienten 2. Ordnung in t-Richtung. Definiert man noch

$$v_{n,\Theta} = \Theta v_{n+1} + (1 - 2\Theta) v_n + \Theta v_{n-1}, \quad 0 \le \Theta \le 1,$$

dann ist

$$\frac{1}{2}(v_{n+\frac{1}{2}} + v_{n-\frac{1}{2}}) = v_{n,\frac{1}{4}}.$$

Mit den Dachfunktionen $\varphi_j,\ j = 1, \ldots, J-1$, als Basis in V_h ist (8.31) äquivalent zu

$$(D_t^2 u_n, \varphi_j) + (au'_{n,\frac{1}{4}}, \varphi_j') = (s_{n,\frac{1}{4}}, \varphi_j), \quad j = 1, \ldots, J-1. \tag{8.32}$$

[1)] DUPONT, J., L^2–estimates for Galerkin methods for second order hyperbolic equations. SIAM J. Numer. Anal. 10(1973), 880–889

Benutzt man wieder die Darstellung $u_n = \sum_{j=1}^{J-1} u_n(x_j)\varphi_j$, dann erhält man als Gleichungssystem für $v_{j,n} = u_n(x_j)$

$$\begin{aligned}&\frac{1}{k^2}\sum_{i=1}^{J-1}(v_{i,n-1} - 2v_{i,n} + v_{i,n+1})(\varphi_i,\varphi_j) + \sum_{i=1}^{J-1} v_{i,(n,\frac{1}{4})}(a\varphi_i',\varphi_j')\\ &= (s_{n,\frac{1}{4}},\varphi_j), \quad j = 1,\ldots,J-1.\end{aligned} \tag{8.33}$$

Für die Wellengleichung mit $a(x) = c^2 > 0$ und $s \equiv 0$ ergibt sich nach Einsetzen der Koeffizienten und Multiplikation mit k^2/h das folgende Gleichungssystem:

$$\begin{aligned}&\frac{1}{6}(v_{j-1,n+1} + 4v_{j,n+1} + v_{j+1,n+1}) - \frac{1}{3}(v_{j-1,n} + 4v_{j,n} + v_{j+1,n})\\ &\quad + \frac{1}{6}(v_{j-1,n-1} + 4v_{j,n-1} + v_{j+1,n-1})\\ &\quad - \lambda^2c^2(v_{j-1,(n,\frac{1}{4})} - 2v_{j,(n,\frac{1}{4})} + v_{j+1,(n,\frac{1}{4})}) \;=\; 0, \quad j = 1,\ldots,J-1.\end{aligned} \tag{8.34}$$

Hierbei bezeichnet $\lambda = \frac{k}{h}$ das zugehörige *Schrittweitenverhältnis*.

Wir bringen alle Terme der n-ten und $(n-1)$-ten Stufe auf die rechte Seite und erhalten

$$\begin{aligned}&\left(\frac{2}{3} + \frac{1}{2}\lambda^2c^2\right)v_{j,n+1} + \left(\frac{1}{6} - \frac{1}{4}\lambda^2c^2\right)(v_{j-1,n+1} + v_{j+1,n+1})\\ &= \left(\frac{4}{3} - \lambda^2c^2\right)v_{j,n} + \left(\frac{1}{3} + \frac{1}{2}\lambda^2c^2\right)(v_{j-1,n} + v_{j+1,n}) - \left(\frac{2}{3} + \frac{1}{2}\lambda^2c^2\right)v_{j,n-1}\\ &\quad - \left(\frac{1}{6} - \frac{1}{4}\lambda^2c^2\right)(v_{j-1,n-1} + v_{j+1,n-1}), \quad j = 1,\ldots,J-1.\end{aligned} \tag{8.35}$$

Für Stabilitätsabschätzungen ist es günstig, das vorliegende *3–Stufen–Verfahren*, d. h. es werden Werte von 3 t-Stufen benötigt, als 2–Stufen–Verfahren zu schreiben. Das geschieht trivialerweise durch die Setzung $w_{j,n} = v_{j,n-1}$. Damit ist (8.35) äquivalent zu folgendem System von 2 Gleichungen:

$$\begin{aligned}&\left(\frac{2}{3} + \frac{1}{2}\lambda^2c^2\right)v_{j,n+1} + \left(\frac{1}{6} - \frac{1}{4}\lambda^2c^2\right)(v_{j-1,n+1} + v_{j+1,n+1})\\ &= \left(\frac{4}{3} - \lambda^2c^2\right)v_{j,n} + \left(\frac{1}{3} + \frac{1}{2}\lambda^2c^2\right)(v_{j-1,n} + v_{j+1,n})\\ &\quad - \left(\frac{2}{3} + \frac{1}{2}\lambda^2c^2\right)w_{j,n} - \left(\frac{1}{6} - \frac{1}{4}\lambda^2c^2\right)(w_{j-1,n} + w_{j+1,n}),\\ &w_{j,n+1} \;=\; v_{j,n}, \quad j = 1,\ldots,J-1.\end{aligned} \tag{8.36}$$

Die zugehörige Matrix erfüllt wieder das starke Zeilensummenkriterium. Für jedes $n = 1, 2, \ldots, N-1$ kann daher das Gleichungssystem mit dem Gaußschen Eliminationsverfahren für Tridiagonalmatrizen oder iterativ gelöst werden. Die Anfangsnäherungen u_0 und u_1 müssen später durch geeignete Approximationen der Anfangsbedingungen (8.21) festgelegt werden.

Mit dem vorwärtsgenommenen Differenzenquotienten erster Ordnung in Zeitrichtung D_t^+ und der L_2–Projektion S_n von $s(., t_n)$ läßt sich für das obige Verfahren die folgende Stabilitätsabschätzung beweisen (s. [Rei85, Thm. 12.15]),

$$\begin{aligned} &\max_{0 \le n \le N-1} (|D_t^+ u_n|_0^2 + |u_{n+\frac{1}{2}}|_1^2) \\ &\le \gamma^2 \left(|D_t^+ u_0|_0^2 + |u_{\frac{1}{2}}|_1^2 + k \sum_{\nu=1}^{N-1} |S_{\nu,\frac{1}{4}}|_0^2 \right), \; n \ge n_0 \; , \end{aligned} \tag{8.37}$$

mit $\gamma^2 = 2e^{4T} \max(1, \alpha_1)/\min(1, \alpha_0)$. Hieraus kann man wieder durch Einsetzen eines geeigneten Fehlerterms die Konvergenz des Finite–Element–Verfahrens mit einer gewissen Konvergenzordnung zeigen, wobei wieder Differenzierbarkeitsvoraussetzungen an die Lösung u gestellt werden müssen (s. [Rei85, Chapt. 13.2]).

Anhang A

Theorie gewöhnlicher Differentialgleichungen

Gegenstände dieses Abschnittes des Anhangs sind

- Existenz und Eindeutigkeit von Systemen von Anfangswertproblemen 1. Ordnung
- Lösungsmethoden für Systeme linearer Differentialgleichungen, auch lineare Differentialgleichungen n-ter Ordnung
- Lösungen von Randwertproblemen für Differentialgleichungen n-ter Ordnung

Die Darstellung erfolgt nach [Wal76].

Wir betrachten Anfangswertprobleme (Abk.: AWPs) und Randwertprobleme (RWPs) der folgenden Form:

(AWP) *Gesucht eine stetig differenzierbare Funktion*

$$y : J := [\xi, \xi + a] \ni x \to \mathbb{K}^n, \quad \mathbb{K} = \mathbb{R} \text{ } bzw. \text{ } \mathbb{K} = \mathbb{C},$$

mit

$$y(\xi) = \eta, \quad y'(x) = f(x, y(x)), \quad x \in J, \tag{A.1}$$

wobei $\eta \in \mathbb{K}^n, \quad f : J \times \mathbb{K}^n \to \mathbb{K}^n$.

(RWP, 2. Ordnung) *Gesucht* $u \in C^2[a, b]$ *mit*

$$u''(x) + p(x)u'(x) + q(x)u(x) = f(x), \quad x \in [a,b], \tag{A.2}$$

$$\alpha_1 u(a) + \alpha_2 u'(a) = \eta_0, \quad \beta_1 u(b) + \beta_2 u'(b) = \eta_1,$$

wobei $p, q, f \in C[a,b]$, $\quad \alpha_1^2 + \alpha_2^2 > 0, \quad \beta_1^2 + \beta_2^2 > 0.$

Allgemeiner betrachten wir noch lineare Differentialgleichungen n-ter Ordnung oder Systeme von n Differentialgleichungen erster Ordnung, bei denen dann n Randbedingungen gefordert werden.

A.1 Existenz und Eindeutigkeit von Lösungen von AWPs

Wir betrachten folgendes AWP

$$y(\xi) = \eta, \quad y'(x) = f(x, y(x)), \quad x \in J, \tag{A.3}$$

wobei die betrachteten Funktionen zunächst reellwertig seien.

Unter einer „Lipschitzbedingung“ $(L \geq 0)$

$$|f(x,y_1) - f(x,y_2)| \leq L|y_1 - y_2|, \quad (x,y_1),(x,y_2) \in S \tag{A.4}$$

in einem „Streifen“ $S := \{(x,y) | \xi \leq x \leq \xi + a, -\infty < y < \infty\}$ gilt der

Satz A.1 (Satz von PICARD-LINDELÖF**)** *Die Funktion $f \in C(S)$ genüge der Lipschitzbedingung (A.4). Dann hat das AWP (A.3) genau eine Lösung y, die in ganz $J = [\xi, \xi + a]$ existiert.*

Der Beweis beruht auf dem Banachschen Fixpunktsatz, indem man zeigt, daß das AWP äquivalent ist zur Lösung der Fixpunktgleichung

$$y = Ty \text{ mit } (Ty)(x) := \eta + \int_\xi^x f(t, y(t))\, dt, \quad x \in J,$$

und der Operator T bezüglich der Maximumnorm eine Kontraktion darstellt. Zusätzlich weiß man, daß die Methode der sukzessiven Approximation eine Folge

$$y_{k+1}(x) = \eta + \int_\xi^x f(t, y_k(t))\, dt, \quad k = 0, 1, 2, \ldots,$$

liefert, die gleichmäßig gegen die Lösung des AWPs konvergiert.

Der obige Existenz- und Eindeutigkeitssatz gilt analog auch für Systeme von Differentialgleichungen. Ist f auf einer Menge D des $J \times \mathbb{R}^n$ definiert, so fragt man beim AWP nach einer Lösung, die durch einen vorgegebenen Punkt $(\xi, \eta) \in D$ geht, d.h. die

$$y_\nu(\xi) = \eta_\nu, \quad \nu = 1, \ldots, n \text{ (oder kurz: } y(\xi) = \eta)$$

und das System von n gewöhnlichen Differentialgleichungen

$$\begin{aligned} y_1' &= f_1(x, y_1, \ldots, y_n) \\ &\vdots \\ y_n' &= f_n(x, y_1, \ldots, y_n) \end{aligned}$$

erfüllt. Gibt es eine Umgebung

$$U := \{(x, z) | \quad |x - \bar{x}| < \delta, |z - \bar{y}(x)| < \delta, \quad x \in J\}$$

mit $(\bar{x}, \bar{y}(x)) \in D \; \forall x \in J$, so daß dort eine Lipschitzbedingung der Form (A.4) — mit Normen anstelle von Beträgen — erfüllt ist, dann gilt der Satz von Picard-Lindelöf in analoger Weise.

Ohne die Voraussetzung der Lipschitzstetigkeit von f bzgl. y gilt der

Satz A.2 (Existenzsatz von PEANO**)** *Ist $f(x, z)$ in einem Gebiet $D \subset J \times \mathbb{R}^n$ stetig und $(\xi, \eta) \in D$, so besitzt das AWP (A.1) mindestens eine Lösung.*

A.2 Lineare Differentialgleichungen

Unter einem *linearen System von n Differentialgleichungen (erster Ordnung)* versteht man das folgende System:

$$\begin{aligned} y_1'(t) &= a_{11}(t)y_1(t) + \ldots + a_{1n}(t)y_n(t) + b_1(t) \\ &\vdots \\ y_n'(t) &= a_{n1}(t)y_1(t) + \ldots + a_{nn}(t)y_n(t) + b_n(t), \end{aligned} \tag{A.5}$$

oder in Matrizenschreibweise

$$y'(t) = A(t)y(t) + b(t),$$

wobei

$$A(t) = (a_{ij}(t))_{i,j}, \quad b(t) = (b_1(t), \ldots, b_n(t))^\top.$$

Die auftretenden Funktionen können reell- oder komplexwertig sein, die unabhängige Variable t ist immer reellwertig.

Aus dem Satz von Picard-Lindelöf erhält man sofort die Existenz und Eindeutigkeit der Lösung eines zu (A.5) gehörigen AWPs, und man hat darüber hinaus eine (Stabilitäts-) Abschätzung für die Lösung:

Satz A.3 (Existenz-, Eindeutigkeits- und Abschätzungssatz) *Die reell- bzw. komplexwertigen Funktionen $A(t), b(t)$ seien stetig in einem (beliebigen) Intervall J, und es sei $\tau \in J$. Dann hat das Anfangswertproblem*

$$y' = A(t)y + b(t), \quad y(\tau) = \eta$$

bei vorgegebenem $\eta \in \mathbb{R}^n$ bzw. $\mathbb{C}^n$ genau eine Lösung $y(t)$. Sie existiert in ganz J. Ist j ein Teilintervall von J, $\tau \in j$ und

$$\|A(t)\| \le L \text{ und } \|b(t)\| \le \delta \text{ in } j, \quad \|\eta\| \le \gamma,$$

so besteht die Abschätzung

$$\|y(t)\| \le \gamma e^{L|t-\tau|} + \frac{\delta}{L}\left(e^{L|t-\tau|} - 1\right) \text{ in } j.$$

Die Lösung $y(t)$ hängt in jedem kompakten Teilintervall $j \subset J$ stetig von $A(t), b(t), \eta$ ab, d.h. zu $\epsilon > 0$ existiert ein $\beta > 0$, so daß

$$\|z(t) - y(t)\| < \epsilon \text{ in } j$$

ist, wenn $z(t)$ eine Lösung des Anfangswertproblems (B, c stetig)

$$z' = B(t)z + c(t), \quad z(\tau) = \zeta$$

und wenn dabei

$$\|B(t) - A(t)\| < \beta, \quad \|b(t) - c(t)\| < \beta \text{ in } j, \quad \|\eta - \zeta\| < \beta$$

ist.

Man nennt das System (A.5) *homogen*, wenn $b(t) \equiv 0$ ist, sonst *inhomogen*. Als Korrolar von Satz A.3 erhält man eine umkehrbar eindeutige, lineare Abbildung zwischen dem $\mathbb{K}^n$ und dem Raum der Lösungen eines homogenen Systems

$$y'(t) = A(t)y(t). \tag{A.6}$$

Satz A.4 *Ist $A(t)$ in J reellwertig (komplexwertig) und stetig, so bilden die reellen (komplexen) Lösungen $y(t)$ der homogenen Gleichung (A.6) einen n-dimensionalen reellen (komplexen) linearen Raum.*

Für festes $\tau \in J$ wird durch

$$\eta \to y(t; \tau, \eta)$$

ein Isomorphismus (umkehrbar eindeutige, lineare Abbildung „auf") zwischen $\mathbb{R}^n$ bzw. $\mathbb{C}^n$ und dem Raum der Lösungen definiert.

Daraus ergeben sich eine Reihe von Folgerungen und Begriffen.

Korollar A.5 *Es gelten die folgenden Aussagen:*

1. *Eine Linearkombination*

   $$y = c_1 y_1 + \ldots + c_k y_k$$

   von Lösungen y_ν von (A.6) stellt wieder eine Lösung dar.

2. *Ist y eine Lösung und $y(t_0) = 0$ für ein $t_0 \in J$, so ist $y \equiv 0$ in J.*

3. *Für $k > n$ sind k Lösungen linear abhängig.*

4. *Es gibt n linear unabhängige Lösungen $y_1, \ldots, y_n$ von (A.6). Jedes solche System heißt* Hauptsystem *oder* Fundamentalsystem *. Jede Lösung y von (A.6) läßt sich mit Hilfe eines Fundamentalsystems eindeutig darstellen als*

   $$y = c_1 y_1 + \ldots + c_n y_n.$$

Die letzte Aussage 4 läßt sich kürzer auch in der Form

$$y = Yc, \quad c \in \mathbb{R}^n$$

schreiben, wobei $Y(t) = (y_1(t), \ldots, y_n(t))$ die zu einem Fundamentalsystem $\{y_1, \ldots, y_n\}$ gehörige *Lösungsmatrix* darstellt. Die n Differentialgleichungen $y_i' = A(t)y_i$, $i = 1, \ldots, n$, lassen sich als Matrizen-Gleichung $Y' = A(t)Y$ schreiben. Die Lösung $Y(t)$ dieser Matrizengleichung ist nach Satz A.4 durch die Vorgabe einer Anfangsbedingung $Y(\tau) = C$ eindeutig bestimmt. Offenbar ist $Y(t)$ genau dann ein Fundamentalsystem, wenn die Matrix C regulär ist. Dann ist auch $Y(t)$ regulär für jedes $t \in J$.

Ein spezielles Fundamentalsystem erhält man aus

$$X(\tau) = E, \quad X' = A(t)X, \tag{A.7}$$

d.h.

$$x_i(\tau) = e_i, \quad x_i' = A(t)x_i, \quad i = 1, \ldots, n, \quad (e_i = i\text{-ter Einheitsvektor}).$$

Mit Hilfe von $X(t)$ kann man durch $y(t) = X(t)\eta$ die Lösung jedes AWP's

$$y(\tau) = \eta, \quad y' = A(t)y$$

erhalten.

Ist $Y(t)$ eine Lösungsmatrix, dann ist auch $Z(t) = Y(t)C$ eine Lösung für eine beliebige Matrix C, denn

$$Z' = Y'C = AYC = AZ.$$

Ist $Y(t)$ ein Fundamentalsystem und C regulär, so ist auch Z ein Fundamentalsystem und jedes Fundamentalsystem läßt sich in der Form $Y(t)C$ mit regulärem C darstellen. Insbesondere gilt mit obigem X und beliebiger Lösungsmatrix $Y(t)$, daß

$$Y(t) = X(t)Y(\tau), \tag{A.8}$$

denn die rechte Seite ist Lösung und der Anfangswert erfüllt $X(\tau)Y(\tau) = EY(\tau) = Y(\tau)$.

Für ein Lösungssystem y_i, $i = 1, \ldots, n$ des homogenen Systems (A.6) bezeichnet $\phi(t) := \det Y(t)$ die zugehörige *Wronski-Determinante*. Mit der *Spur* von A,

$$\operatorname{sp} A(t) := a_{11}(t) + \ldots + a_{nn}(t)$$

gilt der folgende Satz:

Satz A.6 *Die Wronski-Determinante genügt der Differentialgleichung*

$$\phi'(t) = (\operatorname{sp} A(t))\phi(t) \text{ in } J;$$

daher ist

$$\phi(t) = \phi(\tau) \exp\left(\int_\tau^t \operatorname{sp}(A(s))\, ds\right), \quad \tau, t \in J.$$

Offenbar ist entweder $\phi(t) \equiv 0$ oder überall $\neq 0$. $\{y_1, \ldots, y_n\}$ ist genau dann ein Fundamentalsystem, wenn $\phi(t) \neq 0$ ist.

Zwischen den Lösungen des inhomogenen Systems (A.5) und denen des homogenen Systems (A.6) besteht für $n = 1$ der folgende Zusammenhang:

Satz A.7 *Im Fall $n = 1$ ergeben sich sämtliche Lösungen der inhomogenen Differentialgleichung (s. (A.5)) als*

$$y(t) = \bar{y}(t) + x(t),$$

wobei $\bar{y}(t)$ eine feste Lösung der inhomogenen Differentialgleichung ist und $x(t)$ alle Lösungen der homogenen Differentialgleichung durchläuft.

Im allgemeinen n-dimensionalen Fall führt eine analoge Überlegung zum Ziel, die man „Methode der Variation der Konstanten“ nennt. Dabei sucht man, ausgehend von einem Fundamentalsystem $\{y_1, \ldots, y_n\}$ des homogenen Systems, Lösungen des inhomogenen Systems durch

$$z(t) = Y(t)v(t),$$

d.h. man variiert den konstanten Vektor v, um aus den Lösungen $Y(t)v$ des homogenen Systems die des inhomogenen Systems zu erhalten.

Satz A.8 *Seien $A(t), b(t)$ stetig in J und $\tau \in J$. Dann besitzt das AWP*

$$y(\tau) = \eta, \quad y' = A(t)y + b(t) \tag{A.9}$$

die eindeutig bestimmte Lösung

$$y(t) = X(t)\eta + X(t) \int_\tau^t X^{-1}(s)b(s)\, ds, \tag{A.10}$$

wobei $X(t)$ das Fundamentalsystem zu (A.6) mit $X(\tau) = E$ ist.

Ist $Y(t)$ ein beliebiges Fundamentalsystem, so gilt nach (A.8) $Y(t) = X(t)Y(t)$, also $Y^{-1}(t) = Y^{-1}(\tau)X^{-1}(t)$ und $X(t)X^{-1}(s) = Y(t)Y^{-1}(s)$. Mit Hilfe von $Y(t)$ lautet die Darstellung (A.10) also

$$y(t) = Y(t)Y^{-1}(\tau)\eta + Y(t) \int_\tau^t Y^{-1}(s)b(s)\, ds. \tag{A.11}$$

A.3 Systeme mit konstanten Koeffizienten

Für das homogene lineare System

$$y' = Ay \tag{A.12}$$

mit konstanter komplexer Matrix $A = (a_{ij})$ erhält man Lösungen durch den Ansatz

$$y(t) = ce^{\lambda t} = (c_1 e^{\lambda t}, \ldots, c_n e^{\lambda t})^\top. \tag{A.13}$$

Einsetzen in (A.12) liefert

$$y' = \lambda c e^{\lambda t} = Ace^{\lambda t},$$

d.h. $y(t)$ in (A.13) ist genau dann eine Lösung von (A.12), wenn

$$Ac = \lambda c, \tag{A.14}$$

d.h. $c(\neq 0)$ ist ein Eigenvektor zum Eigenwert λ der Matrix A. Eigenwerte ergeben sich als Nullstellen des charakteristischen Polynoms $\det(A - \lambda E)$. Jedes Polynom n-ten Grades besitzt n reelle oder komplexe Nullstellen der Vielfachheit nach gezählt.

Satz A.9 *(komplexer Fall) Seien λ, c, A komplex, $c \neq 0$. Die Funktion $y(t) = ce^{\lambda t}$ ist genau dann eine Lösung von (A.12), wenn λ Eigenwert und c ein zugehöriger Eigenvektor der Matrix A ist. Die Lösungen $y_i(t) = e^{\lambda_i t} c_i$, $i = 1, \ldots, p$, sind genau dann linear unabhängig, wenn die (Eigen-)Vektoren c_i linear unabhängig sind; insbesondere sind sie linear unabhängig, wenn alle Eigenwerte $\lambda_1, \ldots, \lambda_p$ verschieden sind.*

Ist A eine reelle Matrix, dann kann diese durchaus komplexe Eigenwerte haben. Aus einem komplexen Eigenwert kann man über Real- und Imaginärteil zwei reelle Lösungen erhalten. Ist $\lambda = \mu + i\nu$ ein komplexer Eigenwert und $c = a + ib$ ein zugehöriger Eigenvektor der reellen Matrix A, so ergeben sich aus der komplexen Lösung $y = ce^{\lambda t}$ zwei reelle Lösungen

$$\begin{aligned} z(t) &= \operatorname{Re} y = e^{\mu t}(a\cos\nu t - b\sin\nu t), \\ z^*(t) &= \operatorname{Im} y = e^{\mu t}(a\sin\nu t + b\cos\nu t). \end{aligned}$$

Besitzt A n linear unabhängige Eigenvektoren, so kann man auf diese Weise ein reelles Fundamentalsystem konstruieren, wenn man noch beachtet, daß die Aufspaltung einer zugehörigen konjugiert-komplexen Lösung $\bar{y} = \bar{c}e^{\bar{\lambda} t}$ auf dieselben zwei reellen Lösungen führt.

Beispiel A.10 Exponentialfunktion

$$e^B := E + B + \frac{B^2}{2!} + \frac{B^3}{3!} + \dots$$

Betrachtet man $B = At$ und leitet nach t ab, so erhält man

$$\frac{d}{dt}e^{At} = Ae^{At}.$$

Also erhält man ein Fundamentalsystem von $y' = Ay$ auch durch

$$X(t) = e^{At} \text{ mit } X(0) = E.$$

Als Folgerungen ergeben sich:

1. $(e^A)^{-1} = e^{-A}$
2. $e^{A(s+t)} = e^{As}e^{At}$
3. $e^{A+\lambda E} = e^{\lambda} \cdot e^A$

Hat man ein Fundamentalsystem des homogenen Systems, so kann man durch die Methode der Variation der Konstanten die Lösungen von zugehörigen inhomogenen Systemen erhalten. Das Anfangswertproblem

$$y' = Ay + b(t), \quad y(\tau) = \eta \quad (A \text{ konstant}) \tag{A.15}$$

besitzt die Lösung

$$y(t) = e^{A(t-\tau)}\eta + \int_\tau^t e^{A(t-s)}b(s)\,ds. \tag{A.16}$$

Es ist nämlich $X(t) = e^{A(t-\tau)}$ das Fundamentalsystem mit $X(\tau) = E$ und wegen $X(t)^{-1} = e^{-A(t-\tau)}$ erhält man aus (A.10) die obige Lösungsdarstellung. □

A.4 Lineare Differentialgleichungen n-ter Ordnung

Eine lineare Differentialgleichung n-ter Ordnung

$$(Lu)(t) := u^{(n)}(t) + a_{n-1}(t)u^{(n-1)}(t) + \ldots + a_0(t)u(t) = b(t) \tag{A.17}$$

ist äquivalent zum System

$$\begin{aligned} y_1' &= y_2 \\ &\vdots \\ y_{n-1}' &= y_n \\ y_n' &= -(a_0 y_1 + \ldots + a_{n-1} y_n) + b(t), \end{aligned} \tag{A.18}$$

oder kurz

$$y' = A(t)y + \underline{b}(t)$$

mit

$$A = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & 1 \\ -a_0 & -a_1 & -a_2 & \ldots & -a_{n-2} & -a_{n-1} \end{pmatrix}$$

$$y = (y_1, \ldots, y_n)^\top = (u, u', \ldots, u^{(n-1)})^\top, \quad \underline{b} = (0, \ldots, b)^\top.$$

Aus Satz A.3 erhält man sofort einen Existenz- und Eindeutigkeitssatz für lineare, inhomogene Differentialgleichungen n-ter Ordnung:

Satz A.11 *Sind die (reell- oder komplexwertigen) Koeffizienten $a_i(t), b(t)$, $i = 0, \ldots, n$, $a_n \equiv 1$ in einem Intervall J stetig und ist $\tau \in J$, so hat das Anfangswertproblem*

$$Lu := \sum_{i=0}^{n} a_i(t)u^{(i)}(t) = b(t), \quad u^{(\nu)}(\tau) = \eta_\nu, \quad \nu = 0, 1, \ldots, n-1,$$

genau eine Lösung. Sie existiert in ganz J und hängt in jedem kompakten Teilintervall von J stetig von den $a_i(t)$ und von $b(t)$ ab.

Für die zugehörigen homogenen Systeme

$$Lu = 0 \tag{A.19}$$

erhält man analoge Aussagen wie in Abschnitt A.2. Die Lösungen von (A.19) bilden also einen n-dimensionalen Vektorraum, und es existieren n linear unabhängige Lösungen $u_1, \ldots, u_n$, die eineindeutig den Anfangswerten $\eta_0, \eta_1, \ldots, \eta_{n-1}$ zugeordnet sind:

$$U'(t) = A(t)U(t), \quad U(\tau) = C, \tag{A.20}$$

wobei $U(t) = (u, u', \ldots, u^{(n-1)})^\top$, C reguläre $n \times n$-Matrix; die $\{u_1, \ldots, u_n\}$ bilden ein Haupt- oder Fundamentalsystem. Die zugehörige Wronski-Determinante

$$W(t) = \begin{vmatrix} u_1 & \ldots & u_n \\ u_1' & \ldots & u_n' \\ \vdots & \ddots & \vdots \\ u_1^{(n-1)} & \ldots & u_n^{(n-1)} \end{vmatrix} \tag{A.21}$$

erfüllt nach Satz A.6 die Differentialgleichung

$$W' = -a_{n-1}(t)W (= \mathrm{sp}(A(t))W) \tag{A.22}$$

und hat also die Darstellung

$$W(t) = W(\tau) \exp\left(-\int_\tau^t a_{n-1}(s)\, ds\right).$$

Ein spezielles Fundamentalsystem erhält man durch $X' = A(t)X$, $X(\tau) = E$, was hier dem folgenden System entspricht:

$$Lx_i = 0, \quad x_i^{(j)}(\tau) = \begin{cases} 1, j = i-1, \\ 0, \text{sonst}, \end{cases} \quad j = 0, \ldots, n-1,\ i = 1, \ldots, n \tag{A.23}$$

Die „Methode der Variation der Konstanten“ ermöglicht die Bestimmung von Lösungen eines gegebenen inhomogenen Systems n-ter Ordnung (vgl. (A.17)) aus einem Fundamentalsystem $u_1, \ldots, u_n$ (vgl. (A.20)) des zugehörigen homogenen Systems. D.h. man sucht Funktionen $c_1, \ldots, c_n$, so daß

$$W(t) = u_1(t)c_1(t) + \ldots + u_n(t)c_n(t)$$

Lösung von (A.17) wird. Mit

$$z = \begin{pmatrix} w \\ w' \\ \vdots \\ w^{(n-1)} \end{pmatrix}, \quad Y(t) = \begin{pmatrix} u_1 & \dots & u_n \\ \vdots & \ddots & \vdots \\ u_1^{(n-1)} & \dots & u_n^{(n-1)} \end{pmatrix}, \quad \underline{b}(t) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ b(t) \end{pmatrix}$$

soll also $z' = A(t)z + \underline{b}(t)$ erfüllt sein. In der Lösungsdarstellung (A.11) benötigt man $\underline{a}(t) = Y^{-1}(t)\underline{b}(t)$, also die Lösung des linearen Gleichungssystems $Y\underline{a} = \underline{b}$. Mit Hilfe der Cramerschen Regel erhält man die Komponenten von $\underline{a}$ hier besonders einfach:

$$a_i = \frac{V_i}{W}, \quad i = i, \dots, n, \text{ wobei } W = \det Y, \tag{A.24}$$

$$V_i = \det \begin{pmatrix} u_1 & \dots & u_{i-1} & 0 & u_{i+1} & \dots & u_n \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ u_1^{(n-1)} & \dots & u_{i-1}^{(n-1)} & b(t) & u_{i+1}^{(n-1)} & \dots & u_n^{(n-1)} \end{pmatrix},$$

also (Entwicklung nach der i-ten Spalte)

$$V_i(t) = (-1)^{n+i} b(t) W_i(t), \tag{A.25}$$

wobei W_i die Wronski-Determinante (von der Ordnung $n-1$) der Funktionen $u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_n$ ist.

Setzt man speziell $z(\tau) = 0$, so erhält man nach (A.11) als Lösung:

$$z(t) = Y(t) \int_\tau^t \underline{a}(s)\, ds.$$

Eine Lösung w von (A.17) erhält man also — w ist die erste Komponente von z — durch

$$w(t) = \sum_{i=1}^{n} u_i(t)(-1)^{n+i} \int_\tau^t \frac{b(s)}{W(s)} W_i(s)\, ds \tag{A.26}$$

Beispiel A.12 (Lineare Differentialgleichungen zweiter Ordnung)

$$u''(t) + a_1(t)u'(t) + a_0(t)u(t) = b(t) \tag{A.27}$$

Ist $\{u_1, u_2\}$ ein Fundamentalsystem, so ist eine Lösung von (A.27) — nämlich die mit $w(\tau) = w'(\tau) = 0$ — gegeben durch

$$w(t) = -u_1(t) \int_\tau^t \frac{b(s)u_2(s)}{W(s)}\, ds + u_2(t) \int_\tau^t \frac{b(s)u_1(s)}{W(s)}\, ds \tag{A.28}$$

□

Beispiel A.13

$$w'' - w' \cos t + w \sin t = \sin t, \quad t \geq 0$$

Ein Fundamentalsystem ist $\{\exp(\sin t)), \exp(\sin t)\Phi(t)\}$ mit $\Phi(t) := \int_0^t \exp(-\sin s)\, ds$ (vgl. [Wal76, §19, IV]). Die zweite Lösung wurde nach dem sog. d'Alembertschen Reduktionsverfahren ermittelt, auf das hier nicht weiter eingegangen werden soll.

Hier ist $a_{n-1}(s) = a_1(s) = -\cos s$,

$$W(t) = W(\tau) \exp\left(\int_\tau^t \cos s\ ds\right) \overset{\tau=0}{=} W(0) \exp(\sin t)$$
$$W(0) = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1, \quad (\text{da } \Phi(0) = 0).$$

Nach (A.28) ist

$$w(t) = -\exp(\sin t) \int_0^t \sin r \int_0^r \exp(-\sin s)\, ds\, dr + \exp(\sin t)\Phi(t) \int_0^t \sin r\ dr$$

und

$$\begin{aligned} \int_0^t \sin r \int_0^r \exp(-\sin s)\, ds\, dr &= \int_0^t \exp(-\sin s) \int_s^t \sin r\ dr\, ds \\ &= -\int_0^t \exp(-\sin s)(\cos t - \cos s)\, ds \\ &= -\Phi(t)\cos t - \exp(-\sin t) + 1, \end{aligned}$$

also

$$w(t) = \int_0^t \exp(\sin t - \sin s)\, ds + 1 - \exp(\sin t)$$

eine Lösung der vorgegebenen inhomogenen Differentialgleichung. □

A.5 Lineare Differentialgleichungen n-ter Ordnung mit konstanten Koeffizienten

Es sei jetzt

$$Lu := \sum_{i=0}^{n} a_i u^{(i)}, \quad a_i = \text{konstant}, \quad a_n = 1 \tag{A.29}$$

Zur Bestimmung eines Fundamentalsystems des homogenen Systems $Lu = 0$ betrachtet man das *charakteristische Polynom*

$$P(\lambda) = \begin{vmatrix} -\lambda & 1 & 0 & \dots & 0 \\ 0 & -\lambda & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & -\lambda & 1 \\ -a_0 & -a_1 & \dots & -a_{n-2} & -a_{n-1} - \lambda \end{vmatrix}$$

Entwicklung nach der letzten Zeile ergibt

$$P(\lambda) = (-1)^n(\lambda^n + a_{n-1}\lambda^{n-1} + \ldots + a_1\lambda + a_0) \tag{A.30}$$

Aus Satz A.9 und den Ausführungen in Abschnitt A.3 erhält man das folgende Ergebnis. Über den Ansatz $u = e^{\lambda t}$ kann man dies allerdings auch ganz elementar beweisen.

Satz A.14 *Ist λ eine k-fache Nullstelle des charakteristischen Polynoms, so entsprechen ihr k Lösungen*

$$e^{\lambda t}, te^{\lambda t}, \ldots, t^{k-1}e^{\lambda t}$$

von $Lu = 0$. Aus den n Nullstellen des charakteristischen Polynoms $P(\lambda)$ (jede mit ihrer Vielfachheit gezählt) ergeben sich auf diese Weise n linear unabhängige Lösungen, also ein Hauptsystem.

Sind die a_i reell, so enthält dieses Hauptsystem, falls komplexe Nullstellen auftreten, komplexe Lösungen. Ein reelles Hauptsystem wird erhalten, indem man zu einer komplexen Nullstelle $\lambda = \mu + i\nu$ von k-ter Ordnung die k Lösungen in Real- und Imaginärteil aufspaltet,

$$t^q e^{\mu t}\cos \nu t, \quad t^q e^{\mu t}\sin \nu t, \quad q = 0, 1, \ldots, k-1$$

(und die k zu $\bar{\lambda}$ gehörenden Lösungen streicht).

Beispiel A.15 (Schwingungen)

Die Differentialgleichung

$$m\ddot{s} + \beta\dot{s} + ks = 0, \quad t > 0 \tag{A.31}$$

beschreibt die Auslenkung $s(t)$ (von der Ruhelage $s = 0$) einer Masse m im Falle einer *gedämpften Schwingung* . Hierbei ist noch $\beta > 0$ ein *Reibungskoeffizient*, $k > 0$ der Koeffizient der elastischen Rückstellkraft und $\dot{s}$ bedeutet $\frac{ds}{dt}$ (entspr. $\ddot{s}$). Es sind zwei Anfangsbedingungen erforderlich: $s(0), \dot{s}(0)$.

Wir setzen $u = s,\ a = \frac{\beta}{2m},\ b = \frac{k}{m}$, d.h.

$$Lu = u'' + 2au' + bu.$$

Die charakteristische Gleichung

$$P(\lambda) = \lambda^2 + 2a\lambda - b = 0$$

hat die beiden Wurzeln $\lambda_{1,2} = -a \pm \sqrt{a^2 - b}$.

Man unterscheidet verschiedene Fälle, für die wir gleich zugehörige Fundamentalsysteme angeben (nach Satz A.14):

1. „Überdämpfung“ : $a^2 > b$

   $$u_1 = \exp(-a + \sqrt{a^2 - b}\, t), \quad u_2 = \exp(-a - \sqrt{a^2 - b}\, t)$$

   In (A.31) bedeutet dieser Fall $\beta^2 > 4km$, und man spricht von *großer Dämpfung* bzw. *Reibung*. Für $a > 0,\ b > 0$ streben beide Lösungen exponentiell gegen 0 für $t \to \infty$.

2. „Kritische Dämpfung“ : $a^2 = b$ (d.h. $\beta^2 = 4km$)

   $$u_1 = \exp(-at), \quad u_2 = t\exp(-at)$$

3. „Unterdämpfung“ : $a^2 < b$ (d.h. $\beta^2 < 4km$)

   $$u_1 = \exp(-at)\cos\left(\sqrt{b - a^2}\, t\right), \quad u_2 = \exp(-at)\sin\left(\sqrt{b - a^2}\, t\right);$$

   Frequenz der gedämpften Schwingung (periodischer Fall)

$$\nu = \frac{1}{2\pi}\sqrt{b-a^2} = \frac{1}{4\pi m}\sqrt{4km-\beta^2}.$$

Betrachtet man die homogene Differentialgleichung (A.31) mit $\beta > 0$, dann spricht man auch von einer *gedämpften freien Schwingung* . Eine *ungedämpfte freie Schwingung* liegt vor, wenn in (A.31) $\beta = 0$ ist. Nach 3. hat man dann ein Fundamentalsystem $\{\cos\nu_0 t, \sin\nu_0 t\}$ mit $\nu_0 = \sqrt{k/m}$.

Die allgemeine Lösung einer ungedämpften freien Schwingung hat also die Gestalt

$$s(t) = \alpha\cos\nu_0 t + \beta\sin\nu_0 t, \quad t > 0,$$

was man auch in der Form

$$s(t) = R\cos(\nu_0 t - \delta)$$

mit $R = \sqrt{\alpha^2+\beta^2}, \delta = \arctan(\beta/\alpha)$ schreiben kann.

Zum Abschluß betrachten wir noch *ungedämpfte erzwungene Schwingungen*, d.h. $\beta = 0$ in (A.31), und als inhomogenen Term hat man auf der rechten Seite $(F_0/m)\cos\omega t$, d.h. wir betrachten

$$\frac{d^2y}{dt^2} + \omega_0^2 y = \frac{F_0}{m}\cos\omega t, \quad t > 0 \tag{A.32}$$

mit $\omega_0^2 (= \nu_0^2) = k/m$.

Fall $\omega \neq \omega_0$: Eine spezielle Lösung von (A.32) erhält man durch (A.28), wobei

$$W(t) = \begin{vmatrix} \cos\omega_0 t & \sin\omega_0 t \\ -\omega_0\sin\omega_0 t & \omega_0\cos\omega_0 t \end{vmatrix} = \omega_0(\cos^2\omega_0 t + \sin^2\omega_0 t) = \omega_0.$$

Anwendung von Additionstheoremen liefert

$$w(t) = \frac{F_0}{m(\omega_0^2-\omega^2)}\cos\omega t,$$

und eine allgemeine Lösung von (A.32) hat die Gestalt

$$y(t) = \alpha\cos\omega_0 t + \beta\sin\omega_0 t + \frac{F_0}{m(\omega_0^2-\omega)}\cos\omega t.$$

Fall $\omega = \omega_0$: Als spezielle Lösung des inhomogenen Systems erhält man

$$w(t) = \frac{F_0 t}{2m\omega_0} \sin \omega_0 t,$$

denn

$$\begin{aligned} w(t) &= -u_1(t) \int_0^t \frac{F_0}{m\omega_0} u_2(s) \cos \omega_0 s \; ds + u_2(t) \int_0^t \frac{F_0}{m\omega_0} u_1(s) \cos \omega_0 s \; ds \\ &= \frac{F_0}{m\omega_0} \left(-\cos \omega_0 t \int_0^t \sin \omega_0 s \cos \omega_0 s \; ds + \sin \omega_0 t \int_0^t (\cos \omega_0 s)^2 \; ds \right) \\ &= \frac{F_0}{m\omega_0} \left(-\cos \omega_0 t \left[\frac{1}{2\omega_0} \sin^2 \omega_0 s \right]_0^t + \sin \omega_0 t \left[\frac{1}{2} s + \frac{1}{4\omega_0} \sin 2\omega_0 s \right]_0^1 \right) \\ &= \frac{F_0}{m\omega_0} \left(-\frac{1}{\omega_0} \cos \omega_0 t \sin^2 \omega_0 + \left[t + \frac{1}{2\omega_0} \sin 2\omega_0 t \right] \sin \omega_0 t \right) \\ &= \frac{F_0}{2m\omega_0} \sin \omega_0 t, \end{aligned}$$

da $\sin 2\gamma = 2 \sin \gamma \cos \gamma$. Jede Lösung einer ungedämpften erzwungenen Schwingung hat also im Fall $\omega = \omega_0$ die Form

$$y(t) = \alpha \cos \omega_0 t + \beta \sin \omega_0 t + \frac{F_0 t}{2m\omega_0} \sin \omega_0 t.$$

Der dritte Term stellt eine Oszillation mit wachsender Amplitude dar, was bei realen Schwingungsproblemen die Zerstörung entsprechender physikalischer Systeme bewirken kann.

□

A.6 Lineare Randwertaufgaben 2. Ordnung

In zahlreichen Anwendungen in Naturwissenschaften und Technik treten Randwertprobleme für (reelle) lineare Differentialgleichungen zweiter Ordnung auf:

$$u'' + a_1(x)u' + a_0(x)u = y(x), \quad a \le x \le b; \tag{A.33}$$

als mögliche Randbedingungen unterscheidet man:

erste Art:	$u(a) = \eta_1,$	$u(b) = \eta_2,$
zweite Art:	$u'(a) = \eta_1,$	$u'(b) = \eta_2,$
dritte Art:	$\alpha_1 u(a) + \alpha_2 u'(a) = \eta_0,$	$\beta_1 u(b) + \beta_2 u'(b) = \eta_1.$

Es gibt auch noch andere Arten von Randbedingungen, z.B. *periodische*:

$$u(a) = u(b), \quad u'(a) = u'(b).$$

Beispiel A.16 $u'' = 0$

FS: $\{1, x\}$; allgemeine Lösung: $u(x) = a_1 x + a_0$ (lineare Funktionen)

RWP erster Art: Immer eindeutig lösbar.

RWP zweiter Art: Keine Lösung für $\eta_0 \neq \eta_1$, unendlich viele Lösungen für $\eta_0 = \eta_1$. □

Anstelle von (A.33) betrachten wir RWPs der folgenden Form:

$$Lu := (p(x)u')' + q(x)u = g(x), \quad \text{in } J = [a, b] \tag{A.34}$$

mit Randbedingungen dritter Art:

$$\begin{aligned} R_0 u &:= \alpha_1 u(a) + \alpha_2 p(a) u'(a) = \eta_0, \\ R_1 u &:= \beta_1 u(b) + \beta_2 p(b) u'(b) = \eta_1. \end{aligned} \tag{A.35}$$

Man nennt (A.34) auch die *selbstadjungierte Form* des Differentialoperators L; man kann (A.33) durch Multiplikation mit $p(x) := \exp\left(\int_0^x a_1(s)\, ds\right)$ immer in die selbstadjungierte Form (A.34) überführen. (A.34), (A.35) heißt „Sturmsche Randwertaufgabe“ (nach Ch. Sturm (1803-1855)). Wir setzen im folgenden voraus, daß

$$p \in C^1(J),\ q, g \in C(J),\ \text{(alles reellwertig)},\ p(x) > 0 \text{ in } J, \text{ und}$$

$$\alpha_1^2 + \alpha_2^2 > 0,\ \beta_1^2 + \beta_2^2 > 0.$$

Für zwei beliebige Funktionen $u, v \in C^2(J)$ gilt die *Lagrange-Identität*

$$vLu - uLv = \{p(x)(u'v - v'u)\}'. \tag{A.36}$$

Daraus folgt (mit Hilfe partieller Integration)

$$\int_a^b (vLu - uLv)\, dx = 0, \tag{A.37}$$

falls $u, v \in C^2(J)$ den homogenen Randbedingungen $R_i u = R_i v = 0,\ i = 0, 1,$ genügen.

Alle Lösungen von (A.34), (A.35) ergeben sich in der Form

$$v = v^* + u,$$

wobei v^* eine feste Lösung ist und u alle Lösungen des homogenen RWPs durchläuft.

Satz A.17 *Es gelten die obigen Voraussetzungen und $\{u_1, u_2\}$ sei ein Fundamentalsystem der homogenen Differentialgleichung $Lu = 0$. Das inhomogene RWP (A.34), (A.35) ist genau dann eindeutig lösbar, wenn*

$$\begin{vmatrix} R_0u_1 & R_0u_2 \\ R_1u_1 & R_1u_1 \end{vmatrix} \neq 0 \tag{A.38}$$

ist. Insbesondere hat das homogene RWP in diesem Fall nur die triviale Lösung.

Bemerkung: *Ist ein Fundamentalsystem der homogenen DGL bekannt, so hat man nach Beispiel A.12 auch eine Lösung v^* der inhomogenen Gleichung $Lu = g$, und die Lösung der RWP reduziert sich auf die Lösung eines linearen Gleichungssystems zur Bestimmung von c_1, c_2 in der Darstellung*

$$v = v^* + c_1u_1 + c_2u_2;$$

Letztere ergeben sich aus den zwei Randbedingungen; unter (A.38) existiert immer eine eindeutige Lösung.

Beispiel A.18

$$u'' + u = 1, \quad 0 \leq x \leq \pi,\ R_0u := u(0) + u'(0) = \eta_0,\ R_1u := u(\pi) = \eta_1$$

FS: $u_1 = \cos x, u_2 = \sin x$,

$$\begin{vmatrix} R_0(\cos x) & R_0(\sin x) \\ R_1(\cos x) & R_1(\sin x) \end{vmatrix} = \begin{vmatrix} 1 & 1 \\ -1 & 0 \end{vmatrix} = 1(\neq 0)$$

Allgemeine Lösung:

$$v(x) = 1 + c_1 \sin x + c_2 \cos x$$

Fall: $\eta_0 = \eta_1 = 0$: $R_0v = 1 + c_2 + c_1 = 0,\ R_1v = 1 - c_2 = 0 \iff c_1 = -2,\ c_2 = 1$

Lösung: $v(x) = 1 + \cos x - 2\sin x$

Fall: $R_0u := u(0) = \eta_0,\ R_1u := u(1) = \eta_1$.

Die Determinante in (A.38) verschwindet; es gibt — neben der trivialen Lösung — unendlich viele Lösungen der Form $u = C \sin x$.

□

Definition A.19 *Eine Funktion $\gamma(x,\xi)$ wird* Grundlösung *der homogenen DGL $Lu = 0$ genannt, wenn folgendes gilt:*

1. *$\gamma(x,\xi)$ ist stetig in $Q := \{a \le x, \xi \le b\}$.*
2. *In jedem der beiden Dreiecke*

$$Q_1 = \{(x,\xi) | a \le \xi \le x \le b\}, \quad Q_2 = \{(x,\xi) | a \le x \le \xi \le b\}$$

   existieren die stetigen partiellen Ableitungen γ_x, γ_{xx}.

3. *Bei festem $\xi \in J$ ist $\gamma(x,\xi)$ als Funktion von x eine Lösung von $L\gamma = 0$ für $x \neq \xi, x \in J$.*
4. *Auf der Diagonalen $x = \xi$ macht die erste Ableitung einen Sprung:*

$$\gamma_x(x+0,x) - \gamma_x(x-0,x) = \frac{1}{p(x)}, \quad a < x < b.$$

Eine Grundlösung ist nicht eindeutig bestimmt, da auch $\tilde{\gamma}(x,\xi) = \gamma(x,\xi) + \alpha(\xi)u(x)$ Grundlösung ist, falls α stetig und u Lösung von $Lu = 0$ ist.

Beispiel A.20

$$u'' = 0 \Rightarrow \gamma(x,\xi) = \frac{1}{2}|x-\xi|$$
$$u'' + \lambda^2 u = 0 \Rightarrow \gamma(x,\xi) = \frac{1}{2\lambda}\sin\lambda|x-\xi|$$

(Beweis durch Verifizieren; zur Konstruktion siehe unten.) □

Satz A.21 *Ist $\gamma(x,\xi)$ eine Grundlösung, so ist*

$$v(x) = \int_a^b \gamma(x,\xi) g(\xi)\, d\xi \tag{A.39}$$

aus $C^2(J)$ und stellt eine Lösung von $Lv = g(x)$ dar.

Beweisskizze: Aufspalten des Integrals von a bis x und von x bis b sowie 2-maliges Differenzieren in Verbindung mit Eigenschaft 4 einer Grundlösung ergibt

$$v'(x) = \int_a^b \gamma_x(x,\xi) g(\xi)\, d\xi,$$
$$v''(x) = \int_a^b \gamma_{xx}(x,\xi) g(\xi)\, d\xi + \frac{g(x)}{p(x)}.$$

Wegen Eigenschaft 3. folgt daraus die Behauptung. □

Definition: $\Gamma(x,\xi)$ *heißt* Greensche Funktion *für die homogene Randwertaufgabe* $Lu = 0$, $R_0u = R_1u = 0$, *wenn gilt:*

1. $\Gamma(x,\xi)$ *ist eine Grundlösung,*

2. $R_0\Gamma = R_1\Gamma = 0 \quad \forall \xi \in J' := (a,b)$.

Satz A.22 *Unter den obigen Voraussetzungen existiert genau eine Greensche Funktion für die homogene Sturmsche Randwertaufgabe, wenn diese nur die triviale Lösung besitzt, d.h. wenn (A.38) gilt. Sie ist symmetrisch,* $\Gamma(x,\xi) = \Gamma(\xi,x)$. *Die eindeutige Lösung der „halbhomogenen" Randwertaufgabe*

$$Lv = g(x), \quad R_0v = R_1v = 0$$

hat die Darstellung

$$v(x) = \int_a^b \Gamma(x,\xi)g(\xi)\,d\xi. \tag{A.40}$$

Beweisskizze: (teilw.) Die Funktion v in (A.40) ist nach Satz A.21 eine Lösung von $Lv = g$. Da man unter dem Integral (nach x) differenzieren darf, kann Integration und R_i vertauscht werden, und wegen Eigenschaft 2 einer Greenschen Funktion sind die homogenen Randbedingungen erfüllt. Der Beweis der Eindeutigkeit und Symmetrie benutzt wesentlich die Beziehung (A.37). □

Die Konstruktion der Greenschen Funktion aus einem Fundamentalsystem $\{u_1,u_2\}$ ist wie folgt möglich: Durch den Ansatz

$$\Gamma(x,\xi) = \sum_{i=1}^{2}(a_i(\xi) \pm b_i(\xi))u_i(x) \quad \begin{cases} \text{„}+\text{“ in } Q_1 \\ \text{„}-\text{“ in } Q_2 \end{cases}, \tag{A.41}$$

die Forderung der Stetigkeit und der Sprungrelation von Γ_x bei $x = \xi$ erhält man ein Gleichungssystem für b_1, b_2,

$$\sum_{i=1}^{2} b_i(\xi)u_i(\xi) = 0, \quad \sum_{i=1}^{2} b_i(\xi)u_i'(\xi) = \frac{1}{2p(\xi)}.$$

Dieses ist eindeutig nach b_1, b_2 auflösbar, da die Wronski-Determinante gerade die Determinante des zugehörigen Gleichungssystems ist. Die $a_i,\ i = 1, 2$, bestimmt man dann aus den Randbedingungen

$$R_0\Gamma = \sum_{i=1}^{2}(a_i(\xi) - b_i(\xi))R_0u_i = 0,$$
$$R_1\Gamma = \sum_{i=1}^{2}(a_i(\xi) + b_i(\xi))R_1u_i = 0;$$

dieses Gleichungssystem ist wegen (A.38) eindeutig lösbar.

Beispiel A.23 $u'' = 0, \quad u(0) = u(1) = 0,$

FS: $\{1, x\}$, $\begin{vmatrix} R_0u_1 & R_0u_2 \\ R_1u_1 & R_1u_2 \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1,\ \Gamma(x,\xi) = \begin{cases} \xi(x-1), 0 \le \xi \le x \le 1, \\ x(\xi-1), 0 \le x \le \xi \le 1. \end{cases}$

□

Will man das inhomogene Randwertproblem (A.34), (A.35) lösen, so sucht man zunächst eine Funktion $\phi \in C^2(J)$, die (nur) die Randbedingungen $R_i\phi = \eta_i,\ i = 0, 1$, erfüllt. Macht man dann für die Lösung u der inhomogenen Randwertaufgabe den Ansatz $u = \phi + v$, so ergibt sich für v:

$$Lv = h := g - L\phi,\ R_0v = R_1v = 0.$$

Dieses ist nach Satz A.22 lösbar.

Bemerkung: *Die bisherige Theorie kann man ohne Schwierigkeiten auf lineare Differentialgleichungen n-ter Ordnung und auf Systeme linearer Differentialgleichungen erster Ordnung ausdehnen. Die zugehörige Greensche Funktion ist dann eine Matrix (vgl. z.B. [Wal76, §26, X ff]).*

Anhang B

Lösungen partieller Differentialgleichungen

In diesem Abschnitt werden wir Modellbeispiele für die drei Typen von partiellen Differentialgleichungen vorstellen, nämlich die parabolischen, hyperbolischen und elliptischen Differentialgleichungen. Für ein elliptisches Randwertproblem auf einem Rechteck wird dann die Methode der Separation der Variablen vorgestellt, um eine analytische Darstellung für die Lösung in einfachen Fällen zu erhalten. Die Darstellung folgt dem Buch von [Zau83].

B.1 Modellbeispiele für Rand- und Anfangswertaufgaben bei partiellen Differentialgleichungen

1. *Elliptische Differentialgleichungen* . Die einfachsten Beispiele erhält man mit Hilfe des *Laplaceschen Differentialoperators*

$$\Delta = \frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_n^2}$$

für Funktionen von n Veränderlichen. Sei G ein Gebiet, also eine offene zusammenhängende Teilmenge des $\mathbb{R}^n$. Damit lautet die sogenannte *Laplace-Gleichung* oder *Potentialgleichung* :

$$\Delta u = 0 \ \text{ in } \ G.$$

Lösungen dieser Gleichung heißen *harmonische Funktionen* . Es gibt unendlich viele Lösungen dieser Gleichung, insbesondere die *harmonischen Polynome*

$$1,\ x_1, \ldots,\ x_n,\ x_j x_k,\ x_j^2 - x_k^2,\ (j \neq k), \ldots\ .$$

Aus der Funktionentheorie weiß man, daß Real- und Imaginärteil holomorpher Funktionen stets harmonische Funktionen zweier reeller Veränderlicher sind.

Die zur Laplace-Gleichung gehörige inhomogene Gleichung

$$-\Delta u = f \quad \text{in} \quad G$$

heißt *Poisson-Gleichung.* Sie beschreibt z. B. elektrostatische Felder der äußeren Ladungsdichte f, eine belastete Membran unter der äußeren Last f, Strömungen inkompressibler Flüssigkeiten u.s.w. Die zugehörige *Eigenwertgleichung*

$$-\Delta u = \lambda u \quad \text{in} \quad G$$

tritt z.B. bei der *schwingenden Saite* ($n = 1$), *schwingenden Membran* ($n = 2$) oder dreidimensionalen schwingenden Kontinua ($n = 3$) auf.

Eindeutig bestimmte Lösungen derartiger partieller Differentialgleichungen entstehen erst durch Hinzunahme von *Randbedingungen* :

1. Dirichlet-Problem: $\quad u = g \quad \text{auf} \quad \partial G$
2. Neumann-Problem: $\quad \dfrac{\partial u}{\partial N} = g \quad \text{auf} \quad \partial G,$
   wobei $\dfrac{\partial}{\partial N}$ die Ableitung in Richtung der äußeren Normalen bezeichnet.
3. Robbin-Problem: $\quad \alpha u + \beta \dfrac{\partial u}{\partial N} = g \quad \text{auf} \quad \partial G.$

Differentialgleichungen mit variablen Koeffizienten haben im Falle zweiter Ordnung etwa die Form

$$\sum_{j,k=1}^{n} a_{jk}(x) \frac{\partial^2 u}{\partial x_j \partial x_k} + \sum_{j=1}^{n} b_j(x) \frac{\partial u}{\partial x_j} + c(x) u = f \quad \text{in} \quad G,$$

wobei die Koeffizienten a_{jk} der *Elliptizitätsbedingung*

$$\sum_{j,k=1}^{n} a_{jk}(x) \xi_j \xi_k > 0,\ x \in G, \quad \text{für} \sum_{j=1}^{n} \xi_j^2 > 0,$$

genügen.

Ein Modellbeispiel für eine elliptische Differentialgleichung vierter Ordnung ist die sogenannte *Plattengleichung*

$$\Delta^2 u = f \quad \text{in} \quad G.$$

Die Randbedingungen der *eingespannten Platte* lauten

$$u = \frac{\partial u}{\partial N} = 0 \quad \text{auf} \quad \partial G.$$

Wir bemerken noch, daß im Falle $n = 1$ die obigen partiellen Differentialgleichungen und zugehörigen Randwertprobleme in gewöhnliche Differentialgleichungen und Zwei-Punkt-Randwertaufgaben übergehen.

2. *Parabolische Differentialgleichungen.* Grundtyp ist hier die sogenannte *Wärmeleitungsgleichung*

$$\frac{\partial u}{\partial t} = \Delta u + f \quad \text{in} \quad G \times (0,T), \quad G \subset \mathbb{R}^n,$$

mit der *Randbedingung*

$$u(x,t) = r(x,t), \quad x \in \partial G, \quad 0 < t < T,$$

und der *Anfangsbedingung*

$$u(x,0) = g(x), \quad x \in G.$$

Diese Gleichung beschreibt z.B. die Wärmeverteilung in einem Draht ($n = 1$), einer Membran ($n = 2$) oder einem homogenen Material ($n = 3$), die zu Beginn eine Temperaturverteilung g haben, im Verlauf der Zeit am Rande auf der Temperatur r gehalten und im Innern mit der Verteilung f geheizt werden. Im Falle $n = 1$ ist einfacher

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + f \quad \text{in} \quad G \times (0,T).$$

3. *Hyperbolische Differentialgleichungen.* Grundtyp dieser partiellen Differentialgleichungen ist die sogenannte *Wellengleichung*

$$\frac{\partial^2 u}{\partial t^2} = \Delta u + f \quad \text{in} \quad G \times (0,T)$$

mit der *Randbedingung*

$$u(x,t) = r(x,t), \quad x \in \partial G, \quad t \in (0,T),$$

und den *Anfangsbedingungen*

$$u(x,0) = g_0(x), \quad \frac{\partial u}{\partial t}(x,0) = g_1(x), \quad x \in G.$$

Diese Aufgabe beschreibt etwa das zeitliche Verhalten einer schwingenden Saite ($n = 1$), Membran ($n = 2$) oder eines elastischen Kontinuums ($n = 3$) unter einer äußeren Anregung f, die am Rande wie r geführt wird und am Anfang die Auslenkung g_0 und Anfangsgeschwindigkeit g_1 aufweist. Für $f = r = 0$ kann man auch hier wieder die Zeitabhängigkeit separieren und Lösungsansätze in Form von Reihenentwicklungen nach Eigenfunktionen des zugehörigen elliptischen Ortsoperators gewinnen.

Ein einfaches Beispiel einer hyperbolischen Differentialgleichung erster Ordnung bildet die Gleichung

$$\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} = f \quad \text{in} \quad G \times (0,T)$$

mit der Anfangsbedingung

$$u(x,0) = g, \quad x \in G.$$

Das Gebiet G ist hier z.B. der ganze $\mathbb{R}^n$. Bei beschränkten Gebieten treten Periodizitätsbedingungen oder ähnliches auf. Von besonderer Bedeutung sind Systeme entsprechender Gleichungen erster Ordnung, wobei dann a durch eine Matrix zu ersetzen ist.

B.2 Methode der Separation der Variablen

Wir betrachten zunächst die *Laplace-Gleichung* auf einem Rechteck

$$u_{xx} + u_{yy} = 0, \quad 0 \le x \le l,\ 0 \le y \le \hat{l}, \tag{B.1}$$

mit Randbedingungen

$$\begin{aligned} &u(x,0) = f(x),\ u(x,\hat{l}) = y(x), \quad 0 \le x \le l, \\ &\alpha_1 u(0,y) - \beta_1 u_x(0,y) = 0,\ \alpha_2 u(l,y) + \beta_2 u_x(l,y) = 0, \quad 0 \le y \le \hat{l}. \end{aligned} \tag{B.2}$$

Die Methode der *Separation der Variablen* besteht darin, einen Ansatz $u(x,y) = M(x)N(y)$ für die Lösung von (B.1) zu machen und M, N so zu bestimmen, daß auch die Randbedingungen erfüllt sind. Einsetzen in (B.1) liefert

$$-\frac{N''(y)}{N(y)} = \frac{M''(x)}{M(x)}. \tag{B.3}$$

Da die linke bzw. rechte Seite von unterschiedlichen Variablen abhängen, müssen beide Seiten notwendigerweise konstant (gleich $-\lambda$) sein, was

$$-M''(x) = \lambda M(x), \quad 0 \le x \le l, \tag{B.4}$$

$$N''(y) = \lambda N(y), \quad 0 \le y \le \hat{l} \tag{B.5}$$

ergibt. Verlangt man von M, daß die homogenen Randbedingungen bei $x = 0$ und $x = l$ erfüllt werden:

$$\alpha_1 M(0) - \beta_1 M'(0) = 0, \quad \alpha_2 M(l) + \beta_2 M'(l) = 0, \tag{B.6}$$

dann erfüllt auch $u(x,y) = M(x)N(y)$ die entsprechende Randbedingung. Offensichtlich ist $M(x) \equiv 0$ eine Lösung, die aber nicht von Interesse ist — da dann auch $u = 0$ wäre.

Das Problem zur Bestimmung von $M(x)$ ist ein Sturm-Liouvillesches Eigenwertproblem für eine gewöhnliche Differentialgleichung zweiter Ordnung. Aus Positivitätseigenschaften des zugehörigen gewöhnlichen Differentialoperators $L = d^2/dx^2$ weiß man, daß es abzählbar viele nichtnegative Eigenwerte λ_k, $k = 1, 2, \ldots$, gibt, für die $\lambda_k \to \infty$ $(k \to \infty)$ gilt. Die zugehörigen Eigenfunktionen M_k sind paarweise orthogonal (bzgl. des L_2-Skalarprodukts) und jede Funktion $\phi \in C^1[0,l]$, die den homogenen Randbedingungen (B.6) genügt, läßt sich in eine absolut und gleichmäßig konvergente Reihe nach den Eigenfunktionen entwickeln:

$$\phi(x) = \sum_{k=0}^{\infty} c_k M_k(x), \tag{B.7}$$

wobei

$$c_k = \int_0^l \phi(x) M_k(x)\, dx \tag{B.8}$$

als zugehörige *Fourier-Koeffizienten* bezeichnet werden.

Beispiel B.1 Wir betrachten speziell die Randbedingungen

$$M(0) = M(\pi) = 0.$$

Die allgemeine Lösung der Gleichung

$$M'' + \lambda M = 0, \quad x \in [0, \pi],$$

ist

$$\begin{aligned} M(x) &= c_1 + c_2 x \quad \text{im Fall } \lambda = 0, \\ M(x) &= c_1 \exp\left(\sqrt{\lambda}\, x\right) + c_2 \exp\left(-\sqrt{\lambda}\, x\right) \quad \text{im Fall } \lambda < 0, \\ M(x) &= c_1 \cos\left(\sqrt{\lambda}\, x\right) + c_2 \sin\left(-\sqrt{\lambda}\, x\right) \quad \text{im Fall } \lambda > 0. \end{aligned}$$

Einsetzen der Randbedingungen zeigt, daß nur der Fall $\lambda > 0$ eine nichttriviale Lösung liefert, für die dann über die Randbedingungen

$$c_1 = 0, \quad \sin\left(\sqrt{\lambda}\, \pi\right) = 0$$

sein muß. Die Eigenwerte sind also $\lambda_k = k^2,\ k = 1, 2, 3$, mit zugehörigen Eigenfunktionen $M_k(x) = \sin kx$.

Kommt man auf $u(x, y) = M(x)N(y)$ zurück, so erhält man für jeden Eigenwert λ_k über (B.5) eine Gleichung für N_k. Nimmt man der Einfachheit halber an, daß $\lambda_k > 0\ \forall k$, dann ergibt sich

$$N_k(y) = a_k \exp\left(\sqrt{\lambda_k}\, y\right) + b_k \exp\left(-\sqrt{\lambda_k}\, y\right), \quad k = 1, 2, \ldots \tag{B.9}$$

Die Konstanten $a_k,\ b_k$ sind noch zu bestimmen.

Superposition der Lösungen $u_k = M_k N_k$ liefert eine Reihe

$$u = \sum_{k=1}^{\infty} u_k = \sum_{k=1}^{\infty} M_k N_k \tag{B.10}$$

als Lösung; hierin müssen jetzt die a_k, b_k so bestimmt werden, daß die Randbedingungen bei $y = 0$ und $y = \hat{l}$ erfüllt werden. Diese Forderung liefert

$$\begin{aligned} u(x, 0) &= \sum_{k=1}^{\infty} [a_k + b_k] M_k(x) = f(x) \\ u(x, \hat{l}) &= \sum_{k=1}^{\infty} \left[a_k \exp\left(\sqrt{\lambda_k}\, \hat{l}\right) + b_k \exp\left(-\sqrt{\lambda_k}\, \hat{l}\right)\right] M_k(x) = g(x) \end{aligned} \tag{B.11}$$

Skalare Multiplikation mit $M_j(x)$ und Ausnutzung der Orthogonalität $(M_k, M_j)_{L_2} = 0,\ k \neq j$ liefert

$$\begin{aligned}(f, M_j)_{L_2} &= \left(\sum_{k=1}^{\infty}(a_k + b_k)M_k, M_j\right)_{L_2} \\ &= \sum_{k=1}^{\infty}(a_k + b_k)\,(M_k, M_j)_{L_2} \\ &= (a_j + b_j)\,(M_j, M_j)_{L_2}\,,\end{aligned}$$

wobei noch angenommen wurde, daß Summation und Integration vertauscht werden dürfen. Folglich ist

$$a_k + b_k = \frac{(f, M_k)_{L_2}}{(M_k, M_k)_{L_2}}, \quad k = 1, 2, \ldots \tag{B.12}$$

Eine analoge Vorgehensweise für die Randbedingung bei $y = \hat{l}$ liefert

$$a_k \exp\left(\sqrt{\lambda_k}\,\hat{l}\right) + b_k \exp\left(-\sqrt{\lambda_k}\,\hat{l}\right) = \frac{(g, M_k)_{L_2}}{(M_k, M_k)_{L_2}}, \quad k = 1, 2, \ldots \tag{B.13}$$

(B.12), (B.13) stellt ein lineares 2×2-Gleichungssystem zur Bestimmung von $a_k,\ b_k$ dar, was eindeutig lösbar ist.

Der Reihenansatz (B.10) führt also zur formalen Lösung

$$u(x, y) = \sum_{k=1}^{\infty}\left[a_k \exp\left(\sqrt{\lambda_k}\,y\right) + b_k \exp\left(-\sqrt{\lambda_k}\,y\right)\right] M_k(x) \tag{B.14}$$

des Randwertproblems für die elliptische Gleichung (B.1). Ist die Reihe konvergent, so nennt man u in (B.14) auch eine *klassische Lösung*. Schwache oder verallgemeinerte Lösungen werden in Abschnitt 2.1 der Vorlesung behandelt; Regularitätsfragen können nur angeschnitten werden. □

Für das Modellbeispiel einer *parabolischen Differentialgleichung* ,

$$u_t = u_{xx}, \quad (x, t) \in G \times (0, T), \quad G := (0, 1)$$

$$u(0, t) = u(1, t) = 0, \quad t \in (0, T),$$

führt die Separation der Zeit- und Ortsveränderlichen auf den Lösungsansatz

$$u(x,t) = v(x) \cdot w(t).$$

Dann wird

$$\frac{\partial u}{\partial t} = u_{xx} \quad \text{zu} \quad v\,w' = wv''$$

oder

$$\frac{w'}{w} = \frac{v''}{v} = -\lambda,$$

d.h. konstant in t und x. Zusammen mit der Randbedingung genügt der Ortsanteil v dann dem Eigenwertproblem (vgl. Beispiel B.1)

$$v'' + \lambda\, v = 0 \;\text{ in }\; G, \quad v = 0 \quad \text{für } x = 0, 1$$

und der Zeitanteil der gewöhnlichen Differentialgleichung

$$w' = -\lambda\, w \quad : \quad w(t) = w(0)\, \mathrm{e}^{-\lambda t}$$

mit den Eigenwerten λ. Durch *Superposition* solcher spezieller Lösungen $u = vw$ entsteht ein Lösungsansatz der Form

$$u(x,t) = \sum_{k=1}^{\infty} w_k(0)\, v_k(x)\, \mathrm{e}^{-\lambda_k t} \quad .$$

Die Konstanten $w_k(0)$ werden nun so bestimmt, daß für $t = 0$ die Anfangsbedingung $u(x,0) = g(x)$ erfüllt ist. Dies ist möglich, solange sich die Anfangsbedingung g nach dem System der Eigenfunktionen von $-\frac{d^2}{dx^2}$ entwickeln läßt. Man sieht hierbei übrigens, daß elliptische Eigenwertprobleme eine wichtige Rolle für die Struktur der Lösungen parabolischer Probleme spielen.

Anhang C

Übungen

1. **Aufgabe:** Seien p, q, $b \in \mathcal{C}(J)$. Zeigen Sie für eine Differentialgleichung 2-ter Ordnung

$$u'' + p(t)u' + q(t)u = b(t) \ , \qquad t \in J = [\tau, \tau + a] \ ,$$

daß die Wronski–Determinante $W(t) = \begin{vmatrix} u_1 & u_2 \\ u_1' & u_2' \end{vmatrix}$ für zwei Lösungen u_1, u_2 der zugehörigen homogenen Differentialgleichung selbst die Differentialgleichung erster Ordnung

$$W'(t) = -p(t)W(t) \ , \qquad t \in J \ ,$$

erfüllt und damit die Darstellung hat,

$$W(t) = W(\tau) \exp\left(-\int_\tau^t p(s)ds\right) \ , \qquad t \in J \ .$$

2. **Aufgabe:** Sei $[a, b]$ ein beschränktes abgeschlossenes Interval der reellen Zahlengeraden. Man bestimme die Lösung u der inhomogenen Randwertaufgabe

$$u(a) = u(b) = 0 \ , \qquad \frac{d^2u}{dx^2}(x) = w(x) \ , \quad x \in [a, b] \ ,$$

für $w \in C[a, b]$ mit der zugehörigen Greenschen Funktion K in der Gestalt

$$u(x) = \int_a^b K(x, y)w(y)dy \ , \qquad x \in [a, b] \ .$$

Hinweis: Man benutze die Konstruktionsvorschrift der Greenschen Funktion nach Vorlesung.

3. **Aufgabe:** Seien z_1, z_2 Lösungen der Differentialgleichung $z'' + pz' + qz = 0$, $t \in [a,b]$, mit

$$z_1(a) = 0 \ , \ z_1'(a) = 1 \ , \quad z_2(b) = 0 \, , \, z_2'(b) = 1 \ .$$

Sei

$$P(t) = \int_a^t p(s)ds \ , \ \text{ und } \ D_0 = z_1(a)\, z_2'(a) - z_2(a) z_1'(a) \ .$$

Man beweise: Wenn die homogene Randwertaufgabe $z'' + pz' + qz = 0$, $z(a) = z(b) = 0$ nur die triviale Lösung hat, dann ist $z_1(b) \neq 0$, $z_2(a) \neq 0$ und die Funktion

$$u(t) = \frac{\alpha}{z_2(a)} z_2(t) + \frac{\beta}{z_1(b)} z_1(t) + \frac{z_2(t)}{D_0} \int_a^t z_1(s) w^1(s) e^{P(s)} ds + \\ + \frac{z_1(t)}{D_0} \int_t^b z_2(s) w^1(s) e^{P(s)} ds$$

für $t \in [a,b]$, ist eindeutig bestimmte Lösung der Randwertaufgabe

$$u(a) = \alpha \ , \quad u(b) = \beta \ , \quad u'' + pu' + qu = w^1 \ .$$

4. **Aufgabe:** Für die Einheitssphäre $\Omega = \{x \in \mathbb{R}^2 : |x| \leq 1\}$ zeige man, daß die Funktion $v(x) = |x|^\alpha$ in $H^1(\Omega)$ liegt, falls $\alpha > 0$ ist.

Hinweis:
Geben Sie partielle Ableitungen von v an (möglichst in Polarkoordinaten), die die Definitionsgleichung der schwachen Ableitungen erfüllen, da es sich um klassische Ableitungen handelt. Es ist zu zeigen, daß die partiellen Ableitungen L^2–integrierbar sind, was man am besten durch Variablensubstitution in Polarkoordinaten zeigt.

5. **Aufgabe:** Man zeige, daß für eine stetige Funktion w auf $[0,1]$ mit

$$V = \{v \, \text{stetig auf} \, [0,1], \ v' \, \text{stückweise stetig und beschränkt}, \, v(0) = v(1) = 0\}$$

aus

$$\int_0^1 wv dx = 0 \qquad \forall v \in V,$$

folgt, daß

$$w(x) = 0 \quad \text{für alle} \quad x \in [0,1].$$

6. **Aufgabe:** Sei H ein prähilbertscher Raum mit Skalarprodukt $(\cdot,\cdot)$ und zugehöriger Norm $\|u\| = (u,u)^{1/2}$. Man beweise die Cauchysche Ungleichung (auch: Schwarzsche Ungleichung):

$$|(v,w)| \le \|v\|\|w\|, \qquad v,w \in H.$$

7. **Aufgabe:** Zeigen Sie, daß für die Einheitssphäre $G = \{x \in \mathbb{R}^n : |x| < 1\}$ und die Funktionen $g_i(x) = x_i/|x|^2, \quad x \in G, \quad i = 1,\ldots,n$, gilt:

   a) $g_i \notin L^2(G)$, falls $n = 1, 2$;

   b) $g_i \in L^2(G)$, falls $n = 3$.

   Hinweis:
   Hier sind Integrale auszurechnen, was wieder durch Variablensubstitution in Polar– bzw. Kugelkoordinaten geschehen kann (für $n = 2$ bzw. 3).

8. **Aufgabe:** Für eine nicht notwendig äquidistante Unterteilung $0 = x_0 < x_1 < \cdots < x_{M+1} = 1$ von $I = [0,1]$ seien die Funktionen $G_i : [0,1] \to \mathbb{R}$ durch die folgenden Variationsprobleme definiert:

$$(G_i', v') = v(x_i) \qquad \forall v \in H_0^1(I), \qquad i = 1,\ldots,M.$$

   Zeigen Sie:

   a) Die Funktionen G_i haben die Darstellung

$$G_i(x) = \begin{cases} (1-x_i)x, & 0 \le x \le x_i, \\ x_i(1-x), & x_i \le x \le 1 \end{cases}$$

   und sind Elemente von $H_0^1(I)$.

   b) Sei u die Lösung des Modellproblems (f gegeben)

$$(D) \quad -u'' = f \quad \text{in} \quad I = (0,1), \quad u(0) = u(1) = 0,$$

   und u_h die FE-Approximation im Raum aller stetigen, stückweise linearen Funktionen mit Nullrandbedingungen. Zeigen Sie mit Hilfe der Funktionen G_i, daß für den Fehler $e_h = u - u_h$ gilt

$$e_h(x_i) = (e_h', G_i') = 0, \quad i = 1,\ldots,M.$$

Hinweis zu a):
Benutzen Sie den Rieszschen Darstellungssatz und die Sobolevsche Ungleichung

$$\|v\|_{0,\infty} \leq C \, \|v\|_{1,2}, \quad v \in H^1(0,1).$$

9. **Aufgabe:** Für das Dreieck K mit den Eckpunkten $(0,0)$, $(h,0)$, $(0,h)$, $h > 0$ und die linearen Basisfunktionen

$$\varphi_0(z_1,z_2) = 1 - z_1/h - z_2/h, \quad \varphi_1(z_1,z_2) = z_1/h, \quad \varphi_2(z_1,z_2) = z_2/h$$

berechne man die „Steifigkeitselementmatrix"

$$\begin{bmatrix} a_K(\varphi_0,\varphi_0) & a_K(\varphi_0,\varphi_1) & a_K(\varphi_0,\varphi_2) \\ & a_K(\varphi_1,\varphi_1) & a_K(\varphi_1,\varphi_2) \\ symm. & & a_K(\varphi_2,\varphi_2) \end{bmatrix}$$

wobei

$$\begin{aligned} a_K(\varphi_i,\varphi_j) &= \int_K \nabla\varphi_i \cdot \nabla\varphi_j dz \; = \\ &= \int_K \left(\frac{\partial\varphi_i}{\partial z_1} \frac{\partial\varphi_j}{\partial z_1} + \frac{\partial\varphi_i}{\partial z_2} \frac{\partial\varphi_j}{\partial z_2} \right) dz_1 dz_2 \; . \end{aligned}$$

Hinweis: Man beachte, daß die Ableitungen der Funktionen φ_i, $i = 0,1,2$, konstant sind, und daß die Fläche des Dreiecks $h^2/2$ beträgt.

10. **Aufgabe:** Geben Sie zur Randwertaufgabe

$$(D) \quad -u'' = f \quad \text{in} \quad (0,1) \; , \quad u(0) = 0 \; , \quad u'(1) = \gamma_1$$

eine geeignete variationelle Formulierung an. Zeigen Sie, daß die zugehörige Bilinearform beschränkt und elliptisch, und daß die zugehörige Linearform $l(v)$ beschränkt ist.

Hinweise:
— Als zugrundeliegende Norm benutzen Sie $\|.\|_1 = \|.\|_{1,2}$.
— Vgl. Johnson (1987), Problem 1.16

11. **Aufgabe:** Mit der Linearform q bestimme man die Koeffizienten β_k in der Darstellung

$$q(p) := \int_a^b p(x)\,dx = \sum_{k=0}^{n} \beta_k p(x_k)\ , \quad p \in \mathcal{P}_1^1(\Delta)\ ,$$

wobei $\mathcal{P}_1^1(\Delta)$ den Raum der stetigen, stückweise linearen Funktionen für ein Gitter $a = x_0 < x_1 < \ldots < x_n = b$ bezeichnet. Man berechne weiter β_k für äquidistante Maschenweiten $h = x_j - x_{j-1}\ , \quad j = 1, \ldots, n$.

Hinweis: Hier sind einfach die Werte $\beta_k = q(p_k)$ für die Dachfunktionen als Basis in $\mathcal{P}_1^1(\Delta)$ auszurechnen.

12. **Aufgabe:** Für jedes $\nu \in \mathbb{N}$ seien $a = x_0^{(\nu)} < x_1^{(\nu)} < \ldots < x_n^{(\nu)} = b$ Gitterpunkte in $[a, b]$. Für eine stetige Funktion $u \in C[a, b]$ ist dann die *stetige, stückweise lineare interpolatorische Approximation* gegeben durch

$$u^{(\nu)}(x) = u\left(x_{j-1}^{(\nu)}\right) \frac{x_j^{(\nu)} - x}{x_j^{(\nu)} - x_{j-1}^{(\nu)}} + u\left(x_j^{(\nu)}\right) \frac{x - x_{j-1}^{(\nu)}}{x_j^{(\nu)} - x_{j-1}^{(\nu)}}\ ,$$

$$x \in I_j^{(\nu)} = \left[x_{j-1}^{(\nu)}\ , x_j^{(\nu)}\right]\ , \quad j = 1, \ldots, n, \quad \nu \in \mathbb{N}\ .$$

Man beweise: Falls

$$h^{(\nu)} = \max_{1 \le j \le n} \left(x_j^{(\nu)} - x_{j-1}^{(\nu)}\right) \to 0 (\nu \to \infty)\ ,$$

dann ist $u^{(\nu)}\ , v = 1, 2, \ldots$, eine gleichgradig gleichmäßig stetige Folge von Funktionen.

Hinweis: $u^{(\nu)}\ , \nu \in \mathbb{N}$, heißt *gleichgradig gleichmäßig stetig*, wenn es zu jedem $\varepsilon > 0$ ein $\delta > 0$ gibt, so daß für alle $x, x' \in [a, b]$, $|x - x'| \le \delta$, und alle $\nu \in \mathbb{N}$ gilt: $|u^{(\nu)}(x) - u^{(\nu)}(x')| \le \varepsilon$.

13. **Aufgabe:** Im Raum $C^r[a, b]$ aller r-mal stetig differenzierbaren Funktionen auf $[a, b]$ sind Halbnormen bzw. Normen definiert durch

$$|u|_{s,\infty} = \max_{a \le x \le b} \left|\frac{d^s u}{dx^s}(x)\right|\ , \quad \|u\|_{r,\infty} = \max_{s=0,\ldots,r} |u|_{s,\infty}\ .$$

Zeigen Sie, daß die Einbettung von $C^1[a, b]$ in $C^0[a, b]$ kompakt ist, wobei $\|.\|_{1,\infty}$ bzw. $\|.\|_{0,\infty}$ die zugrundeliegenden Normen darstellen.

Hinweis: Man verwende den Satz von Arzela–Ascoli.

14. **Aufgabe:** Zeigen Sie für eine Unterteilung $\Delta : x_0 = a < x_1 < \ldots < x_n = b$, daß die Linearformen

$$g_{js} := \tfrac{1}{2}(\delta^s_{j+0} - \delta^s_{j-0}),$$

$$s = 0, \ldots, r, \quad j = 1, \ldots, n-1,$$

$$\delta^s_{j+0} u = \frac{d^s u}{dx^s}(x_j + 0), \quad \delta^s_{j-0} u = \frac{d^s u}{dx^s}(x_j - 0),$$

linear unabhängig auf $\mathcal{P}_r(\Delta)$ sind.

Hinweis: Zum Nachweis der linearen Unabhängigkeit eines Systems von Linearformen genügt es, ein System von Funktionen $\{u_{kt}\} \subset \mathcal{P}_r(\Delta)$ anzugeben, so daß $\{g_{js}\}, \{u_{kt}\}$ biorthonormal sind. Wählen Sie dafür

$$u_{kt}(x) = \frac{2}{t!}(x - x_k)^t_+, \quad x \in \bigcup_{j=1}^{n} I_j,$$

wobei $(x - x_k)_+ = \max(x - x_k, 0) = \begin{cases} x - x_k, & x \geq x_k, \\ 0, & \text{sonst} \end{cases}$.

15. **Aufgabe:**

a) Bestimmen Sie die baryzentrischen Koordinaten von

$$\underline{e}_0 = (0, 0, \ldots, 0)\,,\ \underline{e}_1 = (1, 0, \ldots, 0), \ldots\,,\ \underline{e}_n = (0, 0, \ldots, 1) \text{ in } \mathbb{R}^n\,.$$

b) Bestimmen Sie das Polynom $p \in \mathcal{P}_1(\mathbb{R}^2)$, das durch die Lagrange–Interpolationsaufgabe

$$p(\underline{e}_0) = 1\,,\ p(\underline{e}_1) = p(\underline{e}_2) = 0$$

definiert ist.

16. **Aufgabe:** Zeigen Sie, daß für die Dimensionen von

$$Q_r(\mathbb{R}^2) = \left\{ p \in \mathcal{P} : p(x) = \sum_{|s|_\infty \leq r} \alpha_s \underline{x}^s\,, \qquad \underline{x} \in \mathbb{R}^2 \right\}$$

gilt

$$\dim Q_r(\mathbb{R}^2) = (r+1)^2$$

Hinweis:
Man verwende vollständige Induktion über r.

17. **Aufgabe:**

(a) Gehen Sie in die Bibliothek (H–C 4303) und stellen Sie fest, wo

- Zentralblatt für Mathematik (Zbl)
- Mathematical Reviews (MR)
- Current Mathematical Publications (CMP)

stehen (Standort angeben, Signatur). Skizzieren Sie kurz den Unterschied zwischen Zbl und MR einerseits und CMP andererseits.

(b) Stellen Sie fest, in welchem Band des Zbl man die aktuelle Klassifikation („Mathematics Subject Classification“) findet, und geben Sie die Klassifikation für folgende Gebiete an:

- gewöhnliche Differentialgleichungen
- partielle Differentialgleichungen
  - elliptische Gleichungen und Systeme
  - parabolische Gleichungen und Systeme
- Funktionalanalysis
- numerische Analyse
  - partielle Differentialgleichungen, Anfangswertprobleme und zeitabhängige Anfangsrandwertprobleme
    - Finite Elemente, Rayleigh–Ritz- und Galerkin-Methoden, finite Methoden
  - partielle Differentialgleichungen, Randwertprobleme
    - Finite Elemente, Rayleigh–Ritz- und Galerkin-Methoden, finite Methoden
    - Boundary element methods
- Geophysik

(c) Geben Sie fünf Publikationen in wissenschaflichen Fachzeitschriften und zwei Bücher aus den Jahren 1994 und 1995 zum Gebiet „Finite Elemente Methode“ an; die Publikationen sollen auch in der Bibliothek vorhanden sein. Welchen der Artikel verstehen Sie nach dem aktuellen Stand der Vorlesung?

18. **Aufgabe:** (Quadraturformeln in $\mathcal{P}_2(\mathbb{R}^2)$)
Die zur Lagrange–Interpolationsaufgabe in $\mathcal{P}_2(\mathbb{R}^2)$ gehörigen Basispolynome $L_{s_1 s_2}$ sind durch

$$L_{s_1 s_2}(\underline{z}) = \prod_{i=0}^{2} \prod_{\ell_i=0}^{s_i-1} \frac{2z_i - \ell_i}{s_i - \ell_i}, \quad \underline{z} = (z_1, z_2),\ s_1 + s_2 \le 2, \tag{C.1}$$

gegeben. Die kanonische Form von Quadraturformeln in $\mathcal{P}_2(\mathbb{R}^2)$ lautet damit

$$\int_K u(\underline{x})\, d\underline{x} = |K| \sum_{|t|\le 2} \alpha_t u(\underline{a}_t)\ , \qquad u \in \mathcal{P}_2(\mathbb{R}^2)\ , \tag{C.2}$$

wobei $K = \mathrm{conv}\{\underline{a}_0, \underline{a}_1, \underline{a}_2\}$, $\{\underline{a}_i\}$ affin unabhängig,

$$\underline{a}_t = \underline{a}_0 + \frac{t_1}{2}\underline{b}_1 + \frac{t_2}{2}\underline{b}_2, \quad t_1 + t_2 \le 2, \quad \underline{b}_i = \underline{a}_i - \underline{a}_0,\ i = 1, 2,$$

$$\alpha_t = \frac{|\det B|}{|K|} \int_E L_t(\underline{z})\, d\underline{z}\ , \quad |t| \le 2.$$

Rechnen Sie mit Hilfe der Formeln 4.(31) der Vorlesung die Koeffizienten α_t aus, und geben Sie damit die Quadraturformel für $u \in \mathcal{P}_2(\mathbb{R}^2)$ an.

19. **Aufgabe:** (Differenzenquotienten in $\mathcal{Q}_2(\mathbb{R}^2)$)
Für $b_j > a_j, j = 1, 2$, sei speziell $c_{j0} = a_j, c_{j2} = b_j, c_{j1} = \frac{a_j + b_j}{2}, j = 1, 2$. Die Stützstellen sind hier gegeben durch $\underline{c}_s, 0 \le s_1, s_2 \le 2$. Die zugehörigen Lagrange Polynome besitzen die Darstellung

$$L_{j0}(x_j) = \frac{x_j - c_{j1}}{c_{j0} - c_{j1}} \cdot \frac{x_j - c_{j2}}{c_{j0} - c_{j2}}$$

$$L_{j1}(x_j) = \frac{x_j - c_{j0}}{c_{j1} - c_{j0}} \cdot \frac{x_j - c_{j2}}{c_{j1} - c_{j2}}$$

$$L_{j2}(x_j) = \frac{x_j - c_{j0}}{c_{j2} - c_{j0}} \cdot \frac{x_j - c_{j1}}{c_{j2} - c_{j1}}$$

mit $x = (x_1, x_2) \in \mathbb{R}^2$. Berechnen Sie die Ableitungen $L'_{j\nu}, L''_{j\nu}, j = 1, 2, \nu = 0, 1, 2$, sowie die kanonische Darstellung der partiellen Ableitungen

$$\frac{\partial u}{\partial x_j}(\underline{c}_{11}) = \sum_{0 \le s_1, s_2 \le 2} \frac{\partial p_s}{\partial x_j}(\underline{c}_{11}) u(\underline{c}_s)\ ,$$

$$\frac{\partial^2 u}{\partial x_j^2}(\underline{c}_{11}) = \sum_{0 \le s_1, s_2 \le 2} \frac{\partial^2 p_s}{\partial x_j^2}(\underline{c}_{11}) u(\underline{c}_s)\ , \qquad u \in \mathcal{Q}_2(\mathbb{R}^2)\ ,$$

bei $\underline{x} = \underline{c}_{11}$, wobei die biorthonormale Basis von Polynomen in $\mathcal{Q}_2(\mathbb{R}^2) = \mathcal{Q}_{22}(\mathbb{R}^2)$ gegeben ist durch

$$p_{00} = L_{10} \cdot L_{20}\ ,\ p_{10} = L_{11} \cdot L_{20}\ ,\ p_{20} = L_{12} \cdot L_{20}\ ,$$
$$p_{01} = L_{10} \cdot L_{21}\ ,\ p_{11} = L_{11} \cdot L_{21}\ ,\ p_{21} = L_{12} \cdot L_{21}\ ,$$
$$p_{02} = L_{10} \cdot L_{22}\ ,\ p_{12} = L_{11} \cdot L_{22}\ ,\ p_{22} = L_{12} \cdot L_{22}\ .$$

20. **Aufgabe:** Zeigen Sie, daß durch

$$|||u||| := \sum_{k=1}^{\ell} |q_k(u)| + |u|_E\ , \qquad u \in E\ ,$$

eine Norm auf E erklärt wird, wobei $|.|_E$ eine Halbnorm auf E mit Nullraum M und $\{q_1, \ldots, q_\ell\}$ ein M–unisolventes System von stetigen Linearformen auf E ist.

Hierbei heißt $\{q_k\} \subset M'$ *M–unisolvent*, wenn die verallgemeinerte Interpolationsaufgabe $q_k(u) = \lambda_k\ ,\ k = 1, \ldots, \ell$, für jede rechte Seite eine eindeutige Lösung $u \in M$ besitzt.

21. **Aufgabe:** Gegeben seien paarweise verschiedene Punkte $x_0, \ldots, x_m \in [a, b]$ und zugehörige Zahlen $r_0, \ldots, r_m \in \mathbb{N}$, für die zugehörige *Hermite–Interpolation* einer Funktion $u \in C^r[a, b]$ $(r := r_0 + \cdots + r_m)$ durch

$$(Pu =)p \in \mathcal{P}_{r-1} : \frac{d^s p}{dx^s}(x_j) = \frac{d^s u}{dx^s}(x_j)\ , \qquad s = 0, \ldots, r_j - 1\ ,\ j = 0, \ldots, m\ ,$$

definiert ist. Betrachtet man die Linearform

$$f(u) = \frac{d^k u}{dx^k}(c)$$

für ein $c \in [a, b]$ und ein $k \in \{0, \ldots, r\}$, dann ergibt die zugehörige interpolatorische Approximation gerade die *Differenzenapproximation*

$$q(u) = q(Pu) = \sum_{j=0}^{m} \sum_{s=0}^{r_j - 1} \gamma_{js}^{(k)} \frac{d^s u}{dx^s}(x_j)$$

mit den Koeffizienten

$$\gamma_{js}^{(k)} = \frac{d^k w_{js}}{dx^k}(c)\ , \qquad s = 0, \ldots, r_j - 1\ ,\ \ j = 0, \ldots, m\ ,$$

und den biorthonormalen Polynomen $w_{js} \in \mathcal{P}_{r-1}$, die durch

$$\frac{d^s w_{rt}}{dx^s}(x_j) = \delta_{st}\delta_{jr}$$

erklärt sind.

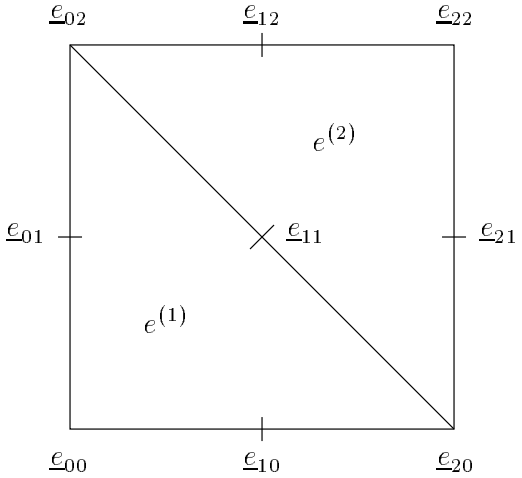
Geben Sie die Differenzenapproximation an für den Fall der Lagrange–Interpolation (d. h. $r_j = 1\,,\ r = m+1$) und der Hermite–Interpolation der Ordnung $r_j = 2$ für alle $j = 0, \ldots, m$. Rechnen Sie die Differenzenapproximationen im Fall der Lagrange–Interpolation für die Fälle

$$k = 1\,,\ r = 2\,; \quad k = 2\,,\ r = 3\,; \quad k = 1\,,\ r = 3$$

explizit aus.

22. **Aufgabe:** Betrachten Sie das Einheitsquadrat, das in 2 Einheitsdreiecke unterteilt ist.

Auf jedem der Dreiecke $e^{(1)}$ und $e^{(2)}$ sei die durch das „Dreieck von Morley“ definierte Hermitesche Interpolationsaufgabe gestellt.



(a) Geben Sie die Lösung u ($= u^{(1)}$ auf $e^{(1)}, = u^{(2)}$ auf $e^{(2)}$) dieser Aufgabe an, wenn u stetig bei $\underline{e}_{02}$ und $\underline{e}_{20}$ und $\partial u/\partial \underline{n}_{11}$ stetig bei $\underline{e}_{11}$ ist, d. h. für $u^{(1)}$ und $u^{(2)}$ werden dieselben Werte bei $\underline{e}_{02}$ und $\underline{e}_{20}$ vorgeschrieben und

$$\frac{\partial u^{(1)}}{\partial \underline{n}_{11}}(\underline{e}_{11}) = \frac{\partial u^{(2)}}{\partial \underline{n}_{11}}(\underline{e}_{11});$$

hierbei bezeichnet $\underline{n}_{11}$ die innere Normale für $e^{(1)}$ beziehungsweise die äußere Normale für $e^{(2)}$ bei $\underline{e}_{11}$.

(b) Ist u stetig auf der Verbindungsstrecke $\overline{\underline{e}_{02}\underline{e}_{20}}$?

23. **Aufgabe:** Sei E ein normierter Teilraum, $M \subset E$ ein endlichdimensionaler Teilraum und $\{q_1, \ldots, q_\ell\}$ ein M–unisolventes System von stetigen Linearformen auf E. Durch

$$|||u||| := \sum_{k=1}^{\ell} |q_k(u)| + |u|_E \,, \qquad u \in E \,,$$

ist dann eine Norm definiert, wenn $|.|_E$ eine Halbnorm auf E mit Nullraum M ist (s. Übung 20). Zeigen Sie: Eine Halbnorm $|.|_E$ mit Nullraum M ist dann und nur dann äquivalent zur Halbnorm $|., M|$, wenn die zugehörige Norm $|||.|||$ äquivalent zur Norm $\|.\|$ von E ist.

Hinweis: Benutzen Sie Lemma 5.1 der Vorlesung.

24. **Aufgabe:** Zeigen Sie die Ungleichungen

$$|u(c)| \le \varepsilon |u|_{1,p} + \varepsilon^{-p'/p} |u|_{0,p}, \quad u \in C^1[a,b],$$

für jede Stelle $c \in [a,b]$, $b > a$, jedes p in $1 \le p \le \infty$, $p' : 1/p + 1/p' = 1$, und jedes ε in $0 < \varepsilon \le \left(\frac{b-a}{2}\right)^{1/p'}$.

Hinweis: Benutzen Sie die Darstellung

$$\begin{aligned} u(c) &= -\int_c^{c+h} \frac{d}{dx}\left\{\frac{c+h-x}{h} u(x)\right\} dx \\ &= \frac{1}{h}\int_c^{c+h} u(x)\,dx - \int_c^{c+h} \frac{c+h-x}{h} u'(x)\,dx \end{aligned}$$

für h in $0 < h \le (b-a)/2$ oder $-(b-a)/2 \le h < 0$.

25. **Aufgabe:** Für die Lagrangesche interpolatorische Approximation $p \in \mathcal{P}_1(I_j)$ für ein Intervall $I_j = [x_{j-1}, x_j]$ gilt die Interpolationsabschätzung

$$|u-p|_{s,q,I_j} \le \sigma h_j^{2-s+\frac{1}{q}-\frac{1}{p}} |u|_{2,p,I_j} \,, \quad u \in C^2(I_j) \,, \;\; s = 0,1 \,, \;\; 1 \le p,q \le \infty \,,$$

wobei

$$\sigma = \sup_{0 \ne y \in \hat{L}^\perp} \frac{|y|_{s,q,[0,1]}}{|y|_{2,p,[0,1]}}$$

$$\hat{L}^\perp = \left\{ u \in C^2[0,1] | u(0) = u(1) = 0 \right\} \,.$$

Zeigen Sie für den Fall $s = 0$:

a) $\sigma \le \frac{1}{8}$ für $p = q = \infty$;

b) $\sigma \le \frac{1}{9}$ für $p = q = 2$;

c) $\sigma \le \frac{1}{12}$ für $q = 1,\ p = \infty$.

Hinweis: Verwenden Sie die Darstellung der Lösung von $y'' = f$ in $[0,1]$, $y(0) = y(1) = 0$, mit Hilfe der Greenschen Funktion (siehe z. B. Walter (1976), § 26).

26. **Aufgabe:** Gegeben seien paarweise verschiedene Punkte $x_0, \ldots, x_m \in [a,b]$ und zugehörige Zahlen $r_0, \ldots, r_m \in \mathbb{N}$, für die zugehörige *Hermite–Interpolation* einer Funktion $u \in C^r[a,b] (r := r_0 + \cdots + r_m)$ durch

$$(Pu =) p \in \mathcal{P}_{r-1} : \frac{d^s p}{dx^s}(x_j) = \frac{d^s u}{dx^s}(x_j) , \qquad s = 0, \ldots, r_j - 1 , \; j = 0, \ldots, m ,$$

definiert ist. Betrachtet man die Linearform

$$f(u) = \frac{d^k u}{dx^k}(c)$$

für ein $c \in [a,b]$ und ein $k \in \{0, \ldots, r\}$, dann ergibt die zugehörige interpolatorische Approximation gerade die *Differenzenapproximation*

$$q(u) = q(Pu) = \sum_{j=0}^{m} \sum_{s=0}^{r_j - 1} \gamma_{js}^{(k)} \frac{d^s u}{dx^s}(x_j)$$

mit den Koeffizienten

$$\gamma_{js}^{(k)} = \frac{d^k w_{js}}{dx^k}(c) , \qquad s = 0, \ldots, r_j - 1 , \; j = 0, \ldots, m ,$$

und den biorthonormalen Polynomen $w_{js} \in \mathcal{P}_{r-1}$, die durch

$$\frac{d^s w_{rt}}{dx^s}(x_j) = \delta_{st} \delta_{jr}$$

erklärt sind.

a) Wenden Sie die Fehlerabschätzung

$$|f(u) - q(u)| \leq \sigma h^{r-k} |u|_{r,\infty} , \qquad u \in C^r[a,b] .$$

auf die Beispiele aus Aufgabe 21 an.

b) Was ändert sich an der Fehlerabschätzung in a), wenn man $u \in H^{r,2}(a,b)$ betrachtet und die Norm $\| \cdot \|_{r,2}$ bzw. die Halbnorm $| \cdot |_{r,2}$ zugrundelegt?

Hinweis zu b): Wenden Sie die Sobolevsche Ungleichung oder Satz 5.23 aus Abschnitt 5.4 der Vorlesung an.

Literaturverzeichnis

[AB84] O. Axelsson und V. A. Barker. *Finite Element Solution of Boundary Value Problems. Theory and Computation.* Academic Press, Orlando, 1984.

[Ada75] R. A. Adams. *Sobolev Spaces.* Academic Press, New York, 1975.

[Agm65] S. Agmon. *Elliptic boundary value problems.* Van Nostrand, New York, 1965.

[Bat82] K.-J. Bathe. *Finite Element Procedures in Engineering Analysis.* Prentice Hall, Englewood Cliffs, 1982.

[Bat86] K.-J. Bathe. *Finite-Elemente-Methoden.* Springer, 1986.

[BCO81] E. B. Becker, G. F. Carey und J. T. Oden. *Finite elements. An introduction.* Prentice-Hall, Englewood Cliffs, 1981.

[Bra91] D. Braess. *Finite Elemente.* Springer, Berlin, 1991.

[Bra92] D. Braess. *Finite Elemente. Theorie, schnelle Löser und Anwendungen in der Elastizitätstheorie.* Springer, Berlin, 1992.

[BS94] S. C. Brunner und L. R. Scott. *The Mathematical Theory of Finite Element Methods.* Springer, New York, 1994.

[Bur87] D. S. Burnett. *Finite Element Analysis.* Addison-Wesley, Reading, 1987.

[Cia78] P. C. Ciarlet. *The finite element method for elliptic problems.* North-Holland, Amsterdam, 1978.

[CL90] P. G. Ciarlet und J. L. Lions, Hrsg. *Handbook of Numerical Analysis. Vol. II: Finite Element Methods (Part 1).* North-Holland, Amsterdam, 1990.

[Fai78] G. Fairweather. *Finite Element Galerkin methods for differential equations.* Lect. N. Pure & Appl. Math., Vol. 34, Dekker, New York, 1978.

[For84] O. Forster. *Analysis 3. Integralrechnung im $\mathbb{R}^n$ mit Anwendungen.* Vieweg, Braunschweig, 1984.

[Fri69] A. Friedman. *Partial differential equations.* Holt, Rinehart & Winston, New York, 1969.

[Gal76] R. H. Gallagher. *Finite-Element-Analysis. Grundlagen.* Springer, Berlin, 1976.

[GR92] C. Großmann und H.-G. Roos. *Numerik partieller Differentialgleichungen.* Teubner, Stuttgart, 1992.

[Gri85] P. Grisvard. *Elliptic Problems in Non-smooth Domains.* Pitman, New York, 1985.

[GRT89] H. Goering, H.-G. Roos und L. Tobiska. *Finite-Element-Methods.* Harri Deutsch, Thun-Frankfurt/M., 1989.

[Joh87] C. Johnson. *Numerical solution of partial differential equations by the finite element method.* Cambridge Univ. Press, Cambridge, 1987.

[Lad85] O. A. Ladyzhenskaya. *The Boundary Value Problems of Mathematical Physics. (Übersetzung aus dem Russischen).* Springer, New York, 1985.

[Mar86] J. T. Marti. *Introduction to Sobolev Spaces and Finite Element Solutions of Elliptic Boundary Value Problems.* Academic Press, New York, 1986.

[Mar89] D. Marsal. *Finite Differenzen und Elemente.* Springer, Berlin, 1989.

[MG80] A. R. Mitchell und D. F. Griffiths. *The finite element method in partial differential equations.* Wiley, Chichester, 1980.

[MM81] T. Meis und U. Marcowitz. *Numerical solution of partial differential equations.* Springer, New York, 1981.

[MM89] U. Meissner und A. Menzel. *Die Methode der Finiten Elemente.* Springer, 1989.

[NdV73] D. H. Norrie und G. de Vries. *The Finite Element Method.* Academic Press, New York, 1973.

[OR76] J. T. Oden und J. N. Reddy. *An introduction to the mathematical theory of finite elements.* Wiley, New York, 1976.

[Rei85] H.-J. Reinhardt. *Analysis of approximation methods for differential and integral equations.* Springer, New York, 1985.

[Rei95] H.-J. Reinhardt. Einführung in die Funktionalanalysis. Skriptum SS 1995. Fachbereich Mathem., Univ.-GH Siegen, 1995.

[Ric86] W. Richter. *Numerische Lösung partieller Differentialgleichungen mit der Finite-Elemente-Methode.* Vieweg, Braunschweig, 1986.

[SB91] B. Szabo und I. Babuska. *Finite element analysis.* Wiley, New York, 1991.

[Sch84] H. R. Schwarz. *Methode der finiten Elemente.* Teubner, Stuttgart, 1984.

[Sch88] H. R. Schwarz. *FORTRAN-Programme zur Methode der finiten Elemente.* Teubner, Stuttgart, 1988.

[Sew85] G. Sewell. *Analysis of a Finite Element Method: PDE/PROTRAN.* Springer, New York, 1985.

[SF73] G. Strang und G. J. Fix. *An analysis of the finite element method.* Prentice-Hall, Englewood Cliffs, 1973.

[SH82] F. Stummel und K. Hainer. *Praktische Mathematik.* Teubner, Stuttgart, 1982.

[Str71] A. H. Stroud. *Approximate calculation of multiple integrals.* Prentice–Hall, Englewood Cliffs, N. J., 1971.

[STW90] A. H. Schatz, V. Thomee und W. L. Wendland. *Mathematical theory of finite and boundary element methods.* DMV Seminar Vol. 15. Birkhäuser, Basel, 1990.

[TGK91] W. Törnig, M. Gipser und B. Kaspar. *Numerische Lösung von partiellen Differentialgleichungen der Technik. Differenzenverfahren, Finite Elemente und die Behandlung großer Gleichungssysteme.* Teubner, Stuttgart, 1991.

[Wal76] W. Walter. *Gewöhnliche Differentialgleichungen. Eine Einführung.* Springer, Berlin, 1976.

[Wal92] W. Walter. *Analysis 2.* Springer, Berlin, 1992.

[Whi73] J. R. Whiteman. *The mathematics of finite elements and applications.* Academic Press, New York, 1973.

[Zau83] E. Zauderer. *Partial differential equations of applied mathematics.* Wiley, New York, 1983.

[ZT89] O. C. Zienkiewicz und R. L. Taylor. *The Finite Element Method. Vol. 1 - Basic Formulation and Linear Problems.* McGraw-Hill, London, 1989.

Index

Ableitung
  schwache, 28
Abschätzung
  quasi–optimale, 14
Anfangswertproblem, 205
Approximation
  hermitesche interpolatorische, 131
  interpolatorische, 121, 165
Assemblierung, 23
Aubin–Nitsche–Trick, 46, 163

baryzentrische Koordinaten, 89
Basis
  duale, 56
Bildraum, 123
Bilinearform, 13
  auf V elliptische, 37
  beschränkte, 37
  gestörte, 173
  positiv definite, 37
  positiv semidefinite, 37
  symmetrische, 37
biorthonormal, 56

Crank–Nicolson–Galerkin–Verfahren, 197, 199

d'Alembertsches Reduktionsverfahren, 217
Dachfunktion, 7
Darstellung
  kanonische, 56, 57
Determinantenabschätzung, 157
Differentialgleichung
  elliptische, 227
  hyperbolische, 200, 229
    erster Ordnung, 230
  parabolische, 229, 233
Differentialgleichung n-ter Ordnung, 214
Differentialgleichung zweiter Ordnung, 216
Differentialoperator
  selbstadjungierte Form, 222
Differenzenapproximation
  Randwertaufgabe, 17
Differenzenoperator
  von positivem Typ, 18
Differenzenquotient, 59, 145
  l-ter Ordnung, 65
  in $\mathcal{Q}_{22}$, 106
  in $\mathcal{Q}_1(\mathbb{R}^2)$, 102
  zentraler
    zweiter Ordnung, 10, 17
direkte Summe, 122
Dreieck von Morley, 113
Dualraum
  algebraischer, 56
  topologischer, 56

Einbettung
  vollstetige, 129
Einheitssimplex, 92
Elastischer Stab, 3
Elastizitätsmodul, 3
Elliptizitätsbedingung, 228
Energie
  aufgebrachte, 4
  elastische, 15
  innere elastische, 4
  potentielle, 4, 15
Energieerhaltung, 3
Energienorm, 13, 33
Eulersche Differentialgleichung, 16

Eulersche Gleichung, 36
Existenzsatz von Peano, 207

5–Punkt–Differenzenapproximation, 24
Fehlerabschätzung, 20
Fehlergleichung, 10, 14, 20
Finite Elemente, 52
    affin äquivalente, 169
    isoparametrische, 193
    reguläre Familie, 169
Finite Elemente Approximation, 32
finites Element, 168
Fourier-Koeffizient, 231
Fouriersches Gesetz, 3
Fundamentalsystem, 209

Galerkin Methode, 9, 13
Galerkin–Verfahren
    halb–diskretes, 202
    halbdiskretes, 196
Gaußsche Quadraturformel, 70, 144
Gesamtsteifigkeitsmatrix, 24, 170
Gitterpunkt, 7
Gleichgewichtszustand, 3
Gleichung
    elliptische
      klassische Lösung, 233
Gramsche Determinante, 60, 68
Gramsche Matrix, 20, 61, 86
Greensche Formel, 28, 31
Greensche Funktion, 225
    Konstruktion, 225
Grundlösung, 224

Halbnorm
    äquivalente, 124
    koerzitive, 129
    stetige, 124
Hamiltonsches Prinzip, 15
harmonische Funktion, 228
harmonisches Polynom, 228
Hauptsystem, 209
Hermite–Interpolation, 66
    der Ordnungen 2, 68
Hilbertraum $L^2(a,b)$, 27
Hookesches Gesetz, 3
Hyperkreis–Fehlerabschätzung, 163

Inkreis, 156
Interpolationsabschätzungen, 140, 142, 143
Interpolationsaufgabe, 76
    hermitesche, 110, 111, 154
    verallgemeinerte, 55, 60
Interpolationsfehler
    Abschätzung, 11
Interpolationsoperator, 123
Inverse Abschätzung, 158
Isomorphismus
    kanonischer, 164

kürzester Abstand, 123
Kegelbedingung, 35
Kettenregel, 154
Koerzitivitätsungleichung, 129
Konditionszahl, 26, 157
Konsistenzordnung, 17
Konvergenzordnung
    optimale, 175, 182
Koordinatensystem
    schiefwinkliges, 87
Kritische Dämpfung, 219

Lagrange Polynom, 64, 98
Lagrange–Interpolation, 64
    in $\mathcal{P}_2^1(\Delta)$, 70
    in $\mathcal{P}_2^1(\Delta)$, 141
Lagrange-Identität, 222
Lagrangesche Interpolationsaufgabe, 86
Laplace–Operator, 15
Laplace-Gleichung, 227, 230
Laplacescher Differentialoperator, 227
Lastvektor, 9
Lemma
    von Bramble–Hilbert, 127
    von Lax–Milgram, 41
lineare Interpolierende, 11
Linienmethode, 197
Lipschitzbedingung, 206
lokale Basis, 9

M–Unisolvenz, 60
Matrix
  Kondition einer, 25
Maximumprinzip
  diskretes, 18
Membran
  elastische, 15
Menge
  M–unisolvente, 60
Methode der Finiten Elemente, 8
Methode der Variation der Konstanten, 211, 215
Minimierungsproblem, 4, 16, 31
Mittelwertoperator, 17
Monom, 82
Multiindex
  Ordnung eines, 82

n–Simplex, 87
9–Punkte–Mittelwertoperator, 24
Newtonsche Darstellung, 65
Normale
  äußere, 28
Nullraum, 123

Orthogonalraum, 122

parabolische Probleme, 195
Parallelogrammidentität, 42
Plattengleichung, 229
Poisson-Gleichung, 228
Polynom
  charakteristisches, 218
Polynom–Splines, 54
Polynome, 82
  kubische, 83
  bikubische, 85
  bilineare, 85
  biquadratische, 85
  lineare, 83
  quadratische, 83
  quartische, 83
  quintische, 83
Potentialgleichung, 227
Prinzip der minimalen potentiellen Energie, 4
Prinzip der virtuellen Arbeit, 4
Projektionsoperator, 123
Punktestern, 17

Quadraturformel, 92, 94, 97, 99–101, 144
  Genauigkeitsgrad, 176
  interpolatorische, 65
  von der Ordnung r, 176
Quotientenraum, 178

Randbedingung, 228
  Dirichlet, 31, 228
  erzwungene, 32
  natürliche, 32
  Neumann, 31, 228
  Robbin, 228
  wesentliche, 32
Randwertaufgabe
  2–Punkt, 2, 12
  schwache Formulierung, 30
  schwache Lösung, 30
  Sturmsche, 222
  zweiter Ordnung, 221
Randwertproblem, 4, 205
  elliptisches
    schwache Form, 40
  erstes
    der Potentialtheorie, 14
  für einen elliptischen Differentialoperator, 36
  schwache Lösung, 40
Raum $\mathcal{P}_r(\Delta)$, 52
Raum $\mathcal{P}_r^m(\Delta)$, 53, 54
Rechteck von Adini, 116
Referenzelement, 168
Regularitätsbedingung, 169
Relation
  charakteristische, 74
Restgliedabschätzung, 123
Restgliedoperator, 123
Ritz Methode, 9
Ritzsches Projektionsverfahren, 19, 45

Satz von Picard-Lindelöf, 206
Satz von Rellich, 152
Schrittweitenverhältnis, 199, 203
Schwerpunkt, 94
Schwingung, 219
    gedämpfte, 219
    gedämpfte freie, 220
    ungedämpfte erzwungene, 220
    ungedämpfte freie, 220
Segmenteigenschaft, 186, 187
Sehnentrapezformel
    summierte, 58, 145
Separation der Variablen, 231
Simpson–Formel, 10
    summierte, 72
Spline–Funktion, 54
    quadratische, 73
Spline-Funktion
    quadratische, 74
Splineapproximation
    quadratische, 142
    lineare, 139
Spur, 210
Stabilitätsabschätzung, 199, 204
Steifigkeitsmatrix, 9
Sturm-Liouvillesches Eigenwertproblem, 231
Superposition, 232, 234
System
    homogenes, 209
System von n Differentialgleichungen, 207

Taylor–Interpolation, 63, 138
Taylorsche Formel, 64
Teilraum $L^{\perp}$, 136
Träger, 9
Triangulierung, 91, 159
    reguläre, 159
    gleichförmige, 24
    quasi–uniforme, 160

Überdämpfung, 219
Umkreis, 156
Unabhängigkeit
    affine, 87, 89
Ungleichung
    Cauchy–Schwarzsche, 11
    Sobolevsche, 148
    verallgemeinerte Poincarésche, 152
    von Poincaré–Friedrichs, 29, 37
Unterdämpfung, 219
Unterteilung, 7

Variation
    erste, 40
Variationsformulierung, 40
Variationsgleichung, 14, 16, 40
Variationsproblem, 4, 16, 31, 40
    endlichdimensionales, 9
Vervollständigung von Räumen, 30
Volumen, 92

Wärmeleitfähigkeit, 4
Wärmeleitung, 3
Wärmeleitungsgleichung, 195, 229
    schwache Lösung, 196
Wellengleichung, 200, 229
    charakteristische Form, 201
    verallgemeinerte, 201
Winkelbedingung, 161
Wronski-Determinante, 210