

Wolfgang Hackbusch

Theorie und Numerik elliptischer Differentialgleichungen

Copyright beim Autor

Inhaltsverzeichnis

Vorwort	vii
Notationen	ix
1 Partielle Differentialgleichungen und ihre Einteilung in Typen	1
1.1 Beispiele	1
1.2 Typeneinteilungen bei Gleichungen zweiter Ordnung	4
1.3 Typeneinteilungen bei Systemen erster Ordnung	5
1.4 Unterschiedliche Eigenschaften der verschiedenen Typen	6
2 Die Potentialgleichung	9
2.1 Problemstellung	9
2.2 Singularitätenfunktion	10
2.3 Mittelwerteigenschaft und Maximumprinzip	13
2.4 Stetige Abhängigkeit von den Randdaten	16
3 Die Poisson-Gleichung	19
3.1 Problemstellung	19
3.2 Lösungsdarstellung mit der Greenschen Funktion	20
3.3 Die Greensche Funktion für die Kugel	24
3.4 Die Neumann-Randwertaufgabe	24
3.5 Die Integralgleichungsmethode	25
4 Differenzenmethode für die Poisson-Gleichung	27
4.1 Einführung: Der eindimensionale Fall	27
4.2 Fünfpunktformel	29
4.3 M-Matrizen, Matrixnormen, positiv definite Matrizen	32
4.4 Eigenschaften der Matrix L_h	38
4.5 Konvergenz	43
4.6 Differenzenverfahren höherer Ordnung	45
4.7 Die Diskretisierung der Neumann-Randwertaufgabe	47
4.7.1 Einseitige Differenz für $\partial u / \partial n$	47
4.7.2 Symmetrische Differenz für $\partial u / \partial n$	50
4.7.3 Symmetrische Differenz für $\partial u / \partial n$ im verschobenen Gitter	51
4.7.4 Beweis des Stabilitätsatzes 4.7.7	52
4.8 Diskretisierung der Poisson-Gleichung im beliebigen Gebiet	57

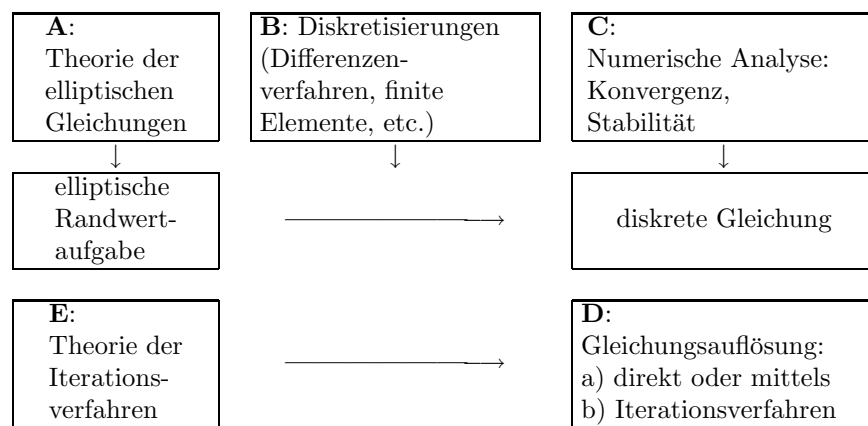
4.8.1	Shortley-Weller-Approximation	57
4.8.2	Interpolation in randnahen Punkten	60
5	Allgemeine Randwertaufgaben	61
5.1	Dirichlet-Randwertaufgaben für lineare Differentialgleichungen zweiter Ordnung	61
5.1.1	Problemstellung	61
5.1.2	Maximumprinzip	62
5.1.3	Eindeutigkeit der Lösung und stetige Abhängigkeit	64
5.1.4	Differenzenverfahren für die allgemeine Differentialgleichung zweiter Ordnung	66
5.1.5	Greensche Funktion	69
5.2	Allgemeine Randbedingungen	70
5.2.1	Formulierung der Randwertaufgabe	70
5.2.2	Differenzenverfahren bei allgemeinen Randbedingungen	71
5.3	Randwertaufgaben höherer Ordnung	74
5.3.1	Die biharmonische Differentialgleichung	74
5.3.2	Allgemeine lineare Differentialgleichung der Ordnung $2m$	74
5.3.3	Diskretisierung der biharmonischen Differentialgleichung	76
6	Exkurs über Funktionalanalysis	79
6.1	Banach-Räume und Hilbert-Räume	79
6.1.1	Normierte Räume	79
6.1.2	Operatoren	80
6.1.3	Banach-Räume	80
6.1.4	Hilbert-Räume	82
6.2	Sobolev-Räume	82
6.2.1	Der Raum $L^2(\Omega)$	82
6.2.2	Die Räume $H^k(\Omega)$ und $H_0^k(\Omega)$	84
6.2.3	Fourier-Transformation und $H^k(\mathbb{R}^n)$	86
6.2.4	$H^s(\Omega)$ für reelles $s \geq 0$	87
6.2.5	Spur- und Fortsetzungssätze	88
6.3	Dualräume	93
6.3.1	Dualraum eines normierten Raumes	93
6.3.2	Adjungierte Operatoren	94
6.3.3	Skalen von Hilbert-Räumen	95
6.4	Kompakte Operatoren	96
6.5	Bilinearformen	98
7	Variationsformulierung	103
7.1	Historische Bemerkungen	103
7.2	Gleichungen mit homogenen Dirichlet-Randbedingungen	104
7.3	Inhomogene Dirichlet-Randbedingung	108
7.4	Natürliche Randbedingungen	109
8	Die Methode der finiten Elemente	115
8.1	Das Ritz-Galerkin-Verfahren	115
8.1.1	Grundlagen	115
8.1.2	Diskussion der diskreten Gleichung	117
8.1.3	Beispiele	118
8.1.4	Diskrete inf-sup-Bedingung	120
8.2	Fehlerabschätzungen	120
8.2.1	Quasioptimalität	120
8.2.2	Konvergenz der Ritz-Galerkin-Lösungen	121
8.2.3	Ritz-Projektion	122
8.2.4	Weitere Stabilitäts- und Fehlerabschätzungen	123
8.3	Finite Elemente	124

8.3.1	Einführung: Lineare Elemente für $\Omega = (a, b)$	124
8.3.2	Lineare Elemente für $\Omega \subset \mathbb{R}^2$	126
8.3.3	Bilineare Elemente für $\Omega \subset \mathbb{R}^2$	128
8.3.4	Quadratische Elemente für $\Omega \subset \mathbb{R}^2$	129
8.3.5	Elemente für $\Omega \subset \mathbb{R}^3$	130
8.3.6	Behandlung von Nebenbedingungen	131
8.4	Fehlerabschätzungen bei Finite-Elemente-Verfahren	132
8.4.1	Vorbereitungen	133
8.4.2	Eigenschaften von Folgen von Finite-Element-Räumen	135
8.4.3	H^1 -Abschätzungen für lineare Elemente	135
8.4.4	L^2 -Abschätzungen für lineare Elemente	137
8.5	Verallgemeinerungen	139
8.5.1	Fehlerabschätzungen für andere Elemente	139
8.5.2	Finite Elemente für Gleichungen höherer Ordnung	139
8.5.3	Finite Elemente für Nichtpolygon-Gebiete	141
8.6	A-posteriori-Fehlerabschätzungen, Adaptivität	143
8.6.1	A-posteriori-Fehlerabschätzungen	143
8.6.2	Effizienz der Finite-Element-Methode	146
8.6.3	Adaptive Finite-Element-Methode	147
8.7	Eigenschaften der Systemmatrix	149
8.7.1	Zusammenhang von \mathbf{L} und L_h	149
8.7.2	Normäquivalenzen und Massematrix	150
8.7.3	Inverse Abschätzung und Kondition von \mathbf{L}	151
8.7.4	Elementmatrizen	153
8.7.5	Positivität, Maximumprinzip	154
8.8	Weitere Hinweise	154
8.8.1	Nichtkonforme Elemente	154
8.8.2	Nichtzulässige Triangulationen	156
8.8.3	Trefftz-Verfahren	156
8.8.4	Finite-Element-Verfahren für singuläre Lösungen	157
8.8.5	Hierarchische Basen	157
8.8.6	Superkonvergenz	158
8.8.7	Die Mörtelmethode ("mortar finite elements")	158
8.8.8	Verwandte Diskretisierungen	160
9	Regularität	161
9.1	Lösungen der Randwertaufgabe in $H^s(\Omega)$, $s > m$	161
9.1.1	Das Regularitätsproblem	161
9.1.2	Regularitätssätze für $\Omega = \mathbb{R}^n$	162
9.1.3	Regularitätssätze für $\Omega = \mathbb{R}_+^n$	168
9.1.4	Regularitätssätze für allgemeines $\Omega \subset \mathbb{R}^n$	170
9.1.5	Regularität bei konvexem Gebiet und Gebieten mit Ecken	173
9.1.6	Regularität im Inneren	175
9.2	Regularitätseigenschaften der Differenzengleichungen	176
9.2.1	Diskrete H^1 -Regularität	176
9.2.2	Konsistenz	180
9.2.3	Optimale Fehlerabschätzungen	185
9.2.4	H_h^2 -Regularität	186
10	Spezielle Differentialgleichungen	189
10.1	Differentialgleichungen mit unstetigen Koeffizienten	189
10.1.1	Formulierung	189
10.1.2	Diskretisierung	190
10.2	Ein singulär gestörtes Problem	191
10.2.1	Die Konvektionsdiffusionsgleichung	191

10.2.2	Stabile Differenzenschemata	193
10.2.3	Finite Elemente	194
11	Eigenwertprobleme elliptischer Operatoren	199
11.1	Formulierung der Eigenwertprobleme	199
11.2	Finite-Element-Diskretisierung	200
11.2.1	Diskretisierung	200
11.2.2	Qualitative Konvergenzresultate	201
11.2.3	Quantitative Konvergenzresultate	204
11.2.4	Komplementäre Probleme	207
11.3	Diskretisierung durch Differenzenverfahren	209
12	Stokes-Gleichungen	215
12.1	Elliptische Differentialgleichungssysteme	215
12.2	Variationsformulierung	217
12.2.1	Schwache Formulierung der Stokes-Gleichungen	217
12.2.2	Sattelpunktprobleme	218
12.2.3	Existenz und Eindeutigkeit der Lösung eines Sattelpunktproblems	220
12.2.4	Lösbarkeit und Regularität des Stokes-Problems	223
12.2.5	Eine V_0 -elliptische Variationsformulierung der Stokes-Gleichung	225
12.3	Gemischte Finite-Element-Methode für das Stokes-Problem	226
12.3.1	Finite-Element-Diskretisierung eines Sattelpunktproblems	226
12.3.2	Stabilitätsbedingungen	227
12.3.3	Stabile Finite-Element-Räume für das Stokes-Problem	228
	Literaturverzeichnis	233
	Index	238

Vorwort

Die partiellen Differentialgleichungen vom elliptischen Typ führen zu einer Reihe von Aufgaben, die in der folgenden Skizze dargestellt werden.



Die *Theorie* der elliptischen Differentialgleichungen (**A**) beschäftigt sich unter anderem mit den Fragen nach der Existenz, Eindeutigkeit und Eigenschaften der Lösung der Randwertaufgaben. Das erste Problem der *Numerik* ist die Beschreibung von *Diskretisierungsverfahren* (**B**), die endlichdimensionale Gleichungen für Näherungen der Lösung ergeben. Der anschließende Teil der Numerik ist die *numerische Analyse* (**C**) der entsprechenden Verfahren. Insbesondere ist zu klären, ob und wie schnell die Näherung gegen die exakte Lösung konvergiert. Die Auflösung der endlichdimensionalen Gleichungen (**D**, **E**) ist im Allgemeinen kein einfaches Problem, da die Zahl der Unbekannten eine Million oder mehr betragen kann. Die Diskussion der Aufgabenbereiche **D**, **E** ist hier ausgespart, da es das Thema der Monographie Hackbusch [42] ist.

Die Beschreibung der Diskretisierungsverfahren und ihre Analyse steht in engem Zusammenhang mit entsprechenden Kapiteln aus der Theorie der elliptischen Differentialgleichungen. Außerdem ist eine fundierte numerische Analyse ohne Grundkenntnisse der Theorie der elliptischen Gleichungen nicht möglich. Damit bietet es sich an, die Theorie und die Numerik gemeinsam zu präsentieren.

Das Buch ist zunächst als eine Einführung in die Behandlung elliptischer Randwertaufgaben gedacht. Es soll den Leser aber auch an weiterführende Literatur zu speziellen Themen und an anwendungsorientierte Aufgaben heranführen. Absichtlich wurden Kapitel, die häufig zu kurz behandelt werden (z.B. Eigenwert-aufgaben) ausführlicher dargestellt.

Die Darstellungen beschränken sich grundsätzlich auf lineare elliptische Gleichungen. Damit ist die Diskussion der für die Anwendungen in der Strömungsmechanik wichtigen Navier-Stokes-Gleichung zwar ausge-

schlossen, aber man findet den Zugang hierzu über die Stokes-Gleichung, die als ein Beispiel eines elliptischen Systems eingehend behandelt wird.

Die aufgeführten Übungsaufgaben, die auch als Bemerkungen ohne Beweis verstanden werden können, sind integrierter Bestandteil der Darstellung. Wird dieses Buch als Grundlage einer Vorlesung benutzt, können sie als Übungen dienen. Der Leser sollte aber auch versuchen, sein Verständnis der Lektüre an den Aufgaben zu testen.

Das Buch entstand aus Vorlesungen an der Ruhr-Universität Bochum und an der Christian-Albrechts-Universität zu Kiel. Es wurden zwei Auflagen vom Teubner-Verlag Stuttgart publiziert ([43]). Anstelle einer dritten Buchauflage wird nun eine überarbeitete und stellenweise erweiterte Version des Buches den interessierten Lesern elektronisch bereitgestellt. Als Papierversion ist darüber hinaus die englische Übersetzung [45] erhältlich.

Der Autor dankt insbesondere Herrn J. Burmeister (Kiel) für die Zuarbeit beim Erstellen der Latex-Version.

Leipzig, Sommer 2005

W. Hackbusch

Erste Fassung: 15.07.2005

Notationen

Lateinisches Alphabet:

$a(\cdot, \cdot)$	Bilinearform (vgl. Seite 98)
$a_\lambda(\cdot, \cdot)$	verschobene Bilinearform (vgl. Seite 201)
$b(\cdot, \cdot)$	Bilinearform in §12.2 (vgl. Seite 218)
b_i	Finite-Element-Basisfunktionen (vgl. Seite 125)
B	unter anderem Symbol für Randdifferentialoperator (vgl. Seiten 70 und 110)
C, C', C_0, \dots	Konstanten
\mathbb{C}	Menge der komplexen Zahlen
$C^0(D)$	Menge der stetigen Funktionen definiert auf D
$C^k(D), C^\infty(D)$	Menge der k -fach bzw. unendlich oft stetig differenzierbaren Funktionen auf D
$C_0^\infty(\Omega)$	C^∞ -Funktionen mit kompaktem Träger in Ω (vgl. Seite 83)
$C^\lambda(D)$	Menge der Hölder-stetigen Funktionen auf D (vgl. Seite 21)
$C^t, C^{k,1}$	Menge von Gebieten mit entsprechenden Glattheitsbedingungen (vgl. Seite 90)
C_E	Koerzivitätskonstante (vgl. Seite 101)
C_S	Stabilitätskonstante (vgl. Seite 115)
$\text{cond}_2(\dots)$	Spektralkondition (vgl. Seite 78)
$d(u, V_N)$	Abstand zwischen Funktion u und Unterraum V_N (vgl. Seite 121)
D^ν	$ \nu $ -fache partielle Ableitung (vgl. (3.2.5b))
\det	Determinantenfunktion
$\text{diag}\{a_{\alpha\alpha} : \alpha \in I\}$	Diagonalmatrix mit Diagonalelementen $a_{\alpha\alpha}$
div	Divergenz(operator)
\mathbf{e}, \mathbf{e}_j	Einheitsvektor
e	in §8 oft Fehler der Finite-Element-Lösung (vgl. Seite 137), in §11: Eigenvektor
e^*	in §11: Eigenvektor des adjungierten Problems (Linkseigenvektor)
e^h, e^{*h}	in §11: Eigenvektor und Linkseigenvektor der Finite-Element-Diskretisierung
$E(\lambda), E^*(\lambda)$	Eigenräume (vgl. Seite 199)
\mathcal{F}	Fourier-Transformation (vgl. Seite 86)
grad	Gradient (auch ∇ ; vgl. Seite 11)
h	Schrittweite des Differenzenverfahrens (vgl. Seite 29), ähnlich für finite Elemente (vgl. Seite 135)
$H^k(\Omega), H^s(\Omega)$	Sobolev-Räume (vgl. Seite 84)
$H_0^k(\Omega), H_0^s(\Omega)$	Sobolev-Räume mit Nullrandbedingung (vgl. Seite (85))
$\mathbf{H}^s(\Omega), \mathbf{H}_0^s(\Omega)$	Sobolev-Räume der vektorwertigen Funktionen (vgl. Seite 217)
I	Einheitsmatrix

I	Indexmenge für Vektoren und Matrizen (vgl. Seite 30)
\mathbb{K}	Körper $\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$
$K_R(\mathbf{x})$	offene Kugel (in \mathbb{R}^n) um \mathbf{x} mit Radius R
\log	natürlicher Logarithmus
L	Differentialoperator (vgl. Seite 5)
L_h	Matrix des Differenzenoperators (vgl. Seite 29)
L_N, L_h	Operator zur Bilinearform auf V_N, V_h (vgl. Seite 117)
\mathbf{L}	Finite-Element-Matrix (vgl. Seite 116)
$L^2(\Omega)$	Menge der quadratintegrablen Funktionen (vgl. Seite 82)
$L_0^2(\Omega)$	Unterraum von $L^2(\Omega)$ (vgl. Seite 217)
$L^\infty(\Omega)$	Menge der beschränkten Funktionen (vgl. Seite 80)
$L(X, Y)$	lineare und beschränkte Abbildungen von X nach Y (vgl. Seite 80)
\vec{n}	Normalenvektor (vgl. Seite 11)
\mathbb{N}	Menge der natürlichen Zahlen $\{1, 2, \dots\}$
\mathbb{N}_0	Menge der natürlichen Zahlen mit Null: $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$
O	Nullmatrix
$\mathcal{O}(\cdot)$	Landau-Symbol
p	in §12 Druckkomponente aus der Stokes-Gleichungen
P	Prolongation der Vektoren in die Finite-Element-Funktionen (vgl. Seite 116)
P_h	Prolongation der Gitterfunktionen in kontinuierliche Funktionen (vgl. Seite 187)
Q_N	orthogonale Projektion auf V_N (vgl. Seite 122)
\mathbb{R}	Menge der reellen Zahlen
\mathbb{R}_+^n	Halbraum von \mathbb{R}^n (vgl. (6.2.18))
R_h, \tilde{R}_h	Restriktion, Abbildung von Funktion in Gitterfunktion (vgl. Seite 43)
$s(\mathbf{x}, \mathbf{y})$	Singularitätenfunktion (vgl. Seite 10)
S_N, S_h	Ritz-Projektion (vgl. Seite 122)
$Spur$	Spur einer Matrix (vgl. Seite 63)
\vec{t}	Tangentialrichtung (vgl. Seite 223)
\mathcal{T}	Finite-Element-Triangulation (vgl. Seite 126)
T, T_i	Dreiecke (oder allgemeinere Elemente) aus \mathcal{T} (vgl. Seite 126)
$Tr(u)$	Träger der Funktion u (vgl. Seite 83)
u	Standardbezeichnung für die Lösung der partielle Differentialgleichung
$u_x, u_y, u_{x_1}, u_{xx}, \dots$	partielle Ableitungen der Funktion u . Kurzschreibweise für $\partial u / \partial x$ usw.
u_h	Gitterfunktion auf $\bar{\Omega}_h$ (vgl. Seite 28)
u^N, u^h	Finite-Element-Lösung aus V_N bzw. V_h (vgl. Seite 116)
\mathbf{u}	Koeffizientenvektor zu Finite-Element-Funktion (vgl. Seite 116)
V	ab §8 Raum, in dem die Bilinearform definiert ist (vgl. Seite 115)
V_N, V_h	Finite-Element-Raum der Dimension N bzw. zur maximalen Schrittweite h (vgl. Seite 115)
W, W_h	Finite-Element-Raum für die Druckkomponente in §12 (vgl. Seite 12.2.4b)
$\mathbf{x} = (x_1, x_2, \dots, x_n)$	unabhängige Variable aus dem \mathbb{R}^n (vgl. Seite 4). Die Komponenten für $n = 2$ sind $\mathbf{x} = (x, y)$ (vgl. Seite 1), für $n = 3$: $\mathbf{x} = (x, y, z)$ (vgl. Seite 10)
x_1, x_2, \dots, x_n	Komponenten von \mathbf{x} im allgemeinen Fall (vgl. Anmerkungen zu \mathbf{x})
x, y, z	siehe \mathbf{x}
\mathbb{Z}	Menge der ganzen Zahlen
$Z = Z(\lambda), Z_h = Z_h(\lambda)$	spezielle Abbildungen aus (11.2.6a,b)

Griechisches Alphabet:

γ	unter anderem Spurabbildung von Funktionen (vgl. Seite 88)
Γ	Rand $\partial\Omega$ des Gebietes Ω (vgl. Seite 9)
Γ_h	Menge der Randgitterpunkte bei Differenzenverfahren (vgl. Seiten 29 und 57)
$\Gamma(x)$	Gamma-Funktion
Δ	Laplace-Operator (vgl. Seite 9).
Δ_h	diskreter Laplace-Operator (vgl. Seite 29)
$\eta(h)$	siehe Lemma 11.2.7 auf Seite 202

λ	unter anderem Eigenwert, insbesondere in §11
λ_h	unter anderem Eigenwert (vgl. Seite 200)
ν	Multiindex aus \mathbb{N}_0^n (vgl. Seite 21)
$\rho(A)$	Spektralradius einer Matrix A (vgl. Seite 34)
φ	häufig verwendet für die Randwerte (vgl. Seite 19)
ω_n	Oberfläche der n -dimensionalen Einheitskugel (vgl. Seite 11)
$\omega(\lambda), \omega_h(\lambda)$	spezielle Größen in §11 (vgl. Seite 201)
Ω	Gebiet der Randwertaufgabe, Teilmenge von \mathbb{R}^n (vgl. Seite 2)
Ω_h	Gitter für Differenzenverfahren (Menge der inneren Punkte) (vgl. (4.1.5a), (4.2.1a) und Seite 51)
$\overline{\Omega}_h$	Gitter für Differenzenverfahren (Menge der inneren und Randpunkte) (vgl. (4.1.5b), (4.2.1c))

Zeichensymbole:

a^*	adjungierte Bilinearform (vgl. Seite 98)
$A \geq B, A \leq B, A > B, A < B$	für Matrizen A, B komponentenweise Ungleichung (vgl. Seite 32)
A^\top	transponierte Matrix zu A
U^\perp	Orthogonalkomplement zu U (vgl. Seite 82)
X'	Dualraum zu X (vgl. Seite 93)
T'	Dualabbildung zu T (vgl. Seite 93)
T^*	adjungierte Abbildung zu T (vgl. Seite 94)
∂X	Rand von X
$\partial/\partial x$	partielle Ableitung nach x
$\partial/\partial n$	Normalenableitung (vgl. Seite 11)
∇u	Gradient von u (vgl. Seite 11)
∂^+, ∂^-	Vorwärts-, Rückwärtsdifferenz (vgl. Seite 27)
∂^0	symmetrische Differenz (vgl. Seite 27)
$\partial_n^-, \partial_n^0$	Differenz in Normalenrichtung (vgl. Seiten 47 und 50)
$\int \dots d\Gamma$	Oberflächenintegration (vgl. Seite 11)
\overline{X}	Abschluss einer Menge X
\hat{u}	Fourier-Transformierte von u (vgl. Seite 86)
$\langle \mathbf{x}, \mathbf{y} \rangle$	Euklidisches Skalarprodukt (vgl. Seite 11)
$(\cdot, \cdot) = (\cdot, \cdot)_X$	Skalarprodukt eines Hilbert-Raumes X (vgl. Seite 82)
$(\cdot, \cdot)_0$	Skalarprodukt $L^2(\Omega)$ (vgl. Seite 82)
$\langle x, x' \rangle_{X \times X'}$	Dualform (vgl. Seite 93)
$ \nu $	Länge des Multiindex ν (vgl. (3.2.5a))
$ \mathbf{x} $	Euklidische Norm des Vektors $\mathbf{x} \in \mathbb{R}^n$ (vgl. (2.2.2))
$ u $	spezielle Bedeutung in §4.3 (vgl. Seite 35)
$ u _s$	Sobolev-Norm in $H^s(\Omega)$ (vgl. Seiten 82 und 88)
$ u _s^\wedge$	über Fourier-Transformation definierte Sobolev-Norm (vgl. (6.2.15), (6.2.16a))
$ u _{k,0}$	Seminorm zu $ u _k$ (vgl. (6.2.9))
$\ \cdot\ _\infty$	Maximumnorm für Funktionen und Vektoren, Zeilensummennorm für Matrizen (vgl. Seiten 16, 36)
$\ \cdot\ _2$	Euklidische Norm und Spektralnrm (vgl. Seite 37)
$\ \cdot\ _X$	Norm zu X (vgl. Seite 79)
$\ T\ _{Y \leftarrow X}$	Operatornorm (vgl. Seite 80)
$u _\Gamma$	Beschränkung der Funktion auf Γ
$\mathbf{1}$	Vektor bestehend aus 1-Komponenten (vgl. Seite 36)
$\begin{bmatrix} & 1 & \\ 1 & -4 & 1 \\ & 1 & \end{bmatrix}$	usw. Sternschreibweise für Differenzenoperator (vgl. Seite 31)

1

Partielle Differentialgleichungen und ihre Einteilung in Typen

Eine gewöhnliche Differentialgleichungen bestimmt eine Funktion, die von nur *einer* Variablen abhängt. Die physikalischen Größen hängen aber im Allgemeinen von drei Raumvariablen und der Zeit ab. Auch wenn die Zeitabhängigkeit für stationäre Prozesse entfällt und sich durch spezielle geometrische Annahmen oft eine Raumdimension einsparen lässt, bleiben noch zwei unabhängige Variablen. Damit ergeben sich *partielle Differentialgleichungen*, d.h. Gleichungen, die die ersten partiellen Ableitungen

$$u_{x_i} = u_{x_i}(x_1, x_2, \dots, x_n) = \partial u(x_1, x_2, \dots, x_n) / \partial x_i \quad (1 \leq i \leq n)$$

oder auch höheren partiellen Ableitungen $u_{x_i x_j}$, usw. enthalten.

Anders als gewöhnliche Differentialgleichungen lassen sich partielle Differentialgleichungen nicht einheitlich analysieren. Vielmehr unterscheidet man drei Typen von Gleichungen, die verschiedene Eigenschaften besitzen und auch unterschiedliche numerische Methoden erfordern.

Bevor die Typenmerkmale definiert werden, sollen zunächst Beispiele partieller Differentialgleichungen vorgeführt werden.

1.1 Beispiele

Alle folgenden Beispiele werden nur zwei unabhängige Variablen x, y enthalten. Die beiden ersten Beispiele sind partielle Differentialgleichungen *erster* Ordnung, da nur erste partielle Ableitungen auftreten.

Beispiel 1.1.1 *Gesucht ist eine Lösung $u(x, y)$ von*

$$u_y(x, y) = 0. \quad (1.1.1)$$

Offenbar muss $u(x, y)$ von y unabhängig sein, d.h. die Lösung hat die Form $u(x, y) = \varphi(x)$. Andererseits ist $u(x, y) = \varphi(x)$ für beliebiges φ eine Lösung von (1.1.1).

Gleichung (1.1.1) ist ein Spezialfall von

Beispiel 1.1.2 *Gesucht ist eine Lösung $u(x, y)$ von*

$$cu_x(x, y) - u_y(x, y) = 0 \quad (c \text{ Konstante}). \quad (1.1.2)$$

u sei eine Lösung. Man führe neue Koordinaten $\xi := x + cy$, $\eta := y$ ein und definiere

$$v(\xi, \eta) := u(x(\xi, \eta), y(\xi, \eta))$$

mit Hilfe von $x(\xi, \eta) = \xi - c\eta$, $y(\xi, \eta) = \eta$. Da $v_\eta = u_x x_\eta + u_y y_\eta$ (Kettenregel) und $x_\eta = -c$, $y_\eta = 1$, folgt $v_\eta(\xi, \eta) = 0$ aus (1.1.2). Diese Gleichung stimmt mit (1.1.1) überein, und Beispiel 1.1.1 zeigt, dass $v(\xi, \eta) = \varphi(\xi)$. Ersetzt man ξ, η wieder durch x, y , erhält man die Darstellung

$$u(x, y) = \varphi(x + cy). \quad (1.1.3)$$

Umgekehrt ist durch (1.1.3) eine Lösung von (1.1.2) gegeben, sobald φ einmal stetig differenzierbar ist.

Um die Lösung einer gewöhnlichen Differentialgleichung $u' - f(u) = 0$ eindeutig festzulegen, braucht man eine Anfangswertvorgabe $u(x_0) = u_0$. Die partielle Differentialgleichung (1.1.2) kann man vervollständigen durch die Anfangswertvorgabe

$$u(x, y_0) = u_0(x) \quad \text{für } x \in \mathbb{R} \quad (1.1.4)$$

auf der Geraden $y = y_0$ (y_0 fest). Der Vergleich von (1.1.3) und (1.1.4) zeigt $\varphi(x + cy_0) = u_0(x)$. Damit ist φ festgelegt: $\varphi(x) = u_0(x - cy_0)$. Die eindeutige Lösung des Anfangswertproblems (1.1.2) und (1.1.4) lautet

$$u(x, y) = u_0(x - c(y_0 - y)). \quad (1.1.5)$$

Die folgenden drei Beispiele sind Differentialgleichungen zweiter Ordnung.

Beispiel 1.1.3 (Potentialgleichung) Sei Ω eine offene Teilmenge des \mathbb{R}^2 . Gesucht ist eine Lösung von

$$u_{xx} + u_{yy} = 0 \quad \text{in } \Omega. \quad (1.1.6)$$

Identifiziert man $(x, y) \in \mathbb{R}^2$ mit der komplexen Zahl $z = x + iy \in \mathbb{C}$, so lässt sich die Menge der Lösungen sofort angeben: Die Real- und Imaginärteile jeder in Ω holomorphen Funktion $f(z)$ sind Lösungen von (1.1.6). Beispiele sind $\operatorname{Re} z^0 = 1$, $\operatorname{Re} z^2 = x^2 - y^2$ und $\operatorname{Re} \log(z - z_0) = \log \sqrt{(x - x_0)^2 + (y - y_0)^2}$, falls $z_0 \notin \Omega$. Zur eindeutigen Festlegung der Lösung braucht man Randwerte $u(x, y) = \varphi(x, y)$ für alle (x, y) auf den Rand $\Gamma = \partial\Omega$ von Ω .

Ein anderer Name für die Potentialgleichung (1.1.6) lautet “Laplace¹-Gleichung”.

Beispiel 1.1.4 (Wellengleichung) Alle Lösungen von

$$u_{xx} - u_{yy} = 0 \quad \text{in } \Omega. \quad (1.1.7)$$

sind gegeben durch

$$u(x, y) = \varphi(x + y) + \psi(x - y), \quad (1.1.8)$$

wobei φ und ψ beliebige, zweifach stetig differenzierbare Funktionen sind. Geeignete Anfangswertvorgaben sind zum Beispiel

$$u(x, 0) = u_0(x), \quad u_y(x, 0) = u_1(x) \quad \text{für } x \in \mathbb{R}, \quad (1.1.9)$$

wobei u_0 und u_1 gegebene Funktionen sind. Setzt man (1.1.8) in (1.1.9) ein, findet man $u_0 = \varphi + \psi$, $u_1 = \varphi' - \psi'$ (φ', ψ' : Ableitungen von φ und ψ) und folgert

$$\varphi' = (u'_0 + u_1)/2, \quad \psi' = (u'_0 - u_1)/2.$$

Hieraus lassen sich φ und ψ bis auf zwei Integrationskonstanten bestimmen. Eine Integrationskonstante kann beliebig festgesetzt werden (z.B. durch $\varphi(0) = 0$), die zweite ist durch $u(0, 0) = u_0(0) = \varphi(0) + \psi(0)$ gegeben.

Übungsaufgabe 1.1.5 Man beweise, dass jede Lösung der Wellengleichung (1.1.7) die Darstellung (1.1.8) besitzt. Hinweis: Man verwende $\xi = x + y$ und $\eta = x - y$ als neue Variablen.

Die nächste Gleichung beschreibt die Wärmeleitung (in einem eindimensionalen Draht von $-\infty$ bis $+\infty$), wenn man u als Temperatur und y als Zeit interpretiert.

¹Pierre-Simon Laplace, geb. am 23. März 1749 in Beaumont-en-Auge, Normandy, gest. am 5. März 1827 in Paris

Beispiel 1.1.6 (Wärmeleitungsgleichung) *Gesucht ist die Lösung von*

$$u_{xx} - u_y = 0 \quad \text{für } x \in \mathbb{R}, y \geq 0. \quad (1.1.10)$$

Der Separationsansatz $u(x, y) = v(x)w(y)$ liefert für jedes $c \in \mathbb{R}$ eine Lösung

$$u(x, y) = \sin(cx) \exp(-c^2 y). \quad (1.1.11a)$$

Eine weitere Lösung von (1.1.10) für $y > 0$ ist

$$u(x, y) = \frac{1}{\sqrt{4\pi y}} \int_{-\infty}^{\infty} u_0(\xi) \exp\left(-\frac{(x-\xi)^2}{4y}\right) d\xi, \quad (1.1.11b)$$

wobei $u_0(\cdot)$ eine beliebige, stetige und beschränkte Funktion ist. Die zu Gleichung (1.1.10) passende Anfangsbedingung enthält anders als (1.1.9) nur eine Vorgabe:

$$u(x, 0) = u_0(x) \quad \text{für } x \in \mathbb{R}. \quad (1.1.12)$$

Die Lösung (1.1.11b), die zunächst nur für $y > 0$ definiert ist, lässt sich stetig in $y = 0$ fortsetzen und erfüllt dort die Anfangswertvorgabe (1.1.12).

Übungsaufgabe 1.1.7 u_0 sei beschränkt in \mathbb{R} und stetig in x . Man beweise, dass die rechte Seite in (1.1.11b) für $y \searrow 0$ gegen $u_0(x)$ strebt. Hinweis: Man zeige zunächst, dass

$$u(x, y) = u_0(x) + \frac{1}{\sqrt{4\pi y}} \int_{-\infty}^{\infty} [u_0(\xi) - u_0(x)] \exp\left(-\frac{(x-\xi)^2}{4y}\right) d\xi,$$

und zerlege das Integral in Integrale über $[x - \delta, x + \delta]$ und $(-\infty, x - \delta) \cup (x + \delta, \infty)$.

Wie bei gewöhnlichen Differentialgleichungen lassen sich Gleichungen höherer Ordnung in Systeme von Gleichungen erster Ordnung umschreiben. Im Folgenden geben wir einige Beispiele an.

Beispiel 1.1.8 *Das Paar (u, v) sei Lösung des Systems*

$$u_x + v_y = 0, \quad v_x + u_y = 0. \quad (1.1.13)$$

Falls u und v zweifach differenzierbar sind, liefert Differentiation von (1.1.13) die Gleichungen $u_{xx} + v_{xy} = 0$ und $v_{xy} + u_{yy} = 0$, die zusammen $u_{xx} - u_{yy} = 0$ ergeben. Damit ist u eine Lösung der Wellengleichung (1.1.7). Gleiches lässt sich für v zeigen.

Beispiel 1.1.9 (Cauchy-Riemann-Differentialgleichungen)² *Erfüllen u und v das System*

$$u_x + v_y = 0, \quad v_x - u_y = 0 \quad \text{in } \Omega \subset \mathbb{R}^2, \quad (1.1.14)$$

so ergibt die gleiche Überlegung wie in Beispiel 1.1.8, dass sowohl u als auch v der Potentialgleichung (1.1.6) genügen.

Beispiel 1.1.10 *Erfüllen u und v das System*

$$u_x + v_y = 0, \quad v_x + u = 0, \quad (1.1.15)$$

so löst v die Wärmeleitungsgleichung (1.1.10).

Ein in der Strömungsmechanik interessantes System findet sich in

²Augustin Louis Cauchy, geb. am 21. August 1789 in Paris, gest. am 23. Mai 1857 in Sceaux (bei Paris)

Beispiel 1.1.11 (Stokes-Gleichungen)³ In dem System

$$u_{xx} + u_{yy} - w_x = 0, \quad (1.1.16a)$$

$$v_{xx} + v_{yy} - w_y = 0, \quad (1.1.16b)$$

$$u_x + v_y = 0, \quad (1.1.16c)$$

bedeuten u und v die Strömungsgeschwindigkeiten in x - und y -Richtung, während w den Druck bezeichnet. Man beachte, dass das System von zweiter Ordnung bezüglich u und v ist, aber keine zweite Ableitung von w auftritt.

1.2 Typeneinteilungen bei Gleichungen zweiter Ordnung

Die allgemeine lineare Differentialgleichung zweiter Ordnung in zwei Variablen lautet

$$a(x, y)u_{xx} + 2b(x, y)u_{xy} + c(x, y)u_{yy} + d(x, y)u_x + e(x, y)u_y + f(x, y)u + g(x, y) = 0. \quad (1.2.1)$$

Definition 1.2.1 (a) Gleichung (1.2.1) heißt elliptisch in (x, y) , falls

$$a(x, y)c(x, y) - b^2(x, y) > 0. \quad (1.2.2a)$$

(b) Gleichung (1.2.1) heißt hyperbolisch in (x, y) , falls

$$a(x, y)c(x, y) - b^2(x, y) < 0. \quad (1.2.2b)$$

(c) Gleichung (1.2.1) heißt parabolisch in (x, y) , falls⁴

$$ac - b^2 = 0 \quad \text{und} \quad \text{Rang} \begin{bmatrix} a & b & d \\ b & c & e \\ b & c & e \end{bmatrix} = 2 \quad \text{in } (x, y). \quad (1.2.2c)$$

(d) Gleichung (1.2.1) heißt elliptisch (hyperbolisch, parabolisch) in $\Omega \in \mathbb{R}^2$, falls sie in allen $(x, y) \in \Omega$ elliptisch (hyperbolisch, parabolisch) ist.

Wenn für verschiedene $(x, y) \in \Omega$ unterschiedliche Fälle vorliegen, spricht man von einer “Gleichung vom gemischten Typ”.

Beispiel 1.2.2 Die Potentialgleichung (1.1.6) ist elliptisch, die Wellengleichung (1.1.7) ist vom hyperbolischen Typ, während die Wärmeleitungsgleichung (1.1.10) parabolisch ist.

Die Typendefinition kann leicht für den Fall verallgemeinert werden, dass mehr als zwei unabhängige Variable x_1, \dots, x_n auftreten. Die allgemeine lineare Differentialgleichung zweiter Ordnung in n Variablen $\mathbf{x} = (x_1, \dots, x_n)$ lautet

$$\sum_{i,j=1}^n a_{ij}(\mathbf{x}) u_{x_i x_j} + \sum_{i=1}^n a_i(\mathbf{x}) u_{x_i} + a(\mathbf{x}) u = f(\mathbf{x}). \quad (1.2.3)$$

Da für zweifach stetig differenzierbare Funktionen $u_{x_i x_j} = u_{x_j x_i}$ gilt, darf in (1.2.3) ohne Beschränkung der Allgemeinheit

$$a_{ij}(\mathbf{x}) = a_{ji}(\mathbf{x}) \quad (1 \leq i, j \leq n) \quad (1.2.4a)$$

angenommen werden. Damit definieren die Koeffizienten $a_{ij}(\mathbf{x})$ eine symmetrische $n \times n$ -Matrix

$$A(\mathbf{x}) = (a_{ij}(\mathbf{x}))_{i,j=1,\dots,n}, \quad (1.2.4b)$$

die demnach nur reelle Eigenwerte besitzt.

³George Gabriel Stokes, geb. am 13. August 1819 in Skreen, County Sligo, Irland, gest. am 1. Februar 1903 in Cambridge

⁴Gelegentlich wird der parabolische Typ nur durch $ac - b^2 = 0$ definiert. Aber die Gleichung $u_{xx}(x, y) + u_x(x, y) = 0$ oder gar die rein algebraische Gleichung $u(x, y) = 0$ möchte man nicht als parabolisch bezeichnen.

Definition 1.2.3 (a) Gleichung (1.2.3) heißt elliptisch in \mathbf{x} , falls alle n Eigenwerte der Matrix $A(\mathbf{x})$ das gleiche Vorzeichen (± 1) besitzen (d.h. falls $A(\mathbf{x})$ positiv oder negativ definit ist).

(b) Gleichung (1.2.3) heißt hyperbolisch in \mathbf{x} , falls $n-1$ Eigenwerte von $A(\mathbf{x})$ gleiches Vorzeichen (± 1) besitzen und ein Eigenwert das entgegengesetzte Vorzeichen hat.

(c) Gleichung (1.2.3) heißt parabolisch in \mathbf{x} , falls ein Eigenwert verschwindet, die übrigen $n-1$ Eigenwerte das gleiche Vorzeichen besitzen und $\text{Rang}(A(\mathbf{x}), \mathbf{a}(\mathbf{x})) = n$, wobei $\mathbf{a}(\mathbf{x}) = (a_1(\mathbf{x}), \dots, a_n(\mathbf{x}))^\top$.

(d) Gleichung (1.2.3) heißt elliptisch in $\Omega \in \mathbb{R}^n$, wenn sie in allen $\mathbf{x} \in \Omega$ elliptisch ist.

Die Definition 1.2.3 macht deutlich, dass die drei genannten Typen keineswegs alle Fälle abdecken. Eine unklassifizierte Gleichung liegt z.B. vor, wenn $A(\mathbf{x})$ zwei positive und zwei negative Eigenwerte besitzt.

Anstelle von (1.2.3) schreibt man auch

$$Lu = f, \quad (1.2.5)$$

wobei

$$L = \sum_{i,j=1}^n a_{ij}(\mathbf{x}) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^n a_i(\mathbf{x}) \frac{\partial}{\partial x_i} + a(\mathbf{x}) \quad (1.2.6)$$

einen linearen Differentialoperator zweiter Ordnung darstellt. Als Hauptteil von L bezeichnet man den Operator

$$L_0 = \sum_{i,j=1}^n a_{ij}(\mathbf{x}) \frac{\partial^2}{\partial x_i \partial x_j},$$

der nur noch die höchsten Ableitungen von L enthält.

Bemerkung 1.2.4 Die Elliptizität bzw. Hyperbolizität der Gleichung (1.2.3) hängt nur vom Hauptteil des Differentialoperators ab.

Übungsaufgabe 1.2.5 (Typinvarianz bei Koordinatenwechsel) Gleichung (1.2.3) sei in $\mathbf{x} \in \Omega$ definiert. Die Transformation $\Phi : \Omega \subset \mathbb{R}^n \rightarrow \Omega' \subset \mathbb{R}^n$ habe in \mathbf{x} eine nichtsinguläre Funktionalmatrix $S = \partial\Phi/\partial\mathbf{x} \in C^1(\Omega)$. Man beweise: Gleichung (1.2.3) ändert ihren Typ in \mathbf{x} nicht, wenn sie in den neuen Koordinaten $\xi = \Phi(\mathbf{x})$ geschrieben wird. Hinweis: Die Matrix $A = (a_{ij})$ wird nach der Transformation zu SAS^\top . Man verwende Bemerkung 1.2.4 und den Trägheitssatz von Sylvester (vgl. Gantmacher [31, p. 273]).

1.3 Typeneinteilungen bei Systemen erster Ordnung

Die Beispiele 1.1.8-10 sind Spezialfälle des allgemeinen linearen Systems erster Ordnung in zwei Variablen:

$$u_x(x, y) - A(x, y) u_y(x, y) + B(x, y) u(x, y) = f(x, y). \quad (1.3.1)$$

Hierbei ist $u = (u_1, \dots, u_m)^\top$ eine Vektorfunktion, und A, B sind $m \times m$ -Matrizen. Anders als in Abschnitt 1.2 kann A nichtsymmetrisch sein und komplexe Eigenwerte besitzen. Falls die Eigenwerte $\lambda_1, \dots, \lambda_m$ reell sind und eine Zerlegung $A = S^{-1}DS$ mit $D = \text{diag}\{\lambda_1, \dots, \lambda_m\}$ existiert, heißt A reell-diagonalisierbar.

Definition 1.3.1 (a) Das System (1.3.1) heißt hyperbolisch in (x, y) , falls $A(x, y)$ reell-diagonalisierbar ist. (b) System (1.3.1) heißt elliptisch in (x, y) , falls kein Eigenwert von $A(x, y)$ reell ist.

Falls A symmetrisch ist oder m verschiedene reelle Eigenwerte besitzt, ist das System hyperbolisch, da diese Bedingungen hinreichend für reelle Diagonalisierbarkeit sind. Insbesondere ist eine einzelne reelle Gleichung stets hyperbolisch.

Gemäß der vorangehenden Bemerkung enthalten die Beispiele 1.1.1 und 1.1.2 hyperbolische Gleichungen. Das System (1.1.13) aus Beispiel 1.1.8 hat die Form (1.3.1) mit

$$A = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$$

Es ist hyperbolisch, da A reell-diagonalisierbar ist:

$$A = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}.$$

Das eng mit der Potentialgleichung (1.1.6) zusammenhängende Cauchy-Riemann-System (1.1.14) ist elliptisch, da es die Form (1.3.1) mit

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

hat und A die Eigenwerte $\pm i$ besitzt.

Das der (parabolischen) Wärmeleitungsgleichung entsprechende System (1.1.15) lässt sich als System (1.3.1) mit

$$A = \begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix}$$

schreiben. Die Eigenwerte ($\lambda_1 = \lambda_2 = 0$) sind zwar reell, aber A ist nicht diagonalisierbar. Damit ist das System (1.1.15) weder hyperbolisch noch elliptisch.

Ein allgemeineres System als (1.3.1) ist

$$A_1 u_x + A_2 u_y + Bu = f. \quad (1.3.2)$$

Falls A_1 regulär ist, ergibt die Multiplikation mit A_1^{-1} die Gestalt (1.3.1) mit $A = -A_1^{-1}A_2$. Sonst muss man das verallgemeinerte Eigenwertproblem $\det(\lambda A_1 + A_2) = 0$ untersuchen. Allerdings kann das System (1.3.2) mit singulären A_1 nicht elliptisch sein, wie aus dem Folgenden hervorgeht (vgl. (1.3.4) mit $\xi_1 = 1$ und $\xi_2 = 0$).

Eine Verallgemeinerung von (1.3.2) auf n unabhängige Variable stellt das System

$$A_1 u_{x_1} + A_2 u_{x_2} + \dots + A_n u_{x_n} + Bu = f \quad (1.3.3)$$

mit $m \times m$ -Matrizen $A_i = A_i(\mathbf{x}) = A_i(x_1, \dots, x_n)$ und $B = B(\mathbf{x})$ dar. Als Spezialfall einer späteren Definition (vgl. Abschnitt 12.1) erhalten wir die

Definition 1.3.2 Das System (1.3.3) heißt elliptisch in \mathbf{x} , falls

$$\det \left(\sum_{i=1}^n \xi_i A_i(\mathbf{x}) \right) \neq 0 \quad \text{für alle } 0 \neq (\xi_1, \dots, \xi_n) \in \mathbb{R}^n. \quad (1.3.4)$$

1.4 Unterschiedliche Eigenschaften der verschiedenen Typen

Die Unterscheidung partieller Differentialgleichungen in verschiedene Typen wäre gegenstandslos, wenn nicht jeder Typ grundlegend andere Eigenschaften hätte. Bei der Diskussion der Beispiele in §1.1 wurde schon erwähnt, dass die Lösung eindeutig bestimmt ist, falls Anfangs- bzw. Randwerte vorgeschrieben werden. In Beispiel 1.1.2 wird die hyperbolische Differentialgleichung (1.1.2) durch die Vorgabe (1.1.4) von u auf der Geraden $y = \text{const}$ ergänzt (siehe Abbildung 1.4.1a). Im Falle der hyperbolischen Wellengleichung (1.1.7) muss auch u_y vorgeschrieben werden (vgl. (1.1.9)), da die Gleichung von zweiter Ordnung ist.

Es genügt auch, die Werte u und u_y auf einem endlichen Intervall $[x_1, x_2]$ vorzugeben, wenn u zusätzlich auf den seitlichen Rändern des Bereiches Ω von Abbildung 1.4.1b vorgeschrieben wird. Diese sogenannte *Anfangsrandwertvorgabe* tritt z.B. bei folgendem physikalischen Problem auf. Eine schwingende Saite wird beschrieben durch die Auslenkung $u(x, t)$ im Punkt $x \in [x_1, x_2]$ zur Zeit t . Die Funktion u erfüllt die Wellengleichung (1.1.7), wobei die Koordinate y der Zeit t entspricht. Zum Anfangszeitpunkt $t = t_0$ ist die Auslenkung $u(x, 0)$ und die Geschwindigkeit $u_t(x, 0)$ für $x_1 < x < x_2$ vorgegeben. Unter der Annahme, dass die Saite an den Randpunkten x_1 und x_2 fest eingespannt ist, erhält man die zusätzlichen Randdaten $u(x_1, t) = u(x_2, t) = 0$ für alle t .

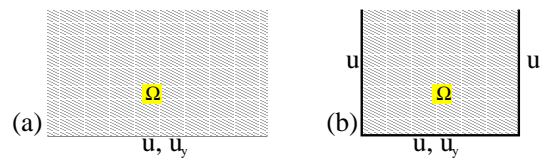


Abb. 1.4.1. (a) Anfangswertvorgabe und (b) Anfangsrandwertvorgabe für hyperbolische Probleme

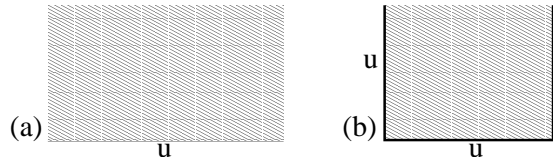


Abb. 1.4.2. (a) Anfangswertvorgabe und (b) Anfangsrandwertvorgabe für parabolische Probleme

Auch für parabolische Gleichungen zweiter Ordnung lassen sich Anfangswert- und Anfangsrandwertaufgaben formulieren (vgl. Abbildung 1.4.2). Allerdings darf als Anfangswert nur $u(x, y_0) = u_0(x)$ vorgeschrieben werden. Eine zusätzliche Vorgabe von $u_y(x, y_0)$ ist nicht möglich, da $u_y(x, y_0) = u_{xx}(x, y_0) = u_0''(x)$ bereits durch die Differentialgleichung (1.1.10) und u_0 festgelegt ist.

Die Wärmeleitungsgleichung (1.1.10) mit der Anfangsrandwertvorgabe

$$\begin{aligned} u(x, t_0) &= u_0(x) && \text{in } [x_1, x_2], \\ u(x_1, t) &= \varphi_1(t), \quad u(x_2, t) = \varphi_2(t) && \text{für } t > t_0 \end{aligned} \quad (1.4.1)$$

(vgl. Abbildung 1.4.2b) beschreibt die Temperatur $u(x, t)$ eines Drahtes, dessen Enden bei $x = x_1$ und $x = x_2$ die Temperaturen $\varphi_1(t)$ und $\varphi_2(t)$ haben. Die Anfangstemperaturverteilung zur Zeit t_0 ist durch $u_0(x)$ gegeben.

Abgesehen von der unterschiedlichen Zahl von Anfangswertvorgaben in den Abbildungen 1.4.1 und 1.4.2 besteht der folgende Unterschied zwischen hyperbolischen und parabolischen Gleichungen:

Bemerkung 1.4.1 Der schraffierte Bereich in den Abbildungen 1.4.1 und 1.4.2 entspricht $t > t_0$ (bzw. $y > y_0$). Bei hyperbolischen Gleichungen lassen sich ebenso Anfangs(rand)wertaufgaben im Bereich $t \leq t_0$ lösen, während parabolische Aufgaben in $t < t_0$ im Allgemeinen keine Lösung besitzen.

Ändert man die parabolische Gleichung $u_t - u_{xx} = 0$ in $u_t + u_{xx} = 0$ ab, so kehrt sich die Orientierung um: Lösungen existieren im Allgemeinen nur für $t \leq t_0$.

Für die Lösung einer elliptischen Gleichung werden Randwerte vorgeschrieben (vgl. Beispiel 1.1.3, Abbildung 1.4.3a,b). Eine Vorgabe wie in Abbildung 1.4.2b würde die Lösung eines elliptischen Problems nicht eindeutig festlegen können, während die Lösung eines parabolischen Problems durch die Randwerte von Abbildung 1.4.3a überbestimmt wäre.

Ein elliptisches Problem mit Vorgaben wie in Abbildung 1.4.1b besitzt im Allgemeinen keine Lösung. Seien z.B. die Bedingungen $u(x, 0) = u(0, y) = u(1, y) = 0$ und $u_y(x, 0) = u_1(x)$ an die Lösung der Potentialgleichung (1.1.6) gestellt, wobei u_1 nicht unendlich oft differenzierbar sei. Würde in $\overline{\Omega} = [0, 1] \times [0, 1]$ eine stetige Lösung u existieren, so ließe sich $u(x, 1)$ in eine Sinus-Reihe entwickeln, und die folgende Aufgabe zeigt, dass u_1 im Widerspruch zur Annahme unendlich oft differenzierbar sein müsste.

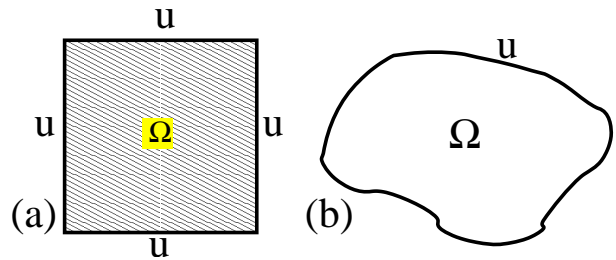


Abb. 1.4.3. Randwertvorgabe für ein elliptisches Problem

Übungsaufgabe 1.4.2 $\varphi \in C^0[0, 1]$ habe die Fourier-Entwicklung $\varphi(x) = \sum_{\nu=1}^{\infty} \alpha_{\nu} \sin(\nu\pi x)$. Man zeige:
a) Die Lösung der Potentialgleichung (1.1.6) im Quadrat $\Omega = (0, 1) \times (0, 1)$ mit Randwerten $u(0, y) = u(x, 0) = u(1, y) = 0$ und $u(x, 1) = \varphi(x)$ ist durch die folgende Reihe gegeben:

$$u(x, y) = \sum_{\nu=1}^{\infty} \frac{\alpha_{\nu}}{\sinh(\nu\pi)} \sin(\nu\pi x) \sinh(\nu\pi y).$$

b) Für $0 \leq x \leq 1$ und $0 \leq y < 1$ ist $u(x, y)$ unendlich oft differenzierbar. Hinweis: $f(x) = \sum \beta_{\nu} \sin(\nu\pi x) \in C^{\infty}[0, 1]$, falls $\lim_{\nu \rightarrow \infty} \beta_{\nu} \nu^k = 0$ für alle $k \in \mathbb{N}$.

Umgekehrt ist es nicht sinnvoll, Randwertvorgaben wie in Abbildung 1.4.3a an ein hyperbolisches Problem zu stellen. Als Beispiel betrachte man die Wellengleichung (1.1.7) in $\overline{\Omega} = [0, 1]$ mit den Randwerten $u(x, 0) = u(0, y) = u(1, y) = 0$ und $u(x, 1/\pi) = \sin(\nu\pi x)$ für $\nu \in \mathbb{N}$. Die Lösung lautet $u(x, y) = \sin(\nu\pi x) \sin(\nu\pi y) / \sin \nu$. Obwohl die Randdaten für alle $\nu \in \mathbb{N}$ betragsmäßig durch eins beschränkt sind, kann die Lösung in $\overline{\Omega}$ beliebig groß werden, da $\sup\{1/\sin \nu : \nu \in \mathbb{N}\} = \infty$. Ein derartiges Randwertproblem nennt man "nicht sachgemäß gestellt" (vgl. Definition 2.4.1).

Übungsaufgabe 1.4.3 *Man beweise: Die Menge $\{\sin \nu : \nu \in \mathbb{N}\}$ ist dicht in $[-1, 1]$.*

Ein weiteres unterscheidendes Merkmal ist die Regularität (Glattheit) der Lösung. Sei u Lösung der Potentialgleichung (1.1.6) in $\Omega \subset \mathbb{R}^2$. Wie in Beispiel 1.1.3 behauptet, ist u Realteil einer in Ω holomorphen Funktion. Da holomorphe Funktionen unendlich oft differenzierbar sind, gilt diese Eigenschaft auch für u .

Im Falle der parabolischen Wärmeleitungsgleichung (1.1.10) mit Anfangswerten $u(x, 0) = u_0$ ist die Lösung u durch (1.1.11b) dargestellt. Für $y > 0$ ist u unendlich oft differenzierbar. Die Glattheit von u_0 spielt bei dieser Überlegung ebensowenig eine Rolle wie die Glattheit der Randwerte im Falle der Potentialgleichung.

Ein vollkommen anderes Resultat findet man für die hyperbolische Wellengleichung (1.1.7). Die Lösung lautet $u(x, y) = \phi(x + y) + \psi(x - y)$, wobei sich ϕ und ψ direkt aus den Anfangsdaten (1.1.9) ergeben. Man prüft nach, dass u k -fach differenzierbar ist, falls u_0 k -fach und u_1 $(k - 1)$ -fach differenzierbar ist.

Wie schon in diesem Abschnitt erwähnt wurde, spielt die Variable y in den hyperbolischen und parabolischen Gleichungen (1.1.1), (1.1.2), (1.1.7), (1.1.10) häufig die Rolle der Zeit. Man bezeichnet deshalb die durch hyperbolische oder parabolische Gleichungen beschriebenen Prozesse als *instationär*. Elliptische Gleichungen, die als Variablen nur Raumkoordinaten enthalten, werden *stationär* genannt. Deutlicher als Definition 1.2.1b,c zeigt Definition 1.2.3b,c die Auszeichnung einer einzigen Variablen (Zeit), die dem Eigenwert $\lambda = 0$ bei der parabolischen und dem Eigenwert mit entgegengesetztem Vorzeichen bei hyperbolischen Gleichungen entspricht.

Der Zusammenhang der verschiedenen Typen wird leichter verständlich, wenn man elliptische Gleichungen in den Variablen x_1, \dots, x_n in Beziehung setzt zu parabolischen und hyperbolischen Gleichungen in den Variablen x_1, \dots, x_n, t .

Bemerkung 1.4.4 *L sei ein Differentialoperator (1.2.6) in den Variablen $\mathbf{x} = (x_1, \dots, x_n)$ und habe elliptischen Typ. L sei so skaliert, dass die Matrix $A(\mathbf{x})$ aus (1.2.4b) nur negative Eigenwerte besitzt. Dann ist*

$$u_t + Lu = 0 \quad (1.4.2)$$

eine parabolische Gleichung für $u(\mathbf{x}, t) = u(x_1, \dots, x_n, t)$. Dagegen ist

$$u_{tt} + Lu = 0 \quad (1.4.3)$$

vom hyperbolischen Typ.

Umgekehrt führen die instationären Probleme (1.4.2) oder (1.4.3) zu der elliptischen Gleichung $Lu = 0$, wenn man Lösungen von (1.4.2) oder (1.4.3) sucht, die von der Zeit t unabhängig sind. Aufgaben vom elliptischen Typ erhält man auch, wenn man Lösungen von (1.4.2) oder (1.4.3) mit Hilfe eines *Separationsansatzes* $u(\mathbf{x}, t) = \varphi(t)v(\mathbf{x})$ sucht. Es ergibt sich

$$\begin{aligned} u(\mathbf{x}, t) &= e^{-\lambda t} v(\mathbf{x}) && \text{im Falle (1.4.2),} \\ u(\mathbf{x}, t) &= e^{\pm i\sqrt{\lambda}t} v(\mathbf{x}) && \text{im Falle (1.4.3),} \end{aligned} \quad (1.4.4)$$

wobei $v(\mathbf{x})$ Lösung der elliptischen *Eigenwertaufgabe*

$$Lv = \lambda v$$

ist (diese Eigenwertaufgaben werden in §11 diskutiert werden).

2

Die Potentialgleichung

2.1 Problemstellung

Die Potentialgleichung aus Beispiel 1.1.3 lautet

$$\Delta u = 0 \quad \text{in } \Omega \subset \mathbb{R}^n, \quad (2.1.1a)$$

wobei $\Delta = \partial^2/\partial x_1^2 + \dots + \partial^2/\partial x_n^2$ der *Laplace-Operator* ist. In der Physik beschreibt Gleichung (2.1.1a) die Potentiale, z.B. das elektrische Potential, wenn Ω keine elektrischen Ladungen enthält, das magnetische Potential bei verschwindender Stromdichte, das Geschwindigkeitspotential usw. Gleichung (2.1.1a) wird auch *Laplace-Gleichung* genannt, da sie von P. S. Laplace in seinem fünfbandigen Werk “*Mécanique céleste*” (geschrieben 1799-1825) beschrieben wurde. Zuerst wurde die Potentialgleichung allerdings 1752 von Euler¹ erwähnt.

Auf den Zusammenhang zwischen der Potentialgleichung für $n = 2$ und der Funktionentheorie ist schon in Beispiel 1.1.3 hingewiesen worden. Der Laplace-Operator ist nicht nur ein Beispiel für einen elliptischen Differentialoperator, sondern sogar der Prototyp (eine sogenannte Normalform). Durch Variablentransformation lässt sich ein elliptischer Differentialoperator zweiter Ordnung so umformen, dass der Hauptteil der Laplace-Operator ist (vgl. Hellwig [46, p. 64f]).

Im Folgenden wird Ω stets ein Gebiet sein.

Definition 2.1.1 (Gebiet) $\Omega \subset \mathbb{R}^n$ heißt *Gebiet*, falls Ω offen und zusammenhängend² ist.

Die Existenz zweiter Ableitungen von u wird nur in Ω , nicht auf dem Rand

$$\Gamma = \partial\Omega$$

von Ω gefordert. Damit eine *Randwertvorgabe*

$$u = \varphi \quad \text{auf } \Gamma \quad (2.1.1b)$$

sinnvoll ist, muss die Stetigkeit von u auf $\overline{\Omega} = \Omega \cup \Gamma$ vorausgesetzt werden.

Die Kombination einer (elliptischen) Differentialgleichung (hier (2.1.1a)) mit einer Randbedingung (hier (2.1.1b)) nennt man *Randwertaufgabe*.

¹Leonhard Euler, geb. 15. April 1707 in Basel, gest. 18. Sept. 1783 in St. Petersburg

² Ω heißt zusammenhängend, wenn es für alle $x, y \in \Omega$ eine stetige Kurve in Ω gibt, die x mit y verbindet, d.h. $\gamma \in C([0, 1])$ mit $\gamma : s \in [0, 1] \mapsto \gamma(s) \in \Omega$ und $\gamma(0) = x, \gamma(1) = y$.

Definition 2.1.2 (harmonisch) *u* heißt harmonisch in Ω , falls *u* zu $C^2(\Omega) \cap C^0(\overline{\Omega})$ gehört und der Potentialgleichung (2.1.1a) genügt.

Dabei bezeichnet $C^0(D)$ [$C^k(D)$, $C^\infty(D)$] die Menge der auf D stetigen [k -fach stetig differenzierbaren bzw. unendlich oft differenzierbaren] Funktionen.

Im Allgemeinen kann man nicht erwarten, dass die Lösung von (2.1.1a,b) in $C^2(\overline{\Omega})$ liegt, wie das folgende Beispiel zeigt.

Beispiel 2.1.3 Sei $\Omega = (0,1) \times (0,1)$ (vgl. Abbildung 1.4.3a). Die Randwerte seien $\varphi(x,y) = x^2$ für $(x,y) \in \Gamma$. Eine Lösung der Randwertaufgabe existiert, gehört aber nicht zu $C^2(\overline{\Omega})$.

Beweis. Die Existenz einer Lösung *u* wird in Satz 7.3.7 behandelt werden. Wäre $u \in C^2(\overline{\Omega})$, so gälte $u_{xx}(x,0) = \varphi_{xx}(x,0) = 2$ für $x \in [0,1]$ und insbesondere $u_{xx}(0,0) = 2$. Aus dem analogen Resultat $u_{yy}(0,0) = \varphi_{yy}(0,0) = 0$ erhielte man $\Delta u(0,0) = 2$ im Widerspruch zu $\Delta u = 0$ in Ω . ■

Im vorliegenden Fall kann man noch $u \in C^1(\overline{\Omega})$ zeigen. Dass auch diese Aussage im Allgemeinen falsch ist, beweist das nächste Beispiel, in dem Ω das sogenannte L-Gebiet aus Abbildung 2.1.1 ist.

Beispiel 2.1.4 Im Gebiet $\Omega = (-\frac{1}{2}, \frac{1}{2}) \times (-\frac{1}{2}, \frac{1}{2}) \setminus [0, \frac{1}{2}) \times [0, \frac{1}{2})$ führe man Polarkoordinaten ein:

$$x = r \cos \varphi, \quad y = r \sin \varphi. \quad (2.1.2)$$

Die Funktion $u(r, \varphi) = r^{2/3} \sin((2\varphi - \pi)/3)$ ($\pi/2 < \varphi < 2\pi$) ist Lösung der Potentialgleichung (2.1.1a) und besitzt glatte Randwerte auf Γ (insbesondere gilt $u = 0$ auf $\Gamma_0 \subset \Gamma$). Trotzdem sind die ersten Ableitungen in $r = 0$ unbeschränkt, d.h. $u \notin C^1(\overline{\Omega})$.

Beweis. Mit u_x, u_y müsste auch $u_r = u_x \cos \varphi + u_y \sin \varphi$ beschränkt sein. Jedoch gilt $u_r = \mathcal{O}(r^{-1/3})$ für $r \rightarrow 0$. Um nachzuprüfen, dass $\Delta u = 0$, benutze man folgende Übungsaufgabe. ■

Übungsaufgabe 2.1.5 Man zeige: a) Bezüglich der Polarkoordinaten (2.1.2) des \mathbb{R}^2 lautet der Laplace-Operator

$$\Delta = \frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} + \frac{1}{r^2} \frac{\partial^2}{\partial \varphi^2}. \quad (2.1.3)$$

b) Bezüglich der drei-dimensionalen Polarkoordinaten

$$x = r \cos \varphi \sin \psi, \quad y = r \sin \varphi \sin \psi, \quad z = r \cos \psi$$

gilt

$$\Delta = \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} + \frac{1}{r^2} \left[\frac{1}{\sin^2 \psi} \frac{\partial^2}{\partial \varphi^2} + \cot \psi \frac{\partial}{\partial \psi} + \frac{\partial^2}{\partial \psi^2} \right]. \quad (2.1.4)$$

Anmerkung. Im allgemeinen n -dimensionalen Fall führt die Transformation auf Polarkoordinaten zu

$$\Delta = \frac{\partial^2}{\partial r^2} + \frac{n-1}{r} \frac{\partial}{\partial r} + \frac{1}{r^2} B,$$

wobei der sogenannte Beltrami-Operator B nur Ableitungen bezüglich der Winkelvariablen enthält.

2.2 Singularitätenfunktion

Als Singularitätenfunktion oder Fundamentallösung bezeichnet man die Funktion

$$s(\mathbf{x}, \mathbf{y}) = \begin{cases} -\frac{1}{\omega_2} \log |\mathbf{x} - \mathbf{y}| & \text{für } n = 2, \\ \frac{|\mathbf{x} - \mathbf{y}|^{2-n}}{(n-2)\omega_n} & \text{für } n > 2, \end{cases} \quad (2.2.1a)$$

von $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, wobei

$$\omega_n = 2 \Gamma(\frac{1}{2})^n / \Gamma(\frac{n}{2}), \quad \text{insbesondere } \omega_2 = 2\pi, \omega_3 = 4\pi, \quad (\Gamma: \text{Gammafunktion}) \quad (2.2.1b)$$

die Oberfläche der n -dimensionalen Einheitskugel ist. Mit

$$|\mathbf{x}| = \sqrt{\sum_{i=1}^n x_i^2} \quad (2.2.2)$$

wird die *Euklidische Norm* des \mathbb{R}^n bezeichnet.

Lemma 2.2.1 Für festes $\mathbf{y} \in \mathbb{R}^n$ löst $s(\mathbf{x}, \mathbf{y})$ bezüglich \mathbf{x} die Potentialgleichung in $\mathbb{R}^n \setminus \{\mathbf{y}\}$.

Beweis. Der Beweis lässt sich direkt durchführen. Einfacher ist es aber, Polarkoordinaten mit \mathbf{y} als Zentrum einzuführen und (1.1.5) anzuwenden, da $s(\mathbf{x}, \mathbf{y})$ nur von $r = |\mathbf{x} - \mathbf{y}|$ abhängt. ■

Zur Vorbereitung des nächsten Satzes muss die Normalenableitung $\partial/\partial n$ eingeführt werden. Sei Ω ein Gebiet mit glattem Rand Γ . Der Vektor $\vec{n}(\mathbf{x}) \in \mathbb{R}^n$ bezeichnet die äußere Normalenrichtung in $\mathbf{x} \in \Gamma$, d.h. \vec{n} ist ein Einheitsvektor senkrecht zur Tangentialhyperebene in \mathbf{x} und zeigt nach außen. Die Normalenableitung von u in $\mathbf{x} \in \Gamma$ ist definiert als

$$\frac{\partial u(\mathbf{x})}{\partial n} = \langle \vec{n}, \nabla u \rangle, \quad (2.2.3a)$$

wobei

$$\nabla u = \text{grad } u = (u_{x_1}, \dots, u_{x_n})^\top \quad (2.2.3b)$$

der *Gradient* von u und

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i y_i \quad (2.2.3c)$$

das Skalarprodukt des \mathbb{R}^n sind. Im Falle der Kugel $K_R(\mathbf{y})$ (vgl. (2.2.7)) ist die Normalenrichtung radial. $\partial u/\partial n$ wird zu $\partial u/\partial r$ bezüglich $r = |\mathbf{x} - \mathbf{y}|$, wenn man Polarkoordinaten mit Zentrum in \mathbf{y} verwendet. Da $\partial s(\mathbf{x}, \mathbf{y})/\partial r = -|\mathbf{x} - \mathbf{y}|^{1-n}/\omega_n$, folgt

$$\frac{\partial s(\mathbf{x}, \mathbf{y})}{\partial n} = -\frac{R^{1-n}}{\omega_n} \quad \text{für alle } \mathbf{x} \in \partial K_R(\mathbf{y}). \quad (2.2.4)$$

Die *erste Greensche³ Formel* (vgl. Green [34])

$$\int_{\Omega} u(\mathbf{x}) \Delta v(\mathbf{x}) \, d\mathbf{x} = - \int_{\Omega} \langle \nabla u(\mathbf{x}), \nabla v(\mathbf{x}) \rangle \, d\mathbf{x} + \int_{\partial\Omega} u(\mathbf{x}) \frac{\partial v(\mathbf{x})}{\partial n} \, d\Gamma \quad (2.2.5a)$$

gilt für $u \in C^1(\overline{\Omega})$, $v \in C^2(\overline{\Omega})$, wenn das Gebiet Ω geeignete Voraussetzungen erfüllt. Dabei bezeichnet $\int_{\partial\Omega} \dots \, d\Gamma$ das Oberflächenintegral.

Gebiete, für die (2.2.5a) gültig ist, heißen *Normalgebiete*. Hinreichende Bedingungen hierfür findet man in Kellogg [49, §IV] und Hellwig [46, p. 11].

Funktionen $u, v \in C^2(\overline{\Omega})$ in einem Normalgebiet Ω erfüllen die *zweite Greensche Formel*

$$\int_{\Omega} u(\mathbf{x}) \Delta v(\mathbf{x}) \, d\mathbf{x} = \int_{\Omega} v(\mathbf{x}) \Delta u(\mathbf{x}) \, d\mathbf{x} + \int_{\partial\Omega} \left[u(\mathbf{x}) \frac{\partial v(\mathbf{x})}{\partial n} - v(\mathbf{x}) \frac{\partial u(\mathbf{x})}{\partial n} \right] \, d\Gamma. \quad (2.2.5b)$$

Satz 2.2.2 Ω sei Normalgebiet, und $u \in C^2(\overline{\Omega})$ sei dort harmonisch. Dann gilt

$$u(\mathbf{y}) = \int_{\partial\Omega} \left[s(\mathbf{x}, \mathbf{y}) \frac{\partial u(\mathbf{x})}{\partial n} - u(\mathbf{x}) \frac{\partial s(\mathbf{x}, \mathbf{y})}{\partial n_x} \right] \, d\Gamma_x \quad \text{für alle } \mathbf{y} \in \Omega. \quad (2.2.6)$$

Die Symbole $\frac{\partial}{\partial n_x}$ und $d\Gamma_x$ beziehen sich dabei auf die Variable \mathbf{x} .

³George Green, geb. im Juli 1793 in Sneinton, Nottingham, gest. am 31. Mai 1841 in Sneinton, Nottingham

Beweis. Mit

$$K_r(\mathbf{y}) = \{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x} - \mathbf{y}| < r\} \quad (2.2.7)$$

bezeichnen wir den Kreis um \mathbf{y} mit Radius r . Da die Singularitätenfunktion $s(\cdot, \mathbf{y})$ in $\mathbf{x} = \mathbf{y}$ nicht differenzierbar ist, kann die Greensche Formel nicht direkt angewandt werden. Sei

$$\Omega_\varepsilon := \Omega \setminus \overline{K_\varepsilon(\mathbf{y})},$$

wobei ε so klein sei, dass $K_\varepsilon(\mathbf{y}) \subset \Omega$. Da Ω_ε wieder Normalgebiet ist, folgt aus $\Delta u = \Delta s = 0$ in Ω_ε (vgl. Lemma 2.2.1) und (2.2.5b) mit $v = s(\cdot, \mathbf{y})$, dass

$$\int_{\partial\Omega_\varepsilon} \left[u(\mathbf{x}) \frac{\partial s(\mathbf{x}, \mathbf{y})}{\partial n_x} - s(\mathbf{x}, \mathbf{y}) \frac{\partial u(\mathbf{x})}{\partial n} \right] d\Gamma_x = 0. \quad (2.2.8a)$$

Es ist $\partial\Omega_\varepsilon = \partial\Omega \cup \partial K_\varepsilon(\mathbf{y})$. In $\mathbf{x} \in \partial K_\varepsilon(\mathbf{y})$ unterscheiden sich jedoch die Normalenrichtungen von $\partial\Omega_\varepsilon$ und $\partial K_\varepsilon(\mathbf{y})$ durch das Vorzeichen. Gleiches gilt für die Normalenableitung, sodass sich das Integral aus (2.2.8a) zerlegen lässt in $\int_{\partial\Omega_\varepsilon} \dots = \int_{\partial\Omega} \dots - \int_{\partial K_\varepsilon(\mathbf{y})} \dots$. Die Behauptung des Satzes wäre bewiesen, wenn $\int_{\partial K_\varepsilon(\mathbf{y})} \dots \rightarrow -u(\mathbf{y})$ für $\varepsilon \rightarrow 0$ gezeigt werden könnte. $\partial u / \partial n$ ist auf $\partial K_\varepsilon(\mathbf{y})$ beschränkt, und $\int_{\partial K_\varepsilon(\mathbf{y})} s(\mathbf{x}, \mathbf{y}) d\Gamma$ strebt wie $\mathcal{O}(\varepsilon |\log \varepsilon|)$ bzw. $\mathcal{O}(\varepsilon)$ gegen null, wie man aus (2.2.1) und $\int_{\partial K_\varepsilon(\mathbf{y})} d\Gamma = \varepsilon^{n-1} \omega_n$ abliest. Somit folgt

$$\int_{\partial K_\varepsilon(\mathbf{y})} s(\mathbf{x}, \mathbf{y}) \frac{\partial u(\mathbf{x})}{\partial n} d\Gamma_x \rightarrow 0 \quad (\varepsilon \rightarrow 0). \quad (2.2.8b)$$

Aus $\int_{\partial K_\varepsilon(\mathbf{y})} d\Gamma = \varepsilon^{n-1} \omega_n$ und (2.2.4) schließt man

$$\int_{\partial K_\varepsilon(\mathbf{y})} u(\mathbf{y}) \frac{\partial s(\mathbf{x}, \mathbf{y})}{\partial n_x} d\Gamma_x = -u(\mathbf{y}). \quad (2.2.8c)$$

Die Stetigkeit von u in \mathbf{y} liefert

$$\left| \int_{\partial K_\varepsilon(\mathbf{y})} [u(\mathbf{x}) - u(\mathbf{y})] \frac{\partial s(\mathbf{x}, \mathbf{y})}{\partial n_x} d\Gamma_x \right| \leq \max_{\mathbf{x} \in \partial K_\varepsilon(\mathbf{y})} |u(\mathbf{x}) - u(\mathbf{y})| \rightarrow 0 \quad \text{für } \varepsilon \rightarrow 0. \quad (2.2.8d)$$

(2.2.8b,c,d) zeigen $\int_{\partial K_\varepsilon(\mathbf{y})} [u \frac{\partial}{\partial n} s - s \frac{\partial}{\partial n} u] d\Gamma \rightarrow -u(\mathbf{y})$ ($\varepsilon \rightarrow 0$), sodass (2.2.8a) den Satz beweist. ■

Jede Funktion der Form

$$\gamma(\mathbf{x}, \mathbf{y}) = s(\mathbf{x}, \mathbf{y}) + \Phi(\mathbf{x}, \mathbf{y}) \quad (2.2.9)$$

heißt *Grundlösung* (der Potentialgleichung) in Ω , wenn für festes $\mathbf{y} \in \Omega$ die Funktion $\Phi(\cdot, \mathbf{y})$ harmonisch in Ω ist und zu $C^2(\overline{\Omega})$ gehört.

Zusatz 2.2.3 *Unter den Voraussetzungen von Satz 2.2.2 gilt für jede Grundlösung in Ω :*

$$u(\mathbf{y}) = \int_{\partial\Omega} \left[\gamma(\mathbf{x}, \mathbf{y}) \frac{\partial u(\mathbf{x})}{\partial n} - u(\mathbf{x}) \frac{\partial \gamma(\mathbf{x}, \mathbf{y})}{\partial n_x} \right] d\Gamma_x \quad (\mathbf{y} \in \Omega). \quad (2.2.10)$$

Beweis. (2.2.5b) impliziert $\int_{\partial\Omega} [\Phi \partial u / \partial n - u \partial \Phi / \partial n] d\Gamma = 0$. ■

Wegen einer möglichen Abschwächung der Bedingung $\Phi = \gamma - s \in C^2(\overline{\Omega})$ auf $\Phi(\cdot, \mathbf{y}) \in C^1(\overline{\Omega}) \cap C^2(\Omega)$ vergleiche man Hellwig [46, p. 34].

Übungsaufgabe 2.2.4 (Green-Funktion auf Kreisen) *Es sei $\Omega = K_R(\mathbf{y})$. Man definiere*

$$\gamma(\mathbf{x}, \xi) = \begin{cases} \frac{1}{(n-2)\omega_n} \left[|\mathbf{x} - \xi|^{2-n} - \left(\frac{|\xi - \mathbf{y}|}{R} |\mathbf{x} - \xi'| \right)^{2-n} \right] & \text{für } n > 2, \\ -\frac{1}{2\pi} \left[\log |\mathbf{x} - \xi| - \log \left(\frac{|\xi - \mathbf{y}|}{R} |\mathbf{x} - \xi'| \right) \right] & \text{für } n = 2, \end{cases} \quad (2.2.11a)$$

mit $\mathbf{x}, \xi \in \Omega$, $\xi' = \mathbf{y} + R^2 |\xi - \mathbf{y}|^{-2} (\xi - \mathbf{y})$ und zeige:

a) γ ist Grundlösung in Ω ,

b) $\gamma(\mathbf{x}, \xi) = \gamma(\xi, \mathbf{x})$,

c) auf der Kugeloberfläche $\Gamma = \partial K_R(\mathbf{y})$ gilt

$$\frac{\partial}{\partial n_\xi} \gamma(\mathbf{x}, \xi) = \frac{\partial}{\partial n_\xi} \gamma(\xi, \mathbf{x}) = -\frac{1}{R\omega_n} \frac{R^2 - |\mathbf{x} - \mathbf{y}|^2}{|\mathbf{x} - \xi|^n} \quad (\xi \in \Gamma). \quad (2.2.11b)$$

2.3 Mittelwerteigenschaft und Maximumprinzip

Definition 2.3.1 (Mittelwerteigenschaft) Eine Funktion u besitzt die Mittelwerteigenschaft in Ω , falls $u \in C^0(\overline{\Omega})$ und falls für alle $\mathbf{x} \in \Omega$ und alle $R > 0$ mit $K_R(\mathbf{x}) \subset \Omega$ die Darstellung

$$u(\mathbf{x}) = \frac{1}{\omega_n R^{n-1}} \int_{\partial K_R(\mathbf{x})} u(\xi) d\Gamma_\xi \quad (2.3.1)$$

gültig ist.

Da $\int_{\partial K_R(\mathbf{x})} d\Gamma = \omega_n R^{n-1}$, ist die rechte Seite in (2.3.1) der über die Kugeloberfläche gemittelte Wert von u . Eine äquivalente Charakterisierung ergibt sich, wenn man über die Kugel $K_R(\mathbf{x})$ mittelt:

Übungsaufgabe 2.3.2 $u \in C^0(\overline{\Omega})$ besitzt die zweite Mittelwerteigenschaft in Ω , wenn

$$u(\mathbf{x}) = \frac{n}{R^n \omega_n} \int_{K_R(\mathbf{x})} u(\xi) d\xi \quad \text{für alle } \mathbf{x} \in \Omega, R > 0 \quad \text{mit } K_R(\mathbf{x}) \subset \Omega.$$

Man zeige, dass die zweite Mittelwerteigenschaft äquivalent zur Mittelwerteigenschaft (2.3.1) ist. Hinweis:

$$\int_{K_R(\mathbf{x})} u(\xi) d\xi = \int_0^R \left(\int_{\partial K_r(\mathbf{x})} u(\xi) d\Gamma_\xi \right) dr.$$

Funktionen mit der Mittelwerteigenschaft erfüllen ein Maximumprinzip, wie es aus der Funktionentheorie für holomorphe Funktionen bekannt ist:

Satz 2.3.3 (Maximum-Minimum-Prinzip) Ω sei ein Gebiet und $u \in C^0(\overline{\Omega})$ sei eine nichtkonstante Funktion, die die Mittelwerteigenschaft besitzt. Dann nimmt u in Ω weder ein Maximum noch Minimum an.

Beweis. 1) Es genügt, den Fall des Maximums zu untersuchen, da ein Minimum von u ein Maximum von $-u$ ist und $-u$ ebenfalls die Mittelwerteigenschaft besitzt.

2) Zur indirekten Beweisführung nehmen wir an, es gäbe ein Maximum in $\mathbf{y} \in \Omega$:

$$u(\mathbf{y}) = M \geq u(\mathbf{x}) \quad \text{für alle } \mathbf{x} \in \Omega.$$

In 3) werden wir $u(\mathbf{y}') = M$ für beliebiges $\mathbf{y}' \in \Omega$ zeigen, d.h. $u \equiv M$ im Widerspruch zur Voraussetzung $u \not\equiv \text{const.}$

3) Beweis von $u(\mathbf{y}') = M$. Sei $\mathbf{y}' \in \Omega$. Da Ω zusammenhängend ist, existiert eine in ω verlaufende Verbindung von \mathbf{y} und \mathbf{y}' , d.h. es gibt ein stetiges $\varphi: [0, 1] \rightarrow \Omega$ mit $\varphi(0) = \mathbf{y}$, $\varphi(1) = \mathbf{y}'$. Wir setzen

$$I := \{s \in [0, 1] : u(\varphi(t)) = M \text{ für alle } 0 \leq t \leq s\}.$$

I enthält mindestens 0 und ist abgeschlossen, da u und φ stetig sind. Damit existiert $s^* = \max\{s \in I\}$, und die Definition von I zeigt $I = [0, s^*]$. In 4) wird $s^* = 1$ nachgewiesen, sodass $\mathbf{y}' = \varphi(1) \in I$ und somit $u(\mathbf{y}') = M$ folgt.

4) Beweis von $s^* = 1$. Die gegenteilige Annahme $s^* < 1$ kann zum Widerspruch geführt werden, indem $u(\mathbf{x}) = M$ in einer Umgebung von $\mathbf{x}^* := \varphi(s^*)$ nachgewiesen wird. Da $\mathbf{x}^* \in \Omega$, existiert $R > 0$ mit $K_R(\mathbf{x}^*) \subset \Omega$. Offenbar folgt $u = M$ in $K_R(\mathbf{x}^*)$, wenn $u = M$ auf $\partial K_r(\mathbf{x}^*)$ für alle $0 < r \leq R$ gezeigt wird.

5) Beweis von $u = M$ auf $\partial K_r(\mathbf{x}^*)$. Gleichung (2.3.1) in \mathbf{x}^* lautet

$$M = u(\mathbf{x}^*) = \frac{1}{\omega_n r^{n-1}} \int_{\partial K_r(\mathbf{x}^*)} u(\xi) d\Gamma_\xi.$$

Generell gilt $u(\xi) \leq M$. Wäre $u(\xi') < M$ für ein $\xi' \in \partial K_r(\mathbf{x}^*)$ und damit $u < M$ in einer Umgebung von ξ' , ergäbe sich auf der rechten Seite ein Mittelwert kleiner als M . Dies beweist $u = M$ auf $\partial K_R(\mathbf{x}^*)$. ■

Einfache Folgerungen aus Satz 2.3.3 enthält der

Zusatz 2.3.4 Ω sei beschränkt. (a) Eine Funktion mit Mittelwerteigenschaft nimmt ihr Maximum und Minimum auf $\partial\Omega$ an. (b) Stimmen zwei Funktionen mit Mittelwerteigenschaft auf dem Rand $\partial\Omega$ überein, so sind sie identisch.

Beweis. (a) Die Extrema werden auf $\bar{\Omega} = \Omega \cup \partial\Omega$ angenommen, da $\bar{\Omega}$ kompakt ist. Ω kommt nach Satz 2.3.3 für ein Extremum nicht in Frage, wenn u nicht auf einer Zusammenhangskomponente von Ω konstant ist. Aber auch in diesem Fall ist die Behauptung offensichtlich.

(b) Erfüllen u und v mit $u = v$ auf $\partial\Omega$ die Mittelwerteigenschaft, so ist diese auch für $w := u - v$ erfüllt. Da $w = 0$ auf $\partial\Omega$, zeigt Teil (a), dass $\max w = \min w = 0$. Also $u = v$ in Ω . ■

Lemma 2.3.5 Harmonische Funktionen besitzen die Mittelwerteigenschaft.

Beweis. Sei u harmonisch in Ω und $\mathbf{y} \in K_R(\mathbf{y}) \subset \Omega$. Wir wenden die Darstellung (2.2.6) für $K_R(\mathbf{y})$ anstelle von Ω an. Der konstante Wert von $s(\mathbf{x}, \mathbf{y})$ auf $\partial K_R(\mathbf{y})$ sei mit $\sigma(R)$ bezeichnet. Wegen (2.2.4) wird Gleichung (2.2.6) zu

$$u(\mathbf{y}) = \sigma(R) \int_{\partial K_R(\mathbf{y})} \frac{\partial u(\xi)}{\partial n} d\Gamma + \frac{1}{\omega_n R^{n-1}} \int_{\partial K_R(\mathbf{y})} u(\xi) d\Gamma.$$

Die Gleichung stimmt mit (2.3.1) überein, wenn das erste Integral verschwindet. Letzteres folgt aus dem folgenden Lemma. ■

Lemma 2.3.6 Sei $u \in C^2(\Omega)$ harmonisch in einem Normalgebiet Ω . Dann gilt

$$\int_{\partial\Omega} \frac{\partial u}{\partial n} d\Gamma = 0. \quad (2.3.2)$$

Beweis. Man setze in der Greenschen Formel (2.2.5a) 1 und u anstelle von u und v ein. ■

Lemma 2.3.5, Satz 2.3.3 und Zusatz 2.3.4 ergeben die nächsten Sätze 2.3.7 und 2.3.8.

Satz 2.3.7 (Maximum-Minimum-Prinzip für harmonische Funktionen) Die Funktion u sei harmonisch im Gebiet Ω und nicht konstant. Dann existiert kein Maximum und kein Minimum in Ω .

Satz 2.3.8 (Eindeutigkeit) Ω sei beschränkt. Eine in Ω harmonische Funktion nimmt ihr Maximum und Minimum auf $\partial\Omega$ an und ist durch die Werte auf $\partial\Omega$ eindeutig bestimmt.

Die Darstellung (2.3.1) von $u(\mathbf{y})$ durch die Werte auf $\partial K_R(\mathbf{y})$ ist ein Spezialfall folgender Formel, die am Ende dieses Abschnittes bewiesen werden wird und für $\mathbf{x} = \mathbf{y}$ die Gleichung (2.3.1) liefert.

Satz 2.3.9 (Poissonsche Integralformel) ⁴ Sei $\varphi \in C^0(\partial K_R(\mathbf{y}))$ und $n \geq 2$. Die Lösung der Randwertaufgabe

$$\Delta u = 0 \quad \text{in } K_R(\mathbf{y}), \quad u = \varphi \quad \text{auf } \partial K_R(\mathbf{y}), \quad (2.3.3)$$

ist gegeben durch die Funktion

$$u(\mathbf{x}) = \frac{R^2 - |\mathbf{x} - \mathbf{y}|^2}{R\omega_n} \int_{\partial K_R(\mathbf{y})} \frac{\varphi(\xi)}{|\mathbf{x} - \xi|^n} d\Gamma_\xi \quad \text{für } \mathbf{x} \in K_R(\mathbf{y}), \quad (2.3.4)$$

die zu $C^\infty(K_R(\mathbf{y})) \cap C^0(\overline{K_R(\mathbf{y})})$ gehört.

Die Mittelwerteigenschaft setzt nur $u \in C^0(\bar{\Omega})$ voraus, während harmonische Funktionen zu $C^2(\Omega) \cap C^0(\bar{\Omega})$ gehören. Um so erstaunlicher ist die folgende Aussage:

Satz 2.3.10 Eine Funktion ist genau dann harmonisch in Ω , wenn sie dort die Mittelwerteigenschaft besitzt.

Beweis. Wegen Lemma 2.3.5 bleibt zu zeigen, dass eine Funktion v mit Mittelwerteigenschaft harmonisch ist. Sei $\mathbf{x} \in K_R(\mathbf{x}) \subset \Omega$ beliebig vorgegeben. Nach Satz 2.3.9 existiert eine in $K_R(\mathbf{x})$ harmonische Funktion u mit

$$\Delta u = 0 \quad \text{in } K_R(\mathbf{y}), \quad u = v \quad \text{auf } \partial K_R(\mathbf{y}).$$

⁴Siméon Denis Poisson, geb. 21. Juni 1781 in Pithiviers, gest. 25. April 1840 in Sceaux

Gemäß Lemma 2.3.5 besitzt u ebenso wie v die Mittelwerteigenschaft, und Zusatz 2.3.4b beweist $u = v$ in $K_R(\mathbf{x})$, d.h. v ist in $K_R(\mathbf{x})$ harmonisch. Da $K_R(\mathbf{x}) \subset \Omega$ beliebig, ist v in Ω harmonisch. ■

Eine wichtige Anwendung des Satzes 2.3.10 ist die folgende Aussage.

Satz 2.3.11 (Harnack)⁵ Sei u_1, u_2, \dots eine Folge in Ω harmonischer Funktionen, die in $\overline{\Omega}$ gleichmäßig konvergieren. Dann ist $u = \lim_{k \rightarrow \infty} u_k$ harmonisch in Ω .

Beweis. Wegen der gleichmäßigen Konvergenz ist die Grenzfunktion stetig: $u \in C^0(\overline{\Omega})$. Der Limesprozess angewandt auf $u_k(\mathbf{x}) = \int_{\partial K_R(\mathbf{x})} u_k(\xi) d\Gamma_\xi / (\omega_n R^{n-1})$ liefert Gleichung (2.3.1) für u ; d.h. u besitzt die Mittelwerteigenschaft. Nach Satz 2.3.10 ist u auch harmonisch in Ω . ■

Die Sätze 2.3.3 und 2.3.7 über das Maximum-Minimum-Prinzip beziehen sich auf globale Extrema. Der Beweis von Satz 2.3.3 schließt noch keine lokalen Extrema im Inneren aus. Er zeigt nur, dass u dann in einem Kreis $K_R(\mathbf{y}) \subset \Omega$ konstant ist. Wie aus der Funktionentheorie bekannt, lässt sich hieraus $u(\mathbf{x}) = M$ in $\overline{\Omega}$ ableiten, wenn u analytisch ist, d.h. in der Umgebung jedes $\mathbf{x} \in \Omega$ eine konvergente Potenzreihenentwicklung gestattet. In der Tat gilt der folgende Satz, dessen Beweis man z.B. in Hellwig [46, p. 93] findet:

Satz 2.3.12 Eine in Ω harmonische Funktion ist dort analytisch.

Es bleibt der Beweis der Poissonschen Integralformel nachzutragen.

Beweis zu Satz 2.3.9. a) Zunächst ist zu zeigen, dass u aus (2.3.4) eine in $K_R(\mathbf{y})$ harmonische Funktion ist, d.h. $\Delta u = 0$ erfüllt. Da der Integrand zweimal stetig differenzierbar und der Integrationsbereich $\Gamma = \partial K_R(\mathbf{y})$ kompakt ist, darf der Laplace-Operator unter das Integralzeichen gezogen werden:

$$\Delta u(\mathbf{x}) = \frac{1}{R\omega_n} \int_{\Gamma} \varphi(\xi) \Delta \left[\left(R^2 - |\mathbf{x} - \mathbf{y}|^2 \right) |\mathbf{x} - \xi|^{-n} \right] d\Gamma_\xi \quad \text{für } \mathbf{x} \in K_R(\mathbf{y}). \quad (2.3.5)$$

Gemäß Übungsaufgabe 2.2.4 gibt es eine Grundlösung $\gamma(\mathbf{x}, \xi)$, sodass

$$\frac{R^2 - |\mathbf{x} - \mathbf{y}|^2}{|\mathbf{x} - \xi|^n R\omega_n} = -\frac{\partial \gamma(\mathbf{x}, \xi)}{\partial n_\xi} = -\frac{\partial \gamma(\xi, \mathbf{x})}{\partial n_\xi} \quad \text{für } \xi \in \Gamma, \mathbf{x} \in K_R(\mathbf{y}). \quad (2.3.6)$$

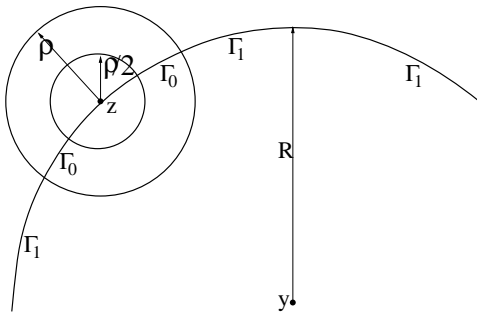


Abb. 2.3.1. Γ und die Kreise $K_\rho(z), K_{\rho/2}(z)$

Aus $\Delta_x \partial \gamma / \partial n_\xi = \frac{\partial}{\partial n_\xi} \Delta_x \gamma(\mathbf{x}, \xi) = 0$ und (2.3.5) schließt man $\Delta u = 0$.

b) Die Darstellung (2.3.4) definiert $u(\mathbf{x})$ zunächst nur für $\mathbf{x} \in K_R(\mathbf{y})$. Es bleibt noch zu zeigen, dass u eine stetige Fortsetzung auf $K_R(\mathbf{y}) = K_R(\mathbf{y}) \cup \Gamma$ besitzt (d.h. $u \in C^0(K_R(\mathbf{y}))$) und dass die stetig fortgesetzten Werte mit den Randwerten φ übereinstimmen:

$$\lim_{K_R(\mathbf{y}) \ni \mathbf{x} \rightarrow \mathbf{z}} u(\mathbf{x}) = \varphi(\mathbf{z}) \quad \text{für } \mathbf{z} \in \Gamma. \quad (2.3.7)$$

Wegen Gleichung (2.3.6) liefert Zusatz 2.2.3 für die Wahl $u \equiv 1$ die Identität

$$\frac{R^2 - |\mathbf{x} - \mathbf{y}|^2}{R\omega_n} = \int_{\Gamma} \frac{d\Gamma_\xi}{|\mathbf{x} - \xi|^n} = 1 \quad \text{für } \mathbf{x} \in K_R(\mathbf{y}). \quad (2.3.8)$$

Sei $\mathbf{z} \in \Gamma$ beliebig. Dank Gleichung (2.3.8) lässt sich schreiben:

$$u(\mathbf{x}) - \varphi(\mathbf{z}) = \frac{R^2 - |\mathbf{x} - \mathbf{y}|^2}{R\omega_n} \int_{\Gamma} \frac{u(\xi) - \varphi(\mathbf{z})}{|\mathbf{x} - \xi|^n} d\Gamma_\xi. \quad (2.3.9a)$$

Wir definieren $\Gamma_0 = \Gamma \cap K_\rho(\mathbf{z})$, $\Gamma_1 = \Gamma / \Gamma_0$ (siehe Abbildung 2.3.1) und zerlegen den Ausdruck (2.3.9a) in $u(\mathbf{x}) - \varphi(\mathbf{z}) = I_0 + I_1$ wobei

$$I_i = \frac{R^2 - |\mathbf{x} - \mathbf{y}|^2}{R\omega_n} \int_{\Gamma_i} \frac{u(\xi) - \varphi(\mathbf{z})}{|\mathbf{x} - \xi|^n} d\Gamma_\xi \quad \text{für } i = 0, 1.$$

⁵Carl Gustav Axel Harnack, 1851-1888, Promotion 1875 an der Universität Erlangen.

Da

$$\left| \int_{\Gamma_0} \frac{u(\xi) - \varphi(\mathbf{z})}{|\mathbf{x} - \xi|^n} d\Gamma_\xi \right| \leq \max_{\xi \in \Gamma_0} |u(\xi) - \varphi(\mathbf{z})| \int_{\Gamma_0} \frac{d\Gamma_\xi}{|\mathbf{x} - \xi|^n} \leq \max_{\xi \in \Gamma_0} |u(\xi) - \varphi(\mathbf{z})| \int_{\Gamma} \frac{d\Gamma_\xi}{|\mathbf{x} - \xi|^n},$$

folgt aus Gleichung (2.3.8), dass

$$I_0 \leq \max_{\xi \in \Gamma_0} |u(\xi) - \varphi(\mathbf{z})|. \quad (2.3.9b)$$

Wegen der Stetigkeit von φ kann $\rho > 0$ so gewählt werden, dass

$$I_0 \leq \varepsilon/2 \quad (2.3.9c)$$

für vorgegebenes $\varepsilon > 0$.

Man setze $C_\varphi := \max_{\xi \in \Gamma} |\varphi(\xi)|$ und wähle $\mathbf{x} \in K_R(\mathbf{y})$ so nah an \mathbf{z} , dass

$$|\mathbf{x} - \mathbf{z}| \leq \delta(\varepsilon) := \frac{\varepsilon}{2} \left(\frac{\rho}{2} \right)^n \frac{1}{4C_\varphi R^{n-1}}$$

und $|\mathbf{x} - \mathbf{z}| \leq \rho/2$. Die letzte Ungleichung impliziert

$$|\mathbf{x} - \xi| \geq \frac{\rho}{2} \quad \text{für } \xi \in \Gamma_1 \quad (\text{siehe Abbildung 2.3.1}).$$

Zusammen mit $R^2 - |\mathbf{x} - \mathbf{y}|^2 = (R + |\mathbf{x} - \mathbf{y}|)(R - |\mathbf{x} - \mathbf{y}|) \leq 2R(R - |\mathbf{x} - \mathbf{y}|) \leq 2R|\mathbf{z} - \mathbf{x}| \leq 2R\delta(\varepsilon)$ ergibt sich

$$|I_1| = \left| \frac{R^2 - |\mathbf{x} - \mathbf{y}|^2}{R\omega_n} \int_{\Gamma_1} \frac{u(\xi) - \varphi(\mathbf{z})}{|\mathbf{x} - \xi|^n} d\Gamma_\xi \right| \leq \frac{2}{\omega_n} \delta(\varepsilon) \frac{2C_\varphi}{(\rho/2)^n} \int_{\Gamma_1} d\Gamma.$$

Aus $\int_{\Gamma_1} d\Gamma \leq \int_{\Gamma} d\Gamma = R^{n-1}\omega_n$ und der Definition von $\delta(\varepsilon)$ folgt

$$|I_1| \leq \varepsilon/2. \quad (2.3.9d)$$

Damit existiert zu jedem $\varepsilon > 0$ ein $\delta(\varepsilon) > 0$, sodass die Ungleichung $|\mathbf{x} - \mathbf{z}| \leq \delta(\varepsilon)$ die Abschätzung $|u(\mathbf{x}) - \varphi(\mathbf{z})| = |I_0 + I_1| \leq |I_0| + |I_1| \leq \varepsilon$ impliziert (vgl. (2.3.9c,d)). Also ist (2.3.7) bewiesen, und die stetige Fortsetzung von u auf $K_R(\mathbf{y})$ führt zu $u \in C^0(K_R(\mathbf{y}))$. ■

Übungsaufgabe 2.3.13 Man beweise: Die durch die Poissonsche Integralformel (2.3.4) definierte Funktion u gehört zu $C^\infty(K_R(\mathbf{y}))$ und löst $\Delta u = 0$ in $K_R(\mathbf{y})$, auch wenn φ nur als eine auf $\Gamma = \partial K_R(\mathbf{y})$ integrierbare Funktion vorausgesetzt wird. Für jeden Stetigkeitspunkt $\mathbf{z} \in \Gamma$ von φ gilt $u(\mathbf{x}) \rightarrow \varphi(\mathbf{z})$ ($K_R(\mathbf{y}) \ni \mathbf{x} \rightarrow \mathbf{z}$).

Übungsaufgabe 2.3.14 Ω sei beschränkt. u_1 und u_2 seien harmonisch in Ω mit Randwerten φ_1 und φ_2 auf $\Gamma = \partial\Omega$. Man beweise: a) $\varphi_1 \leq \varphi_2$ auf Γ impliziert $u_1 \leq u_2$ in Ω . b) Ist außerdem Ω zusammenhängend und gilt $\varphi_1(\mathbf{x}) < \varphi_2(\mathbf{x})$ in mindestens einem Punkt $\mathbf{x} \in \Gamma$, so folgt $u_1 < u_2$ überall in Ω .

2.4 Stetige Abhängigkeit von den Randdaten

Definition 2.4.1 Eine abstrakte Aufgabe der Form

$$A(x) = y, \quad x \in X, y \in Y,$$

heißt sachgemäß gestellt, wenn sie für alle $y \in Y$ eine eindeutige Lösung $x \in X$ besitzt und diese stetig von y abhängt.

Es ist wichtig, eine mathematische Fragestellung als sachgemäß gestellt zu erkennen, da andernfalls prinzipielle Schwierigkeiten bei der numerischen Lösung zu erwarten sind. Im Falle der Randwertaufgabe (2.1.1a,b) sind $X \subset C^2(\Omega) \cap C^0(\overline{\Omega})$ der Raum der in Ω harmonischen Funktionen und $Y = C^0(\Gamma)$ die Menge der stetigen Randdaten auf $\Gamma = \partial\Omega$. Die Topologien von X und Y sind durch die *Supremumsnormen* gegeben:

$$\|u\|_\infty := \sup_{\mathbf{x} \in \Omega} |u(\mathbf{x})| \quad \text{bzw.} \quad \|\varphi\|_\infty := \sup_{\mathbf{x} \in \Gamma} |\varphi(\mathbf{x})|. \quad (2.4.1)$$

Die Frage der Existenz einer Lösung von (2.1.1a,b) muss noch zurückgestellt werden (siehe §7). Dagegen ist die Eindeutigkeit schon in Satz 2.3.8 bestätigt worden, wenn Ω beschränkt ist. Dass die Beschränktheit von Ω nicht ohne Weiteres gestrichen werden kann, zeigt

Beispiel 2.4.2 Die Funktionen

$$\begin{aligned}
u(x_1, x_2) &= x_1 && \text{in } \Omega = (0, \infty) \times \mathbb{R}, \\
u(x_1, x_2) &= \log(x_1^2 + x_2^2) && \text{in } \Omega = \mathbb{R}^2 \setminus \overline{K_1(0)}, \\
u(x_1, x_2) &= \sin(x_1) \sinh(x_2) && \text{in } \Omega = (0, \pi) \times (0, \infty)
\end{aligned} \tag{2.4.2}$$

erfüllen ebenso wie die triviale Lösung $u = 0$ die Randwertaufgabe $\Delta u = 0$ in Ω , $u = 0$ auf $\Gamma = \partial\Omega$.

Für beschränktes Ω hängen die harmonischen Funktionen (Lösungen von (2.1.1a,b)) nicht nur stetig, sondern auch Lipschitz-stetig⁶ von den Randdaten ab:

Satz 2.4.3 Ω sei beschränkt. Sind u^I und u^II Lösungen von

$$\Delta u^I = \Delta u^{II} = 0 \quad \text{in } \Omega, \quad u^I = \varphi^I \text{ und } u^{II} = \varphi^{II} \quad \text{auf } \Gamma = \partial\Omega,$$

so gilt

$$\|u^I - u^{II}\|_\infty \leq \|\varphi^I - \varphi^{II}\|_\infty. \tag{2.4.3}$$

Beweis. $v := u^I - u^{II}$ ist Lösung von $\Delta v = 0$ in Ω und $v = \varphi^I - \varphi^{II}$ auf Γ . Nach Satz 2.3.8 nimmt v sein Maximum und Minimum auf Γ an:

$$-\|\varphi^I - \varphi^{II}\|_\infty \leq v(\mathbf{x}) \leq \|\varphi^I - \varphi^{II}\|_\infty \quad \text{für alle } \mathbf{x} \in \overline{\Omega}.$$

Die Definition (2.4.1) von $\|\cdot\|_\infty$ liefert (2.4.3). ■

Gilt $\|\varphi_n - \varphi\|_\infty \rightarrow 0$ für eine Folge von Randwerten, so zeigt Satz 2.4.3, dass die zugehörigen Lösungen $\|u_n - u\|_\infty \rightarrow 0$ erfüllen. Der folgende Satz besagt, dass die Existenz einer Lösung von $\Delta u = 0$ in Ω , $u = \varphi$ auf Γ nicht vorausgesetzt zu werden braucht.

Satz 2.4.4 Ω sei beschränkt. $\varphi_n \in C^0(\Gamma)$ sei eine Folge von Randdaten, die gleichmäßig gegen φ konvergieren: $\|\varphi_n - \varphi\|_\infty \rightarrow 0$. u_n seien die Lösungen von $\Delta u_n = 0$ in Ω , $u_n = \varphi_n$ auf Γ . Dann konvergieren die Funktionen gleichmäßig in $\overline{\Omega}$ gegen ein $u \in C^2(\Omega) \cap C^0(\overline{\Omega})$, und u löst die Randwertaufgabe $\Delta u = 0$ in Ω , $u = \varphi$ auf Γ .

Beweis. Da $\|\cdot\|_\infty \leq \|\varphi_n - \varphi_m\|_\infty$ nach Satz 2.4.3, ist die Folge u_n Cauchy-konvergent. Da $C^0(\overline{\Omega})$ vollständig ist, konvergiert u_n gleichmäßig gegen ein $u \in C^0(\overline{\Omega})$. Gemäß Satz 2.3.11 ist u harmonisch (d.h. $\Delta u = 0$). Offensichtlich ist auch $u = \varphi$ auf Γ erfüllt. ■

Eine für die numerische Behandlung ebenso wichtige Frage wird in der Literatur seltener behandelt: Hängt die Lösung auch stetig von der Gestalt des Randes Γ ab? Abbildung 2.4.1 zeigt Gebiete Ω' und Ω'' , die Ω approximieren. Ein Polygonebiet, wie z.B. Ω'' tritt bei der Methode der finiten Elemente auf (siehe §8.5.3).

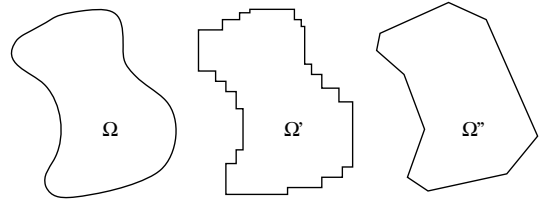


Abb. 2.4.1. Approximation von Ω durch Ω' und Ω''

Sei Ω_n eine Folge von Gebieten mit $\Gamma_n = \partial\Omega_n$. Wir sagen, dass $\Gamma_n \rightarrow \Gamma$, falls $\text{dist}(\Gamma_n, \Gamma) \rightarrow 0$. Dabei ist $\text{dist}(\Gamma_n, \Gamma) := \sup\{\text{dist}(\mathbf{x}, \Gamma) : \mathbf{x} \in \Gamma_n\}$, $\text{dist}(\mathbf{x}, \Gamma) := \inf\{|\mathbf{x} - \mathbf{y}| : \mathbf{y} \in \Gamma\}$.

Weiter ist zu erklären, wann $\varphi_n \in C^0(\Gamma_n)$ gleichmäßig gegen $\varphi \in C^0(\Gamma)$ konvergiert. Wir definieren die gleichmäßig Konvergenz für φ_n und für u_n :

$\varphi_n \in C^0(\Gamma_n)$ konvergiert gleichmäßig gegen $\varphi \in C^0(\Gamma)$, falls für jedes $\varepsilon > 0$ Zahlen $N(\varepsilon)$ und $\delta(\varepsilon) > 0$ existieren, sodass folgende Implikation gilt:

$$n \geq N(\varepsilon), \quad \mathbf{x} \in \Gamma, \quad |\mathbf{x} - \mathbf{y}| \leq \delta(\varepsilon) \implies |\varphi_n(\mathbf{y}) - \varphi(\mathbf{x})| \leq \varepsilon. \tag{2.4.4a}$$

$u_n \in C^0(\Omega_n)$ konvergiert gleichmäßig gegen $u \in C^0(\Omega)$, falls

$$\lim_{n \rightarrow \infty} [\sup\{|u_n(\mathbf{x}) - u(\mathbf{x})| : \mathbf{x} \in \Omega_n \cap \Omega\}] = 0. \tag{2.4.4b}$$

⁶Rudolf Otto Sigismund Lipschitz, geb. 14. Mai 1832 in Königsberg, gest. 7. Okt. 1903 in Bonn

Bemerkung 2.4.5 a) K sei eine kompakte Menge (d.h. abgeschlossen und beschränkt) mit $\Gamma \subset K$, $\Gamma_n \subset K$ für alle n . $\phi_n \in C^0(K)$ konvergiere gleichmäßig auf K gegen ϕ . Wenn $\varphi_n = \phi_n$ auf Γ_n und $\varphi = \phi$ auf Γ , so ist (2.4.4a) erfüllt.

b) Es gelte $\Omega_n \subset \Omega$ für alle n , und \bar{u}_n sei folgende (unstetige) Fortsetzung von u_n auf $\bar{\Omega}$: $\bar{u}_n = u_n$ auf $\bar{\Omega}_n$, $\bar{u}_n = u$ auf $\bar{\Omega} \setminus \bar{\Omega}_n$. Dann ist (2.4.4b) äquivalent mit der gleichmäßigen Konvergenz $\bar{u}_n \rightarrow u$ auf $\bar{\Omega}$.

Satz 2.4.6 Seien Ω beschränkt, $\Omega_n \subset \Omega$ und $\Gamma_n \rightarrow \Gamma$. Die in Ω_n harmonischen Funktionen u_n seien Lösungen von

$$\Delta u_n = 0 \quad \text{in } \Omega_n, \quad u_n = \varphi_n \quad \text{auf } \Gamma_n. \quad (2.4.5a)$$

$\varphi_n \in C^0(\Gamma_n)$ konvergiere im Sinne von (2.4.4a) gleichmäßig gegen $\varphi \in C^0(\Gamma)$. Dann gelten die folgenden Aussagen: (a) Wenn eine Lösung $u \in C^2(\Omega) \cap C^0(\bar{\Omega})$ von

$$\Delta u = 0 \quad \text{in } \Omega, \quad u = \varphi \quad \text{auf } \Gamma. \quad (2.4.5b)$$

existiert, gilt $u_n \rightarrow u$ im Sinne von (2.4.4b).

(b) Ist umgekehrt $u_n \rightarrow u \in C^0(\bar{\Omega})$ im Sinne von (2.4.4b) erfüllt, ist u die Lösung von (2.4.5b).

Beweis. a) Die Fortsetzung \bar{u}_n sei wie in Bemerkung 2.4.5b definiert. Da u gleichmäßig stetig auf $\bar{\Omega}$ ist, existiert $\delta_u(\varepsilon) > 0$ für alle $\varepsilon > 0$, sodass

$$|u(\mathbf{x}) - u(\mathbf{y})| \leq \frac{\varepsilon}{2}, \quad \text{wenn } |\mathbf{x} - \mathbf{y}| \leq \delta_u(\frac{\varepsilon}{2}). \quad (2.4.6a)$$

Man setze $\delta^*(\varepsilon) := \min\{\delta_u(\frac{\varepsilon}{2}), \delta(\frac{\varepsilon}{2})\}$ mit der Schranke δ aus (2.4.4a). Wegen $\Gamma_n \rightarrow \Gamma$ existiert $N_\Gamma(\varepsilon)$, sodass $\text{dist}(\Gamma_n, \Gamma) \leq \delta^*(\varepsilon)$ für $n \geq N_\Gamma(\varepsilon)$. Für

$$n \geq N^*(\varepsilon) := \max\{N_\Gamma(\varepsilon), N(\varepsilon/2)\} \quad (N \text{ aus (2.4.4a)})$$

auf alle $\mathbf{x} \in \bar{\Omega}$ wollen wir $|\bar{u}_n(\mathbf{x}) - u(\mathbf{x})| \leq \varepsilon$ zeigen. Für $\mathbf{x} \in \bar{\Omega}_n$ ist die Abschätzung trivial wegen $\bar{u}_n(\mathbf{x}) = u_n(\mathbf{x})$. Für alle $\mathbf{x} \in \bar{\Omega}_n \subset \bar{\Omega}$ gilt dagegen

$$|\bar{u}_n(\mathbf{x}) - u(\mathbf{x})| = |u_n(\mathbf{x}) - u(\mathbf{x})| \leq \max_{\mathbf{x} \in \Gamma_n} |u_n(\mathbf{x}) - u(\mathbf{x})| \quad (2.4.6b)$$

(vgl. Satz 2.3.8), weil $u_n - u$ harmonisch in ω_n ist. Es bleibt $|u_n(\mathbf{x}) - u(\mathbf{x})|$ für $\mathbf{x} \in \Gamma_n$ abzuschätzen. Zu $\mathbf{x} \in \Gamma_n$ mit $n \geq N^*(\varepsilon)$ existiert $\mathbf{y} \in \Gamma$ mit $|\mathbf{x} - \mathbf{y}| \leq \delta(\varepsilon/2)$. Somit erhalten wir

$$\begin{aligned} |u_n(\mathbf{x}) - u(\mathbf{x})| &= |\varphi_n(\mathbf{x}) - u(\mathbf{x})| \leq |\varphi_n(\mathbf{x}) - \varphi(\mathbf{y})| + |\varphi(\mathbf{y}) - u(\mathbf{x})| \\ &\leq |\varphi_n(\mathbf{x}) - \varphi(\mathbf{y})| + |u(\mathbf{y}) - u(\mathbf{x})| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \end{aligned}$$

aus (2.4.4a) und (2.4.6a). Da $\mathbf{x} \in \Gamma_n$ beliebig, folgt $|u_n - u| \leq \varepsilon$ auf Γ_n , und (2.4.6b) beweist die gleichmäßige Konvergenz $\bar{u}_n \rightarrow u$ auf $\bar{\Omega}$. Nach Bemerkung 2.4.5b ist (2.4.4b) erfüllt.

b) Sei $K \subset \Omega$ eine kompakte Menge. Da $\Gamma_n \rightarrow \Gamma$, gibt es ein $N(K)$, sodass $K \subset \Omega_n$ für $n \geq N(K)$. Damit konvergiert $\{u_n : n \geq N(K)\}$ im üblichen Sinne gleichmäßig auf K gegen u , sodass sich Satz 2.3.11 anwenden lässt: u ist in K harmonisch. Da $K \subset \Omega$ beliebig gewählt werden konnte, folgt $u \in C^2(\Omega)$. Nach Voraussetzung gilt bereits $u \in C^0(\bar{\Omega})$. Die Annahme der Randwerte $u = \varphi$ auf Γ folgert man aus $\varphi_n \rightarrow \varphi$ und $\Gamma_n \rightarrow \Gamma$. ■

In Satz 2.4.4 konnte die Existenz der Lösung u von (2.4.5b) schon aus $\varphi_n \rightarrow \varphi$ abgeleitet werden. Dieser Schluss ist im Falle von $\Omega_n \neq \Omega$ nicht möglich, wie das folgende Beispiel zeigt.

Sei $\Omega_n := K_1(0) \setminus \overline{K_{1/n}(0)} \subset \Omega := K_1(0) \setminus \{0\} \subset \mathbb{R}^2$. Die Ränder sind $\Gamma_n = \partial K_1(0) \cup \partial K_{1/n}(0)$, $\Gamma = \partial K_1(0) \cup \{0\}$ und erfüllen $\Gamma_n \rightarrow \Gamma$. Die Randwerte

$$\varphi = \varphi_n = 0 \quad \text{auf } \partial K_1(0), \quad \varphi_n = 1 \quad \text{auf } \partial K_{1/n}(0), \quad \varphi(0,0) = 1$$

genügen der Bedingung $\varphi_n \rightarrow \varphi$ (vgl. (2.4.4a) und Bemerkung 2.4.5a)). Die Lösungen u_n von (2.4.5a) lassen sich explizit angeben:

$$u_n(\mathbf{x}) = \log(|\mathbf{x}|) / \log(1/n).$$

Offenbar gilt punktweise $u_n(\mathbf{x}) \rightarrow u(\mathbf{x}) := 0$, aber $u = 0$ erfüllt weder (2.4.4b) noch die Randwertaufgabe (2.4.5b). Im Umkehrschluss folgert man aus Satz 2.4.6a die

Bemerkung 2.4.7 Die Potentialgleichung hat in $\Omega = K_1(0) \setminus \{0\} \subset \mathbb{R}^2$ keine Lösung $u \in C^2(\Omega) \cap C^0(\bar{\Omega})$, die die Randwerte $u(\mathbf{x}) = 0$ auf $\partial K_1(0)$ und $u(\mathbf{x}) = 1$ in $\mathbf{x} = 0$ annimmt.

3

Die Poisson-Gleichung

3.1 Problemstellung

Die *Poisson-Gleichung* lautet

$$\Delta u = f \quad \text{in } \Omega \quad (3.1.1a)$$

mit vorgegebenem $f \in C^0(\Omega)$. In der physikalischen Interpretation ist f der Quellterm [z.B. die Ladungsdichte im Falle eines elektrischen Potentials u]. Zur eindeutigen Festlegung der Lösung braucht man eine Randwertvorgabe, zum Beispiel die Dirichlet¹-Bedingung

$$u = \varphi \quad (3.1.1b)$$

Definition 3.1.1 u heißt klassische Lösung der Randwertaufgabe (3.1.1a,b), falls $u \in C^2(\Omega) \cap C^0(\overline{\Omega})$ die Gleichungen (3.1.1a,b) punktweise erfüllt.

Bis zur Einführung schwacher Lösungen in §7 werden wir mit “Lösung” stets die “klassische Lösung” meinen.

Die Lösung des Randwertproblems (3.1.1a,b) wird im Allgemeinen nicht mehr die Mittelwerteigenschaft und das Maximumprinzip erfüllen. Wohl aber gelten diese Eigenschaften für Differenzen zweier Lösungen u_1 und u_2 der Poisson-Gleichung, da $\Delta(u_1 - u_2) = f - f = 0$. Damit folgt sofort die Eindeutigkeit der Aufgabe (3.1.1a,b), und der Satz 2.4.3 lässt sich übertragen.

Satz 3.1.2 Ω sei beschränkt. (a) Die Lösung von (3.1.1a) ist eindeutig bestimmt. (b) Sind u^I und u^II Lösungen der Poisson-Gleichung zu Randwerten φ^I und φ^II so gilt

$$\|u^I - u^II\|_\infty \leq \|\varphi^I - \varphi^II\|_\infty. \quad (3.1.2)$$

Beweis. (b) Der Beweis des Satzes 2.4.3 kann wörtlich wiederholt werden.

(a) Sind u^I und u^II zwei Lösungen von (3.1.1a), so zeigt (3.1.2), dass $\|u^I - u^II\|_\infty \leq \|\varphi^I - \varphi^II\|_\infty = 0$. Also $u^I = u^II$. ■

Ebenso lassen sich die Sätze 2.4.4 und 2.4.6 sinngemäß übertragen.

¹Johann Peter Gustav Lejeune Dirichlet, am 13 Februar 1805 in Düren (damals Frankreich) geboren, am 5 May 1859 in Göttingen gestorben.

3.2 Lösungsdarstellung mit der Greenschen Funktion

Lemma 3.2.1 Die Lösung von (3.1.1a) gehöre zu $C^2(\overline{\Omega})$, wobei Ω ein Normalgebiet sei. Dann gestattet u die Darstellung

$$u(\mathbf{x}) = - \int_{\Omega} \gamma(\xi, \mathbf{x}) f(\xi) d\xi + \int_{\Gamma} \left[\gamma(\xi, \mathbf{x}) \frac{\partial}{\partial n} u(\xi) - u(\xi) \frac{\partial}{\partial n_{\xi}} \gamma(\xi, \mathbf{x}) \right] d\Gamma_{\xi} \quad (3.2.1)$$

für jede Grundlösung γ aus (2.2.9).

Der Beweis ist der gleiche wie in Satz 2.2.2 bzw. Zusatz 2.2.3. Der Term $\int_{\Omega_{\varepsilon}} \gamma \Delta u d\xi$ mit $\Omega_{\varepsilon} = \Omega \setminus \overline{K_{\varepsilon}(\mathbf{x})}$ wird zu $\int_{\Omega_{\varepsilon}} \gamma f d\xi$. Da die Singularität von $\gamma(\xi, \mathbf{x})$ bei $\xi = \mathbf{x}$ integrierbar ist, konvergiert $\int_{\Omega_{\varepsilon}} \gamma f d\xi$ gegen $\int_{\Omega} \gamma f d\xi$, wenn $\varepsilon \rightarrow 0$.

Übungsaufgabe 3.2.2 a) Sei $\Omega \subset \mathbb{R}^n$ beschränkt, $\mathbf{x}_0 \in \Omega$, $f \in C^0(\overline{\Omega} \setminus \{\mathbf{x}_0\})$ und $|f(\mathbf{x})| \leq C |\mathbf{x} - \mathbf{x}_0|^{-s}$ für $s < n$. Man zeige: $\int_{\Omega} f(\mathbf{x}) d\mathbf{x}$ existiert als uneigentliches Integral.

b) Sei $\Omega \subset \mathbb{R}^n$ beschränkt, und $\mathbf{x}_0(\xi) \in \Omega$ hänge stetig von $\xi \in D$ (D kompakt) ab. $f(\mathbf{x}, \xi)$ sei stetig in $(\mathbf{x}, \xi) \in \overline{\Omega} \times D$ mit $\mathbf{x} \neq \mathbf{x}_0(\xi)$ und erfülle $|f(\mathbf{x}, \xi)| \leq C |\mathbf{x} - \mathbf{x}_0(\xi)|^{-s}$, $s < n$. Man beweise, dass $F(\xi) := \int_{\Omega} f(\mathbf{x}, \xi) d\mathbf{x}$ stetig ist: $F \in C^0(D)$.

Im Randintegral von (3.2.1) läßt sich $u(\xi)$ durch $\varphi(\xi)$ ersetzen (vgl. (3.1.1b)). Dagegen ist $\partial u / \partial n$ auf Γ unbekannt und darf auch nicht beliebig vorgegeben werden, da die Randwerte (3.1.1b) die Lösung schon eindeutig festlegen (vgl. Satz 3.1.2). Um $\int_{\Gamma} \gamma \partial u / \partial n d\xi$ zum Verschwinden zu bringen, muss man die Grundlösung so wählen, dass $\gamma(\xi, \mathbf{x}) = 0$ für $\xi \in \Gamma, \mathbf{x} \in \Omega$.

Definition 3.2.3 Eine Grundlösung g aus (2.2.9) heißt Greensche Funktion (erster Art), wenn $g(\xi, \mathbf{x}) = 0$ für alle $\xi \in \Gamma, \mathbf{x} \in \Omega$.

Die Existenz einer Greenschen Funktion hängt eng mit der Lösbarkeit der Randwertaufgabe für die Potentialgleichung zusammen:

Bemerkung 3.2.4 Die Greensche Funktion existiert, wenn für alle $\mathbf{x} \in \Omega$ die Randwertaufgabe $\Delta \Phi = 0$ in Ω und $\Phi = -s(\cdot, \mathbf{x})$ auf Γ eine Lösung $\Phi \in C^2(\overline{\Omega})$ besitzt.

Die vorhergehenden Überlegungen ergeben den

Satz 3.2.5 Ω sei Normalgebiet. Die Randwertaufgabe (3.1.1a) besitze eine Lösung $u \in C^2(\overline{\Omega})$. Die Existenz der Greenschen Funktion erster Art sei vorausgesetzt. Dann läßt sich u explizit darstellen durch

$$u(\mathbf{x}) = - \int_{\Omega} g(\xi, \mathbf{x}) f(\xi) d\xi - \int_{\Gamma} \varphi(\xi) \frac{\partial}{\partial n_{\xi}} g(\xi, \mathbf{x}) d\Gamma_{\xi}. \quad (3.2.2)$$

Im Folgenden wollen wir den Schluss umkehren. Die Existenz der Greenschen Funktion sei vorausgesetzt. Repräsentiert dann die durch Gleichung (3.2.2) definierte Funktion u die klassische Lösung der Randwertaufgabe (3.1.1a,b)? Hierzu ist insbesondere nachzuweisen, dass $u \in C^2(\Omega)$ und $\Delta u = f$. Zunächst ist noch nicht einmal offensichtlich, dass die durch Gleichung (3.2.2) definierte Funktion $u(\mathbf{x})$ stetig von \mathbf{x} abhängt, denn die Definition einer Grundlösung $\gamma(\xi, \mathbf{x})$ verlangt nicht die Stetigkeit bezüglich des zweiten Argumentes \mathbf{x} . Trotzdem ist die Greensche Funktion $g(\xi, \mathbf{x})$ bezüglich \mathbf{x} aus $C^2(\Omega \setminus \{\xi\})$, wie das folgende Resultat zeigt (vgl. [56, p. 67]).

Übungsaufgabe 3.2.6 Ω sei Normalgebiet. Die Greensche Funktion für Ω existiere, und für festes $\mathbf{y} \in \Omega$ gelte $g(\cdot, \mathbf{y}) \in C^2(\overline{\Omega} \setminus \{\mathbf{y}\})$ (schwächere Voraussetzungen sind möglich!). Man beweise

$$g(\mathbf{x}, \mathbf{y}) = g(\mathbf{y}, \mathbf{x}) \quad \text{für } \mathbf{x}, \mathbf{y} \in \Omega. \quad (3.2.3)$$

Hinweis: Man wende die Greensche Formel (2.2.5b) mit $\Omega_{\varepsilon} = \Omega \setminus [K_{\varepsilon}(\mathbf{x}') \cup K_{\varepsilon}(\mathbf{x}'')]$, $\mathbf{x}', \mathbf{x}'' \in \Omega$, $u(\mathbf{x}) := g(\mathbf{x}, \mathbf{x}')$, $v(\mathbf{x}) := g(\mathbf{x}, \mathbf{x}'')$ an und benutze (2.2.10).

Will man die Aussage des Satzes 3.2.5 umkehren, stößt man auf die überraschende Schwierigkeit, präzise Bedingungen an den Quellterm f zu stellen. Notwendig für $u \in C^2(\overline{\Omega})$ ist die natürliche Forderung $f \in C^0(\overline{\Omega})$. Diese ist aber nicht ausreichend, wie der folgende Satz zeigt, dessen Beweis am Ende dieses Abschnittes nachgetragen wird.

Satz 3.2.7 *Auch wenn der Rand Γ und die Randwerte φ hinreichend glatt sind und die Greensche Funktion existiert, gibt es Funktionen $f \in C^0(\overline{\Omega})$, denen keine Lösungen $u \in C^2(\overline{\Omega})$ entsprechen.*

Satz 3.2.7 zeigt, dass Gleichung (3.2.2) für $f \in C^0(\overline{\Omega})$ keineswegs eine klassische Lösung darzustellen braucht. Eine ausreichende Bedingung an f ist dagegen die Hölder²-Stetigkeit.

Definition 3.2.8 (Hölder-Stetigkeit) $f \in C^0(\overline{\Omega})$ heißt Hölder-stetig in $\overline{\Omega}$ zum Exponenten $\lambda \in (0, 1)$, falls eine Konstante $C = C(f)$ existiert, sodass

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq C |\mathbf{x} - \mathbf{y}|^\lambda \quad \text{für } \mathbf{x}, \mathbf{y} \in \overline{\Omega}. \quad (3.2.4a)$$

Wir schreiben $f \in C^\lambda(\overline{\Omega})$ und definieren die Norm $\|f\|_{C^\lambda(\overline{\Omega})}$ mit Hilfe der kleinsten Konstanten C , die (3.2.4a) und $|f(\mathbf{x})| \leq C$ erfüllt:

$$\|f\|_{C^\lambda(\overline{\Omega})} := \max \left\{ \sup \left\{ \frac{|f(\mathbf{x}) - f(\mathbf{y})|}{|\mathbf{x} - \mathbf{y}|^\lambda} : \mathbf{x}, \mathbf{y} \in \overline{\Omega}, \mathbf{x} \neq \mathbf{y} \right\}, \|f\|_\infty \right\} \quad (3.2.4b)$$

$f \in C^k(\overline{\Omega})$ heißt k -fach Hölder-stetig differenzierbar in $\overline{\Omega}$ (zum Exponenten λ), falls $D^\nu f \in C^\lambda(\overline{\Omega})$ für alle $|\nu| \leq k$. Dabei sind ν ein Multiindex vom Betrag $|\nu|$,

$$\nu = (\nu_1, \dots, \nu_n) \quad \text{mit } \nu_i \in \mathbb{N}_0, \quad |\nu| = \nu_1 + \dots + \nu_n, \quad (3.2.5a)$$

und D^ν ein $|\nu|$ -facher partieller Ableitungsoperator:

$$D^\nu = \frac{\partial^{|\nu|}}{\partial x_1^{\nu_1} \partial x_2^{\nu_2} \dots \partial x_n^{\nu_n}}. \quad (3.2.5b)$$

Die k -fach Hölder-stetig differenzierbaren Funktionen bilden den linearen Raum $C^{k+\lambda}(\overline{\Omega})$ mit der Norm

$$\|f\|_{C^{k+\lambda}(\overline{\Omega})} = \max \left\{ \|D^\nu f\|_{C^\lambda(\overline{\Omega})} : |\nu| \leq k \right\}. \quad (3.2.4c)$$

Wenn $s = k + \lambda$, wird auch $C^s(\overline{\Omega})$ für $C^{k+\lambda}(\overline{\Omega})$ geschrieben. Die k -fach Lipschitz-stetig differenzierbaren Funktionen $f \in C^{k,1}(\overline{\Omega})$ ergeben sich für die Wahl $\lambda = 1$ in (3.2.4a,b). Der Vollständigkeit halber sei hinzugefügt, dass

$$\|f\|_{C^k(\overline{\Omega})} = \max \{ \|D^\nu f\|_\infty : |\nu| \leq k \} \quad (3.2.4d)$$

als Norm in $C^k(\overline{\Omega})$ für ganzzahliges $k \geq 0$ gewählt wird.

Übungsaufgabe 3.2.9 a) f heißt lokal Hölder-stetig in Ω , falls für jedes $\mathbf{x} \in \Omega$ eine Umgebung $K_\varepsilon(\mathbf{x})$ existiert, sodass $f \in C^\lambda(K_\varepsilon(\mathbf{x}) \cap \Omega)$. Man beweise: Ist $\overline{\Omega}$ kompakt, so folgt $f \in C^\lambda(\overline{\Omega})$ aus der lokalen Hölder-Stetigkeit in $\overline{\Omega}$. Man formuliere und beweise entsprechende Aussagen für $C^{k+\lambda}(\overline{\Omega})$ und $C^{k,1}(\overline{\Omega})$.

b) Sei $s > 0$. Man zeige $|\mathbf{x}|^s \in C^s(\overline{K_R(0)})$ falls $s \notin \mathbb{N}$, sonst $|\mathbf{x}|^s \in C^{s-1,1}(\overline{K_R(0)})$. Hinweis: $1-t^s \leq (1-t)^s$ für $0 \leq t \leq 1$, $s \geq 0$.

Die Funktion u aus Gleichung (3.2.2) kann in $u_1 + u_2$ zerlegt werden, wobei $u_1 = -\int_\Omega g f d\xi$ und $u_2 = -\int_\Gamma \varphi \partial g / \partial n d\Gamma$. u ist Lösung der Randwertaufgabe (3.1.1a), falls wir zeigen können, dass u_1 und u_2 Lösungen von

$$\Delta u_1 = f \quad \text{in } \Omega, \quad u_1 = 0 \quad \text{auf } \Gamma, \quad \Delta u_2 = 0 \quad \text{in } \Omega, \quad u_2 = \varphi \quad \text{auf } \Gamma$$

sind.

²Otto Ludwig Hölder, geb. am 22. Dez. 1859 in Stuttgart, gest. am 29. Aug. 1937 in Leipzig

Satz 3.2.10 Falls die Greensche Funktion existiert und geeignete Bedingungen erfüllt, ist

$$u(\mathbf{x}) = - \int_{\Gamma} \varphi(\xi) \frac{\partial}{\partial n_{\xi}} g(\xi, \mathbf{x}) d\Gamma_{\xi}. \quad (3.2.6)$$

klassische Lösung von $\Delta u = 0$ in Ω und $u = \varphi$ auf Γ .

Der Beweis verläuft im Prinzip wie für Satz 2.3.9 (vgl. Leis [56, p. 69]).

Satz 3.2.11 Die Greensche Funktion $g(\cdot, \mathbf{x}) \in C^2(\Omega \setminus \{\mathbf{x}\})$, $\mathbf{x} \in \Omega$, existiere, und es sei $f \in C^{\lambda}(\overline{\Omega})$. Dann ist

$$u(\mathbf{x}) = - \int_{\Omega} f(\xi) g(\xi, \mathbf{x}) d\xi \quad (3.2.7)$$

eine klassische Lösung von $\Delta u = f$ in Ω und $u = 0$ auf Γ .

Beweis. Die Randbedingung $u(\mathbf{x}) = 0$ für $\mathbf{x} \in \Gamma$ folgt leicht aus $g(\mathbf{x}, \xi) = 0$ und (3.2.3). Die Eigenschaft $u \in C^1(\overline{\Omega})$ und die Darstellung $u_{x_i}(\mathbf{x}) = - \int_{\Omega} f(\xi) g_{x_i}(\xi, \mathbf{x}) d\xi$ ergeben sich aus der

Übungsaufgabe 3.2.12 Seien $\Omega \subset \mathbb{R}^n$ beschränkt und $A := \{(\xi, \mathbf{x}) \in \overline{\Omega} \times \overline{\Omega} : \xi \neq \mathbf{x}\}$. Für die Ableitungen von f nach \mathbf{x} gelte $D_{\mathbf{x}}^{\nu} f \in C^0(A)$ und $|D_{\mathbf{x}}^{\nu} f(\xi, \mathbf{x})| \leq C |\mathbf{x} - \xi|^{-s}$ mit $s < n$ für alle $|\nu| \leq k$. Man beweise $F(\mathbf{x}) := \int_{\Omega} f(\xi, \mathbf{x}) d\xi \in C^k(\overline{\Omega})$ und $D^{\nu} F(\mathbf{x}) = \int_{\Omega} D_{\mathbf{x}}^{\nu} f(\xi, \mathbf{x}) d\xi$, $|\nu| \leq k$.

Zum Beweis von $u \in C^2(\overline{\Omega})$ lässt sich dieser Schluss nicht wiederholen, denn $g_{x_i x_j} = \mathcal{O}(|\xi - \mathbf{x}|^{-n})$ hat eine nicht-integrierbare Singularität. Wir schreiben die erste Ableitung in der Form

$$u_{x_i}(\mathbf{x}) = - \int_{\Omega} [f(\xi) - f(\mathbf{x})] g_{x_i}(\xi, \mathbf{x}) d\xi - f(\mathbf{x}) \int_{\Omega} g_{x_i}(\xi, \mathbf{x}) d\xi. \quad (3.2.8a)$$

Sei $\partial F(\mathbf{x})$ der Differenzenquotient $(F(\mathbf{x}^{\varepsilon}) - F(\mathbf{x}))/\varepsilon$ mit $x_j^{\varepsilon} = x_j + \varepsilon$ und $x_i^{\varepsilon} = x_i$ ($i \neq j$). Die Produktregel $\partial_j(FG)(\mathbf{x}) = G(\mathbf{x})\partial_j F(\mathbf{x}) + F(\mathbf{x}^{\varepsilon})\partial_j G(\mathbf{x})$ angewandt auf Gleichung (3.2.8a) liefert

$$\partial_j u_{x_i}(\mathbf{x}) = - \int_{\Omega} [f(\xi) - f(\mathbf{x}^{\varepsilon})] \partial_j g_{x_i}(\xi, \mathbf{x}) d\xi - f(\mathbf{x}^{\varepsilon}) \partial_j \frac{\partial}{\partial x_i} \int_{\Omega} g(\xi, \mathbf{x}) d\xi.$$

Da $[f(\xi) - f(\mathbf{x})]g_{x_i x_j}(\xi, \mathbf{x}) = \mathcal{O}(|\xi - \mathbf{x}|^{\lambda-n})$ integrierbar ist, ergibt der Grenzprozess $\varepsilon \rightarrow 0$ die Darstellung

$$u_{x_i x_j}(\mathbf{x}) = - \int_{\Omega} [f(\xi) - f(\mathbf{x})] g_{x_i x_j}(\xi, \mathbf{x}) d\xi - f(\mathbf{x}) \partial_j \frac{\partial^2}{\partial x_i \partial x_j} \int_{\Omega} g(\xi, \mathbf{x}) d\xi. \quad (3.2.8b)$$

Gleichung (3.2.8b) impliziert $\Delta u = \int_{\Omega} [f(\xi) - f(\mathbf{x})] \Delta g d\xi - f(\mathbf{x}) \Delta \int_{\Omega} g d\xi = -f(\mathbf{x}) \Delta \int_{\Omega} g(\xi, \mathbf{x}) d\xi$, sodass $\Delta \int_{\Omega} g d\xi = -1$ zu zeigen bleibt. Man wähle $K_R(\mathbf{z})$ so, dass $\mathbf{x} \in K_R(\mathbf{z}) \subset \Omega$. Die Greensche Funktion hat die Form (2.2.9): $g = s + \Phi$. Die ersten beiden Terme in

$$\int_{\Omega} g(\xi, \mathbf{x}) d\xi = \int_{\Omega \setminus K_R(\mathbf{z})} g(\xi, \mathbf{x}) d\xi + \int_{K_R(\mathbf{z})} \Phi(\xi, \mathbf{x}) d\xi + \int_{K_R(\mathbf{z})} s(\xi, \mathbf{x}) d\xi$$

sind harmonisch in $K_R(\mathbf{z})$, sodass $\Delta \int_{K_R(\mathbf{z})} s(\xi, \mathbf{x}) d\xi = -1$ zu beweisen ist. $\sigma(r)$ sei durch $s(\zeta, \mathbf{x}) = \sigma(|\zeta - \mathbf{x}|)$ erklärt (vgl. (2.2.1)). Für festes $r > 0$ sei

$$v(\mathbf{x}) := \frac{1}{\omega_n r^{n-1}} \int_{\partial K_r(\mathbf{z})} s(\xi, \mathbf{x}) d\Gamma_{\xi} \quad (3.2.8c)$$

gesetzt. Für alle $\mathbf{x} \notin \partial K_r(\mathbf{z})$ (d.h. $|\mathbf{x} - \mathbf{z}| \neq r$) ist v harmonisch, weil $s(\zeta, \mathbf{x})$ auf $\partial K_r(\mathbf{z})$ nicht-singulär ist und $\Delta_{\mathbf{x}} s(\zeta, \mathbf{x}) = 0$ erfüllt. Da $s(\cdot, \mathbf{x})$ in $K_r(\mathbf{z})$ für $r < |\mathbf{z} - \mathbf{x}|$ harmonisch ist, gilt die Mittelwerteigenschaft (2.3.1), die sich jetzt als

$$v(\mathbf{x}) = s(\mathbf{z}, \mathbf{x}) = \sigma(|\mathbf{z} - \mathbf{x}|) \quad \text{für } |\mathbf{z} - \mathbf{x}| > r \quad (3.2.8d)$$

schreibt. Gemäß Übungsaufgabe 3.2.2b ist $v(\mathbf{x})$ stetig in \mathbb{R}^n , sodass auch

$$v(\mathbf{x}) = \sigma(r) \quad \text{für } |\mathbf{z} - \mathbf{x}| = r \quad (3.2.8e)$$

gilt. v ist in $K_r(\mathbf{x})$ harmonisch mit den konstanten Randwerten (3.2.8e). Die eindeutige Lösung lautet daher

$$v(\mathbf{x}) = \sigma(r) \quad \text{für } |\mathbf{z} - \mathbf{x}| \leq r \quad (3.2.8f)$$

Die Gleichungen (3.2.8c,d,f) ergeben

$$\int_{\partial K_r(\mathbf{z})} s(\xi, \mathbf{x}) d\Gamma_\xi = \omega_n r^{n-1} \sigma(\max\{r, |\mathbf{z} - \mathbf{x}|\})$$

und wegen $0 < |\mathbf{z} - \mathbf{x}| < R$

$$\begin{aligned} \int_{K_r(\mathbf{z})} s(\xi, \mathbf{x}) d\xi &= \int_0^R \int_{\partial K_r(\mathbf{z})} s(\xi, \mathbf{x}) d\Gamma_\xi dr \\ &= \omega_n \int_0^{|\mathbf{z}-\mathbf{x}|} r^{n-1} \sigma(r) dr + \omega_n \int_{|\mathbf{z}-\mathbf{x}|}^R r^{n-1} \sigma(r) dr \\ &= \omega_n \frac{|\mathbf{z}-\mathbf{x}|^n}{n} \sigma(|\mathbf{z}-\mathbf{x}|) + \omega_n \frac{r^n}{n} \sigma(r) \Big|_{|\mathbf{z}-\mathbf{x}|}^R - \omega_n \int_{|\mathbf{z}-\mathbf{x}|}^R \frac{r^n}{n} \sigma'(r) dr \\ &= \frac{\omega_n}{n} R^n \sigma(R) - \omega_n \int_{|\mathbf{z}-\mathbf{x}|}^R \frac{r^n}{n} \left(-\frac{r^{1-n}}{\omega_n} \right) dr = \frac{\omega_n}{n} R^n \sigma(R) + \int_{|\mathbf{z}-\mathbf{x}|}^R \frac{r}{n} dr \\ &= \frac{\omega_n}{n} R^n \sigma(R) + \frac{R^2}{2n} - \frac{|\mathbf{z}-\mathbf{x}|^2}{2n}. \end{aligned}$$

Unabhängig von R , n und \mathbf{z} ergibt sich hieraus $\Delta \int_{K_R(\mathbf{z})} s(\zeta, \mathbf{x}) d\zeta = -1$. ■

Die Sätze 3.2.10 und 3.2.11 ergeben den

Satz 3.2.13 *Unter den Voraussetzungen der Sätze 3.2.10 und 3.2.11 stellt (3.2.2) die klassische Lösung der Randwertaufgabe (3.1.1a) dar.*

Abschließend seien zwei Ungleichungen für die Greensche Funktion als Übung gestellt:

Übungsaufgabe 3.2.14 *a) In Ω möge die Greensche Funktionen g existieren. Man zeige*

$$0 \leq g(\mathbf{x}, \mathbf{y}) \leq s(\mathbf{x}, \mathbf{y}) \quad \text{für } \mathbf{x}, \mathbf{y} \in \Omega \subset \mathbb{R}^d, \quad d \geq 3. \quad (3.2.9)$$

Welche Ungleichung entsteht für $n = 2$?

b) Seien g_1, g_2 Greensche Funktionen in $\Omega_1 \subset \Omega_2$. Man beweise

$$g_1(\mathbf{x}, \mathbf{y}) \leq g_2(\mathbf{x}, \mathbf{y}) \quad \text{für } \mathbf{x}, \mathbf{y} \in \Omega_1 \subset \Omega_2. \quad (3.2.10)$$

Hinweis: Übungsaufgabe 2.3.14.

Beweis des Satzes 3.2.7. Aufgrund eines späteren Satzes (Satz 6.1.13) folgt Satz 3.2.7 aus Satz 3.2.15: ■

Satz 3.2.15 *Die Lösung u hängt nicht stetig von f ab, wenn die Supremumsnorm (2.4.1) als Norm in $C^0(\overline{\Omega})$ und jene aus (3.2.4d) als Norm in $C^2(\overline{\Omega})$ verwendet werden.*

Beweis. Sei $\Omega = K_1(0) \subset \mathbb{R}^2$ und $\varphi = 0$. Der Kreis Ω ist ein Normalgebiet, für das die Greensche Funktion bekannt ist (vgl. Satz 3.3.1). Gemäß Satz 3.2.11 existieren Lösungen $u^n \in C^2(\overline{\Omega})$ von $\Delta u^n = f_n$ in Ω , $u^n = 0$ auf Γ für die Funktionen

$$f_n(\mathbf{x}) = \frac{x_2^2 - x_1^2}{r^2} \rho_n(r), \quad r := |\mathbf{x}|, \quad \rho_n(r) := \min \left\{ nr, 1 / \left| \log \frac{r}{2} \right| \right\},$$

die zu $C^0(\overline{\Omega})$ gehören und gleichmäßig beschränkt sind: $\|f_n\|_\infty = 1/\log 2$. Nach Satz 3.2.11 gilt $u^n(\mathbf{x}) = - \int_\Omega g(\xi, \mathbf{x}) f_n(\xi) d\xi$. Da $|f_n(\xi)| \leq n |\xi|$, folgt aus Übungsaufgabe 3.2.12

$$u_{x_1 x_1}^n(\mathbf{x}) = - \int_\Omega g_{x_1 x_1}(\xi, \mathbf{x}) f_n(\xi) d\xi = - \int_\Omega \Phi_{x_1 x_1}(\xi, \mathbf{x}) f_n(\xi) d\xi - \int_\Omega s_{x_1 x_1}(\xi, \mathbf{x}) f_n(\xi) d\xi \quad \text{in } \mathbf{x} = 0,$$

wobei $g = \Phi + s$. Das erste Integral ist beschränkt, weil $\Phi \in C^2(\overline{\Omega})$. Die Ableitung der Singularitätenfunktion lautet $s_{x_1 x_1}(\xi, 0) = (\xi_1^2 - \xi_2^2)/|\xi|^4$. Die spezielle Wahl von f_n liefert für $\mathbf{x} = 0$

$$u_{x_1 x_1}^n(0) = - \int_{\Omega} \Phi_{x_1 x_1}(\xi, 0) f_n(\xi) d\xi - \int_{\Omega} [\xi_1^2 - \xi_2^2]^2 |\xi|^{-6} \rho_n(|\xi|) d\xi.$$

Das Oberflächenintegral $K := \int_{\partial K_r(0)} [\xi_1^2 - \xi_2^2]^2 |\xi|^{-5} d\xi > 0$ ist von $r \in (0, 1]$ unabhängig, sodass das zweite Integral die Form $I_n := K \int_0^1 r^{-1} \rho_n(r) dr$ annimmt. Da $\int_{\varepsilon}^1 [r \log(\frac{r}{\varepsilon})]^{-1} dr$ für $\varepsilon \rightarrow 0$ divergiert, erhalten wir $I_n \rightarrow \infty$ für $n \rightarrow \infty$. Wegen $\|u^n\|_{C^2(\overline{\Omega})} \geq |u_{x_1 x_1}^n(0)|$, ist die Abbildung $f \mapsto u$ nicht beschränkt, also nicht stetig. ■

3.3 Die Greensche Funktion für die Kugel

Satz 3.3.1 Die Greensche Funktion für die Kugel $K_R(\mathbf{y})$ ist durch die Funktion aus (2.2.11a) gegeben. Für $f \in C^\lambda(\overline{\Omega})$ und $\varphi \in C^{2+\lambda}(\Gamma)$ mit $0 < \lambda < 1$ definiert die Darstellungsformel (3.2.2) eine Lösung $u \in C^{2+\lambda}(\overline{\Omega})$ der Randwertaufgabe $\Delta u = f$ in Ω , $u = \varphi$ auf Γ .

Den Beweis des Satzes liefert ein Resultat von Schauder³, das im Satz 9.1.24 zitiert wird.

Im Falle $n = 2$ lässt sich der \mathbb{R}^2 mit \mathbb{C} vermöge $(x, y) \leftrightarrow z = x + iy$ identifizieren. Die folgenden Überlegungen basieren auf

Übungsaufgabe 3.3.2 Die Abbildung $\Phi : z = x + iy \in \Omega \mapsto \zeta = \xi + i\eta = \Phi(z) \in \Omega'$ sei holomorph. Man zeige

$$\Delta_z u(\Phi(z)) = |\Phi'|^2 \Delta_\zeta u(\zeta), \quad \Phi' = \xi_x + i\eta_x \quad (3.3.1)$$

für $u \in C^2(\Omega')$.

Gleichung (3.3.1) zeigt insbesondere, dass eine holomorphe Koordinatentransformation harmonische Funktionen wieder in harmonische Funktionen abbildet. Ein beliebiges, einfach zusammenhängendes Gebiet mit mindestens zwei Randpunkten lässt sich dank des Riemannschen Abbildungssatzes durch eine konforme Abbildung $\Phi_{z_0} : z \in \Omega \mapsto \Phi_{z_0}(z) \in K_1(0)$ so auf den Einheitskreis abbilden, dass $\Phi_{z_0}(z_0) = 0$ für jedes vorgegebene $z_0 \in \Omega$. Sei $g(\zeta, \zeta')$ die Greensche Funktion in $K_1(0)$. Man prüft nach, dass $G(z, z_0) := g(\Phi_{z_0}(z), 0)$ wieder Grundlösung ist. $z \in \partial\Omega$ impliziert $\Phi_{z_0}(z) \in \partial K_1(0)$, d.h. $G(z, z_0) = 0$. Also ist $G(z, z_0)$ Greensche Funktion in Ω . Dies beweist

Satz 3.3.3 Sei $\Omega \subset \mathbb{R}^2$ einfach zusammenhängend mit mindestens zwei Randpunkten. Dann existiert die Greensche Funktion erster Art für Ω .

Die explizite Darstellung verschiedener Greenscher Funktionen findet man z.B. bei [89, p. 348]. Von numerischem Interesse ist gelegentlich die Tatsache, dass man durch konforme Abbildungen störende (z.B. einspringende) Ecken in Ω beseitigen kann (vgl. [33, p. 70]).

Beispiel 3.3.4 Sei Ω das L -Gebiet aus Beispiel 2.1.4. Wähle $\Phi(z) = z^{2/3} : \Omega \mapsto \Omega'$. Φ ist konform in Ω . Die Winkelseiten $\Gamma_0 \subset \partial\Omega$ (vgl. Abbildung 2.1.1) werden in eine einzige Strecke abgebildet, sodass Ω' keine einspringende Ecke mehr hat. Der Poisson-Gleichung $\Delta u = f$ in Ω entspricht die Gleichung $\Delta v(\zeta) = \frac{9}{4} |\zeta| f(\zeta)$ in Ω' .

3.4 Die Neumann-Randwertaufgabe

In (3.1.1b) wie auch (2.1.1b) wurden die Randwerte $u = \varphi$ auf Γ vorgeschrieben. Diese sogenannte Dirichlet-Bedingung oder “Randbedingung erster Art” ist nicht die einzige Möglichkeit. Eine Alternative ist die Neumann⁴-Bedingung

$$\frac{\partial}{\partial n} u(\mathbf{x}) = \varphi(\mathbf{x}) \quad \text{auf } \Gamma. \quad (3.4.1)$$

³Juliusz Pawel Schauder, geb. am 21. Sept. 1899 in Lvov (Galizien, Österreich), gest. im Sept. 1943 in Lvov.

⁴Carl Gottfried Neumann, am 7. Mai 1832 in Königsberg geboren und am 27. März 1925 in Leipzig gestorben.

In der Physik tritt diese zweite Randbedingung, wie sie auch genannt wird, häufiger auf als die Dirichlet-Bedingung. Ist z.B. u das Geschwindigkeitspotential eines Gases, so bedeutet $\partial u / \partial n = 0$, dass sich das Gas am Rande γ nur tangential bewegen kann (d.h. die Wand nicht durchdringen kann).

Bis auf Ausnahmefälle ist die Randwertaufgabe $Lu = f$ in Ω und $\partial u / \partial n = \varphi$ auf Γ eindeutig lösbar. Ein Ausnahmefall liegt aber gerade für $L = \Delta$ vor:

Satz 3.4.1 Ω sei Normalgebiet. Die Poisson-Gleichung $\Delta u = f$ mit der Neumann-Randbedingung (3.4.1) kann nur lösbar sein, wenn

$$\int_{\Gamma} \varphi(\xi) d\Gamma_{\xi} = \int_{\Omega} f(\mathbf{x}) d\mathbf{x}. \quad (3.4.2)$$

Existiert eine Lösung u , so ist auch $u + c$ für jede Konstante $c \in \mathbb{C}$ eine Lösung.

Beweis. i) Man wiederhole den Beweis von Lemma 2.3.6 für $\Delta u = f$.

ii) Offenbar erfüllt $u + c$ ebenfalls die Gleichungen. ■

Später (in Beispiel 7.4.8) werden wir zeigen, dass die Neumann-Randwertaufgabe für die Poisson-Gleichung genau dann eine Lösung besitzt, wenn (3.4.2) erfüllt ist, und dass sich zwei Lösungen nur um eine Konstante unterscheiden können.

In der Darstellung (3.2.1) treten sowohl die Werte $u(\xi)$, $\xi \in \Gamma$, als auch die Normalenableitung $\partial u / \partial n$ auf. Die Greensche Funktion erster Art war so gewählt, dass $g(\xi, \mathbf{x}) = 0$ für $\xi \in \Gamma$. Im Falle der zweiten Randbedingung (3.4.1) macht man den Ansatz $\partial \gamma(\xi, \mathbf{x}) / \partial n_{\xi} = c$ (c : Konstante), d.h.

$$\frac{\partial}{\partial n_{\xi}} \Phi(\xi, \mathbf{x}) = c - \frac{\partial}{\partial n_{\xi}} s(\xi, \mathbf{x})$$

für $\Phi = \gamma - s$. Zusatz 2.2.3 mit $u \equiv 1$ und $\gamma = s$ liefert

$$\int_{\Gamma} \frac{\partial}{\partial n_{\xi}} \Phi(\xi, \mathbf{x}) d\Gamma_{\xi} = cL + 1 \quad \text{mit } L := \int_{\Gamma} d\Gamma.$$

Da Φ harmonisch sein muss (d.h. $f := \Delta \Phi = 0$), ist gemäß Gleichung (3.4.2) $cL + 1 = 0$ eine notwendige Bedingung für die Existenz von Φ . Damit lautet die Bedingung an die Greensche Funktion zweiter Art für die Potentialgleichung

$$\frac{\partial}{\partial n_x} \gamma(\xi, \mathbf{x}) = -1 / \int_{\Gamma} d\Gamma.$$

Damit wird der Term $\int_{\Gamma} u \partial \gamma / \partial n d\Gamma$ in (3.2.1) zu $\text{const} \cdot \int_{\Gamma} u d\Gamma$. Da u nur bis auf eine Konstante bestimmt ist (vgl. Satz 3.4.1), kann man diese Konstante durch die zusätzliche Bedingung $\int_{\Gamma} u d\Gamma = 0$ festlegen. Damit ergibt sich folgendes Resultat, wenn wir g für γ schreiben:

$$u(\mathbf{x}) = - \int_{\Omega} f(\xi) g(\xi, \mathbf{x}) d\xi + \int_{\Gamma} \varphi(\xi) g(\xi, \mathbf{x}) d\Gamma_{\xi}.$$

Die Greensche Funktion zweiter Art für die Kugel $K_R(0) \subset \mathbb{R}^3$ ist bei Leis [56, p. 79] angegeben.

3.5 Die Integralgleichungsmethode

In der Darstellung (3.2.1) der Poisson-Lösung kann die Singularitätenfunktion s insbesondere als γ gewählt werden. Setzt man außerdem die vorgegebenen Neumann-Daten φ aus (3.4.1) ein, erhält man

$$u(\mathbf{x}) = - \int_{\Gamma} k(\mathbf{x}, \xi) u(\xi) d\Gamma_{\xi} + g(\mathbf{x}) \quad \text{für } \mathbf{x} \in \Omega \quad (3.5.1)$$

mit der Kernfunktion $k(\mathbf{x}, \xi) := \partial s(\xi, \mathbf{x}) / \partial n_{\xi}$ und der folgenden Funktion g :

$$g(\mathbf{x}) := g_1(\mathbf{x}) + g_2(\mathbf{x}), \quad g_1(\mathbf{x}) := \int_{\Gamma} s(\mathbf{x}, \xi) \varphi(\xi) d\Gamma_{\xi}, \quad g_2(\mathbf{x}) := \int_{\Omega} s(\xi, \mathbf{x}) f(\xi) d\xi.$$

Die rechte Seite von Gleichung (3.5.1) mit der noch unbekannten Randbelegung $u(\xi)$, $\xi \in \Gamma$, kann man als Lösungsansatz verwenden:

$$\Phi(\mathbf{x}) = - \int_{\Gamma} k(\mathbf{x}, \xi) u(\xi) d\Gamma_{\xi} + g(\mathbf{x}). \quad (3.5.2)$$

Der erste Summand der rechten Seite in (3.5.2) heißt *Doppelschichtpotential* (Dipolpotential), g_1 ist ein *Einfachschichtpotential*, während g_2 ein *Volumenpotential* darstellt.

Für jedes $u \in C^0(\Gamma)$ ist Φ aus (3.5.2) eine Lösung der Poisson-Gleichung (3.1.1a) in Ω . Φ ist aber auch für Argumente \mathbf{x} im Außenraumgebiet $\mathbb{R}^n \setminus \Omega$ definiert. Eine genauere Untersuchung der Kernfunktion $k(\mathbf{x}, \zeta)$ zeigt, dass diese für glatte Ränder Γ nur schwach singular ist. Damit ist Φ auch für $\mathbf{x} \in \Gamma$ erklärt. Die nunmehr auf dem gesamten \mathbb{R}^n definierte Funktion Φ ist in den Punkten des Randes unstetig. In $\mathbf{x}_0 \in \Gamma$ existieren sowohl der innere Grenzwert $\Phi_{-}(\mathbf{x}_0)$ für $\Omega \ni \mathbf{x} \rightarrow \mathbf{x}_0$ als auch der äußere Grenzwert $\Phi_{+}(\mathbf{x}_0)$ für $\mathbb{R}^n \setminus \Omega \ni \mathbf{x} \rightarrow \mathbf{x}_0$. Als drittes existiert der Funktionswert $\Phi(\mathbf{x}_0)$ von (3.5.2). Ihr Zusammenhang ist durch die folgenden Sprungrelationen gegeben (vgl. Hackbusch [44, Satz 8.2.8]):

$$\Phi_{+}(\mathbf{x}_0) - \Phi_{-}(\mathbf{x}_0) = -u(\mathbf{x}_0) \quad \text{für } \mathbf{x}_0 \in \Gamma, \quad (3.5.3a)$$

$$\Phi_{+}(\mathbf{x}_0) + \Phi_{-}(\mathbf{x}_0) = 2\Phi(\mathbf{x}_0) \quad \text{für } \mathbf{x}_0 \in \Gamma, \quad (3.5.3b)$$

Damit der Ansatz (3.5.2) die Lösung u aus (3.5.1) ergibt, müssen die (aus dem Inneren fortgesetzten) Randwerte Φ_{-} mit der unter dem Integral eingesetzten Funktion u übereinstimmen: $\Phi_{-} = u$. Die Gleichungen (3.5.3a,b) löst man nach Φ_{-} auf: $\Phi_{-}(\mathbf{x}_0) = \Phi(\mathbf{x}_0) + \frac{1}{2}u(\mathbf{x}_0)$. Die Randbedingung $\Phi_{-} = u$ führt somit auf

$$u(\mathbf{x}_0) = 2\Phi(\mathbf{x}_0) = -2 \int_{\Gamma} k(\mathbf{x}, \xi) u(\xi) d\Gamma_{\xi} + 2g(\mathbf{x}) \quad \text{für } \mathbf{x}_0 \in \Gamma. \quad (3.5.4)$$

Gleichung (3.5.4) nennt man eine Fredholmsche Integralgleichung zweiter Art für die unbekannte Funktion $u \in C^0(\Gamma)$. Das ursprüngliche Neumann-Randwertproblem (3.1.1a), (3.4.1) und die Integralgleichung (3.5.4) sind in folgendem Sinne äquivalent:

- (a) Ist u die Lösung des Neumann-Randwertproblems, so erfüllen die Randwerte $u(\zeta)$, $\zeta \in \Gamma$, die Integralgleichung (3.5.4).
- (b) Ist $u \in C^0(\Gamma)$ eine Lösung der Integralgleichung (3.5.4), so liefert die Darstellung (3.5.1) die Lösung von (3.1.1a), (3.4.1) im gesamten Gebiet Ω .

Die Integralgleichung (3.5.4) ist nur ein Beispiel mehrerer möglicher Integralgleichungsformulierungen, die sich hinsichtlich der Kernfunktion $k(\mathbf{x}, \xi)$ unterscheiden. Da k von der Singularitätenfunktion abgeleitet wird, ist diese Methode auf alle Differentialgleichungen mit bekannter Singularitätenfunktion anwendbar.

Die Überführung einer Randwertaufgabe in eine Integralgleichung und die anschließende Lösung der Integralgleichung nennt man *Integralgleichungsmethode*. Sie erlaubt zum Beispiel einen neuen Zugang zu Existenzaussagen, indem man die Lösbarkeit von (3.5.4) nachweist.

Zur numerischen Lösung der Integralgleichung (3.5.4) gibt es mehrere Zugänge. Wie bei der Finite-Element-Methode (siehe §8) ersetzt man u durch eine Ansatzfunktion eines n -dimensionalen Raumes V_n , z.B. durch eine stückweise lineare Funktion, die über einer Zerlegung der Oberfläche Γ in Dreiecksfacetten definiert ist. Fordert man dann die Gleichheit (3.5.4) für alle \mathbf{x}_0 , die den n Ecken der Dreiecke entsprechen, erhält man die Kollokationsmethode. Integriert man dagegen die Gleichung (3.5.4) mit den Basisfunktionen $b_j \in V_n$, ergibt sich die Galerkin-Methode, die direkt dem Vorgehen in §8 entspricht. Beide Diskretisierungsverfahren werden unter dem Namen "Randelementmethode" (oft abgekürzt als BEM für "boundary element method") zusammengefasst.

Anders als bei der Finite-Element-Methode in §8 hat man bei der Randelementmethode kompliziertere Integrationen durchzuführen (Oberflächen als Integrationsbereich, singuläre Integranden), die aber trotzdem numerisch effizient durchführbar sind.

Literatur zur Integralgleichungs- und Randelementmethode findet man z.B. bei Hackbusch [44, §§7-9], Kreß [54] und Sauter-Schwab [72].

4

Differenzenmethode für die Poisson-Gleichung

4.1 Einführung: Der eindimensionale Fall

Bevor Differenzenverfahren für die Poisson-Gleichung entwickelt werden, wollen wir zunächst an die Diskretisierung gewöhnlicher Differentialgleichungen erinnern. Die Differentialgleichung zweiter Ordnung

$$a(x)u''(x) + b(x)u'(x) + c(x)u(x) = f(x)$$

kann durch Anfangswerte $u(x_1) = u_1$, $u'(x_1) = u'_1$ oder Randwerte $u(x_1) = u_1$, $u(x_2) = u_2$ ergänzt werden. Die gewöhnlichen Anfangswertaufgaben entsprechen den hyperbolischen und parabolischen Anfangswertproblemen, während man ein gewöhnliches Randwertproblem als elliptische Randwertaufgabe in einer Variablen auffassen kann. Insbesondere ist

$$-u''(x) = f(x) \quad \text{in } x \in (0, 1), \quad (4.1.1a)$$

$$u(0) = \varphi_0, \quad u(1) = \varphi_1, \quad (4.1.1b)$$

als die eindimensionale Poisson-Gleichung $-\Delta u = f$ im Gebiet $\Omega = (0, 1)$ mit Dirichlet-Werten auf dem Rand $\Gamma = \{0, 1\}$ anzusehen.

Die Differenzenverfahren sind dadurch charakterisiert, dass Ableitungen durch Differenzenquotienten (dividierte Differenzen) ersetzt werden, die im Folgenden nur “Differenzen” genannt werden. Die erste Ableitung $u'(x)$ kann durch verschiedene (sogenannte “erste”) Differenzen approximiert werden, z.B. durch

- die *Vorwärts- oder rechtsseitige Differenz*

$$(\partial^+ u)(x) := [u(x+h) - u(x)]/h, \quad (4.1.2a)$$

- die *Rückwärts- oder linksseitige Differenz*

$$(\partial^- u)(x) := [u(x) - u(x-h)]/h, \quad (4.1.2b)$$

(rechts- und linksseitige Differenzen werden gemeinsam als *einseitige Differenzen* bezeichnet),

- die *symmetrische Differenz*

$$(\partial^0 u)(x) := [u(x+h) - u(x-h)]/(2h), \quad (4.1.2c)$$

wobei $h > 0$ die *Schrittweite* genannt wird.

Eine naheliegende *zweite Differenz* für $u''(x)$ ist

$$(\partial^- \partial^+ u)(x) := [u(x+h) - 2u(x) + u(x-h)]/h^2. \quad (4.1.2d)$$

Man bezeichnet $\partial^+, \partial^-, \partial^0, \partial^+ \partial^-$ auch als Differenzenoperatoren. $\partial^- \partial^+$ lässt sich als Produkt $\partial^- \circ \partial^+$ oder auch $\partial^+ \circ \partial^-$ deuten, d.h. $(\partial^+ \partial^-)u(x) = \partial^+(\partial^- u(x))$.

Lemma 4.1.1 Sei $[x-h, x+h] \subset \overline{\Omega}$. Es gilt

$$(\partial^+ u)(x) = u'(x) + hR \quad \text{mit } |R| \leq \frac{1}{2} \|u\|_{C^2(\overline{\Omega})}, \quad \text{falls } u \in C^2(\overline{\Omega}), \quad (4.1.3a)$$

$$(\partial^0 u)(x) = u'(x) + h^2 R \quad \text{mit } |R| \leq \frac{1}{6} \|u\|_{C^3(\overline{\Omega})}, \quad \text{falls } u \in C^3(\overline{\Omega}), \quad (4.1.3b)$$

$$(\partial^- \partial^+ u)(x) = u''(x) + h^2 R \quad \text{mit } |R| \leq \frac{1}{12} \|u\|_{C^4(\overline{\Omega})}, \quad \text{falls } u \in C^4(\overline{\Omega}), \quad (4.1.3c)$$

Beweis. Der Beweis wird nur für (4.1.3c) vorgeführt. Setzt man die Taylor¹-Formel

$$u(x \pm h) = u(x) \pm hu'(x) + h^2 u''(x)/2 \pm h^3 u'''(x)/6 + h^4 R_4 \quad (4.1.4a)$$

$$\text{mit } R_4 = \frac{h^{-4}}{3!} \int_x^{x+h} u''''(\xi)(x \pm h - \xi)^3 d\xi = \frac{1}{4!} u''''(\xi \pm \vartheta h) \quad (4.1.4b)$$

und $\vartheta \in (0, 1)$ in Gleichung (4.1.2d) ein, folgt (4.1.3c) wegen $R = [u''''(x + \vartheta_1 h) + u''''(x - \vartheta_2 h)]/24$. ■

Wir ersetzen $\Omega = (0, 1)$ und $\overline{\Omega} = [0, 1]$ durch die *Gitter*

$$\Omega_h = \{h, 2h, \dots, (n-1)h = 1-h\}, \quad (4.1.5a)$$

$$\overline{\Omega}_h = \{0, h, 2h, \dots, 1-h, 1\} \quad (4.1.5b)$$

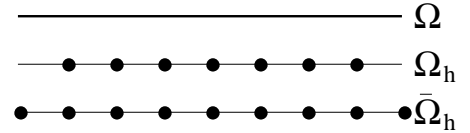


Abb. 4.1.1. Gitter für $h = 1/8$

der *Schrittweite* $h = 1/n$ (vgl. Abbildung 4.1.1). Für $x \in \Omega_h$ enthält $\partial^- \partial^+ u(x)$ nur die Werte von u bei $x, x \pm h \in \Omega_h$. Unter der Voraussetzung, dass die Lösung u von Gleichung (4.1.1a,b) zu $C^4(\overline{\Omega})$ gehört, liefert (4.1.3c) die Gleichungen

$$-\partial^- \partial^+ u(x) = f(x) + \mathcal{O}(h^2) \quad \text{für } x \in \Omega_h. \quad (4.1.6)$$

Streicht man in Gleichung (4.1.6) den Restterm $\mathcal{O}(h^2)$, ergibt sich

$$-\partial^- \partial^+ u_h(x) = h^{-2} [-u_h(x-h) + 2u_h(x) - u_h(x+h)] = f(x) \quad \text{für } x \in \Omega_h. \quad (4.1.7a)$$

Diesen $n-1$ Gleichungen stehen $n+1$ Unbekannte $\{u_n(x) : x \in \overline{\Omega}_h\}$ gegenüber. Die fehlenden zwei Gleichungen ergeben sich aus der Randbedingung (4.1.1b):

$$u_h(0) = \varphi_0, \quad u_h(1) = \varphi_1. \quad (4.1.7b)$$

u_h ist eine auf $\overline{\Omega}_h$ definierte *Gitterfunktion*. Ihre Einschränkung auf Ω_h ergibt den Vektor

$$u_h = \begin{pmatrix} u_h(h) \\ u_h(2h) \\ \vdots \\ u_h(1-h) \end{pmatrix}.$$

Eliminiert man in (4.1.7a) die Komponenten $u_h(0)$ und $u_h(1)$ mit Hilfe von Gleichung (4.1.7b), bleibt das Gleichungssystem

$$L_h u_h = q_h \quad (4.1.8a)$$

¹Brook Taylor, geb. 18. Aug. 1685 in Edmonton, Middlesex, England, gest. 29. Dez. 1731 in Somerset House, London

mit

$$L_h = h^{-2} \begin{bmatrix} 2 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 2 \end{bmatrix}, \quad q_h = \begin{bmatrix} f(h) + h^{-2}\varphi_0 \\ f(2h) \\ f(3h) \\ \vdots \\ f(1-2h) \\ f(1-h) + h^{-2}\varphi_1 \end{bmatrix}. \quad (4.1.8b)$$

4.2 Fünfpunktformel

Wir wählen zunächst das Einheitsquadrat

$$\Omega = \{(x, y) : 0 < x < 1, 0 < y < 1\}$$

als zugrundeliegendes Gebiet. Allgemeinere Gebiete werden in §4.8 diskutiert werden. Bei der Diskretisierung wird Ω ersetzt durch das Gitter

$$\Omega_h = \{(x, y) \in \Omega : x/h, y/h \in \mathbb{Z}\} \quad (4.2.1a)$$

zur äquidistanten² Schrittweite $h = 1/n$ ($n \in \mathbb{N}$) (vgl. Abbildung 4.2.1). Die diskreten Randpunkte bilden die Menge

$$\Gamma_h = \{(x, y) \in \Gamma : x/h, y/h \in \mathbb{Z}\}. \quad (4.2.1b)$$

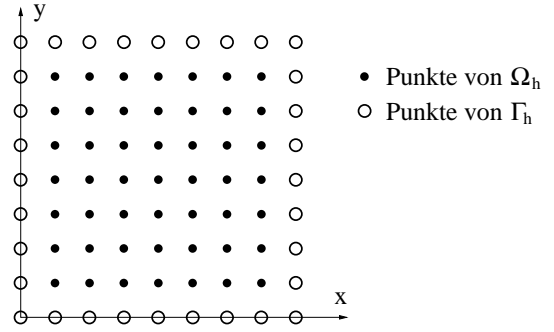


Abb. 4.2.1. Zweidimensionales Gitter

Wie in (4.1.1b) setzen wir

$$\overline{\Omega}_h = \Omega_h \cup \Gamma_h = \{(x, y) \in \overline{\Omega} : x/h, y/h \in \mathbb{Z}\} \quad (4.2.1c)$$

In der Poisson-Gleichung

$$-\Delta u \equiv -u_{xx} - u_{yy} = f \quad \text{in } \Omega, \quad (4.2.2a)$$

$$u = \varphi \quad \text{auf } \Gamma \quad (4.2.2b)$$

lassen sich die doppelten Ableitungen u_{xx} und u_{yy} jeweils durch die Differenz (4.1.3c) in x - bzw. y -Richtung ersetzen:

$$\begin{aligned} (\Delta_h u)(x, y) &:= (\partial_x^- \partial_x^+ + \partial_y^- \partial_y^+) u(x, y) \\ &= h^{-2} [u(x-h, y) + u(x+h, y) + u(x, y-h) + u(x, y+h) - 4u(x, y)]. \end{aligned} \quad (4.2.3)$$

Da auf der rechten Seite von (4.2.3) die Funktion u an fünf Punkten ausgewertet wird, heißt Δ_h auch *Fünfpunktformel*. Die Diskretisierung der Randwertaufgabe (4.2.2a,b) mit Hilfe von Δ_h führt auf die Differenzgleichungen

$$-\Delta_h u_h(\mathbf{x}) = f(\mathbf{x}) \quad \text{für } \mathbf{x} \in \Omega_h, \quad (4.2.4a)$$

$$u_h(\mathbf{x}) = \varphi(\mathbf{x}) \quad \text{für } \mathbf{x} \in \Gamma_h. \quad (4.2.4b)$$

Durch (4.2.4a,b) erhält man eine Gleichung pro Gitterpunkt $\mathbf{x} = (x, y) \in \overline{\Omega}_h$ und damit eine Gleichung pro Komponente der Gitterfunktion $u_h = (u_h(\mathbf{x}))_{\mathbf{x} \in \overline{\Omega}_h}$. Anders als im eindimensionalen Fall gibt es keine natürliche Anordnung der Gitterpunkte, sodass man nicht sofort zu einer Matrixdarstellung wie in (4.1.8b) gelangt. Die einzig natürliche Indizierung von u_n ist die durch $\mathbf{x} \in \Omega_h$ oder das Paar $(i, j) \in \mathbb{N}^2$ mit $\mathbf{x} = (x, y) = (ih, jh)$. Die Matrixelemente seien durch

$$L_{h, \mathbf{x}\xi} := L_{\mathbf{x}\xi} := \begin{cases} 4h^{-2} & \text{falls } \mathbf{x} = \xi \in \Omega_h, \\ -h^{-2} & \text{falls } \mathbf{x} \in \Omega_h, \xi \in \overline{\Omega}_h, \mathbf{x} - \xi = \begin{pmatrix} 0 \\ \pm h \end{pmatrix} \text{ oder } \mathbf{x} - \xi = \begin{pmatrix} \pm h \\ 0 \end{pmatrix}, \\ 0 & \text{sonst} \end{cases} \quad (4.2.5)$$

²Die im Folgenden beschriebenen Differenzenverfahren besitzen stets äquidistante Schrittweiten. Eine Variante mit allgemeineren Gittern sind die sogenannten Box-Schemata (vgl. §8.8.3). Sie sind andererseits eng verwandt mit Finite-Element-Verfahren.

erklärt. Für $\mathbf{x} = \xi$ ist $L_{\mathbf{x}\mathbf{x}}$ ein Diagonalelement; im zweiten Falle (wenn $L_{\mathbf{x}\xi} = -h^{-2}$) heißen \mathbf{x} und ξ *benachbart*. Eliminiert man die Komponenten $u_h(\mathbf{x})$ für alle $\mathbf{x} \in \Gamma_h$ mit Hilfe von Gleichung (4.2.4b), nimmt Gleichung (4.2.4a) die folgende Form an:

$$\sum_{\xi \in \Omega_h} L_{\mathbf{x}\xi} u_h(\xi) = q_h(\mathbf{x}) \quad \text{für } \mathbf{x} \in \Omega_h, \quad (4.2.6a)$$

$$\text{wobei } q_h := f_h + \varphi_h, \quad f_h(\mathbf{x}) := f(\mathbf{x}), \quad \varphi_h(\mathbf{x}) := - \sum_{\xi \in \Gamma_h} L_{\mathbf{x}\xi} u_h(\xi) = - \sum_{\xi \in \Gamma_h} L_{\mathbf{x}\xi} \varphi(\xi). \quad (4.2.6b)$$

Zum Beweis spalte man die Summe

$$-\Delta_h u_h(\mathbf{x}) = \sum_{\xi \in \Omega_h} L_{\mathbf{x}\xi} u_h(\xi)$$

in $\sum_{\xi \in \Omega_h} \dots$ und $\sum_{\xi \in \Gamma_h} \dots$ auf. Die zweite Teilsumme ist $-\varphi_h$ aus (4.2.6b) und wird auf die rechte Seite der Gleichung gebracht.

Bemerkung 4.2.1 f_h ist die Beschränkung von f auf das Gitter Ω_h . Für alle randfernen Punkte gilt $\varphi_h(\mathbf{x}) = 0$; dabei heißt $\mathbf{x} \in \Omega_h$ randfern, falls alle Nachbarn $\mathbf{x} \pm (0, h)$ und $\mathbf{x} \pm (h, 0)$ zu Ω_h gehören. Im Falle homogener Randwerte $\varphi = 0$ ist $q_h = f_h$.

Das Gleichungssystem (4.2.6a) schreibt man wieder in der Form (4.1.8a):

$$L_h u_h = q_h,$$

wobei die Matrix

$$L_h = (L_{\mathbf{x}\xi})_{\mathbf{x}, \xi \in \Omega_h} \quad (4.2.5')$$

und die Gitterfunktionen $u_h = (u_h(\mathbf{x}))_{\mathbf{x} \in \Omega_h}$ und $q_h = (q_h(\mathbf{x}))_{\mathbf{x} \in \Omega_h}$ durch ihre Komponenten beschrieben sind (vgl. (4.2.5), (4.2.6b)). Genaugenommen ist L_h keine Matrix, sondern eine lineare Abbildung, da die Indizes $\mathbf{x} \in \Omega_h$ nicht *angeordnet* sind.

Übungsaufgabe 4.2.2 a) Eine $N \times N$ -Matrix P heißt Permutationsmatrix, falls $w := Pv$ für alle $v \in \mathbb{R}^N$ die Koeffizienten $w_i = v_{\pi(i)}$, $1 \leq i \leq N$, besitzt, wobei π eine Permutation der Indizes $\{1, \dots, N\}$ ist. Man zeige: P ist unitär, d.h. $P^{-1} = P^\top$.

b) I sei eine Indexmenge mit N Elementen. Die "Matrix"-Koeffizienten $A_{\alpha\beta}$ ($\alpha, \beta \in I$) seien gegeben. Der Anordnung $\alpha_1, \dots, \alpha_N$ der Indizes entspricht die $N \times N$ -Matrix $A = (a_{ij})_{i,j=1,\dots,N}$ mit $a_{ij} = A_{\alpha_i, \alpha_j}$. Sei \hat{A} die Matrix, die zu einer zweiten Anordnung $\hat{\alpha}_1, \dots, \hat{\alpha}_N$ von I gehört. Man beweise: Es gibt eine Permutationsmatrix P , sodass $\hat{A} = PAP^\top$.

c) Sei $A_{\alpha\beta} = A_{\beta\alpha}$ für alle $\alpha, \beta \in I$. Man zeige: Für jede Anordnung der Indizes ist die zugehörige Matrix symmetrisch.

Eine mögliche lineare Aufzählung der Indizes $\mathbf{x} \in \Omega_h$ ist durch die *lexikographische Anordnung*

$$\begin{array}{ccccccc} (h, h), & (2h, h), & (3h, h), & \dots & (1-h, h), \\ (h, 2h), & (2h, 2h), & (3h, 2h), & \dots & (1-h, 2h), \\ & & & \vdots & \\ (h, 1-h), & (2h, 1-h), & (3h, 1-h), & \dots & (1-h, 1-h), \end{array} \quad (4.2.7)$$

gegeben. Allgemein geht der Punkt $\mathbf{x} = (x_1, \dots, x_d)$ dem Punkt $\mathbf{y} = (y_1, \dots, y_d)$ in der lexikographischen Anordnung voraus (d.h. " $\mathbf{x} < \mathbf{y}$ "), wenn für ein $j \in \{1, \dots, d\}$ die Bedingungen $x_i = y_i$ (für alle $i > j$) und $x_j < y_j$ zutreffen. Jede Zeile in (4.2.7) entspricht einer sogenannten x -Linie im Gitter Ω_h (vgl. Abbildung 4.2.1). Ein Vektor u_h , dessen $(n-1)^2$ Komponenten in der Reihenfolge (4.2.7) aufgezählt sind, zerfällt daher in $n-1$ Blöcke (sogenannte x -Blöcke). Die Blockzerlegung der Vektoren erzeugt eine Blockzerlegung der Matrix L_h , die in (4.2.8) angegeben ist.

Übungsaufgabe 4.2.3 a) Bei lexikographischer Nummerierung der Gitterpunkte hat die Matrix L_h die Gestalt

$$L_h = h^{-2} \begin{bmatrix} T & -I & & & \\ -I & T & -I & & \\ & \ddots & \ddots & \ddots & \\ & & -I & T & -I \\ & & & -I & T \end{bmatrix}, \quad T = \begin{bmatrix} 4 & -1 & & & \\ -1 & 4 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 4 & -1 \\ & & & -1 & 4 \end{bmatrix}, \quad (4.2.8)$$

wobei T eine $(n-1) \times (n-1)$ -Matrix ist und L_h $(n-1)^2$ Blöcke enthält. I ist die $(n-1) \times (n-1)$ -Einheitsmatrix.

b) Ω sei das Rechteck $\Omega = (0, a) \times (0, b) = \{(x, y) : 0 < x < a, 0 < y < b\}$. Die Schrittweite h erfülle die Bedingungen $a = nh$ und $b = mh$. Man zeige: Die Diskretisierung (4.2.4a,b) in dem entsprechenden Gitter Ω_h führt zu einer Matrix, die ebenfalls die Gestalt (4.2.8) besitzt. Dabei enthält L_h aber $(m-1)^2$ Blöcke der Größe $(n-1) \times (n-1)$.

Eine weitere, häufig verwendete Anordnung ist das *Schachbrettmuster*. Hierzu teilt man Ω_h in die “weißen” und “schwarzen” Felder auf:

$$\begin{aligned} \Omega_h^w &:= \{(x, y) : (x+y)/h \text{ ungerade}\}, \\ \Omega_h^s &:= \{(x, y) : (x+y)/h \text{ gerade}\}. \end{aligned} \quad (4.2.9)$$

Man nummeriert zunächst die weißen Felder $(x, y) \in \Omega_h^w$ lexikographisch durch, danach diejenigen von Ω_h^s . Die Zerlegung (4.2.9) induziert eine Zerlegung der Vektoren in 2 Blöcke und eine Zerlegung der Matrix L_h in $2 \cdot 2 = 4$ Blöcke.

Übungsaufgabe 4.2.4 Bezüglich der Schachbrettanordnung nimmt die Matrix L_h die Form

$$L_h = h^{-2} \begin{bmatrix} \begin{array}{c|c} 4 & \\ \hline & A \end{array} & \begin{array}{c|c} & \\ \hline & 4 \end{array} \\ \hline \begin{array}{c|c} & A^\top \\ \hline & \end{array} & \begin{array}{c|c} 4 & \\ \hline & \end{array} \end{bmatrix} \quad (4.2.10)$$

an, wobei A eine im Allgemeinen rechteckige Blockmatrix ist, weil für gerades n Ω_h^w und Ω_h^s verschieden viele Punkte enthalten.

Die vollständige $(n-1)^2 \times (n-1)^2$ -Matrix L_h aus (4.2.8) oder (4.2.10) wird weder für die theoretischen Untersuchungen des Gleichungssystems $L_h u_h = q_h$ noch für ihre numerische Auflösung benötigt. Alle im Folgenden betrachteten Eigenschaften von L_h sind invariant gegenüber einer Umnummerierung der Gitterpunkte. Numerische Verfahren zur Auflösung von $L_h u_h = q_h$ benutzen zwar implizit eine Anordnung der Gitterpunkte (ausgenommen spezielle Algorithmen für Parallelrechner), sie verwenden aber nie die vollständige $(n-1)^2 \times (n-1)^2$ -Matrix L_h . Jeder brauchbare Algorithmus muss berücksichtigen, dass L_h *schwach besetzt* ist, d.h. wesentlich mehr Null- als Nichtnullelemente besitzt.

Im Folgenden werden wir wieder zur Indizierung mittels $\mathbf{x} \in \Omega_h$ zurückkehren. Trotzdem werden wir von der durch (4.2.5) definierten Matrix L_h sprechen.

Der Differenzenoperator Δ_h wird auch durch den “Stern”

$$\Delta_h = h^{-2} \begin{bmatrix} & 1 & & \\ 1 & -4 & 1 & \\ & 1 & & \end{bmatrix} \quad (4.2.11)$$

beschrieben. Die allgemeine Definition eines Differenzensternes (mit variablen Koeffizienten) lautet

$$\begin{bmatrix} \vdots & & & \\ c_{-1,1}(x, y) & c_{0,1}(x, y) & c_{1,1}(x, y) & \\ \dots & c_{-1,0}(x, y) & c_{0,0}(x, y) & c_{1,0}(x, y) & \dots \\ c_{-1,-1}(x, y) & c_{0,-1}(x, y) & c_{1,-1}(x, y) & \\ \vdots & & & \end{bmatrix} u_h(x, y) := \sum_{i,j} c_{ij}(x, y) u_h(x + ih, y + jh), \quad (4.2.12)$$

wobei nicht angeschriebene Koeffizienten Nullen darstellen. Die Summation erstreckt sich über alle angegebenen Koeffizienten.

Achtung. Der Stern (4.2.11) stellt keine Untermatrix von L_h dar! Die Koeffizienten des Sterns erscheinen in jeder Zeile von L_h (die Koeffizienten entfallen, wenn $(x + ih, y + jh)$ nicht mehr zu Ω_h gehört).

Bemerkung 4.2.5 Man beachte, dass der Differenzenoperator $-\Delta_h$ nicht mit der Matrix L_h gleichgesetzt werden kann, da Δ_h keine Informationen über Art und Ort der Randbedingungen enthält. Wir werden L_h als die zu einem Differenzenstern (4.2.12) gehörende Matrix bezeichnen, wenn das Gleichungssystem $L_h u_h = q_h$ aus den Differenzengleichungen in $\mathbf{x} \in \Omega_h$ nach Elimination der Dirichlet-Randwerte $u_h(\mathbf{x}) = \varphi(\mathbf{x})$ ($\mathbf{x} \in \Gamma_h$) entsteht. Auch wenn der Vektor u_h in $L_h u_h = q_h$ nur Komponenten $u_h(\mathbf{x})$ für innere Gitterpunkte $\mathbf{x} \in \Omega_h$ enthält, wird u_h gelegentlich mit der Gitterfunktion auf u_h gleichgesetzt, die auf Γ_h die vorgeschriebenen Randwerte (4.2.4b) annimmt.

4.3 M-Matrizen, Matrixnormen, positiv definite Matrizen

Die Elemente der Matrix A werden im Folgenden mit $a_{\alpha\beta}$ ($\alpha, \beta \in I$) bezeichnet. Damit treten A und die Indexmenge I an die Stelle von L_h und Ω_h . Wir schreiben

$$A \geq B, \quad \text{falls } a_{\alpha\beta} \geq b_{\alpha\beta} \quad \text{für alle } \alpha, \beta \in I,$$

und definieren analog $A \leq B$, $A > B$, $A < B$. Mit O wird die Nullmatrix bezeichnet.

Definition 4.3.1 A heißt M-Matrix, wenn

$$a_{\alpha\alpha} > 0 \quad \text{für alle } \alpha, \beta \in I, \quad a_{\alpha\beta} \leq 0 \quad \text{für alle } \alpha \neq \beta, \quad (4.3.1a)$$

$$A \text{ nichtsingulär und } A^{-1} \geq O. \quad (4.3.1b)$$

Die Ungleichungen (4.3.1a) lassen sich sofort für L_h nachweisen (vgl. (4.2.5)). Dagegen brauchen wir noch Kriterien und Hilfssätze, um (4.3.1b) zu beweisen.

Ein Index $\alpha \in I$ heißt *direkt mit* $\beta \in I$ *verbunden*, falls $a_{\alpha\beta} \neq 0$. Der Index $\alpha \in I$ heißt *mit* $\beta \in I$ *verbunden*, falls eine "Verbindung" (Kette von direkten Verbindungen)

$$\alpha = \alpha_0, \alpha_1, \dots, \alpha_k = \beta \quad \text{mit} \quad a_{\alpha_{i-1}\alpha_i} \neq 0 \quad (1 \leq i \leq k) \quad (4.3.2)$$

existiert. Die Indexmenge I zusammen mit den direkten Verbindungen bildet den *Graphen* (Matrixgraphen) von A (vgl. Abbildung 4.3.1). Häufig hat A eine symmetrische Struktur, d.h. $a_{\alpha\beta} \neq 0$ gilt genau dann, wenn $a_{\beta\alpha} \neq 0$. In diesem Falle ist α mit β genau dann (direkt) verbunden, falls β mit α (direkt) verbunden ist.

$$\text{Matrix: } A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \quad \text{Graph: } \begin{array}{ccc} & \nearrow \boxed{2} & \\ \boxed{1} & & \downarrow \boxed{3} \\ & \nwarrow \boxed{3} & \end{array} \quad \left(\begin{array}{ll} \boxed{1}, \boxed{2}, \boxed{3}: & \text{Indizes} \\ \longrightarrow: & \text{direkte Verbindung} \end{array} \right)$$

Abb. 4.3.1. Graph der irreduziblen Matrix A

Definition 4.3.2 (irreduzibel) Eine Matrix A heißt irreduzibel, falls jedes $\alpha \in I$ mit jedem $\beta \in I$ verbunden ist.

Im Falle der Matrix $A = L_h$ aus (4.2.5') sind zwei Indizes $\mathbf{x}, \mathbf{y} \in \Omega_h$ genau dann direkt verbunden, wenn $\mathbf{y} = \mathbf{x}$ oder wenn \mathbf{y} Nachbar von \mathbf{x} . Offensichtlich lassen sich beliebige $\mathbf{x}, \mathbf{y} \in \Omega_h$ durch eine Kette $\mathbf{x} = \mathbf{x}(0), \mathbf{x}(1), \dots, \mathbf{x}(k) = \mathbf{y}$ von benachbarten Punkten verbinden. Also ist L_h irreduzibel.

Übungsaufgabe 4.3.3 Man beweise: A ist genau dann irreduzibel, wenn es keine Anordnung der Indizes gibt, sodass die entstehende Matrix die Gestalt $A = \begin{bmatrix} A_{11} & A_{12} \\ O & A_{22} \end{bmatrix}$ hat, wobei A_{11} und A_{22} quadratische $n_1 \times n_1$ - bzw. $n_2 \times n_2$ -Matrizen ($n_1 \geq 1, n_2 \geq 1$) und A_{12} eine $n_1 \times n_2$ -Untermatrix sind.

Die wichtige Frage, ob $A = L_h$ nichtsingulär ist, lässt sich als Spezialfall der folgenden Aussage behandeln.

Kriterium 4.3.4 (Gerschgorin)³ Mit $K_r(\mathbf{z})$ sei der offene Kreis $\{\zeta \in \mathbb{C} : |\mathbf{z} - \zeta| < r\}$ um \mathbf{z} und mit $\overline{K_r(\mathbf{z})} := \{\zeta \in \mathbb{C} : |\mathbf{z} - \zeta| \leq r\}$ der abgeschlossene Kreis bezeichnet. (a) Alle Eigenwerte von A liegen in

$$\bigcup_{\alpha \in I} \overline{K_{r_\alpha}(a_{\alpha\alpha})} \quad \text{mit} \quad r_\alpha = \sum_{\beta \in I \setminus \{\alpha\}} |a_{\alpha\beta}|.$$

(b) Falls A irreduzibel ist, liegen die Eigenwerte sogar in

$$\left(\bigcup_{\alpha \in I} K_{r_\alpha}(a_{\alpha\alpha}) \right) \cup \left(\bigcap_{\alpha \in I} \partial K_{r_\alpha}(a_{\alpha\alpha}) \right).$$

Beweis. (a) Seien λ ein Eigenwert von A und u ein zugehöriger Eigenvektor, der ohne Beschränkung der Allgemeinheit $\|u\|_\infty = 1$ erfülle, wobei

$$\|u\|_\infty := \max\{|u_\alpha| : \alpha \in I\} \quad (4.3.3)$$

die Maximumnorm ist. Es gibt (mindestens) einen Index $\gamma \in I$ mit $|u_\gamma| = 1$.

Hilfsbehauptung 1: $|u_\gamma| = 1$ impliziert

$$|\lambda - a_{\gamma\gamma}| \leq \sum_{\beta \in I \setminus \{\gamma\}} |a_{\gamma\beta}| |u_\beta| \leq \sum_{\beta \in I \setminus \{\gamma\}} |a_{\gamma\beta}| = r_\gamma. \quad (*)$$

Aus (*) folgt $\lambda \in \overline{K_{r_\gamma}(a_{\gamma\gamma})}$ und damit die Behauptung. Zum Beweis der Hilfsbehauptung verwende man die zum Index γ gehörende Gleichung aus $Au = \lambda u$:

$$\lambda u_\gamma = \sum_{\beta \in I} a_{\gamma\beta} u_\beta, \quad \text{d.h.} \quad (\lambda - a_{\gamma\gamma}) u_\gamma = \sum_{\beta \in I \setminus \{\gamma\}} a_{\gamma\beta} u_\beta.$$

Aus $|u_\gamma| = 1$ ergibt sich $|\lambda - a_{\gamma\gamma}| = |(\lambda - a_{\gamma\gamma})u_\gamma| \leq |\sum_{\beta \neq \gamma} a_{\gamma\beta} u_\beta|$. Indem man den Betrag in die Summe zieht und $|u_\beta| \leq \|u\|_\infty = 1$ benutzt, folgt (*).

(b) A sei irreduzibel und λ ein beliebiger Eigenwert von A mit zugehörigem Eigenvektor u , der wieder durch $\|u\| = 1$ normiert sei. Der Fall $\lambda \in \bigcup_{\alpha \in I} K_{r_\alpha}(a_{\alpha\alpha})$ führt sofort zur Behauptung. Sei deshalb $\lambda \notin \bigcup_{\alpha \in I} K_{r_\alpha}(a_{\alpha\alpha})$ angenommen.

Hilfsbehauptung 2: Sei $a_{\gamma\beta} \neq 0$ (d.h. γ direkt mit β verbunden). $|u_\gamma| = 1$ und $|\lambda - a_{\gamma\gamma}| = r_\gamma$ implizieren $|u_\beta| = 1$ und $|\lambda - a_{\beta\beta}| = r_\beta$.

Teil (a) beweist die Existenz eines $\gamma \in I$ mit $|u_\gamma| = 1$ und $|\lambda - a_{\gamma\gamma}| \leq r_\gamma$. Nach Annahme muss $|\lambda - a_{\gamma\gamma}| = r_\gamma$ gelten, sodass die Hilfsbehauptung 2 auf γ anwendbar ist. Da A irreduzibel ist, existiert für ein beliebiges $\beta \in I$ eine Verbindung (2) von γ mit β : $\alpha_0 = \gamma, \alpha_1, \dots, \alpha_k = \beta, a_{\alpha_{i-1}\alpha_i} \neq 0$. Hilfsbehauptung 2 zeigt

$$|u_{\alpha_i}| = 1 \quad \text{und} \quad |\lambda - a_{\alpha_i\alpha_i}| = r_{\alpha_i} \quad \text{für alle } i = 0, \dots, k;$$

insbesondere $\lambda \in \partial K_{r_\beta}(a_{\beta\beta})$ für $\beta = \alpha_k$. Da β beliebig gewählt wurde, folgt $\lambda \in \bigcap \partial K_{r_\alpha}(a_{\alpha\alpha})$, und die Behauptung ist bewiesen.

Beweis der Hilfsbehauptung 2. Neben der Ungleichungskette (*) gilt $|\lambda - a_{\gamma\gamma}| = r_\gamma$, sodass alle Ungleichungen in (*) zu Gleichungen werden. Insbesondere muss

$$\sum_{\beta \in I \setminus \{\gamma\}} |a_{\gamma\beta}| |u_\beta| = \sum_{\beta \in I \setminus \{\gamma\}} |a_{\gamma\beta}|$$

gelten. Da $|u_\beta| \leq \|u\|_\infty = 1$, muss sogar für jeden Summanden $|a_{\gamma\beta}| |u_\beta| = |a_{\gamma\beta}|$ erfüllt sein. $a_{\gamma\beta} \neq 0$ impliziert daher $|u_\beta| = 1$. Die Anwendung der Hilfsbehauptung 1 auf β liefert $|\lambda - a_{\beta\beta}| \leq r_\beta$. Die Annahme $\lambda \notin \bigcup_{\alpha \in I} K_{r_\alpha}(a_{\alpha\alpha})$ beweist $|\lambda - a_{\beta\beta}| = r_\beta$. ■

³Semjon Aranowitsch Gerschgorin, geb. am 24. August 1901 in Pruzhany, Belarus, gest. am 30. Mai 1933

Übungsaufgabe 4.3.5 Sei $I_\alpha := \{\beta \in I : \text{es existiert eine Verbindung (4.3.2) von } \alpha \text{ zu } \beta\}$. Man zeige: Die Eigenwerte von A liegen in

$$\bigcup_{\alpha \in I} \left\{ K_{r_\alpha}(a_{\alpha\alpha}) \cup \bigcap_{\beta \in I_\alpha} \partial K_{r_\beta}(a_{\beta\beta}) \right\}.$$

Definition 4.3.6 (Diagonaldominanz) (a) A heißt diagonaldominant, falls

$$\sum_{\beta \in I \setminus \{\alpha\}} |a_{\alpha\beta}| < |a_{\alpha\alpha}| \quad (4.3.4a)$$

für alle $\alpha \in I$. (b) A heißt irreduzibel diagonaldominant, falls A irreduzibel ist, Ungleichung (4.3.4a) für mindestens einen Index $\alpha \in I$ gilt und

$$\sum_{\beta \in I \setminus \{\alpha\}} |a_{\alpha\beta}| \leq |a_{\alpha\alpha}| \quad \text{für alle } \alpha \in I. \quad (4.3.4b)$$

Man beachte, dass eine Matrix, die irreduzibel und diagonaldominant ist, zwar irreduzibel diagonaldominant ist, die Umkehrung jedoch nicht zu gelten braucht.

Die Matrix L_h aus §4.2 ist nicht diagonaldominant, aber irreduzibel diagonaldominant, denn L_h ist irreduzibel und erfüllt (4.3.4b). In allen randnahen Punkten - das sind solche $\mathbf{x} \in \Omega_h$, die einen Randpunkt $\mathbf{y} \in \Gamma_h$ als Nachbarn haben - gilt dagegen (4.3.4a): $\sum_{\beta \neq \alpha} |a_{\alpha\beta}| \leq 3h^{-2} < 4h^{-2} = a_{\alpha\alpha}$.

Der *Spektralradius* $\rho(A)$ einer Matrix A ist durch den größten Betrag aller Eigenwerte gegeben:

$$\rho(A) := \max\{|\lambda| : \lambda \text{ Eigenwert von } A\}. \quad (4.3.5)$$

Im Folgenden spalten wir A in

$$A = D - B, \quad D := \text{diag}\{a_{\alpha\alpha} : \alpha \in I\}, \quad (4.3.6a)$$

auf, wobei D der Diagonalanteil von A ist:

$$d_{\alpha\alpha} = a_{\alpha\alpha}, \quad d_{\alpha\beta} = 0 \quad \text{für } \alpha \neq \beta. \quad (4.3.6b)$$

$B := D - A$ ist der Außerdiagonalanteil:

$$b_{\alpha\alpha} = 0, \quad b_{\alpha\beta} = -a_{\alpha\beta} \quad \text{für } \alpha \neq \beta. \quad (4.3.6c)$$

Kriterium 4.3.7 Es gelte (4.3.6a-c). Hinreichende Bedingungen für

$$\rho(D^{-1}B) < 1 \quad (4.3.7)$$

sind die *Diagonaldominanz* oder die *irreduzible Diagonaldominanz* von A .

Beweis. 1) Die Koeffizienten von $C := D^{-1}B$ lauten

$$c_{\alpha\beta} = -a_{\alpha\beta}/a_{\alpha\alpha} \quad (\alpha \neq \beta), \quad c_{\alpha\alpha} = 0.$$

Aus der Diagonaldominanz (4.3.4a) folgt $r_\alpha := \sum_{\beta \neq \alpha} |c_{\alpha\beta}| < 1$ für alle $\alpha \in I$. Auf Grund des Gerschgorin-Kriteriums 4.3.4a liegen alle Eigenwerte λ von C in $\bigcup_{\alpha \in I} \overline{K_{r_\alpha}(c_{\alpha\alpha})} = \bigcup_{\alpha \in I} \overline{K_{r_\alpha}(0)}$, sodass $|\lambda| \leq \max r_\alpha < 1$ und damit auch $\rho(C) = \rho(D^{-1}B) < 1$ folgen.

2) Ist A irreduzibel diagonaldominant, so gilt $r_\beta \leq 1$ für alle $\beta \in I$ und $r_\alpha < 1$ für mindestens ein α . Gemäß Kriterium 4.3.4b liegen alle Eigenwerte von C in $\bigcup_{\beta \in I} K_{r_\beta}(0) \cup (\bigcap_{\beta \in I} \partial K_{r_\beta}(0))$. Diese Menge liegt in $K_1(0)$, falls $\bigcap_{\beta \in I} \partial K_{r_\beta}(0) \subset K_1(0)$. Sei zunächst angenommen, dass alle r_β übereinstimmen: $r_\beta = r$ für alle β . Da $r_\alpha < 1$ für ein $\alpha \in I$, folgt $r < 1$ und $\bigcap_{\beta \in I} \partial K_{r_\beta}(0) = \partial K_r(0) \subset K_1(0)$. Sind aber nicht alle r_β gleich, so ist $\bigcap_{\beta \in I} \partial K_{r_\beta}(0)$ leer. In beiden Fällen gilt daher $\lambda \in K_1(0)$, und (4.3.7) ist bewiesen. ■

Übungsaufgabe 4.3.8 a) Man definiere folgende Abschwächung der irreduziblen Diagonaldominanz. A erfülle die Ungleichungen (4.3.4b), und für alle $\beta \in I$ existiere eine Verbindung (4.3.2) zu einem Index $\alpha \in I$, für den die echte Ungleichung (4.3.4a) zutrifft. Man beweise, dass auch unter dieser Voraussetzung $\rho(D^{-1}B) < 1$ gilt. Hinweis: Man benutze Übungsaufgabe 4.3.5.

b) Man zeige: Die geometrische Reihe $S = \sum_{\nu=0}^{\infty} C^{\nu}$ konvergiert genau dann, wenn $\rho(C) < 1$. Es gilt $S = (I - C)^{-1}$. Hinweis: Man stelle C in der Form QRQ^{\top} dar (Q : unitär, R : obere Dreiecksmatrix) und zeige $\|C^{\nu}\|_{\infty} \leq K[\rho(C)]^{\nu}$.

c) u sei Vektor. Wir definieren $|u|$ als Vektor (!) mit den Koeffizienten $|u|_{\alpha} := |u_{\alpha}|$. Für zwei Vektoren wird $v \leq w$ geschrieben, falls $v_{\alpha} \leq w_{\alpha}$ ($\alpha \in I$). Man zeige:

- 1) $AB \geq O$, falls $A \geq O$ und $B \geq O$; $AB > O$, falls $A > O$ und $B > O$;
- 2) $AD > O$, falls $A > O$ und $D \geq O$ nichtsinguläre Diagonalmatrix;
- 3) $Av \leq Aw$, falls $A \geq O$ und $v \leq w$; $\|v\|_{\infty} \leq \|w\|_{\infty}$, falls $0 \leq v \leq w$;
- 4) $Au \leq |Au| \leq A|u|$, falls $A \geq O$.

Die Bedeutung der Ungleichung (4.3.7) ergibt sich aus

Lemma 4.3.9 A erfülle (4.3.1a). D und B seien durch (4.3.6a-c) definiert. A ist genau dann eine M-Matrix, wenn $\rho(D^{-1}B) < 1$.

Beweis. 1) $C := D^{-1}B$ erfülle $\rho(C) < 1$. Dann konvergiert die geometrische Reihe $S := \sum_{\nu=0}^{\infty} C^{\nu}$ (vgl. Übungsaufgabe 4.3.8b). Aus $D^{-1} \geq O$ und $B \geq O$ schließt man $C \geq O$, $C^{\nu} \geq O$ und $S \geq O$. Da $I = S(I - C) = SD^{-1}(D - B) = SD^{-1}A$, hat A die Inverse $A^{-1} = SD^{-1}$. $D^{-1} \geq O$ und $S \geq O$ ergeben $A^{-1} \geq O$. Damit ist auch (4.3.1b) gezeigt, d.h. A ist M-Matrix.

2) A sei eine M-Matrix. Zu einem Eigenwert λ von $D^{-1}B$ wähle man einen Eigenvektor $u \neq 0$. Gemäß Übungsaufgabe 4.3.8c gilt

$$|\lambda| |u| = |\lambda u| = |D^{-1}Bu| \leq D^{-1}B |u|.$$

Wegen $A^{-1}D \geq O$ (vgl. (4.3.1a,b)) erhält man $-A^{-1}DD^{-1}B|u| \leq -A^{-1}D|\lambda||u|$, sodass

$$|u| = A^{-1}(D - B)|u| = A^{-1}D(I - D^{-1}B)|u| \leq A^{-1}D|u| - A^{-1}D|\lambda||u| = (1 - |\lambda|)A^{-1}D|u|$$

folgt. Für $|\lambda| \geq 1$ ergäbe sich die Ungleichung $|u| \leq 0$, d.h. $u = 0$ im Widerspruch zur Annahme $u \neq 0$. Hieraus folgt $|\lambda| < 1$ für jeden Eigenwert von $C = D^{-1}B$, also $\rho(D^{-1}B) < 1$. ■

Kriterium 4.3.7 und Lemma 4.3.9 ergeben das

Kriterium 4.3.10 Ist eine Matrix A mit der Eigenschaft (4.3.1a) diagonaldominant oder irreduzibel diagonaldominant, so ist A eine M-Matrix.

Satz 4.3.11 Eine irreduzible M-Matrix A hat eine (elementweise) positive Inverse: $A^{-1} > O$.

Beweis. Seien $\alpha, \beta \in I$ beliebig gewählt. Es existiert eine Verbindung (4.3.2): $\alpha = \alpha_0, \alpha_1, \dots, \alpha_k = \beta$. Sei $C := D^{-1}B$. Da $c_{\alpha_{i-1}\alpha_i} > 0$, folgt

$$(C^k)_{\alpha\beta} = \sum_{\gamma_1, \dots, \gamma_{k-1} \in I} c_{\alpha\gamma_1} c_{\gamma_1\gamma_2} \cdots c_{\gamma_{k-1}\beta} \geq c_{\alpha\alpha_1} c_{\alpha_1\alpha_2} \cdots c_{\alpha_{k-1}\beta} > 0.$$

Gemäß Lemma 4.3.9 gilt $\rho(C) < 1$, sodass $S := \sum_{\nu=0}^{\infty} C^{\nu}$ konvergiert. Da $S_{\alpha\beta} \geq (C^k)_{\alpha\beta} > 0$ und $\alpha, \beta \in I$ beliebig, ist $S > O$ bewiesen. Die Behauptung ergibt sich aus $A^{-1} = SD^{-1} > O$ (vgl. Beweis des Lemmas 4.3.9). ■

Im Folgenden wollen wir Normabschätzungen für A^{-1} herleiten.

Definition 4.3.12 V sei ein linearer Raum (Vektorraum) über dem Körper der reellen Zahlen ($\mathbb{K} := \mathbb{R}$) oder der komplexen Zahlen ($\mathbb{K} := \mathbb{C}$). $\|\cdot\|$ heißt Norm in V , falls

$$\|u\| = 0 \quad \text{nur für } 0 \neq u \in V, \quad (4.3.8a)$$

$$\|u + v\| \leq \|u\| + \|v\| \quad \text{für alle } u, v \in V, \quad (4.3.8b)$$

$$\|\lambda u\| = |\lambda| \|u\| \quad \text{für alle } \lambda \in \mathbb{K}, u \in V. \quad (4.3.8c)$$

Beispiel. Sei $V = R^{\#I}$ ($\#I :=$ Anzahl der Elemente der Indexmenge I). Die in (4.3.3) definierte Maximumnorm erfüllt die Normaxiome (4.3.8a-c).

Sieht man die Elemente $u \in V$ als Vektoren an, nennt man $\|\cdot\|$ auch Vektornorm. Aber auch die Matrizen bilden einen linearen Raum. Im letzten Falle nennt man $\|\cdot\|$ eine Matrixnorm. Eine spezielle Klasse von Matrixnormen enthält die

Definition 4.3.13 (zugeordnete Matrixnorm) V sei Vektorraum mit Vektornorm $\|\cdot\|$. Dann bezeichnet man

$$\| \| A \| \| := \sup\{\|Au\| / \|u\| : 0 \neq u \in V\} \quad (4.3.9)$$

als die der Vektornorm $\|\cdot\|$ zugeordnete Matrixnorm.

Übungsaufgabe 4.3.14 $\| \| \cdot \| \|$ sei durch (4.3.9) definiert. Man zeige: (a) $\| \| \cdot \| \|$ ist eine Norm; (b) es gilt

$$\| \| AB \| \| \leq \| \| A \| \| \| \| B \| \|, \quad (4.3.10a)$$

$$\| \| I \| \| = 1 \quad (I: \text{Einheitsmatrix}), \quad (4.3.10b)$$

$$\|Au\| \leq \| \| A \| \| \|u\| \quad \text{für alle } u \in V, \quad (4.3.10c)$$

$$\| \| A \| \| \geq \rho(A). \quad (4.3.10d)$$

Es ist üblich, die zugeordnete Matrixnorm mit dem gleichen Normsymbol zu bezeichnen, d.h. statt $\| \| \cdot \| \|$ wird $\|\cdot\|$ geschrieben und Indizes “xyz”, die in $\|\cdot\|_{xyz}$ eine spezielle Vektornorm charakterisieren, werden in der gleichen Form $\|\cdot\|_{xyz}$ für die zugeordnete Matrixnorm übernommen.

Beispiel. Die der Maximumnorm $\|\cdot\|_\infty$ (vgl. (4.3.3)) zugeordnete Matrixnorm heißt *Zeilensummennorm* und wird ebenfalls mit $\|\cdot\|_\infty$ bezeichnet. Sie hat die explizite Darstellung

$$\|A\|_\infty = \max_{\alpha \in I} \left\{ \sum_{\beta \in I} |a_{\alpha\beta}| \right\}. \quad (4.3.11)$$

Übungsaufgabe 4.3.15 a) Man beweise (4.3.11). b) Für Matrizen $O \leq B \leq C$ gilt $\|B\|_\infty \leq \|C\|_\infty$.

Im nächsten Satz bezeichnen wir mit $\mathbf{1}$ den Vektor, der nur Einsen als Komponenten besitzt:

$$\mathbf{1}_\alpha = 1 \quad \text{für alle } \alpha \in I.$$

Zur Schreibweise $v \leq w$ vergleiche man Übungsaufgabe 4.3.8c.

Satz 4.3.16 A sei eine M -Matrix, und es gebe einen Vektor w mit $Aw \geq \mathbf{1}$. Dann gilt $\|A^{-1}\|_\infty \leq \|w\|_\infty$.

Beweis. Wie im Beweis von Lemma 4.3.9 sei $|u|$ der Vektor mit den Komponenten $|u_\alpha|$. Für jedes u ist $|u| \leq \|u\|_\infty \mathbf{1} \leq \|u\|_\infty Aw$. Da $A^{-1} \geq O$, erhalten wir

$$|A^{-1}u| \leq A^{-1}|u| \leq \|u\|_\infty A^{-1}Aw = \|u\|_\infty w$$

(vgl. Übungsaufgabe 4.3.8c) und $\|A^{-1}u\|_\infty / \|u\|_\infty \leq \|w\|_\infty$. Definition 4.3.13 liefert $\|A^{-1}\|_\infty \leq \|w\|_\infty$. ■

Wie man mit Hilfe einer majorisierenden Matrix abschätzen kann, zeigt

Satz 4.3.17 A und A' seien M -Matrizen mit $A' \geq A$. Dann gelten die Ungleichungen

$$O \leq A'^{-1} \leq A^{-1} \quad \text{und} \quad \|A'^{-1}\|_\infty \leq \|A^{-1}\|_\infty. \quad (4.3.12)$$

Beweis. $A'^{-1} \leq A^{-1}$ folgt aus $A^{-1} - A'^{-1} = A^{-1}(A' - A)A'^{-1}$ und $A^{-1} \geq O$, $A' - A \geq O$, $A'^{-1} \geq O$. Der Rest ergibt sich aus Übungsaufgabe 4.3.15b. ■

Übungsaufgabe 4.3.18 Man beweise (4.3.12) unter den folgenden schwächeren Voraussetzungen: A ist M -Matrix, A' erfüllt (4.3.1a) und $A' \geq A$. Hinweis: Man wiederhole die Überlegungen aus dem ersten Teil des Beweises von Lemma 4.3.9 mit den zu A' gehörenden Matrizen D' und B' .

Übungsaufgabe 4.3.19 *B sei eine Hauptuntermatrix von A, d.h. es gibt eine Teilmenge $I' \subset I$, sodass B durch die Koeffizienten $b_{\alpha\beta} = a_{\alpha\beta}$ ($\alpha, \beta \in I'$) definiert ist. Man beweise: Ist A eine M-Matrix, so auch B und es gilt $0 \leq (B^{-1})_{\alpha\beta} \leq (A^{-1})_{\alpha\beta}$ für alle $\alpha, \beta \in I'$. Hinweis: Man wende Übungsaufgabe 4.3.18 auf die folgende Matrix A' an: $a'_{\alpha\beta} = a_{\alpha\beta}$ für $\alpha, \beta \in I'$, $a'_{\alpha\alpha} = a_{\alpha\alpha}$ für $\alpha \in I \setminus I'$ und $a'_{\alpha\beta} = 0$ sonst.*

Eine weitere, wohlbekannte Vektornorm ist die *Euklidische Norm*

$$\|u\|_2 := \sqrt{c \sum_{\alpha \in I} |u_\alpha|^2} \quad (4.3.13)$$

mit fester Skalierungskonstanten $c > 0$ (z.B. führt die Wahl $c = h^2$ im Zusammenhang mit den Gitterfunktionen aus §4.2 dazu, dass $c \sum_{\alpha \in I}$ eine Näherung für die Integration \int_Ω darstellt). Die $\|\cdot\|_2$ zugeordnete Matrixnorm ist unabhängig vom Faktor c. Sie heißt *Spektralnorm* und wird ebenfalls mit $\|\cdot\|_2$ bezeichnet. Der Name leitet sich aus der folgenden Charakterisierung ab.

Übungsaufgabe 4.3.20 *Man beweise: (a) Für symmetrische Matrizen gilt $\|A\|_2 = \rho(A)$ (vgl. (4.3.5)). (b) Für jede (reelle) Matrix gilt*

$$\|A\|_2 = \sqrt{\rho(A^\top A)} = [\text{maximaler Eigenwert von } A^\top A]^{1/2}.$$

(c) Für jede Matrix gilt $\|A\|_2^2 \leq \|A\|_2 \|A^\top\|_2$. Hinweis: (b) und (4.3.10d).

Beim Beweis der Übungsaufgabe macht man Gebrauch vom *Skalarprodukt*

$$\langle u, v \rangle = c \sum_{\alpha \in I} u_\alpha v_\alpha \quad (4.3.14a)$$

(c wie in (4.3.13)) und seinen Eigenschaften

$$\langle u, u \rangle = \|u\|_2^2, \quad \langle Au, v \rangle = \langle u, A^\top v \rangle, \quad |\langle u, v \rangle| \leq \|u\|_2 \|v\|_2. \quad (4.3.14b)$$

Dabei ist der Fall $K = \mathbb{R}$ zugrundegelegt: Alle Matrizen und Vektoren seien reell. (Im Falle $K = \mathbb{C}$ ist v_α in (4.3.14a) durch \bar{v}_α und A^\top in (4.3.14b) durch A^H zu ersetzen).

Definition 4.3.21 (positiv definit) *Eine Matrix A heißt positiv definit, falls sie symmetrisch ist und*

$$\langle Au, u \rangle > 0 \quad \text{für alle } u \neq 0. \quad (4.3.15)$$

Übungsaufgabe 4.3.22 *Man beweise: (a) Eine symmetrische Matrix ist genau dann positiv definit, wenn alle Eigenwerte positiv sind.*

(b) Alle Hauptuntermatrizen einer positiv definiten Matrix sind positiv definit (vgl. Übungsaufgabe 4.3.19).

(c) Die Diagonalelemente $a_{\alpha\alpha}$ einer positiv definiten Matrix sind positiv.

(d) A heißt positiv semidefinit, falls (4.3.15) mit \leq anstelle von $>$ gilt. Eine positiv semidefinite Matrix A besitzt genau eine positiv semidefinite Quadratwurzel $B = A^{1/2}$, die $B^2 = A$ erfüllt. Ist A positiv definit, so auch $A^{1/2}$.

Eine Folgerung aus der Übungsaufgabe 4.3.22a ist

Lemma 4.3.23 *Eine positiv definite Matrix ist nichtsingulär und besitzt eine positiv definite Inverse.*

Die Eigenschaft “ A^{-1} positiv definit” ist weder hinreichend noch notwendig für die Eigenschaft “ $A^{-1} \geq 0$ ” einer M-Matrix. In beiden Fällen ist jedoch die (irreduzible) Diagonaldominanz hinreichend (vgl. Kriterium 4.3.10 und das nachfolgende Kriterium).

Kriterium 4.3.24 *Ist eine symmetrische Matrix mit positiven Diagonalelementen diagonaldominant oder irreduzibel diagonaldominant, so ist sie positiv definit.*

Beweis. Da $r_\alpha < a_{\alpha\alpha}$ bzw. $r_\alpha \leq a_{\alpha\alpha}$, schneiden die Gerschgorin-Kreise aus Kriterium 4.3.4 nicht die Halbachse $(-\infty, 0]$, sodass alle Eigenwerte positiv sein müssen. Gemäß Übungsaufgabe 4.3.22a ist A positiv definit. ■

Lemma 4.3.25 λ_{\min} und λ_{\max} seien der minimale bzw. maximale Eigenwert einer positiv definiten Matrix A . Dann gilt

$$\|A\|_2 = \lambda_{\max}, \quad \|A^{-1}\|_2 = \lambda_{\min}. \quad (4.3.16)$$

Beweis. Übungsaufgabe 4.3.20a zeigt $\|A\|_2 = \rho(A)$ und $\|A^{-1}\|_2 = \rho(A^{-1})$. Aus (4.3.5) ergeben sich die Gleichungen $\rho(A) = \lambda_{\max}$ und $\rho(A^{-1}) = 1/\lambda_{\min}$, da $\lambda_{\min} > 0$. ■

4.4 Eigenschaften der Matrix L_h

Satz 4.4.1 Die in (4.2.5) definierte Matrix L_h (Fünfpunktformel) hat die folgenden Eigenschaften:

$$L_h \text{ ist } M\text{-Matrix}, \quad (4.4.1a)$$

$$L_h \text{ ist positiv definit}, \quad (4.4.1b)$$

$$\|L_h\|_{\infty} \leq 8h^{-2}, \quad \|L_h^{-1}\|_{\infty} \leq 1/8, \quad (4.4.1c)$$

$$\|L_h\|_2 \leq 8h^{-2} \cos^2\left(\frac{\pi h}{2}\right) < 8h^{-2}, \quad (4.4.1d)$$

$$\|L_h^{-1}\|_2 \leq \frac{1}{8} h^2 \sin^{-2}\left(\frac{\pi h}{2}\right) = \frac{1}{2\pi^2} + \mathcal{O}(h^2) \leq 1/16. \quad (4.4.1e)$$

Beweis. a) In §4.3 wurde bereits erwähnt, dass L_h irreduzibel diagonaldominant ist und den Ungleichungen (4.3.1a) genügt. Gemäß Kriterium 4.3.10 ist L_h eine M-Matrix.

b) Da L_h symmetrisch und irreduzibel diagonaldominant ist, folgt (4.4.1b) aus Kriterium 4.3.24.

c) $\|L_h\| \leq 8h^{-2}$ liest man aus (4.2.5) und (4.3.11) ab. Zur Abschätzung von L_h^{-1} wende man Satz 4.3.16 mit $w(x, y) = x(1-x)/2$ an. Es gilt $L_h w \geq \mathbf{1}$ [sogar $(L_h w)(x, y) = 1$ bis auf $y = h$ und $y = 1-h$] und $\|w\|_{\infty} \leq w(1/2, y) = 1/8$.

d) Die Ungleichungen (4.4.1d,e) ergeben sich aus Lemma 4.3.25 und dem nachfolgenden Lemma. ■

Lemma 4.4.2 Die $(n-1)^2$ Eigenvektoren von L_h sind $u^{\nu\mu}$ ($1 < \nu, \mu < n-1$):

$$u^{\nu\mu}(x, y) = \sin(\nu\pi x) \sin(\mu\pi y) \quad ((x, y) \in \Omega_h). \quad (4.4.2a)$$

Die zugehörigen Eigenwerte sind

$$\lambda_{\nu\mu} = 4h^{-2} \left(\sin^2\left(\frac{\nu\pi h}{2}\right) + \sin^2\left(\frac{\mu\pi h}{2}\right) \right), \quad 1 \leq \nu, \mu \leq n-1. \quad (4.4.2b)$$

Beweis. Seien Ω_h^{1D} das eindimensionale Gitter (4.1.5a) und $u^{\nu}(x) := \sin(\nu\pi x)$. Für jedes $x \in \Omega_h^{1D}$ gilt

$$\partial^- \partial^+ u^{\nu}(x) = h^{-2} [\sin(\nu\pi(x-h)) + \sin(\nu\pi(x+h)) - 2\sin(\nu\pi x)] = 2h^{-2} \sin(\nu\pi x) [\cos(\nu\pi h) - 1]$$

wegen $\sin(\nu\pi(x \pm h)) = \sin(\nu\pi x) \cos(\nu\pi h) \pm \cos(\nu\pi x) \sin(\nu\pi h)$. Die Identität $1 - \cos \xi = 2 \sin^2(\xi/2)$ liefert

$$-\partial^- \partial^+ u^{\nu}(x) = 4h^{-2} \sin^2\left(\frac{\nu\pi h}{2}\right) u^{\nu}(x), \quad x \in \Omega_h^{1D}. \quad (4.4.2c)$$

Sei L_h^{1D} die Matrix (4.1.8b). Man beachte, dass $(\partial^- \partial^+ u)(h)$ im Gegensatz zu $(L_h^{1D} u)(h)$ noch den Randwert $u(0)$ enthält; ebenso hängt $(\partial^- \partial^+ u)(1-h)$ von $u(1)$ ab. Da $u(0) = \sin(0) = 0$ und $u(1) = \sin(\nu\pi) = 0$, gilt jedoch $L_h^{1D} u^{\nu} = -\partial^- \partial^+ u^{\nu}$, und (4.4.2c) überträgt sich:

$$L_h^{1D} u^{\nu} = 4h^{-2} \sin^2\left(\frac{\nu\pi h}{2}\right) u^{\nu}, \quad 1 \leq \nu \leq n-1. \quad (4.4.2d)$$

Die zweidimensionale Gitterfunktion $u^{\nu\mu}$ aus (4.4.2a) lässt sich als (Tensor-)Produkt $u^{\nu}(x)u^{\mu}(y)$ schreiben. $(L_h u^{\nu\mu})(x, y)$ stimmt mit der Summe $u^{\mu}(y)(L_h^{1D} u^{\nu})(x) + u^{\nu}(x)(L_h^{1D} u^{\mu})(y)$ überein, sodass (4.4.2b) aus (4.4.2d) folgt. ■

Im Weiteren wollen wir die Parallelen zwischen den Eigenschaften der Poisson-Gleichung (4.2.2) und der diskreten Fünfpunktformel (4.2.4a,b) aufzeigen.

Das Analogon der Mittelwerteigenschaft (2.3.1) ist die Gleichung

$$u_h(x, y) = \frac{1}{4} [u_h(x-h, y) + u_h(x+h, y) + u_h(x, y-h) + u_h(x, y+h)] \quad (4.4.3)$$

Aus (4.2.3) und (4.2.4a) mit $f = 0$ ergibt sich die

Bemerkung 4.4.3 Die Lösung u_h der diskreten Potentialgleichung (4.2.4a) mit $f = 0$ erfüllt Gleichung (4.4.3) in allen Gitterpunkten $(x, y) \in \Omega_h$.

Wie im kontinuierlichen Falle leitet man aus der Mittelwerteigenschaft (4.4.3) das Maximum-Minimum-Prinzip ab.

Bemerkung 4.4.4 Sei u_h eine nichtkonstante Lösung der diskreten Potentialgleichung (4.2.4a) (mit $f = 0$). Die Extrema $\max\{u_h(\mathbf{x}) : \mathbf{x} \in \overline{\Omega_h}\}$ und $\min\{u_h(\mathbf{x}) : \mathbf{x} \in \overline{\Omega_h}\}$ werden nicht auf Ω_h , sondern auf Γ_h angenommen.

Beweis. Wäre u_h maximal in $(x, y) \in \Omega_h$, müssen wegen Gleichung (4.4.3) alle Nachbarpunkte $(x \pm h, y)$ und $(x, y \pm h)$ den gleichen Wert tragen. Da sich je zwei Punkte durch eine Kette von Nachbarpunkten verbinden lassen, folgt hieraus $u_h = \text{const}$ im Widerspruch zur Voraussetzung. ■

Der letzte Beweis benutzt indirekt, dass L_h irreduzibel ist. Die Irreduzibilität von L_h entspricht der Voraussetzung in Satz 2.3.7, dass Ω ein Gebiet ist, also zusammenhängend ist.

Die Übertragung der Sätze 2.4.3 und 3.1.2 lautet wie folgt.

Satz 4.4.5 a) Seien u_h^1 und u_h^2 zwei Lösungen von (4.2.4a): $-\Delta_h u_h^i = f$ zu verschiedenen Randwerten $u_h^i = \varphi^i$ ($i = 1, 2$). Dann gilt

$$\|u_h^1 - u_h^2\|_\infty \leq \max_{\mathbf{x} \in \Gamma_h} |\varphi^1(\mathbf{x}) - \varphi^2(\mathbf{x})|, \quad (4.4.4a)$$

$$u_h^1 \leq u_h^2 \text{ in } \Omega_h, \text{ falls } \varphi^1 \leq \varphi^2 \text{ auf } \Gamma_h. \quad (4.4.4b)$$

b) Eine Lösung u_h von $-\Delta_h u_h = f \geq 0$ mit Randwerten $u_h = \varphi \geq 0$ erfüllt im gesamten Gitter $u_h \geq 0$.

Beweis. a) Sei $w_h := u_h^2 - u_h^1$. 1) Im Falle von $\varphi^1 \leq \varphi^2$ hat man $w_h \geq 0$ auf Γ_h und $-\Delta_h w_h = 0$. Bemerkung 4.4.4 beweist $w_h = \text{const} \geq 0$ oder $w_h > 0$ und damit (4.4.4b).

a2) Sei M die rechte Seite von (4.4.4a). $-M \leq w_h \leq M$ auf Γ_h impliziert $-M \leq w_h \leq M$ auf Ω_h .

b) $f \geq 0$, $\varphi \geq 0$ implizieren $q_h \geq 0$ in (4.2.6b). Die M-Matrixeigenschaft (4.4.1a) zeigt $u_h = L_h^{-1} q_h \geq 0$. ■

Das diskrete Analogon der Greenschen Funktion $g(\mathbf{x}, \xi)$ ist durch $h^{-2} L_h^{-1}$ gegeben. Sei δ_ξ der skalierte Einheitsvektor

$$\delta_\xi(\mathbf{x}) = \begin{cases} h^{-2} & \text{falls } \mathbf{x} = \xi, \\ 0 & \text{falls } \mathbf{x} \neq \xi, \end{cases} \quad (\mathbf{x}, \xi \in \Omega_h). \quad (4.4.5a)$$

Die Spalte der Matrix $h^{-2} L_h^{-1}$ zum Index $\xi \in \Omega_h$ ist durch

$$g_h(\cdot, \xi) := L_h^{-1} \delta_\xi \quad (\xi \in \Omega_h) \quad (4.4.5b)$$

gegeben. Für festes $\xi \in \Omega_h$ ist $g_h(\cdot, \xi)$ eine auf Ω_h definierte Gitterfunktion. Der Definitionsbereich wird auf $\overline{\Omega_h} \times \overline{\Omega_h}$ erweitert:

$$\bar{g}_h(\mathbf{x}, \xi) = \begin{cases} g_h(\mathbf{x}, \xi) & \text{falls } \mathbf{x}, \xi \in \Omega_h, \\ 0 & \text{falls } \mathbf{x} \in \Gamma_h \text{ oder } \xi \in \Gamma_h \end{cases} \quad (\mathbf{x}, \xi \in \overline{\Omega_h}). \quad (4.4.5c)$$

$g_h(\mathbf{x}, \xi)$ sind die Koeffizienten von $h^{-2} L_h^{-1} : g_h(\mathbf{x}, \xi) = h^{-2} (L_h^{-1})_{\mathbf{x}\xi}$. Die Symmetrie von L_h beweist die

Bemerkung 4.4.6 $\bar{g}_h(\mathbf{x}, \xi) = \bar{g}_h(\xi, \mathbf{x})$ für alle $\mathbf{x}, \xi \in \overline{\Omega_h}$ (vgl. (3.2.3)).

An die Darstellung (3.2.7) erinnert die

Bemerkung 4.4.7 Die Lösung u_h des Gleichungssystems (4.2.4a) mit Randwerten $\varphi = 0$ lautet

$$u_h(\mathbf{x}) = h^2 \sum_{\xi \in \Omega_h} \bar{g}_h(\mathbf{x}, \xi) f(\xi) \quad \text{für } \mathbf{x} \in \overline{\Omega_h}. \quad (4.4.6)$$

Gleichung (4.4.6) ist die komponentenweise Darstellung der Gleichung $u_h = L_h^{-1} f_h$. Der Faktor h^2 kompensiert h^{-2} aus (4.4.5a). Er wurde eingeführt, damit die Summation $h^2 \sum$ in (4.4.6) das Integral \int_{Ω} approximiert. Da in §3.2 die Poisson-Gleichung $\Delta u = f$, hier jedoch $-\Delta u = f$ betrachtet wird, unterscheiden sich die rechten Seiten in (3.2.7) und (4.4.6) durch das Vorzeichen.

Auch die diskrete Greensche Funktion ist positiv (vgl. (3.2.9)):

Bemerkung 4.4.8 $0 < g_h(\mathbf{x}, \xi) \leq h^{-2}/8$ für $\mathbf{x}, \xi \in \Omega_h$.

Beweis. Die obere Schranke folgt aus $g_h(\mathbf{x}, \xi) \leq \|g_h(\cdot, \xi)\|_{\infty} \leq \|L_h^{-1}\|_{\infty} \|\delta_{\xi}\|_{\infty}$, $\|\delta_{\xi}\|_{\infty} = h^{-2}$ und (4.4.1c). $g_h > 0$ schließt man aus $L_h^{-1} > 0$ (vgl. Satz 4.3.11). ■

Die Abschätzung $g_h(\mathbf{x}, \xi) \leq h^{-2}/8$ ist zu pessimistisch und kann erheblich verbessert werden.

Lemma 4.4.9 Die diskrete Greensche Funktion g_h aus (4.4.5b) erfüllt die Abschätzung

$$0 < g_h(\mathbf{x}, \xi) \leq \frac{\log 2}{4 \log 3} \left(1 - \frac{\log(|\mathbf{x} - \xi|^2 + h^2/2)}{\log 2} \right) \leq \frac{\log \frac{2}{h}}{\log 9} \quad \text{für alle } \mathbf{x}, \xi \in \overline{\Omega}_h. \quad (4.4.7)$$

Die Abschätzung durch $\mathcal{O}(|\log h|)$ spiegelt die logarithmische Singularität der Singularitätenfunktion $s(\mathbf{x}, \xi) = -\log(|\mathbf{x} - \xi|^2)/4\pi$ wider.

Beweis. Zum Beweis der Ungleichung (4.4.7) definiere man

$$s_h(\mathbf{x}, \xi) := \frac{\log 2}{4 \log 3} \left(1 - \frac{\log(|\mathbf{x} - \xi|^2 + h^2/2)}{\log 2} \right). \quad (4.4.8a)$$

a) Zunächst soll $s_h(\mathbf{x}, \xi) \geq 0$ für alle $\mathbf{x} \in \overline{\Omega}_h$, $\xi \in \Omega_h$ gezeigt werden. Da $\xi \in \Omega_h$ ein innerer Gitterpunkt ist, folgt $|x_i - \xi_i| \leq 1 - h$ ($i = 1, 2$) und damit $|\mathbf{x} - \xi|^2 \leq 2(1 - h)^2$. Da $1/2$ die größtmögliche Schrittweite ist, variiert h in $(0, 1/2]$. Hierfür gilt

$$|\mathbf{x} - \xi|^2 + h^2/2 \leq 2(1 - h)^2 + \frac{1}{2}h^2 < 2,$$

und für s_h aus (4.4.8a) folgt die Zwischenbehauptung:

$$s_h(\mathbf{x}, \xi) > \frac{\log 2}{4 \log 3} \left(1 - \frac{\log 2}{\log 2} \right) = 0 \quad \text{für alle } \mathbf{x} \in \overline{\Omega}_h, \xi \in \Omega_h.$$

b) Wir setzen $u_h(\mathbf{x}) := s_h(\mathbf{x}, 0)$, wobei formal alle $\mathbf{x} \in \{(\nu h, \mu h) : \nu, \mu \in \mathbb{Z}\}$ zugelassen werden. Die Fünfpunktformel angewandt auf u_h liefert

$$-\Delta_h u_h = \frac{1}{4 \log 3} \Delta_h \log(|\cdot|^2 + h^2/2). \quad (4.4.8b)$$

Für die Auswertung am Ursprung $\mathbf{x} = \mathbf{0}$ beachte man $\log(|\mathbf{x}|^2 + h^2/2) = \log(h^2/2)$ für $\mathbf{x} = \mathbf{0}$ und $\log(|\mathbf{x}'|^2 + h^2/2) = \log(h^2 + h^2/2) = \log(3h^2/2) = \log 3 + \log(h^2/2)$ für alle Nachbargitterpunkten \mathbf{x}' von $\mathbf{x} = \mathbf{0}$, sodass gemäß (4.4.8b)

$$-(\Delta_h u_h)(\mathbf{0}) = \frac{h^{-2}}{4 \log 3} (-4 \log(h^2/2) + 4 \log(3h^2/2)) = h^{-2} \frac{-\log(h^2/2) + \log(3h^2/2)}{\log 3} = h^{-2}. \quad (4.4.8c)$$

c) Für die Auswertung von $-\Delta_h u_h$ in allgemeinen Gitterpunkten $\mathbf{x} = (x, y)$ erhält man

$$\begin{aligned} -\Delta_h u_h(\mathbf{x}) &= \frac{h^{-2}}{4 \log 3} \left(\begin{aligned} &\log\left(x^2 + (y+h)^2 + \frac{h^2}{2}\right) \\ &+ \log\left((x-h)^2 + y^2 + \frac{h^2}{2}\right) - 4 \log\left(x^2 + y^2 + \frac{h^2}{2}\right) + \log\left((x+h)^2 + y^2 + \frac{h^2}{2}\right) \\ &+ \log\left(x^2 + (y-h)^2 + \frac{h^2}{2}\right) \end{aligned} \right) \\ &= \frac{h^{-2}}{4 \log 3} \log \frac{\left((x+h)^2 + y^2 + \frac{h^2}{2}\right) \left((x-h)^2 + y^2 + \frac{h^2}{2}\right) \left(x^2 + (y+h)^2 + \frac{h^2}{2}\right) \left(x^2 + (y-h)^2 + \frac{h^2}{2}\right)}{\left(x^2 + y^2 + \frac{h^2}{2}\right)^4}. \end{aligned}$$

Zum Nachweis von $-\Delta_h u_h \geq 0$ ist

$$\left((x+h)^2 + y^2 + \frac{h^2}{2}\right) \left((x-h)^2 + y^2 + \frac{h^2}{2}\right) \left(x^2 + (y+h)^2 + \frac{h^2}{2}\right) \left(x^2 + (y-h)^2 + \frac{h^2}{2}\right) \geq \left(x^2 + y^2 + \frac{h^2}{2}\right)^4$$

zu zeigen. Diese Ungleichung folgt aber aus der Identität

$$\begin{aligned} & \left((x+h)^2 + y^2 + \frac{h^2}{2}\right) \left((x-h)^2 + y^2 + \frac{h^2}{2}\right) \left(x^2 + (y+h)^2 + \frac{h^2}{2}\right) \left(x^2 + (y-h)^2 + \frac{h^2}{2}\right) - \left(x^2 + y^2 + \frac{h^2}{2}\right)^4 \\ &= 5h^8 + 4h^6 x^2 + 4h^6 y^2 + 16h^4 x^2 y^2 \geq 0. \end{aligned}$$

d) Wegen $s_h(\mathbf{x}, \xi) = u_h(\mathbf{x} - \xi)$ folgt aus den Teilen b) und c), dass $-(\Delta_h s_h(\cdot, \xi))(\mathbf{x}) \geq 0$ und $-(\Delta_h s_h(\cdot, \mathbf{x}))(\mathbf{x}) = h^{-2}$. Für $\phi_\xi := s_h(\cdot, \xi) - g_h(\cdot, \xi)$ erhält man

$$-\Delta_h \phi_\xi(\mathbf{x}) = -\Delta_h [s_h(\cdot, \xi) - g_h(\cdot, \xi)](\mathbf{x}) \quad \begin{cases} = h^{-2} - h^{-2} = 0 & \text{für } \mathbf{x} = \xi \in \Omega_h, \\ \geq 0 - 0 = 0 & \text{für } \mathbf{x} \neq \xi \in \Omega_h. \end{cases}$$

Am Rand (d.h. für $\mathbf{x} \in \Gamma_h$) gilt nach Teil a)

$$\phi_\xi(\mathbf{x}) = s_h(\mathbf{x}, \xi) - g_h(\mathbf{x}, \xi) = s_h(\mathbf{x}, \xi) \geq 0.$$

Satz 4.4.5b zeigt $\phi_\xi \geq 0$ in Ω_h , also $g_h(\mathbf{x}, \xi) \leq s_h(\mathbf{x}, \xi)$, womit (4.4.7) bewiesen ist. ■

Sei φ_h wie in (4.2.6b) definiert. Die Lösung der diskreten Potentialgleichung

$$-\Delta_h u_h = 0 \quad \text{in } \Omega_h, \quad u_h = \varphi \quad \text{auf } \Gamma_h$$

ist durch $u_h := L_h^{-1} \varphi_h$ gegeben (wenn man die zunächst nur auf Ω_h definierte Gitterfunktion durch φ auf Γ_h fortsetzt; vgl. Bemerkung 4.2.5). Die Darstellung mit Hilfe von g_h lautet

$$u_h(\mathbf{x}) = h^2 \sum_{\xi \in \Omega_h} g_h(\xi, \mathbf{x}) \varphi(\xi) \quad (\mathbf{x} \in \Omega_h).$$

Da $\varphi_h(\mathbf{x})$ in allen randfernen Punkten \mathbf{x} verschwindet, genügt es, die Summe über die randnahen Punkte zu erstrecken. Summiert man über die benachbarten Randpunkte von

$$\Gamma'_h := \{\xi \in \Gamma_h : \xi \text{ ist kein Eckpunkt } (0,0), (1,0), (0,1), (1,1)\}$$

statt über die randnahen Punkte, so ergibt die Definition von φ_h die Darstellung

$$u_h(\mathbf{x}) = -h \sum_{\xi \in \Gamma'_h} \partial_n^- \bar{g}_h(\xi, \mathbf{x}) \varphi(\xi) \quad (\mathbf{x} \in \Omega_h), \quad (4.4.9)$$

wobei ∂_n^- die Rückwärtsdifferenz bezüglich der Normalenrichtung \vec{n} ist:

$$\partial_n^- \bar{g}_h(\xi, \mathbf{x}) = h^{-1} [\bar{g}_h(\xi, \mathbf{x}) - \bar{g}_h(\xi - h\vec{n}, \mathbf{x})]$$

(man beachte $\bar{g}_h(\xi, \mathbf{x}) = 0$ für $\xi \in \Gamma'_h \subset \Gamma_h$). Das Argument $\xi - h\vec{n}$ durchläuft alle randnahen Punkte.

Bemerkung 4.4.10 Gleichung (4.4.9) entspricht der Darstellung in Satz 3.2.10. Die Summation $h \sum_{\xi \in \Gamma'_h}$ approximiert das dortige Integral \int_Γ .

Abschließend soll die Abschätzung der Lösung $u_h = L_h^{-1} q_h$ durch $\|u_h\|_\infty \leq \|L_h^{-1}\|_\infty \|q_h\|_\infty \leq \|q_h\|_\infty / 8$ (vgl. (4.4.1c)) genauer betrachtet werden. Gemäß (4.2.6b) enthält $q_h = f_h + \varphi_h$ die rechte Seite $f_h(\mathbf{x}) = f(\mathbf{x})$ der diskreten Poisson-Gleichung und die in φ_h versteckten Randwerte $\varphi(\mathbf{x})$, $\mathbf{x} \in \Gamma_h$. Der folgende Satz gibt eine Abschätzung an, in der diese Anteile separiert sind.

Satz 4.4.11 $q_h = f_h + \varphi_h$ sei gemäß (4.2.6b) aus f und φ konstruiert. Die diskrete Lösung $u_h = L_h^{-1} q_h$ der Poisson-Randwertaufgabe lässt sich abschätzen durch

$$\|u_h\|_\infty \leq \|L_h^{-1}\|_\infty \max_{\mathbf{x} \in \Omega_h} |f(\mathbf{x})| + \max_{\xi \in \Gamma'_h} |\varphi(\xi)| \leq \frac{1}{8} \max_{\mathbf{x} \in \Omega_h} |f(\mathbf{x})| + \max_{\xi \in \Gamma'_h} |\varphi(\xi)|. \quad (4.4.10)$$

Beweis. Setze $u'_h := L_h^{-1}f_h$ und $u''_h := L_h^{-1}\varphi_h$. Die Abschätzung des ersten Summanden in $u_h = u'_h + u''_h$ ergibt den ersten Summanden in (4.4.10). Zur Abschätzung von u''_h verwende man Ungleichung (4.4.4a) mit $u_h^1 = u''_h$, $\varphi^1 = \varphi$ und $u_h^2 = 0$, $\varphi^2 = 0$. ■

Die entsprechende Ungleichung

$$\|u_h\|_\infty \leq C \|f\|_\infty + \|\varphi\|_\infty \quad (4.4.11)$$

für die Lösung der Randwertaufgabe (4.2.2a,b) wurde bisher noch nicht erwähnt, wird aber in allgemeinerem Zusammenhang in §5.1.3 bewiesen werden.

Die Maximumnorm $\|\cdot\|_\infty$ lässt sich in (4.4.10) durch die Euklidischen Normen

$$\|u_h\|_{2,\Omega_h} := \sqrt{h^2 \sum_{\mathbf{x} \in \Omega_h} |u_h(\mathbf{x})|^2}, \quad \|\varphi\|_{2,\Gamma'_h} := \sqrt{h^2 \sum_{\xi \in \Gamma'_h} |\varphi(\xi)|^2}$$

ersetzen. Dabei entsprechen sich jeweils $h^2 \sum_{\Omega_h}$ und \int_Ω einerseits und $h \sum_{\Gamma'_h}$ und \int_Γ andererseits.

Satz 4.4.12 *Unter den Voraussetzungen des Satzes 4.4.11 gilt*

$$\|u_h\|_{2,\Omega_h} \leq \|L_h^{-1}\|_2 \|f_h\|_{2,\Omega_h} + \frac{1}{\sqrt{2}} \|\varphi\|_{2,\Gamma'_h} \leq \frac{1}{16} \|f_h\|_{2,\Omega_h} + \frac{1}{\sqrt{2}} \|\varphi\|_{2,\Gamma'_h}. \quad (4.4.12)$$

Beweis. 1) Es genügt, den Fall der Potentialgleichung (d.h. $f = 0$) zu betrachten. Die Beschränkung von φ auf Γ'_h erzeuge die Gitterfunktion $\phi_h : \phi_h(\mathbf{x}) = \varphi(\mathbf{x})$ für $\mathbf{x} \in \Gamma'_h$. Die Abbildung $\phi_h \rightarrow u_h = L_h^{-1}\varphi_h$ sei durch die rechteckige Matrix A vermittelt: $u_h = A\phi_h$. Gemäß Gleichung (4.4.9) lauten die Koeffizienten von A

$$a_{\mathbf{x}\xi} = -h\partial_n^- \bar{g}_h(\xi, \mathbf{x}) = g_h(\xi - h\vec{n}, \mathbf{x}) = g_h(\mathbf{x}, \xi - h\vec{n}) \quad \text{für } \mathbf{x} \in \Omega_h, \xi \in \Gamma'_h.$$

Da $A \geq 0$ (vgl. Bemerkung 4.4.8), erhält man die Zeilensummennorm $\|A\|_\infty := \max_{\mathbf{x}} \sum_{\xi} a_{\mathbf{x}\xi}$ als $\|A\phi_h\|_\infty$ für die Wahl $\phi_h(\mathbf{x}) = 1$ in allen $\mathbf{x} \in \Gamma'_h$. Die Lösung $v_h = A\phi_h$ lautet dann $v_h = \mathbf{1}$ (warum?), sodass

$$\|A\|_\infty = \|A\phi_h\|_\infty = \|\mathbf{1}\|_\infty = 1$$

folgt.

2) Die Spaltensummen von A sind $s(\xi) := \sum_{\mathbf{x} \in \Omega_h} a_{\mathbf{x}\xi} = \sum_{\mathbf{x} \in \Omega_h} g_h(\mathbf{x}, \xi - h\vec{n})$ für $\xi \in \Gamma'_h$. Die Gitterfunktion $v_h := h^{-2}L_h^{-1}\mathbf{1}$ hat in den randnahen Punkten $\xi - h\vec{n}$ die Werte

$$s(\xi) = v_h(\xi - h\vec{n}) \quad (\xi \in \Gamma'_h, \vec{n} \text{ Normalenrichtung})$$

wie sich aus Bemerkung 4.4.7 ergibt. Sei $\xi = (\xi_1, \xi_2) \in \Gamma'_h$ ein Punkt des linken oder rechten Randes (d.h. $\xi_1 = 0$ oder 1). Wie im Beweis von (4.4.1c) erwähnt, gilt $w_h(x, y) \geq \mathbf{1}$ für $w_h(x, y) := x(1-x)/2$. Da $L_h^{-1} \geq 0$, ist auch $w_h \geq L_h^{-1}\mathbf{1}$ und somit $v_h < h^{-2}w_h$. Insbesondere folgt die Abschätzung

$$s(\xi) = v_h(\xi - h\vec{n}) \leq h^{-2}w_h(\xi - h\vec{n}) = h^{-2}h(1-h)/2 \leq h^{-1}/2$$

im randnahen Punkt $\xi - h\vec{n} = (h, \xi_2)$ oder $(1-h, \xi_2)$. Für einen Punkt ξ aus dem unteren oder oberen Rand erhält man die gleiche Abschätzung, wenn man $w_h(x, y) := y(1-y)/2$ verwendet. Da die Spaltensummen $s(\xi)$ die Zeilensummen von A^\top sind, haben wir

$$\|A^\top\|_\infty = \max\{s(\xi) : \xi \in \Gamma'_h\} \leq h^{-1}/2$$

bewiesen.

3) Es ist $\|A^\top A\|_2 = \rho(A^\top A) \leq \|A^\top A\|_\infty \leq \|A^\top\|_\infty \|A\|_\infty \leq h^{-1}/2$ (vgl. Übungsaufgabe 4.3.20a, (4.3.10d), (4.3.10a)), sodass die Lösung $u_h = L_h^{-1}\varphi_h = A\phi_h$ folgender Abschätzung genügt:

$$\begin{aligned} \|u_h\|_{2,\Omega_h}^2 &= h^2 \sum_{\mathbf{x} \in \Omega_h} |u_h(\mathbf{x})|^2 = h^2 \sum_{\mathbf{x} \in \Omega_h} |(A\phi_h)(\mathbf{x})|^2 = h^2 \sum_{\mathbf{x} \in \Gamma'_h} \phi_h(\mathbf{x}) (A^\top A\phi_h)(\mathbf{x}) \\ &\leq h^2 \|A^\top A\|_2 \sum_{\mathbf{x} \in \Gamma'_h} \phi_h^2(\mathbf{x}) \leq \frac{h}{2} \sum_{\mathbf{x} \in \Gamma'_h} \phi_h^2(\mathbf{x}) = \frac{1}{2} \|\phi_h\|_{2,\Gamma'_h}^2. \end{aligned}$$

Da $\|\varphi\|_{2,\Gamma'_h} = \|\phi_h\|_{2,\Gamma'_h}$, folgt die Behauptung. ■

4.5 Konvergenz

U_h sei der Vektorraum der Gitterfunktionen auf $\overline{\Omega}_h$. Die diskrete Lösung $u_h \in U_h$ und die kontinuierliche Lösung $u \in C^0(\overline{\Omega})$ lassen sich wegen der verschiedenen Definitionsbereiche nicht unmittelbar vergleichen. Bei Differenzenverfahren ist es üblich, beide Funktionen auf dem Gitter $\overline{\Omega}_h$ zu vergleichen. Dazu hat man die Lösung u mittels einer “*Restriktion*”

$$u \mapsto R_h u \in U_h \quad (u: \text{kontinuierliche Lösung}) \quad (4.5.1)$$

in U_h abzubilden. Im Folgenden werden wir R_h als Beschränkung auf $\overline{\Omega}_h$ wählen:

$$(R_h u)(\mathbf{x}) = u(\mathbf{x}) \quad \text{für alle } \mathbf{x} \in \overline{\Omega}_h. \quad (4.5.2)$$

Der Grenzübergang $h \rightarrow 0$ wird wie folgt präzisiert. Sei $H \subset \mathbb{R}_+$ eine Teilmenge mit Häufungspunkt null: $0 \in \overline{H}$. Zum Beispiel bilden die bisher betrachteten Schrittweiten die Menge $H = \{1/n : n \in \mathbb{N}\}$. Für jedes $h \in H$ sei U_h mit einer Norm $\|\cdot\|_h$ versehen.

Definition 4.5.1 (Konvergenz) Die diskreten Lösungen $u_h \in U_h$ konvergieren (bezüglich des Normensystems $\|\cdot\|_h$, $h \in H$) gegen u , falls

$$\|u_h - R_h u\|_h \rightarrow 0. \quad (4.5.3a)$$

Konvergenz der Ordnung k liegt vor, wenn

$$\|u_h - R_h u\|_h = \mathcal{O}(h^k). \quad (4.5.3b)$$

Der Nachweis der Konvergenz wird üblicherweise mit Hilfe der Begriffe “Stabilität” und “Konsistenz” geführt. Ungleichungen der Form $\|u_h - R_h u\|_h \leq \dots$, wie sie in diesem Abschnitt auftreten, nennt man auch *Fehlerabschätzungen*, da $u_h - R_h u$ der *Diskretisierungsfehler* ist.

Die Diskretisierung $\{L_h : h \in H\}$ heißt *stabil* bezüglich $\|\cdot\|_\infty$, falls

$$\sup_{h \in H} \|L_h^{-1}\|_\infty < \infty. \quad (4.5.4)$$

Für die in §4.2 definierte Diskretisierung ist die Stabilität in (4.4.1c) bezüglich der Zeilensummennorm und in (4.4.1d) bezüglich der Spektralnorm nachgewiesen worden. Die Gitterfunktion f_h in $-\Delta_h u_h = f_h$ ist die Restriktion

$$f_h = \tilde{R}_h f \quad (4.5.5a)$$

von f , wobei in (4.2.6b) \tilde{R}_h als die Beschränkung auf Ω_h gewählt wurde:

$$(\tilde{R}_h f)(\mathbf{x}) = f(\mathbf{x}) \quad \text{für alle } \mathbf{x} \in \Omega_h. \quad (4.5.5b)$$

Die Schreibweise \tilde{R}_h soll deutlich machen, dass eine Wahl $\tilde{R}_h \neq R_h$ möglich ist.

Die Formulierung der Konsistenz wird einfacher, wenn man anstelle der Matrix L_h den zugrundeliegenden Differenzenoperator $-\Delta_h$ betrachtet (vgl. Bemerkung 4.2.5). Es sei D_h ein allgemeiner Differenzenoperator, d.h. $-D_h u_h = f_h$ diskretisiere $-\Delta u = f$. Die durch das Tripel (D_h, R_h, \tilde{R}_h) beschriebene Diskretisierung heißt bezüglich $\|\cdot\|_\infty$ konsistent von der Ordnung k (konsistent zum Laplace-Operator), falls

$$\|D_h R_h u - \tilde{R}_h \Delta u\|_\infty \leq K h^k \|u\|_{C^{k+2}(\overline{\Omega})} \quad (4.5.6)$$

für alle $u \in C^{k+2}(\overline{\Omega})$. Dabei ist die Konstante K von h und u unabhängig.

Bemerkung 4.5.2 R_h und \tilde{R}_h seien durch (4.5.2) und (4.5.5b) gegeben. Die Fünfpunktformel Δ_h ist konsistent von der Ordnung 2: Abschätzung (4.5.6) gilt mit $k = 2$ und $K = 1/6$.

Beweis. Die Entwicklung (4.1.3c) läßt sich in x - und y -Richtung anwenden und liefert

$$\Delta_h R_h u(x, y) = \Delta u(x, y) + h^2(R_x + R_y) \quad \text{mit } |R_x|, |R_y| \leq \frac{1}{12} \|u\|_{C^4(\overline{\Omega})}. \quad (4.5.7)$$

■

Satz 4.5.3 Die Diskretisierung (D_h, R_h, \tilde{R}_h) sei konsistent von der Ordnung k . Die zum Differenzenoperator D_h gehörende Matrix L_h (vgl. Bemerkung 4.2.5) sei stabil bezüglich $\|\cdot\|_\infty$. Dann ist das Verfahren konvergent von der Ordnung k , falls $u \in C^{k+2}(\bar{\Omega})$.

Beweis. $w_h := u_h - R_h u$ erfüllt die Differenzengleichungen

$$-D_h w_h = D_h u_h - D_h R_h u = f_h + D_h R_h u = \tilde{R}_h f + D_h R_h u = D_h R_h u - \tilde{R}_h \Delta u \quad \text{in } \Omega_h.$$

Wegen der homogenen Randbedingung $w_h = 0$ auf Γ_h gilt $w_h = L_h^{-1}(D_h R_h u - \tilde{R}_h \Delta u)$ (vgl. Bemerkung 4.2.1). (4.5.4) und (4.5.6) liefern (4.5.3b). ■

Setzt man die konkreten Werte $\|L_h^{-1}\|_\infty \leq 1/8$ und $K = 1/6$ ein, ergibt sich der

Zusatz 4.5.4 Die kontinuierliche Lösung des Randwertproblems (4.2.2a,b) gehöre zu $C^4(\bar{\Omega})$. u_h sei die in (4.2.4a,b) definierte diskrete Lösung. Dann konvergiert u_h von zweiter Ordnung gegen u :

$$\|u_h - R_h u\|_\infty \leq \frac{h^2}{48} \|u\|_{C^4(\bar{\Omega})}. \quad (4.5.8)$$

Die Voraussetzung $u \in C^4(\bar{\Omega})$ läßt sich abschwächen:

Zusatz 4.5.5 Unter der Voraussetzung $u \in C^{3,1}(\bar{\Omega})$ gilt ebenfalls

$$\|u_h - R_h u\|_\infty \leq \frac{h^2}{48} \|u\|_{C^{3,1}(\bar{\Omega})}. \quad (4.5.8')$$

Beweis. Der Restterm R_4 aus (4.1.4b) kann auch als

$$R_4 = h^{-4} \int_x^{x \pm h} [u'''(\xi) - u'''(x)] \frac{(x \pm h - \xi)^2}{2!} d\xi$$

geschrieben werden. Die Lipschitz-Abschätzung $|u'''(\xi) - u'''(x)| \leq |\xi - x| \|u\|_{C^{3,1}(\bar{\Omega})}$ liefert $R_4 \leq \frac{\|u\|_{C^{3,1}(\bar{\Omega})}}{4!}$, sodass in (4.5.7), (4.5.6) und (4.5.8) die Norm $\|u\|_{C^4(\bar{\Omega})}$ durch $\|u\|_{C^{3,1}(\bar{\Omega})}$ ersetzt werden kann. ■

Schwächt man dagegen $u \in C^{3,1}(\bar{\Omega})$ durch $u \in C^s(\bar{\Omega})$ mit $2 < s < 4$ weiter ab, erhält man eine schwächere Konvergenzordnung.

Zusatz 4.5.6 Unter der Voraussetzung $u \in C^s(\bar{\Omega})$, $2 < s \leq 4$, konvergiert u_h von der Ordnung $s - 2$:

$$\|u_h - R_h u\|_\infty \leq K_s h^{s-2} \|u\|_{C^s(\bar{\Omega})}, \quad (4.5.8'')$$

wobei $K_s := 1/[2s(s-1)]$ für $2 \leq s \leq 3$ und $K_s := 1/[2s(s-1)(s-2)]$ für $3 \leq s \leq 4$.

Der Beweis ergibt sich aus

Übungsaufgabe 4.5.7 Man zeige $\|\Delta_h R_h u - \tilde{R}_h \Delta u\|_\infty \leq 8K_s h^{s-2} \|u\|_{C^s(\bar{\Omega})}$, $2 < s \leq 4$.

Wenn auch die Konvergenzbeweise einfach sind, so sind die Resultate doch unbefriedigend. Wie schon aus Beispiel 2.1.3 hervorgeht, erfüllt die kontinuierliche Lösung der Randwertaufgabe (4.2.2a,b) im Allgemeinen nicht einmal $u \in C^2(\bar{\Omega})$, obwohl mindestens $u \in C^s(\bar{\Omega})$ mit $s > 2$ in Zusatz 4.5.6 für die Konvergenz benötigt wird. Zu stärkeren Resultaten führt eine Analyse, die in §9.2 behandelt wird. Dass auch unter geringeren Voraussetzungen Fehler der Größenordnung $\mathcal{O}(h^2)$ auftreten, zeigt das folgende

Beispiel 4.5.8 Löst man die Differenzengleichung (4.2.4a,b) für $-\Delta u = 1$ in $\Omega = (0,1) \times (0,1)$ und $u = 0$ auf Γ , so erhält man im Mittelpunkt $x = y = 1/2$ die Werte $u_h(1/2, 1/2)$, die in der ersten Spalte der Tabelle 1 wiedergegeben sind. Die exakte Lösung $u(1/2, 1/2) = 0.0736713 \dots$ ergibt sich aus einer Darstellung, die man in Beispiel 8.1.12 findet. Die Quotienten $\varepsilon_h/\varepsilon_{2h}$ der Fehler $\varepsilon_h = u(1/2, 1/2) - u_h(1/2, 1/2)$ approximieren $1/4$. Dies beweist $u_h(1/2, 1/2) = u(1/2, 1/2) + \mathcal{O}(h^2)$, obwohl $u \notin C^2(\bar{\Omega})$. In $(1/2, 1/2)$ hat u_h außerdem

die asymptotische Entwicklung $u_h(1/2, 1/2) = u(1/2, 1/2) + h^2 e(1/2, 1/2) + \mathcal{O}(h^4)$. Den von h unabhängigen Fehlerterm $e(1/2, 1/2)$ eliminiert man durch die Richardson⁴-Extrapolation

$$u_{h,2h}(1/2, 1/2) := \frac{1}{3} [4u_h(1/2, 1/2) - u_{2h}(1/2, 1/2)].$$

Die extrapolierten Werte sind schon für $h = 1/16$ sehr genau (vgl. letzte Spalte der Tabelle 4.5.1).

h	$u_h(\frac{1}{2}, \frac{1}{2})$	$u_{h,2h}(\frac{1}{2}, \frac{1}{2})$	ε_h	Quotient	$\varepsilon_{h,2h}$	Quotient
1/8	0.0727826		8.89 ₁₀₋₄			
1/16	0.07344576	0.0736668	2.26 ₁₀₋₄	0.250	4.5 ₁₀₋₆	
1/32	0.0736147373	0.07367106	5.66 ₁₀₋₅	0.251	2.7 ₁₀₋₇	0.06
1/64	0.0736571855	0.07367133	1.41 ₁₀₋₅	0.249		

Tabelle 4.5.1. Differenzenlösung zu Beispiel 4.5.8 ($\varepsilon_{h,[2h]} := |u(\frac{1}{2}, \frac{1}{2}) - u_{h,[2h]}(\frac{1}{2}, \frac{1}{2})|$)

4.6 Differenzenverfahren höherer Ordnung

Die Fünfpunktformel (4.2.11) ist von zweiter Ordnung. Auch wenn die Lösung u zu $C^s(\bar{\Omega})$ mit $s > 4$ gehört, ergibt sich keine bessere Schranke für $\Delta_h R_h u - R_h \Delta u$ als $\mathcal{O}(h^2)$. Eine naheliegende Methode, Differenzenverfahren höherer Ordnung zu konstruieren, ist die folgende. Als Ansatz für eine Diskretisierung der zweiten Ableitung u'' wählt man

$$(D_h u_h)(x) = h^{-2} \sum_{\nu=-k}^k c_\nu u_h(x + \nu h).$$

Die Taylor-Entwicklung liefert

$$(D_h R_h u_h)(x) = \sum_{\mu=0}^{2k} a_\mu h^{\mu-2} u^{(\mu)}(x) + \mathcal{O}(h^{2k-1}), \quad a_\mu = \frac{1}{\mu!} \sum_{\nu=-k}^k c_\nu \nu^\mu.$$

Die $2k+1$ Gleichungen $a_0 = a_1 = a_3 = a_4 = \dots = a_{2k} = 0$ und $a_2 = 1$ bilden ein lineares System für die $2k+1$ unbekannten Koeffizienten c_ν . Für $k=1$ erhält man die übliche Differenzenformel (4.1.3c), für $k=2$ ergibt sich eine Differenz vierter Ordnung:

$$h^2 (D_h u_h)(x) = -\frac{1}{12} [u_h(x-2h) + u_h(x+2h)] + \frac{4}{3} [u_h(x-h) + u_h(x+h)] - \frac{5}{2} u_h(x).$$

Wendet man diese Näherung für u'' auf die x - und y -Koordinaten an, ergibt sich für $-\Delta$ der Differenzenstern

$$\frac{h^{-2}}{12} \begin{bmatrix} & & 1 & & \\ & & -16 & & \\ 1 & -16 & 60 & -16 & 1 \\ & & -16 & & \\ & & 1 & & \end{bmatrix} \quad (4.6.1)$$

(vgl. (4.2.12)). Das Differenzenschema (4.6.1) ist zwar von vierter Ordnung, bereitet aber Schwierigkeiten in den randnahen Punkten. Um zum Beispiel in $(h, h) \in \Omega_h$ die Differenzenformel aufzustellen, benötigt man Werte $u_h(-h, h)$ und $u_h(h, -h)$ außerhalb von $\bar{\Omega}_h$ (vgl. Abbildung 4.6.1). Eine Möglichkeit wäre, das Schema (4.6.1) nur in randfernen Punkten zu verwenden und in randnahen Punkten die Fünfpunktformel (4.2.11) zu benutzen.

Die eben beschriebenen Komplikationen treten nicht auf, wenn man sich auf kompakte Neunpunktformeln beschränkt; als solche bezeichnet man Differenzenverfahren (4.2.12), die durch

$$c_{\alpha\beta} \neq 0 \quad \text{nur für } -1 \leq \alpha, \beta \leq 1$$

⁴Lewis Fry Richardson, geb. 11. Okt. 1881 in Newcastle upon Tyne, gest. 30. Sept. 1953 in Kilmun (Schottland)

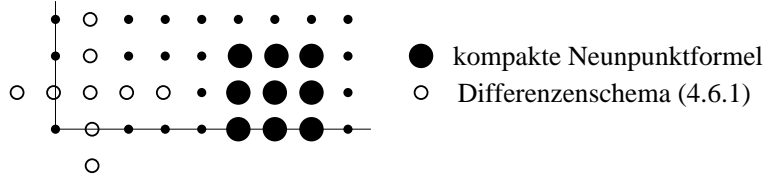


Abb. 4.6.1. Differenzenschema (4.6.1) und kompakte Neunpunktformel

charakterisiert sind (vgl. Abbildung 4.6.1). Ein Ansatz mit den 9 freien Parametern $c_{\alpha,\beta}$ ($-1 \leq \alpha, \beta \leq 1$) liefert allerdings ein negatives Resultat: Es gibt keine kompakte Neunpunktformel mit $D_h u = -\Delta u + \mathcal{O}(h^3)$. In diesem Sinne ist die Fünfpunktformel schon optimal. Dennoch lassen sich Neunpunktverfahren vierter Ordnung erreichen, wenn man auch die rechte Seite f_h des Gleichungssystems (4.2.6a,b) geeignet wählt. Wendet man das kompakte Neunpunktschema

$$D_h := \frac{h^{-2}}{6} \begin{bmatrix} -1 & -4 & -1 \\ -4 & 20 & -4 \\ -1 & -4 & -1 \end{bmatrix} \quad (4.6.2)$$

auf $u \in C^8(\bar{\Omega})$ an, ergibt die Taylor-Entwicklung

$$-D_h u = -\Delta u - \frac{h^2}{12} \Delta^2 u - \frac{h^4}{360} \left[\frac{\partial^4}{\partial x^4} + 4 \frac{\partial^4}{\partial x^2 \partial y^2} + \frac{\partial^4}{\partial y^4} \right] \Delta u + \mathcal{O}(h^6). \quad (4.6.3)$$

Wesentlich ist, dass sich der Fehlerterm durch $-\Delta u$ und damit durch f ausdrücken lässt. Für die spezielle Wahl der Restriktion \tilde{R}_h durch

$$f_h = \tilde{R}_h f := \frac{1}{6} \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 4 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} f, \quad (4.6.4)$$

d.h.

$$f_h(x, y) = (\tilde{R}_h f)(x, y) := \frac{1}{12} [f(x-h, y) + f(x+h, y) + f(x, y-h) + f(x, y+h) + 8f(x, y)] \quad (4.6.4')$$

erhält man die Entwicklung

$$f_h(x, y) = f(x, y) + \frac{h^2}{12} \Delta f(x, y) + \mathcal{O}(h^4), \quad (4.6.5)$$

die wegen $f = -\Delta u$ bis auf $\mathcal{O}(h^4)$ mit (4.6.3) übereinstimmt. Die Matrix L_h des Gleichungssystems, das sich nach Elimination der Randwerte $u_h(\mathbf{x}) = \varphi(\mathbf{x})$, $\mathbf{x} \in \Gamma_h$, ergibt, hat die Koeffizienten

$$L_{\mathbf{x}\xi} = h^{-2} \begin{cases} 20/6 & \text{falls } \mathbf{x} = \xi, \\ -1/6 & \text{falls } \mathbf{x} - \xi = (\pm h, \pm h) \text{ oder } \mathbf{x} - \xi = (\pm h, \mp h), \\ -4/6 & \text{falls } \mathbf{x} - \xi = (\pm h, 0) \text{ oder } \mathbf{x} - \xi = (0, \pm h), \\ 0 & \text{sonst.} \end{cases} \quad (4.6.6)$$

Die rechte Seite des Gleichungssystems $L_h u_h = q_h$ ist

$$q_h := f_h + \varphi_h, \quad f_h = \tilde{R}_h f \text{ gemäß (4.6.4')}, \quad \varphi_h := \sum_{\xi \in \Gamma_h} L_{\mathbf{x}\xi} \varphi(\xi). \quad (4.6.7)$$

Die Diskretisierung (D_h, R_h, \tilde{R}_h) mit D_h aus (4.6.2), R_h aus (4.5.2), \tilde{R}_h aus (4.6.4') heißt *Mehrstellenverfahren* (vgl. Collatz [25]).

Übungsaufgabe 4.6.1 D_h und L_h seien durch (4.6.2) bzw. (4.6.6) definiert. Man beweise:

- L_h ist eine M-Matrix;
- $\|L_h^{-1}\|_{\infty} \leq 1/8$ (Stabilität), $\|L_h\| \leq 20h^{-2}/3$,
- $\|D_h R_h u - R_h \Delta u\|_{\infty} \leq -\frac{11h^4}{180} \|u\|_{C^6(\bar{\Omega})}$, falls $u \in C^6(\bar{\Omega})$ (Konsistenz).

Satz 4.6.2 (Konvergenz des Mehrstellenverfahrens) Ω_h sei wie in §4.2 definiert. u_h sei die Lösung des Mehrstellenverfahrens $L_h u_h = q_h$ (L_h aus (4.6.6), q_h aus (4.6.7)). Die Lösung u der Randwertaufgabe (4.2.2a,b) gehöre zu $C^6(\bar{\Omega})$ oder $C^{5,1}(\bar{\Omega})$. Dann gilt die Fehlerabschätzung

$$\|u_h - R_h u\|_\infty \leq \frac{11 h^4}{1440} \|u\|_{C^6(\bar{\Omega})} \quad \text{bzw.} \quad \|u_h - R_h u\|_\infty \leq \frac{11 h^4}{1440} \|u\|_{C^{5,1}(\bar{\Omega})} \quad (4.6.8a)$$

Im Falle der Potentialgleichung (d.h. $f = 0$) gilt sogar

$$\|u_h - R_h u\|_\infty \leq K h^6 \|u\|_{C^8(\bar{\Omega})}, \quad \|u_h - R_h u\|_\infty \leq K h^6 \|u\|_{C^{7,1}(\bar{\Omega})}, \quad (4.6.8b)$$

falls $u \in C^8(\bar{\Omega})$ bzw. $u \in C^{7,1}(\bar{\Omega})$.

Beweis. (4.6.8a) ergibt sich wie in Satz 4.5.3 aus Übungsaufgabe 4.6.1b,c. Im Falle $f = 0$ verschwindet auch der $\mathcal{O}(h^4)$ -Term in (4.6.3). ■

4.7 Die Diskretisierung der Neumann-Randwertaufgabe

Die Dirichlet-Randbedingung $u(\mathbf{x}) = \varphi(\mathbf{x})$ wurde direkt im Differenzenverfahren verwendet, eine Diskretisierung erübrigte sich. Anders ist es bei der Neumann-Randwertaufgabe

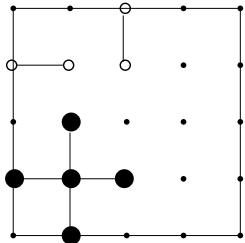
$$-\Delta u = f \quad \text{in } \Omega = (0,1) \times (0,1), \quad \frac{\partial u}{\partial n} = \varphi \quad \text{auf } \Gamma. \quad (4.7.1)$$

Die Normalenableitung, die explizit

$$\begin{aligned} \frac{\partial u}{\partial n} &= -u_y \quad \text{für } \mathbf{x} = (x_1, 0) \in \Gamma, & \frac{\partial u}{\partial n} &= u_y \quad \text{für } \mathbf{x} = (x_1, 1) \in \Gamma, \\ \frac{\partial u}{\partial n} &= -u_x \quad \text{für } \mathbf{x} = (0, x_2) \in \Gamma, & \frac{\partial u}{\partial n} &= u_x \quad \text{für } \mathbf{x} = (1, x_2) \in \Gamma \end{aligned} \quad (4.7.2)$$

lautet, muss ebenso wie der Laplace-Operator durch eine Differenz ersetzt werden. Wir werden drei unterschiedliche Diskretisierungen untersuchen.

4.7.1 Einseitige Differenz für $\partial u / \partial n$



● Diskretisierung der Poisson-Gleichung

○ Diskretisierung der Neumann-Bedingung

Die Poisson-Gleichung führt auf die $(n-1)^2 = (1/h-1)^2$ Gleichungen

$$(-\Delta_h u_h)(\mathbf{x}) = f(\mathbf{x}) \quad \text{für alle } \mathbf{x} \in \Omega_h, \quad (4.7.3)$$

die die Werte von $u_h(\mathbf{x})$ für alle $\mathbf{x} \in \bar{\Omega}'_h$ benötigen, wobei

$$\bar{\Omega}'_h := \Omega_h \cup \Gamma'_h, \quad \Gamma'_h := \Gamma_h \setminus \{(0,0), (0,1), (1,0), (1,1)\}.$$

Um weitere $4(n-1)$ Gleichungen für $\{u_h(\mathbf{x}) : \mathbf{x} \in \Gamma'_h\}$ zu erhalten, ersetzen wir in allen $\mathbf{x} \in \Gamma'_h$ die Normalenableitung $\partial u / \partial n = \varphi$ durch die Rückwärtsdifferenz

$$(\partial_n^- u_h)(\mathbf{x}) = \frac{1}{h} [u_h(\mathbf{x}) - u_h(\mathbf{x} - h\vec{n})] = \varphi(\mathbf{x}) \quad \text{für } \mathbf{x} \in \Gamma'_h. \quad (4.7.4)$$

Setzt man die entsprechenden Normalenrichtungen n für die vier Quadratseiten ein, erhält man

$$\left. \begin{aligned} \frac{1}{h} [u_h(x, 0) - u_h(x, h)] &= \varphi(x, 0) \\ \frac{1}{h} [u_h(x, 1) - u_h(x, 1-h)] &= \varphi(x, 1) \end{aligned} \right\} \quad \text{für } x = h, 2h, \dots, 1-h, \quad (4.7.4')$$

$$\left. \begin{aligned} \frac{1}{h} [u_h(0, y) - u_h(h, y)] &= \varphi(0, y) \\ \frac{1}{h} [u_h(1, y) - u_h(1-h, y)] &= \varphi(1, y) \end{aligned} \right\} \quad \text{für } y = h, 2h, \dots, 1-h.$$

Die Gleichungen (4.7.3) und (4.7.4) ergeben $(n+1)^2 - 4$ Gleichungen für ebenso viele Unbekannte.

Übungsaufgabe 4.7.1 Nach einer Umskalierung der Gleichungen (4.7.4) zu $h^{-1}\partial_n^- u_h(\mathbf{x}) = h^{-1}\varphi(\mathbf{x})$, $\mathbf{x} \in \Gamma'_h$, bilden diese Gleichungen zusammen mit (4.7.3) ein System $L_h u_h = q_h$. Man zeige: L_h ist symmetrisch und erfüllt (4.3.1a).

Wie beim Dirichlet-Problem können die Variablen $u_h(\mathbf{x})$, $\mathbf{x} \in \Gamma'_h$, mit Hilfe von (4.7.4) aus (4.7.3) eliminiert werden. Beispielsweise wird Gleichung (4.7.3) im randnahen Punkt (h, y) zu

$$\frac{1}{h^2} [3u_h(h, y) - u_h(h, y - h) - u_h(h, y + h) - u_h(2h, y)] = f(h, y) + \frac{1}{h}\varphi(0, y).$$

Der Stern $h^{-2} \begin{bmatrix} -1 & -1 \\ -1 & 4 & -1 \end{bmatrix}$ wird demnach zu

$$h^{-2} \begin{bmatrix} & -1 \\ 0 & 3 & -1 \\ & -1 \end{bmatrix}, \quad h^{-2} \begin{bmatrix} & -1 \\ -1 & 3 & 0 \\ & -1 \end{bmatrix}, \quad h^{-2} \begin{bmatrix} & 0 \\ -1 & 3 & -1 \\ & -1 \end{bmatrix}, \quad h^{-2} \begin{bmatrix} & -1 \\ -1 & 3 & -1 \\ & 0 \end{bmatrix}$$

in der Nähe des linken, rechten, oberen bzw. unteren Randes. In den Eckpunkten sind sogar zwei Randwerte zu ersetzen, sodass z.B. in $(h, h) \in \Omega_h$ der Stern $h^{-2} \begin{bmatrix} 0 & -1 \\ 0 & 2 & -1 \end{bmatrix}$ lautet. Abgesehen vom Sonderfall $h = 1/2$ erhält man für die $(n-1)^2$ Werte $u_h(\mathbf{x})$, $\mathbf{x} \in \Omega_h$, das Gleichungssystem

$$L_h u_h = q_h \quad \text{mit} \quad (4.7.5a)$$

$$L_{\mathbf{x}\mathbf{x}} = \begin{cases} 4/h^2 & \text{falls } \mathbf{x} \in \Omega_h \text{ randfern,} \\ 2/h^2 & \text{falls } \mathbf{x} \in \{(h, h), (h, 1-h), (1-h, h), (1-h, 1-h)\}, \\ 3/h^2 & \text{sonst,} \end{cases} \quad (4.7.5b)$$

$$L_{\mathbf{x}\xi} = \begin{cases} -1/h^2 & \text{falls } \mathbf{x}, \xi \in \bar{\Omega}_h \text{ benachbart sind,} \\ 0 & \text{sonst für } \mathbf{x} \neq \xi, \end{cases}$$

$$q_h = f_h + \varphi_h, \quad f_h(\mathbf{x}) = (\tilde{R}_h f)(\mathbf{x}) := f(\mathbf{x}), \quad \varphi_h(\mathbf{x}) = -h \sum_{\xi \in \Gamma_h} L_{\mathbf{x}\xi} \varphi(\xi). \quad (4.7.5c)$$

Bemerkung 4.7.2 (a) L_h ist symmetrisch und erfüllt die Vorzeichenbedingung (4.3.1a). (b) Bei lexikographischer Anordnung der Gitterpunkte von Ω_h hat L_h die Gestalt

$$L_h = h^{-2} \begin{bmatrix} T-I & -I & & & \\ -I & T & -I & & \\ & \ddots & \ddots & \ddots & \\ & & -I & T & -I \\ & & & -I & T-I \end{bmatrix}, \quad T = \begin{bmatrix} 3 & -1 & & & \\ -1 & 4 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 4 & -1 \\ & & & -1 & 3 \end{bmatrix}.$$

Die Matrix L_h ist singulär, denn das System $L_h u_h = q_h$ ist im Allgemeinen ebenso wie das kontinuierliche Randwertproblem (4.7.1) nicht lösbar. Das Analogon des Satzes 3.4.1 lautet wie folgt.

Satz 4.7.3 Das Gleichungssystem (4.7.5a) ist genau dann lösbar, wenn

$$-h^2 \sum_{\mathbf{x} \in \Omega_h} f(\mathbf{x}) = h \sum_{\mathbf{x} \in \Gamma'_h} \varphi(\mathbf{x}). \quad (4.7.6)$$

Je zwei Lösungen von (4.7.5a) können sich nur um eine Konstante unterscheiden: $u_h^1 - u_h^2 = c\mathbf{1}$, $c \in \mathbb{R}$.

Beweis. Offenbar gilt $L_h \mathbf{1} = 0$, d.h. $\mathbf{1} \in \text{Kern}(L_h)$. Aus Satz 4.7.4 wird $\dim(\text{Kern}(L_h)) = 1$ folgen. Dies beweist

$$\text{Kern}(L_h) = \{c\mathbf{1} : c \in \mathbb{R}\} \quad (4.7.7)$$

und damit den zweiten Teil der Behauptung. (4.7.5a) ist genau dann lösbar, wenn das Skalarprodukt $\langle v, q_h \rangle$ für alle $v \in \text{Kern}(L_h^\top) = \text{Bild}(L_h)^\perp$ verschwindet. Wegen $L_h^\top = L_h$ und (4.7.7) ist

$$\langle \mathbf{1}, q_h \rangle = 0, \quad \text{d.h.} \quad \sum_{\mathbf{x} \in \Omega_h} q_h(\mathbf{x}) = 0 \quad (4.7.6')$$

hinreichend und notwendig. Nach Definition (4.7.5c) stimmen (4.7.6) und (4.7.6') überein. ■

Die Bedingung (4.7.6) sei erfüllt. Das System (4.7.5a) kann wie folgt gelöst werden. Man wähle ein beliebiges $\mathbf{x}_0 \in \Omega_h$ und normiere die (nur bis auf eine Konstante bestimmte) Lösung u_h durch

$$u_h(\mathbf{x}_0) = 0. \quad (4.7.8)$$

Sei \hat{u}_h der Vektor u_h ohne die Komponente $u_h(\mathbf{x}_0)$. \hat{L}_h sei die Untermatrix von L_h , in der man die Zeile und Spalte zum Index \mathbf{x}_0 gestrichen hat. Ebenso sei \hat{q}_h konstruiert. Dann ist

$$\hat{L}_h \hat{u}_h = \hat{q}_h \quad (4.7.9)$$

ein System mit $(n-1)^2 - 1$ Gleichungen und Unbekannten.

Satz 4.7.4 *Das Gleichungssystem (4.7.9) ist lösbar; insbesondere ist \hat{L}_h eine symmetrische M-Matrix. Unter der Voraussetzung (4.7.6) ergibt $\hat{u}_h = \hat{L}_h^{-1} \hat{q}_h$ ergänzt durch (4.7.8) die Lösung u_h des Systems (4.7.5a).*

Beweis. 1) Als Hauptuntermatrix von L_h ist \hat{L}_h symmetrisch. Auch in $\Omega_h \setminus \{\mathbf{x}_0\}$ lassen sich je zwei Gitterpunkte durch eine Kette von Nachbarpunkten verbinden, sodass \hat{L}_h irreduzibel ist. Für alle $\mathbf{x} \in \Omega_h \setminus \{\mathbf{x}_0\}$ gilt (4.3.4b), in Nachbarpunkten von \mathbf{x}_0 sogar (4.3.4a), sodass \hat{L}_h irreduzibel diagonaldominant ist. Gemäß Kriterium 4.3.10 ist \hat{L}_h eine M-Matrix, also nichtsingulär.

2) Ist u_h Lösung von (4.7.5a), lässt sich ohne Beschränkung der Allgemeinheit (4.7.8) annehmen, sodass u_h beschränkt auf $\Omega_h \setminus \{\mathbf{x}_0\}$ auch Gleichung (4.7.9) löst und mit der eindeutigen Lösung \hat{u}_h übereinstimmen muss. ■

Als Folgerung aus Satz 4.7.4 erhält man $\text{Rang}(L_h) \geq \text{Rang}(\hat{L}_h) = (n-1)^2 - 1$, d.h. $\dim(\text{Kern}(L_h)) = 1$ und somit (4.7.7).

Eine weitere Möglichkeit zur Lösung der Gleichung (4.7.5a) ist der Übergang zu einem erweiterten Gleichungssystem

$$\bar{L}_h \bar{u}_h = \bar{q}_h \quad \text{mit} \quad (4.7.10a)$$

$$\bar{L}_h = \begin{bmatrix} L_h & \mathbf{1} \\ \mathbf{1}^\top & 0 \end{bmatrix}, \quad \bar{u}_h = \begin{bmatrix} u_h \\ \lambda \end{bmatrix}, \quad \bar{q}_h = \begin{bmatrix} q_h \\ \sigma \end{bmatrix}, \quad (4.7.10b)$$

wobei σ beliebig vorgegeben werden kann.

Satz 4.7.5 *Gleichung (4.7.10a) ist stets lösbar. Ist die letzte Komponente λ der Lösung \bar{u}_h gleich null, so ist Bedingung (4.7.6) erfüllt und u_h stellt jene Lösung des Systems (4.7.5a) dar, die durch $\mathbf{1}^\top u_h = \sum_{\mathbf{x} \in \Omega_h} u_h(\mathbf{x}) = \sigma$ normiert ist. Gilt dagegen $\lambda \neq 0$, lässt sich u_h als Lösung von $L_h u_h = \tilde{q}_h$ interpretieren, wobei $\tilde{q}_h = q_h - \lambda \mathbf{1}$ zu $\tilde{f}(\mathbf{x}) := f(\mathbf{x}) - \lambda$ gehört, und \tilde{f} und φ der Bedingung (4.7.6) genügen.*

Beweis. $\mathbf{1}$ ist linear unabhängig von den Spalten von L_h , sodass $\text{Rang}[L_h, \mathbf{1}] = \text{Rang}(L_h) + 1 = (n-1)^2$. Ebenso ist $(\mathbf{1}^\top, 0)$ linear unabhängig von den Zeilen von $[L_h, \mathbf{1}]$, sodass $\text{Rang}(\bar{L}_h) = (n-1)^2 + 1$, d.h. \bar{L}_h ist nichtsingulär. Die weiteren Aussagen liest man aus (4.7.10b) ab. ■

Empfehlung. Man sollte entweder Gleichung (4.7.10a,b) verwenden oder aber Gleichung (4.7.9), nachdem man zuvor explizit q_h durch $\tilde{q}_h := q_h - (\mathbf{1}^\top q_h / \mathbf{1}^\top \mathbf{1}) \mathbf{1}$ ersetzt hat.

Zur Begründung dieser Empfehlung beachte man, dass die Lösbarkeitsbedingung des kontinuierlichen Problems $\int_\Omega f d\mathbf{x} + \int_\Gamma \varphi d\Gamma = 0$ lautet und keinesfalls die diskrete Lösbarkeitsbedingung (4.7.6) impliziert. Für glatte Funktionen f und φ lässt sich Gleichung (4.7.6) bis auf einen Rest $\mathcal{O}(h)$ zeigen. Somit ist es im Allgemeinen unumgänglich f_h und q_h durch $f_h - \lambda \mathbf{1}$ und $q_h - \lambda \mathbf{1}$ ($\lambda = \mathbf{1}^\top q_h / \mathbf{1}^\top \mathbf{1}$) zu ersetzen. Im Falle der Gleichung (4.7.10a,b) wird diese Korrektur implizit durchgeführt. Wird dagegen Gleichung (4.7.9) ohne jede Korrektur verwendet, so lässt sich die entstehende Lösung interpretieren als Lösung von $L_h u_h = \tilde{q}_h$ mit $\tilde{q}_h(\mathbf{x}) = q_h(\mathbf{x})$ für $\mathbf{x} \neq \mathbf{x}_0$ und $\tilde{q}_h(\mathbf{x}_0) := -\sum_{\mathbf{x} \neq \mathbf{x}_0} q_h(\mathbf{x})$. Das heißt, auch hier wird implizit eine Korrektur von q_h vorgenommen, nur wird die Korrektur nicht wie zuvor auf alle Komponenten verteilt, sondern auf $q_h(\mathbf{x}_0)$ konzentriert. Wenn Gleichung (4.7.6) bis auf $\mathcal{O}(h)$ erfüllt ist, unterscheiden sich $q_h(\mathbf{x}_0)$ und $\tilde{q}_h(\mathbf{x}_0)$ um $\mathcal{O}(h^{-1})$. Demgemäß enthält die Lösung \hat{u}_h der Gleichung (4.7.9) eine Singularität im Punkte \mathbf{x}_0 .

Satz 4.7.6 (Konvergenz) Sei $u \in C^{3,1}(\bar{\Omega})$ Lösung des Neumann-Problems (4.7.1). $\bar{u}_h = \begin{pmatrix} u_h \\ \lambda \end{pmatrix}$ sei die Lösung der Gleichung (4.7.10a). Dann gibt es ein $c \in \mathbb{R}$ und von u und h unabhängige Konstanten C, C' , sodass

$$|\lambda| \leq C'h \|u\|_{C^{0,1}(\bar{\Omega})}, \quad \|u_h - R_h u - c\mathbf{1}\|_\infty \leq C \left[h \|u\|_{C^{1,1}(\bar{\Omega})} + h^2 \|u\|_{C^{3,1}(\bar{\Omega})} \right]. \quad (4.7.11)$$

Beweis. 1) Es ist $\lambda = \mathbf{1}^\top q_h / \mathbf{1}^\top \mathbf{1} = (h^2 \sum_{\Omega_h} f(\mathbf{x}) + h \sum_{\Gamma'_h} \varphi(\mathbf{x})) h^{-2} (n-1)^2$. Der erste Faktor besteht wegen $\int_\Omega f d\mathbf{x} + \int_\Gamma \varphi d\Gamma = 0$ nur aus dem Quadratfehler $\mathcal{O}(h \|u\|_{C^{0,1}(\bar{\Omega})})$.

2) Gemäß Satz 4.7.5 ist u_h Lösung von $L_h u_h = \tilde{q}_h := q_h - \lambda \mathbf{1}$. Dies entspricht den Differenzengleichungen

$$-\Delta_h u_h = \tilde{f}_h := f_h - \lambda \mathbf{1} \quad \text{in } \Omega_h, \quad \partial_n^- u_h = \varphi \quad \text{auf } \Gamma'_h.$$

Die Differenz $w_h := u_h - R_h u$ erfüllt

$$-\Delta_h w_h = -\Delta_h u_h + \Delta_h R_h u = \Delta_h R_h u + f_h - \lambda \mathbf{1} = \Delta_h R_h u - \tilde{R}_h \Delta u - \lambda \mathbf{1} =: c_h \quad \text{in } \Omega_h, \quad (4.7.12a)$$

$$\partial_n^- w_h = \partial_n^- u_h - \partial_n^- R_h u = \varphi - \partial_n^- R_h u = \frac{\partial u}{\partial n} - \partial_n^- R_h u =: \psi \quad \text{auf } \Gamma'_h. \quad (4.7.12b)$$

Die Konsistenzfehler betragen

$$\|c_h\|_\infty \leq \frac{1}{6} h^2 \|u\|_{C^{3,1}(\bar{\Omega})} \quad (\text{vgl. Bemerkung 4.5.2}),$$

$$|\psi(\mathbf{x})| \leq \frac{1}{2} h \|u\|_{C^{1,1}(\bar{\Omega})} + |\lambda| \quad (\text{vgl. Lemma 4.1.1}).$$

Da die Lösung w_h existiert, ist die Bedingung (4.7.6) für c_h und ψ erfüllt: $h^2 \sum_{\Omega_h} c_h(\mathbf{x}) + h \sum_{\Gamma'_h} \psi(\mathbf{x}) = 0$. $\tilde{w}_h := w_h - c\mathbf{1}$ mit $c = \mathbf{1}^\top w_h / \mathbf{1}^\top \mathbf{1}$ ist die durch $\mathbf{1}^\top w_h = 0$ normierte Lösung von (4.7.12a,b). Die Anwendung des folgenden Satzes 4.7.7 auf Gleichung (4.7.12a, b) liefert Ungleichung (4.7.11) (vgl. Bemerkung 4.5.2). ■

Satz 4.7.7 (Stabilität) Bedingung (4.7.6) sei erfüllt. Die Lösung u_h von (4.7.3), (4.7.4) sei mittels $\mathbf{1}^\top u_h = 0$ normiert. Dann gibt es von u und h unabhängige Konstanten C_1, C_2 , sodass

$$\|u_h\|_\infty \leq C_1 \max_{\mathbf{x} \in \Omega_h} |f(\mathbf{x})| + C_2 \max_{\mathbf{x} \in \Gamma'_h} |\varphi(\mathbf{x})|. \quad (4.7.13)$$

Der Beweis dieses Satzes, der dem Satz 4.4.11 entspricht, wird in §4.7.4 nachgetragen.

4.7.2 Symmetrische Differenz für $\partial u / \partial n$

Wie aus Satz 4.7.6 hervorgeht, verursacht die einseitige Differenz ∂_n^- den Fehlerterm $\mathcal{O}(h)$. Es liegt nahe, ∂_n^- durch eine symmetrische Differenz zu ersetzen. Hierzu wird die Fünfpunktdiskretisierung in allen Punkten $\mathbf{x} \in \bar{\Omega}_h = \Omega_h \cup \Gamma_h$ (vgl. (4.2.1c)) aufgestellt:

$$-\Delta_h u_h = f_h := \tilde{R}_h f \quad \text{in } \bar{\Omega}_h, \quad (4.7.14a)$$

wobei \tilde{R}_h die Beschränkung auf $\bar{\Omega}_h$ ist. Die symmetrische Differenz ∂_n^0 ist definiert durch

$$(\partial_n^0 u_h)(\mathbf{x}) := \frac{1}{2h} [u_h(\mathbf{x} + h\vec{n}) - u_h(\mathbf{x} - h\vec{n})] = \varphi(\mathbf{x}) \quad \text{für } \mathbf{x} \in \Gamma_h. \quad (4.7.14b)$$

Den Eckpunkten werden dabei zwei Normalenrichtungen zugeordnet, sodass für die Eckpunkte je zwei Gleichungen der Form (4.7.14b) aufgestellt werden können. In der Ecke $\mathbf{x} = (0,0)$ hat man zum Beispiel die Normalen $\vec{n} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, $\vec{n} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$. Die entsprechenden Gleichungen (4.7.14b) führen auf unterschiedliche Werte

$$\varphi(0+,0) = \lim_{x \searrow 0} \varphi(x,0) \quad \text{und} \quad \varphi(0,0+) = \lim_{y \searrow 0} \varphi(0,y).$$

Für $\mathbf{x} \in \Gamma_h$ benötigt die Differenzenformel (4.7.14a) die Werte in Punkten $\mathbf{x} + h\vec{n}$ außerhalb von $\bar{\Omega}_h$. Diese lassen sich mit Hilfe von (4.7.14b) eliminieren, sodass ein Gleichungssystem $L_h u_h = q_h$ für die $(n+1)^2$ Komponenten $u_h(\mathbf{x})$, $\mathbf{x} \in \bar{\Omega}_h$, übrigbleibt.

Übungsaufgabe 4.7.8 a) Bei lexikographischer Anordnung der Gitterpunkte von $\overline{\Omega}_h$ hat L_h die Gestalt

$$L_h = h^{-2} \begin{bmatrix} T & -2I & & & \\ -I & T & -I & & \\ & \ddots & \ddots & \ddots & \\ & & -I & T & -I \\ & & & -2I & T \end{bmatrix}, \quad T = \begin{bmatrix} 4 & -2 & & & \\ -1 & 4 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 4 & -1 \\ & & & -2 & 4 \end{bmatrix}.$$

b) L_h ist nicht symmetrisch, aber $D_h L_h$ mit $D_h = \text{diag}\{d(x)d(y)\}$, $d(0) = d(1) = 1/2$, $d(\cdot) \equiv 1$ sonst, ist symmetrisch.

Die sinnngemäße Übertragung des Satzes 4.7.3 lautet:

Satz 4.7.9 Gleichung (4.7.14a,b) ist genau dann lösbar, wenn $\mathbf{1}^\top D_h q_h = 0$ für D_h aus Übungsaufgabe 4.7.8b. Je zwei Lösungen können sich nur um eine Konstante unterscheiden. Die Formulierung von $\mathbf{1}^\top D_h q_h = 0$ mit Hilfe von f und φ lautet

$$-h^2 \sum_{(x,y) \in \overline{\Omega}_h} d(x)d(y)f(x,y) = 2h \sum_{(x,y) \in \Gamma_h} d(x)d(y)\varphi(x,y), \quad (4.7.15)$$

wobei in der zweiten Summe der Summand für Eckpunkte doppelt auftritt und die beiden (verschiedenen) Grenzwerte von φ berücksichtigt.

Bemerkung 4.7.10 Die Summen in (4.7.15) sind summierte Trapezformeln⁵. Aus $\int_{\Omega} f d\mathbf{x} + \int_{\Gamma} \varphi d\Gamma = 0$ folgt daher Gleichung (4.7.15) bis auf einen Rest $\mathcal{O}(h^2 \|u\|_{C^{1,1}(\overline{\Omega})})$.

Ohne Schwierigkeiten lassen sich die Sätze 4.7.4 und 4.7.5 übertragen. Der Konvergenzsatz 4.7.6 wird zu

Satz 4.7.11 Sei $u \in C^{3,1}(\overline{\Omega})$ eine Lösung von (4.7.1). $\bar{u}_h = \begin{pmatrix} u_h \\ \chi \end{pmatrix}$ sei Lösung von $L_h \bar{u}_h = \begin{pmatrix} D_h q_h \\ 0 \end{pmatrix}$ mit der Matrix $\bar{L}_h = \begin{bmatrix} D_h L_h & \mathbf{1} \\ \mathbf{1}^\top & 0 \end{bmatrix}$. Es liegt Konvergenz der Ordnung 2 vor:

$$|\lambda| \leq C' h^2 \|u\|_{C^{1,1}(\overline{\Omega})}, \quad \|u_h - R_h u - c\mathbf{1}\|_{\infty} \leq C h^2 \|u\|_{C^{3,1}(\overline{\Omega})}. \quad (4.7.16)$$

Beweis. Der Beweis ist im Wesentlichen der gleiche wie für Satz 4.7.6. Eine zusätzliche, technische Schwierigkeit besteht darin, dass der Konsistenzfehler $\Delta_h R_h u - R_h \Delta u$ auch in $\mathbf{x} \in \Gamma_h$ zu bestimmen ist, obwohl u nur in $\overline{\Omega}$ definiert ist. Statt wie in (4.7.12a,b) die Differenzengleichung und die Randdiskretisierung getrennt zu behandeln, analysiere man direkt die Gleichungen $L_h u_h = q_h$, aus denen die Werte $u_h(\mathbf{x} + h\vec{n})$ außerhalb von $\overline{\Omega}$ bereits eliminiert worden sind. ■

4.7.3 Symmetrische Differenz für $\partial u / \partial n$ im verschobenen Gitter

Verschiebt man das bisherige Gitter um $h/2$ in x - und y -Richtung, ergibt sich das Gitter

$$\Omega_h := \left\{ (x,y) \in \Omega : \frac{x}{h} - \frac{1}{2} \in \mathbb{Z}, \frac{y}{h} - \frac{1}{2} \in \mathbb{Z} \right\}$$

aus Abbildung 4.7.2. Die randnahen Punkte von Ω_h haben den Abstand $h/2$ von Γ . Wir setzen

$$\Gamma_h := \left\{ (x,y) \in \Gamma : \frac{x}{h} - \frac{1}{2} \in \mathbb{Z} \text{ oder } \frac{y}{h} - \frac{1}{2} \in \mathbb{Z} \right\}$$

(vgl. Abbildung 4.7.2). Jedem randnahen Punkt $\mathbf{x} - h\vec{n}/2$ ($\mathbf{x} \in \Gamma_h$) entspricht ein außenliegender Nachbar $\mathbf{x} + h\vec{n}/2$. Die Diskretisierung der Neumann-Aufgabe (4.7.1) lautet

$$-\Delta_h u_h(\mathbf{x}) = f(\mathbf{x}) \quad \text{in } \mathbf{x} \in \Omega_h, \quad (4.7.17a)$$

$$h^{-1} [u_h(\mathbf{x} + h\vec{n}/2) - u_h(\mathbf{x} - h\vec{n}/2)] = \varphi(\mathbf{x}) \quad \text{für } \mathbf{x} \in \Gamma'_h. \quad (4.7.17b)$$

⁵vgl. Stoer [77, §3.1].

Die Differenz (4.7.17b) ist symmetrisch bezüglich des Randpunktes \mathbf{x} und stimmt trotzdem mit der Rückwärtsdifferenz ∂_n^- im Gitterpunkt $\mathbf{x} + h\vec{n}/2$ überein.

Bemerkung 4.7.12 Nach Elimination der Werte $u_h(\mathbf{x} + h\vec{n}/2)$, $\mathbf{x} \in \Gamma_h$, ergibt sich ein Gleichungssystem $L_h u_h = q_h$, wobei Bemerkung 4.7.2 auch auf diese Matrix L_h zutrifft. Anders als in §4.7.1 hat L_h die Größe $n^2 \times n^2$.

Die Sätze 4.7.3, 4.7.4, 4.7.5, 4.7.7 gelten sinngemäß. Satz 4.7.6 gilt mit Ungleichung (4.7.16) anstelle von (4.7.11).

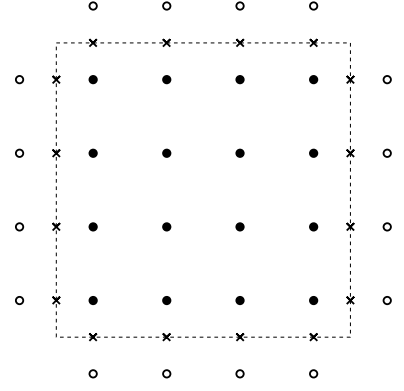


Abb. 4.7.2. Verschobenes Gitter, •: innerer Punkt von Ω_h , o: äußerer Gitterpunkt, x: Randpunkt auf Γ_h

4.7.4 Beweis des Stabilitätsatzes 4.7.7

Im Falle von Dirichlet-Randwerten folgt die Stabilitätsaussage des Satzes 4.4.11 sofort aus dem Maximumprinzip und der Abschätzung von L_h^{-1} . Dagegen ist die entsprechende Aussage des Satzes 4.7.7 für Neumann-Randwerte nicht so einfach nachzuweisen. In der Literatur findet man nur schwächere Abschätzungen, die auf der rechten Seite der Gleichung (4.7.13) einen zusätzlichen Faktor $|\log h|$ enthalten. Dieser Faktor ist nur dann nicht zu vermeiden, wenn man die Gleichung (4.7.9), $\hat{L}_h \hat{u}_h = \hat{q}_h$, verwendet, wenn die Bedingung (4.7.6) nicht erfüllt ist.

Die diskrete Greensche Funktion⁶ (zweiter Art) $g_h(\mathbf{x}, \xi)$, $\mathbf{x} \in \Omega_h \cup \Gamma_h$, $\xi \in \Omega_h$, sei durch

$$-\Delta_h g_h(\mathbf{x}, \xi) = \delta(\mathbf{x}, \xi) := \begin{cases} h^{-2} & \text{für } \mathbf{x} = \xi \\ 0 & \text{sonst} \end{cases} - \begin{cases} h^{-1}/(4-4h) & \text{falls } \mathbf{x} \text{ randnah} \\ 0 & \text{falls } \mathbf{x} \text{ randfern} \end{cases}, \quad (4.7.18a)$$

$$(\partial_n^- g_h)(\mathbf{x}, \xi) = 0 \quad \text{für } \mathbf{x} \in \Gamma_h' \quad (4.7.18b)$$

definiert, wobei Δ_h und ∂_n^- auf \mathbf{x} wirken. g_h existiert, da $\sum_{\mathbf{x} \in \Omega_h} \delta(\mathbf{x}, \xi) = 0$ die Lösbarkeitsbedingung (4.7.6') beweist.

Lemma 4.7.13 Für beliebiges q_h mit $\mathbf{1}^\top q_h = 0$ (vgl. (4.7.6')) stellt

$$u_h(\mathbf{x}) := h^2 \sum_{\xi \in \Omega_h} q_h(\xi) g_h(\mathbf{x}, \xi) \quad (4.7.19)$$

eine Lösung von $L_h u_h = q_h$ dar.

Beweis. In randfernen $\mathbf{x} \in \Omega_h$ ist $(L_h u_h)(\mathbf{x}) = -\Delta_h u_h(\mathbf{x}) = q_h(\mathbf{x})$. In randnahen $\mathbf{x} \in \Omega_h$ hat man

$$(L_h u_h)(\mathbf{x}) = (-\Delta_h u_h)(\mathbf{x}) = q_h(\mathbf{x}) - \frac{1/h}{4-4h} \sum_{\xi \in \Omega_h} q_h(\xi) = q_h(\mathbf{x})$$

wegen $\partial_n^- u_h = 0$ und (4.7.6'). ■

Satz 4.7.14 Die Gleichungen (4.7.18a,b) bestimmen g_h nur bis auf eine Konstante. Die Greensche Funktion $g_h(\mathbf{x}, \xi)$ kann so gewählt werden, dass

$$|g_h(\mathbf{x}, \xi)| \leq C [1 + |\log(|\mathbf{x} - \xi| + h)|] \quad \text{für } \mathbf{x}, \xi \in \Omega_h. \quad (4.7.20)$$

Diese Ungleichung entspricht der Abschätzung (4.4.7) im Dirichlet-Fall. Bevor Satz 4.7.14 bewiesen wird, soll gezeigt werden, dass hieraus Satz 4.7.7 folgt.

Beweis zu Satz 4.7.7. Analog zu $\int_\Omega [1 + |\log|\mathbf{x} - \xi||] d\mathbf{x} \leq K_1$ und $\int_\Gamma [1 + |\log|\mathbf{x} - \xi||] d\Gamma_x \leq K_2$ ergibt sich

$$h^2 \sum_{\xi \in \Omega_h} [1 + |\log(|\mathbf{x} - \xi| + h)|] \leq K_1', \quad h \sum_{\mathbf{x} \in \Gamma_h'} [1 + |\log(|\mathbf{x} - \xi| + h)|] \leq K_2'.$$

⁶Die diskrete Greensche Funktion aus (4.4.5b,c) gehörte zur Dirichlet-Randwertaufgabe. Wie in §3.4 im Falle der kontinuierlichen Greenschen Funktion muss für das Neumann-Problem eine Modifikation verwendet werden.

Aus (4.7.19), (4.7.20) und $q_h = f_h + \varphi_h$ (vgl. (4.7.5c)) folgt daher die Abschätzung

$$|u_h(\mathbf{x})| \leq K_1 \|f_h\|_\infty + hK_2 \|\varphi_h\|_\infty \leq K_1 \max_{\mathbf{x} \in \Omega_h} |f(\mathbf{x})| + 2K_2 \max_{\mathbf{x} \in \Gamma'_h} |\varphi(\mathbf{x})|.$$

Ist $\tilde{u}_h := u_h - \mathbf{1}^\top u_h / \mathbf{1}^\top \mathbf{1}$ die durch $\mathbf{1}^\top \tilde{u}_h = 0$ normierte Lösung von $L_h u_h = q_h$, so überlegt man sich, dass

$$\|\tilde{u}_h\|_\infty \leq 2 \inf \{\|u_h - c\mathbf{1}\|_\infty : c \in \mathbb{R}\} \leq 2 \|\tilde{u}_h\|_\infty$$

sodass Ungleichung (4.7.11) und Satz 4.7.7 bewiesen sind. \blacksquare

Es bleibt Satz 4.7.14 zu beweisen. Die Konstruktion der im folgenden Lemma postulierten Funktion G_h ist Gegenstand der nächsten Seiten.

Lemma 4.7.15 *e sei einer der Einheitsvektoren $(1, 0)$ oder $(0, 1)$. $\xi \in \Omega_h$ sei derart, dass auch $\xi + h\mathbf{e} \in \Omega_h$. Für alle diese \mathbf{e} und ξ existiere eine Funktion $G_h(\mathbf{x}) = G_h(\mathbf{x}; \xi, \mathbf{e})$ mit den Eigenschaften*

$$-\Delta_h G_h(\mathbf{x}) = h^{-2} \begin{cases} 1 & \text{falls } \mathbf{x} = \xi \\ -1 & \text{falls } \mathbf{x} = \xi + h\mathbf{e} \\ 0 & \text{sonst} \end{cases} \quad \text{in } \Omega_h, \quad (4.7.21a)$$

$$\partial_n^- G_h = 0 \quad \text{auf } \Gamma'_h, \quad (4.7.21b)$$

$$|G_h(\mathbf{x})| \leq hC' / (|\mathbf{x} - \xi| + h), \quad (4.7.21c)$$

wobei C' nicht von \mathbf{e} und ξ abhängt. Dann gilt Satz 4.7.14 (und somit auch Satz 4.7.7).

Beweis. Für $\xi, \xi' \in \Omega_h$ sei $g_h(\mathbf{x}, \xi, \xi')$ definiert als Lösung von

$$\partial_n^- g_h(\cdot, \xi, \xi') = 0 \quad \text{auf } \Gamma'_h, \quad -\Delta_h g_h(\cdot, \xi, \xi') = h^{-2} \begin{cases} 1 & \text{falls } \mathbf{x} = \xi \\ -1 & \text{falls } \mathbf{x} = \xi' \\ 0 & \text{sonst} \end{cases} \quad \text{in } \Omega_h.$$

Für $\xi' = \xi + h\mathbf{e}$ stimmt $g_h(\cdot, \xi, \xi')$ mit G_h aus Lemma 4.7.15 überein. Für beliebige ξ, ξ' findet man eine Verbindung $\xi = \xi^0, \xi^1, \dots, \xi^\ell = \xi'$ mit $\xi^{k+1} - \xi^k = \pm h\mathbf{e}^k$, $\mathbf{e}^k = (1, 0)$ oder $(0, 1)$. Da $g_h(\cdot, \xi, \xi') = \sum_{k=1}^\ell g_h(\cdot, \xi^{k-1}, \xi^k)$, liefert (4.7.21c) die Abschätzung

$$|g_h(\mathbf{x}, \xi, \xi')| \leq h \sum_{k=1}^\ell C' / (|\mathbf{x} - \xi^{k-1}| + h).$$

Betrachtet man zunächst den Fall $x_1 = \xi_1 = \xi'_1$, $x_2 \leq \xi_2 < \xi'_2$ und verwendet man

$$\sum_{k=k_1}^{k_2} 1/k \leq \text{const} \cdot (1 + \log(k_2 h) - \log(k_1 h)),$$

so erhält man

$$|g_h(\mathbf{x}, \xi, \xi')| \leq C'' [1 + |\log(|\mathbf{x} - \xi| + h)| + |\log(|\mathbf{x} - \xi'| + h)|].$$

Diese Abschätzung ergibt sich auch im allgemeinen Fall, nur hat man die Verbindung ξ^k so zu wählen, dass $|\mathbf{x} - \xi^k| \geq \min\{|\mathbf{x} - \xi|, |\mathbf{x} - \xi'|\}$. Da $h \sum_{\xi' \in \Gamma'_h} |\log(|\mathbf{x} - \xi'| + h)| \leq \text{const}$ für alle $\mathbf{x} \in \Omega_h$, erfüllt

$$g_h(\mathbf{x}, \xi) := \frac{h}{4 - 4h} \sum_{\xi' \in \Gamma'_h} g_h(\mathbf{x}, \xi, \xi')$$

die Ungleichung (4.7.20). Wegen $\sum_{\Gamma'_h} 1 = (4 - 4h)/h$ sind auch die Gleichungen (4.7.18a,b) erfüllt, d.h. $g_h(\mathbf{x}, \xi)$ ist die Greensche Funktion. \blacksquare

Die in Lemma 4.7.15 benötigte Funktion G_h soll explizit konstruiert werden. Grundbaustein ist die diskrete Singularitätenfunktion $s_h(\mathbf{x}, \xi)$ auf dem Gitter Q_h des \mathbb{R}^2 :

$$Q_h := \{(x, y) \in \mathbb{R}^2 : x/h, y/h \in \mathbb{Z}\}.$$

Lemma 4.7.16 Die durch $s_h(\mathbf{x}, \xi) := \sigma_h(\mathbf{x} - \xi)$ und

$$\sigma_h(\mathbf{x}) := \frac{1}{16\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{e^{i(x_1\eta_1 + x_2\eta_2)/h} - 1}{\sin^2(\eta_1/2) + \sin^2(\eta_2/2)} d\eta_1 d\eta_2 = \frac{1}{8\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{\sin^2((x_1\eta_1 + x_2\eta_2)/(2h))}{\sin^2(\eta_1/2) + \sin^2(\eta_2/2)} d\eta_1 d\eta_2$$

für alle $\mathbf{x}, \xi \in Q_h$ definierte Singularitätenfunktion hat die Eigenschaft $-\Delta_h s_h(\mathbf{x}, \xi) = h^{-2}$ für $\mathbf{x} = \xi$ und $-\Delta_h s_h = 0$ sonst.

Beweis. Sei $e(\mathbf{x}, \eta) := \exp(i(x_1\eta_1 + x_2\eta_2)/h)$. Man beachte, dass

$$-\Delta_h(e(\mathbf{x}, \eta) - 1) = 4h^{-2}e(\mathbf{x}, \eta)[\sin^2(\eta_1/2) + \sin^2(\eta_2/2)]$$

$$\text{und } \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e(\mathbf{x}, \eta) d\eta = \begin{cases} 4\pi^2 & \text{für } \mathbf{x} = 0 \\ 0 & \text{für } \mathbf{x} \neq 0 \end{cases}.$$

Für Multiindizes $\alpha = (\alpha_1, \alpha_2) \in \mathbb{N}^2$ mit $\alpha_1, \alpha_2 \geq 0$ definiert man die partiellen Differenzenoperatoren der Ordnung $|\alpha| = \alpha_1 + \alpha_2$ durch

$$\partial^\alpha = (\partial_x^+)^{\alpha_1} (\partial_y^+)^{\alpha_2}. \quad (4.7.22)$$

Ausgehend von der Darstellung in Lemma 4.7.16 beweist Thomée [82] das

Lemma 4.7.17 Es gilt $|(\partial^\alpha \sigma_h)(\mathbf{x})| \leq C(|\mathbf{x}| + h)^{-|\alpha|}$ für alle $\mathbf{x} \in Q_h$.

Zur Konstruktion der Funktion G_h aus Lemma 4.7.15 wählen wir ohne Beschränkung der Allgemeinheit $\mathbf{e} = (1, 0)$ und halten $\xi \in \Omega_h$ mit $\xi + h\mathbf{e} \in \Omega_h$ fest. Die Funktion

$$G'_h(\mathbf{x}) := h(\partial_x^- \sigma_h)(\mathbf{x} - \xi) = \sigma_h(\mathbf{x} - \xi) - \sigma_h(\mathbf{x} - \xi - h\mathbf{e}) \quad (4.7.23)$$

erfüllt zwar (4.7.21a), nicht aber die Randbedingung (4.7.21b). Die Symmetrisierungsoperatoren S_x und S_y seien definiert durch

$$S_x u_h(x, y) := \frac{1}{2} [u_h(x, y) - u_h(h - x, y)], \quad S_y u_h(x, y) := \frac{1}{2} [u_h(x, y) - u_h(x, h - y)]$$

für $(x, y) \in Q_h$. Die Funktion

$$G''_h := 4S_x S_y G'_h \quad (4.7.24)$$

ist symmetrisch bezüglich der Spiegelachsen $y = h/2$ und $x = h/2$. Daher gilt (4.7.21b) am linken und unteren Rand:

$$\partial_n^- G''_h(0, y) = \partial_n^- G''_h(x, 0) = 0. \quad (4.7.25)$$

Außerdem erfüllt G''_h ebenso wie G'_h die Bedingung (4.7.21a). Für jedes $\beta \in \mathbb{N}^2$ definieren wir den Operator P_β durch

$$\begin{aligned} (P_\beta u_h)(x, y) \\ = \frac{1}{4} [u_h(x + \beta_1 L, y + \beta_2 L) + u_h(x - \beta_1 L, y + \beta_2 L) + u_h(x + \beta_1 L, y - \beta_2 L) + u_h(x - \beta_1 L, y - \beta_2 L)], \end{aligned}$$

wobei $L = 2 - 2h$, und setzen

$$G_h^\beta(\mathbf{x}) := (P_\beta G''_h)(\mathbf{x}) - (P_\beta G''_h)(0). \quad (4.7.26)$$

Lemma 4.7.18 Für $\beta = (\beta_1, \beta_2)$ mit $\|\beta\|_\infty \geq 2$ gilt

$$|\partial^\alpha G_h^\beta(\mathbf{x})| \leq hK/|\beta|^3 \quad \text{für alle } |\alpha| \leq 2, \mathbf{x} \in \Omega_h. \quad (4.7.27)$$

Dabei ist K unabhängig von der Wahl der $\xi, \xi + h\mathbf{e} \in \Omega_h$.

Beweis. Definitionsgemäß gilt

$$G_h^\beta(0, 0) = 0. \quad (4.7.28a)$$

Der Operator P_β erhält die Symmetrie: $u_h = S_x u_h$ impliziert $P_\beta u_h = S_x P_\beta u_h$. Daher gelten die Identitäten $G_h^\beta = S_x G_h^\beta = S_y G_h^\beta$ und haben zur Folge, dass

$$\partial_x^+ G_h^\beta(0, 0) = \partial_y^+ G_h^\beta(0, 0) = 0. \quad (4.7.28b)$$

$G_h^\beta(\mathbf{x})$ ist Linearkombination von $h\partial_x^- \sigma_h(\mathbf{x} - \tilde{\xi})$ für verschiedene $\tilde{\xi}$ mit $|\mathbf{x} - \tilde{\xi}| + h \geq K'|\beta|$, wenn $\mathbf{x} \in \Omega_h$. Gemäß Lemma 4.7.17 ist

$$\left| \partial^\alpha G_h^\beta(\mathbf{x}) \right| \leq hCK''/|\beta|^3 \quad \text{für alle } |\alpha| \leq 2, \mathbf{x} \in \Omega_h, \quad (4.7.28c)$$

sodass (4.7.27) für $|\alpha| = 2$ folgt. Sei $|\alpha| = 1$. $\partial^\alpha G_h(\mathbf{x})$ kann man in der Form

$$\partial^\alpha G_h^\beta(0) + \sum_{k=1}^{\ell} \left[\partial^\alpha G_h^\beta(\mathbf{x}^k) - \partial^\alpha G_h^\beta(\mathbf{x}^{k-1}) \right] \quad \text{mit } \mathbf{x}^0 = \mathbf{x}, \mathbf{x}^\ell = \mathbf{x}, \|\mathbf{x}^k - \mathbf{x}^{k-1}\|_\infty = h$$

schreiben. Jeder Summand hat die Gestalt $\pm h\partial^\gamma G_h^\beta$ mit $|\gamma| = 2$, sodass (4.7.28b,c) die Abschätzung (4.7.28d) ergibt:

$$\left| \partial^\alpha G_h^\beta(\mathbf{x}) \right| \leq 2hCK''/|\beta|^3 \quad \text{für } |\alpha| = 1, \mathbf{x} \in \Omega_h. \quad (4.7.28d)$$

Ebenso folgert man aus (4.7.28a) und (4.7.28d) die Ungleichung (4.7.27) für $|\alpha| = 0$, womit das Lemma 4.7.18 bewiesen ist. ■

Da $\sum_{\beta \in \mathbb{Z}^2 \setminus (0,0)} |\beta|^{-3} < \infty$, existiert die unendliche Summe

$$G_h(\mathbf{x}) := \sum_{\beta \in \mathbb{Z}^2} G_h^\beta(\mathbf{x}). \quad (4.7.29)$$

Da $-\Delta_h G_h^\beta(\mathbf{x}) = 0$ in Ω_h für alle $\beta \neq (0,0)$, erfüllt auch G_h die Gleichung (4.7.21a). Wie schon beim Beweis von (4.7.28b) erwähnt, gilt $G_h = S_x G_h = S_y G_h$, sodass auch G_h der Gleichung (4.7.25) genügt: $\partial_n^- G_h = 0$ am linken und unteren Rand. Der Nachweis von $\partial_n^- G_h = 0$ an den anderen Rändern erfordert das

Lemma 4.7.19 G_h ist L -periodisch: $G_h(x, y) = G_h(x + L, y) = G_h(x, y + L)$.

Beweis. Sei $G_h''(x + \beta_1 L, y + \beta_2 L) - G_h''(\beta_1 L, \beta_2 L)$ durch $\gamma_\beta(x, y)$ abgekürzt. Definition (4.7.29) besagt, dass⁷

$$G_h = \lim_{k \rightarrow \infty} \sum_{|\beta_1| \leq k} \sum_{|\beta_2| \leq k} \gamma_\beta. \quad (4.7.30a)$$

γ_β kann man als Summe über die Differenzen

$$\begin{aligned} & \text{und} \quad \begin{aligned} & h\partial_h^+ G_h''(x + \beta_1 L, \beta_2 L + \nu h) && \text{mit } \nu h = 0 \text{ bis } \nu h = y \\ & h\partial_h^+ G_h''(\beta_1 L + \nu h, \beta_2 L) && \text{mit } \nu h = 0 \text{ bis } \nu h = x \end{aligned} \end{aligned}$$

schreiben. Für $|\beta_1| \geq 2$ ist der Abstand der Argumente von ξ und $\xi + h\mathbf{e}$ stets $\leq (|\beta_1| - 1)h$. Jeder Summand ist daher gemäß Lemma 4.7.17 durch $\mathcal{O}(h^2/(|\beta_1| - 1)^2) = \mathcal{O}(h^2/\beta_1^2)$ beschränkt. γ_β als Summe dieser Größen ist abschätzbar durch

$$|\gamma_\beta| \leq \mathcal{O}(h^2/\beta_1^2) \quad \text{für } |\beta_1| \geq 2. \quad (4.7.30b)$$

Wir wollen nun zeigen, dass auch die Darstellung (4.7.30c) gilt:

$$G_h = \lim_{k \rightarrow \infty} \sum_{|\beta_1 - 1| \leq k} \sum_{|\beta_2 - 1| \leq k} \gamma_\beta = \lim_{k \rightarrow \infty} \sum_{1 - k \leq \beta_1 \leq 1 + k} \sum_{|\beta_2 - 1| \leq k} \gamma_\beta. \quad (4.7.30c)$$

⁷Die Partialsummen $\sum_{|\beta_1| \leq k} \sum_{|\beta_2| \leq k} G_h^\beta(\mathbf{x})$ und $\sum_{|\beta_1| \leq k} \sum_{|\beta_2| \leq k} \gamma_\beta(\mathbf{x})$ sind identisch. Man beachte, dass die Summe (4.7.29) unbedingt konvergiert. Dagegen konvergiert $\sum \gamma_\beta$ nur bedingt. (4.7.30a) beschreibt die Konvergenz von $\sum \gamma_\beta$ für eine spezielle Summationsreihenfolge. Dies erklärt, warum die Identität in (4.7.30c) keine Trivialität ist.

Die Summen (4.7.30a/c) unterscheiden sich um

$$D_h := \sum_{|\beta_2| \leq k} [\gamma_{(1+k, \beta_2)} - \gamma_{(-k, \beta_2)}].$$

Die beiden β_1 -Werte $1+k$ und $-k$ erfüllen $|\beta_1| \leq k \leq 2$ für $k \leq 2$. Der Betrag der Summe D_h ist nach (4.7.30b) durch $\sum_{|\beta_2| \leq k} \mathcal{O}(h/k^2) = (2k+1)\mathcal{O}(h/k^2) = \mathcal{O}(h/k)$ beschränkt, sodass die Grenzwerte der Doppelsummen in (4.7.30a) und (4.7.30c) übereinstimmen. Umbenennungen von β_1 in $\beta_1 - 1$ überführt (4.7.30c) in

$$G_h(x, y) = \lim_{k \rightarrow \infty} \sum_{-k \leq \beta_1, \beta_2 \leq k} [G_h''(x + \beta_1 L + L, y + \beta_2 L) - G_h''(\beta_1 L + L, \beta_2 L)]. \quad (4.7.30d)$$

Der erste Teil der Identität

$$\begin{aligned} \sum_{\beta_1, \beta_2 = -k}^k [G_h''(\beta_1 L, \beta_2 L) - G_h''(\beta_1 L + L, \beta_2 L)] &= \sum_{\beta_2 = -k}^k [G_h''(-kL, \beta_2 L) - G_h''(kL + L, \beta_2 L)] \\ &= \sum_{\beta_2 = -k}^k [G_h''(h + kL, \beta_2 L) - G_h''(kL + L, \beta_2 L)] \end{aligned} \quad (4.7.30e)$$

ist elementar, der zweite ergibt sich aus der Symmetrie $G_h'' = S_x G_h''$. Wie oben lassen sich die Summanden in der letzten Summe durch $\mathcal{O}(h/k^2)$ abschätzen, sodass (4.7.30e) für $k \rightarrow \infty$ verschwindet. Zusammen mit (4.7.30d) erhält man

$$\begin{aligned} G_h(x, y) &= \lim_{k \rightarrow \infty} \sum_{-k \leq \beta_1, \beta_2 \leq k} [G_h''(x + \beta_1 L + L, y + \beta_2 L) - G_h''(\beta_1 L, \beta_2 L)] \\ &= \lim_{k \rightarrow \infty} \sum_{-k \leq \beta_1, \beta_2 \leq k} \gamma_\beta(x + L, y) = G_h(x + L, y). \end{aligned}$$

Der Nachweis von $G_h(x, y) = G_h(x, y + L)$ ist analog. ■

Beweis von Satz 4.7.14. Da $L = 2 - 2h$, ergeben die Symmetrie $G_h = S_x G_h$ und die Periodizität

$$G_h(1, y) = G_h(h - 1, y) = G_h(h - 1 + L, y) = G_h(1 - h, y),$$

d.h. $\partial_n^- G_h(1, y) = 0$ auf dem oberen Rand $(1, y) \in \Gamma_h'$. Ebenso zeigt man $\partial_n^- G_h(x, 1) = 0$ auf dem rechten Rand. Damit ist auch (4.7.21b) nachgewiesen. Es bleibt, (4.7.21c) zu zeigen:

$$|G_h(\mathbf{x})| \leq hC/(|\mathbf{x} - \xi| + h) \quad \text{in } \Omega_h.$$

Die G_h^β für $\|\beta\|_\infty \leq 1$ sind Linearkombinationen von $h\partial_x^- \sigma_h(\mathbf{x} - \tilde{\xi})$ für $\tilde{\xi} = \xi$ und andere $\tilde{\xi} \notin \Omega_h$, die durch S_x , S_y und P_β erzeugt werden. Für alle $\tilde{\xi}$ gilt $|\mathbf{x} - \tilde{\xi}| \geq |\mathbf{x} - \xi|$, sodass aus Lemma 4.7.17

$$|G_h^\beta(\mathbf{x})| \leq hC'/(|\mathbf{x} - \xi| + h) \quad \text{für } \|\beta\|_\infty \leq 1, \mathbf{x} \in \Omega_h$$

folgt. Andererseits zeigt Lemma 4.7.18, dass

$$\left| \sum_{\|\beta\|_\infty \geq 2} G_h^\beta(\mathbf{x}) \right| \leq h \sum_{\|\beta\|_\infty \geq 2} \frac{K}{|\beta|^3} = hK' \leq hK''/(|\mathbf{x} - \xi| + h).$$

Somit sind die Voraussetzungen des Lemmas 4.7.15 erfüllt: G_h mit (4.7.21a-c) existiert. Damit sind die Sätze 4.7.14 und 4.7.7 bewiesen. ■

4.8 Diskretisierung der Poisson-Gleichung im beliebigen Gebiet

4.8.1 Shortley-Weller-Approximation

Die Randwertaufgabe

$$-\Delta u = f \quad \text{in } \Omega, \quad u = \varphi \quad \text{auf } \Gamma \quad (4.8.1)$$

sei für ein beliebiges Gebiet Ω gegeben. Legt man ein quadratisches Gitter der Schrittweite h über Ω , erhält man

$$\Omega_h := \{(x, y) \in \Omega : x/h \in \mathbb{Z} \text{ und } y/h \in \mathbb{Z}\} \quad (4.8.2)$$

als Menge der Gitterpunkte. Dagegen hat man die Menge Γ_h der Randpunkte anders als im Falle des Quadrates zu definieren. Der linke Nachbarpunkt von $(x, y) \in \Omega_h$ lautet $(x - h, y)$. Falls die Verbindungsstrecke $\{(x - \vartheta h, y) : \vartheta \in (0, 1]\}$ nicht vollständig in Ω liegt, gibt es einen Randpunkt

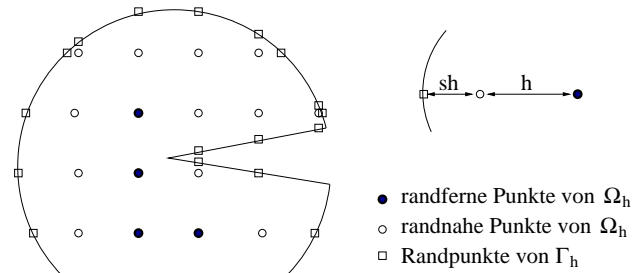


Abb. 4.8.1. Ω_h und Γ_h

Falls die Verbindungsstrecke $\{(x - \vartheta h, y) : \vartheta \in (0, 1]\}$ nicht vollständig in Ω liegt, gibt es einen Randpunkt

$$(x - sh, y) \in \Gamma \quad \text{mit } (x, y) \in \Omega_h, \quad s \in (0, 1], \quad (x - \vartheta h, y) \in \Omega \text{ für alle } \vartheta \in [0, s), \quad (4.8.3a)$$

der nun anstelle von $(x - h, y)$ als linker Nachbarpunkt von (x, y) bezeichnet wird (vgl. Abbildung 4.8.1). Entsprechend können rechte, untere und obere Nachbarpunkte Randpunkte der folgenden Gestalt sein:

$$(x + sh, y) \in \Gamma \quad \text{mit } (x, y) \in \Omega_h, \quad s \in (0, 1], \quad (x + \vartheta h, y) \in \Omega \text{ für alle } \vartheta \in [0, s), \quad (4.8.3b)$$

$$(x, y - sh) \in \Gamma \quad \text{mit } (x, y) \in \Omega_h, \quad s \in (0, 1], \quad (x, y - \vartheta h) \in \Omega \text{ für alle } \vartheta \in [0, s), \quad (4.8.3c)$$

$$(x, y + sh) \in \Gamma \quad \text{mit } (x, y) \in \Omega_h, \quad s \in (0, 1], \quad (x, y + \vartheta h) \in \Omega \text{ für alle } \vartheta \in [0, s). \quad (4.8.3d)$$

Wir setzen

$$\Gamma_h := \{\text{Randpunkte, die eine der Bedingungen (4.8.3a-d) erfüllen}\} \quad (4.8.4)$$

Ein Gitterpunkt $(x, y) \in \Omega_h$, der einen Nachbarn aus Γ_h besitzt, heißt *randnah*. Alle anderen Punkte von Ω_h heißen *randfern*. Wie aus Abbildung 4.8.1 hervorgeht, kann ein Punkt randnah sein, obwohl $(x \pm h, y)$ und $(x, y \pm h)$ zu Ω_h gehören (wenn nämlich nicht die gesamte Verbindungsstrecke in Ω verläuft).

Übungsaufgabe 4.8.1 Für ein konvexes Gebiet Ω zeige man

$$\Gamma_h = \{(x, y) \in \Gamma : x/h \in \mathbb{Z} \text{ oder } y/h \in \mathbb{Z}\}.$$

Warum gilt diese Charakterisierung nicht im Allgemeinen ohne die Konvexitätsvoraussetzung?

Will man die zweite Ableitung $u''(x)$ mit Hilfe der Werte von u bei $x' < x < x''$ approximieren, so kann man die Newtonschen⁸ dividierten Differenzen verwenden:

$$u''(x) = 2 \left[\frac{u(x'') - u(x)}{x'' - x} - \frac{u(x) - u(x')}{x - x'} \right] / (x'' - x') + \text{Rest}. \quad (4.8.5)$$

Übungsaufgabe 4.8.2 Man zeige: (a) Taylor-Entwicklung liefert

$$|\text{Rest}| \leq \frac{1}{3} \frac{(x'' - x)^2 + (x - x')^2}{x'' - x'} \|u\|_{C^3([x', x''])} \leq \frac{\max\{x'' - x, x - x'\}}{3} \|u\|_{C^3([x', x''])} \quad (4.8.6)$$

für den Rest aus Gleichung (4.8.5), falls $u \in C^3([x', x''])$.

(b) In Gleichung (4.8.6) lässt sich die Norm von $C^3([x', x''])$ durch die von $C^{2,1}([x', x''])$ ersetzen.

(c) Falls $x'' = x + h$ und $x' = x - h$, stimmt die Differenz in (4.8.5) mit der üblichen zweiten Differenz $\partial^- \partial^+ u(x)$ überein.

⁸Sir Isaac Newton, geb. 4. Jan. 1643 in Woolsthorpe, gest. 31. März 1727 in London

Um die Differenzengleichung für $-\Delta u = f$ in $(x, y) \in \Omega_h$ aufzustellen, verwenden wir wieder die vier Nachbarpunkte

$$(x - s_\ell h, y), \quad (x + s_r h, y), \quad (x, y - s_u h), \quad (x, y + s_o h) \in \Omega_h \cup \Gamma_h, \quad 0 < s_* \leq 1,$$

die oben definiert wurden. Für randferne Punkte gilt $s_* = 1$; für randnahe Punkte (x, y) liegt mindestens ein Nachbar auf Γ_h , und der entsprechende Abstand $s_* h$ kann kleiner als h sein. Gleichung (4.8.5) mit $x' = x - s_\ell h$ und $x'' = x + s_r h$ liefert eine Approximation für u_{xx} . Analog kann u_{yy} durch eine dividierte Differenz ersetzt werden. Insgesamt erhält man das *Differenzenschema von Shortley und Weller* [76]:

$$-D_h u(x, y) := \frac{1}{h^2} \left\{ \begin{aligned} & - \frac{2}{s_o(s_o + s_u)} u(x, y + s_o h) \\ & - \frac{2}{s_\ell(s_\ell + s_r)} u(x - s_\ell h, y) + \left(\frac{2}{s_\ell s_r} + \frac{2}{s_u s_o} \right) u(x, y) - \frac{2}{s_r(s_\ell + s_r)} u(x + s_r h, y) \\ & - \frac{2}{s_u(s_o + s_u)} u(x, y - s_u h) \end{aligned} \right\} \quad (4.8.7)$$

Bemerkung 4.8.3 Sobald $s_\ell = s_r = s_u = s_o = 1$, stimmt D_h mit der Standard-Fünfpunktformel Δ_h überein.

Das diskrete Randwertproblem erhält die Gestalt

$$-D_h u_h = f_h := \tilde{R}_h f \quad \text{auf } \Omega_h \quad \text{mit} \quad (\tilde{R}_h f)(\mathbf{x}) = f(\mathbf{x}), \quad (4.8.8a)$$

$$u_h = \varphi \quad \text{auf } \Gamma_h. \quad (4.8.8b)$$

Die fünf Koeffizienten in der rechten Seite der Gleichung (4.8.7) definieren die Matrixelemente $L_{\mathbf{x}\xi}$ für $\xi = \mathbf{x}$ und für die vier Nachbarn ξ von \mathbf{x} . Sonst wird $L_{\mathbf{x}\xi} = 0$ gesetzt. Die rechte Seite des Gleichungssystems

$$L_h u_h = q_h, \quad (4.8.9a)$$

in der u_h als Gitterfunktion auf Ω_h (nicht $\Omega_h \cup \Gamma_h$) aufgefasst wird, ist durch

$$q_h = f_h + \varphi_h, \quad \varphi_h(\mathbf{x}) := - \sum_{\xi \in \Gamma_h} L_{\mathbf{x}\xi} \varphi(\xi) \quad (4.8.9b)$$

gegeben. Wieder gilt $\varphi_h(\mathbf{x}) = 0$ für randferne $\mathbf{x} \in \Omega_h$.

Satz 4.8.4 Ω sei beschränkt und in einem Streifen $(x_0, x_0 + d) \times \mathbb{R}$ oder $\mathbb{R} \times (y_0, y_0 + d)$ der Breite d enthalten. Für die zur Shortley-Weller-Diskretisierung gehörende Matrix L_h gilt:

- a) L_h ist im Allgemeinen nicht symmetrisch;
- b) L_h ist eine M-Matrix mit

$$\|L_h^{-1}\|_\infty \leq d^2/8. \quad (4.8.10)$$

Beweis. a) Sei $\mathbf{x} = (x, y) \in \Omega_h$ randnah wegen $(x - s_\ell h, y) \in \Gamma_h$, aber der Nachbar $\mathbf{x}' = (x + h, y) \in \Omega_h$ randfern. Dann gilt $L_{\mathbf{x}\mathbf{x}'} = -2h^{-2}/(1 + s_\ell) \neq -h^{-2} = L_{\mathbf{x}'\mathbf{x}}$, sobald $s_\ell < 1$. Anders als in Übungsaufgabe 4.7.1 lässt sich im Allgemeinen keine Skalierung finden, sodass $D_h L_h$ (D_h diagonal) symmetrisch wird.

b) L_h braucht nicht notwendig irreduzibel und damit irreduzibel diagonaldominant zu sein. Aber die schwächere Bedingung aus Übungsaufgabe 4.3.8 ist erfüllt und beweist die M-Matrix-Eigenschaft.

c) Zum Beweis von (4.8.10) wird Satz 4.3.16 verwendet. Liegt Ω im Streifen $(x_0, x_0 + d) \times \mathbb{R}$, wählen wir $w_h(x, y) := R_h w$, $w := (x - x_0)(x_0 + d - x)/2$. Der Rest in Gleichung (4.8.5) enthält nur dritte Ableitungen, die für w verschwinden. Daher stimmt $D_h w_h$ mit $\Delta w = -1$ überein:

$$-D_h u_h = \mathbf{1} \quad \text{in } \Omega_h, \quad w_h \geq 0 \quad \text{auf } \Gamma_h.$$

Das entsprechende Gleichungssystem lautet $L_h w_h = q_h := f_h + \varphi_h$ mit $f_h = \mathbf{1}$ und $q_h \geq 0$. Daher gilt $L_h w_h \geq \mathbf{1}$ und Satz 4.3.16 beweist $\|L_h^{-1}\|_\infty \leq \|w_h\|_\infty \leq (d/2)^2/2 = d^2/8$. ■

Übungsaufgabe 4.8.5 Man beweise das Analogon der Abschätzung (4.4.10):

$$\|u_h\|_\infty \leq \|L_h^{-1}\|_\infty \|f_h\|_\infty + \max_{\xi \in \Gamma_h} |\varphi(\xi)| \leq \frac{d^2}{8} \max_{\mathbf{x} \in \Omega_h} |f(\mathbf{x})| + \max_{\xi \in \Gamma_h} |\varphi(\xi)|.$$

Mit (4.8.10) ist die Stabilität nachgewiesen. Dagegen ist die Konsistenzordnung nur noch 1, denn in randnahen Punkten hat

$$c_h := D_h R_h u - \tilde{R}_h \Delta u$$

wegen der nichtsymmetrischen Differenz nur die Größenordnung $\mathcal{O}(h^1)$. Dabei ist R_h die Beschränkung auf $\Omega_h \cup \Gamma_h$. \tilde{R}_h wurde in (4.8.8a) definiert. Wir wollen zeigen, dass trotzdem Konvergenz der zweiten Ordnung vorliegt. Die Differenz $w_h := u_h - R_h u$ der diskreten Lösung $u_h = L_h^{-1} q_h$ und der Lösung $u \in C^{3,1}(\bar{\Omega})$ von (4.8.1) erfüllt

$$-D_h w_h = -D_h u_h + D_h R_h u = \tilde{R}_h f + D_h R_h u = D_h R_h u - \tilde{R}_h \Delta u = c_h \quad \text{in } \Omega_h, \quad (4.8.11a)$$

$$w_h = 0 \quad \text{auf } \Gamma_h, \quad (4.8.11b)$$

sodass $w_h = L_h^{-1} c_h$ folgt. c_h lässt sich schreiben als $c_h^x + c_h^y$, wobei c_h^x (c_h^y) der Diskretisierungsfehler der x -Differenz (y -Differenz) ist. c_h^x wird wiederum in $c_h^x = c_h^{x,1} + c_h^{x,2}$ aufgespalten:

$$c_h^{x,2}(x, y) := \begin{cases} c_h^x(x, y), & \text{falls } s_\ell = s_r = 1 \\ 0 & \text{sonst} \end{cases}, \quad c_h^{x,1} := c_h^x - c_h^{x,2}.$$

Analog definiert man $c_h^{y,1}$ und $c_h^{y,2}$ und setzt

$$c_h^1 = c_h^{x,1} + c_h^{y,1}, \quad c_h^2 = c_h^{x,2} + c_h^{y,2}, \quad w_h^i := L_h^{-1} c_h^i \quad (i = 1, 2).$$

Die Fehler c_h^2 werden durch (4.5.7) beschrieben:

$$\|w_h^2\|_\infty \leq \|L_h^{-1}\|_\infty \|c_h^2\|_\infty, \quad \|c_h^2\|_\infty \leq \frac{1}{6} h^2 \|u\|_{C^{3,1}(\bar{\Omega})}. \quad (4.8.12a)$$

Mit $K := \frac{1}{3} h^3 \|u\|_{C^{2,1}(\bar{\Omega})}$ definiere man

$$v_h = K \mathbf{1} \quad \text{in } \Omega_h, \quad v_h = 0 \quad \text{auf } \Gamma_h, \quad \tilde{c}_h := L_h v_h.$$

In randfernen $\mathbf{x} \in \Omega_h$ ist $\tilde{c}_h(\mathbf{x}) = 0$, in randnahen $\mathbf{x} \in \Omega_h$ gilt dagegen

$$\tilde{c}_h(\mathbf{x}) = K \sum_{\xi \in \Omega_h} L_{\mathbf{x}\xi} = -K \sum_{\xi \in \Gamma_h} L_{\mathbf{x}\xi} \quad (\mathbf{x} \in \Omega_h \text{ randnah}).$$

Man betrachte zum Beispiel den Fall $\mathbf{x} = (x, y) \in \Omega_h$ randnah, $\xi = (x - s_\ell h, y) \in \Gamma_h$. Gemäß Übungsaufgabe 4.8.2a,b gehört zur x -Differenz der Fehler

$$|c_h^{x,1}(\mathbf{x})| \leq \frac{h}{3} \frac{s_r^2 + s_\ell^2}{s_r + s_\ell} \|u\|_{C^{2,1}(\bar{\Omega})} = h^{-2} \frac{s_r^2 + s_\ell^2}{s_r + s_\ell} K \leq K \frac{2h^{-2}}{s_\ell(s_r + s_\ell)} = -K L_{\mathbf{x}\xi}.$$

Die analoge Abschätzung für $c_h^{y,1}(\mathbf{x})$ ergibt $|c_h^1(\mathbf{x})| \leq \tilde{c}_h(\mathbf{x})$. Wegen $c_h^1(\mathbf{x}) = \tilde{c}_h(\mathbf{x}) = 0$ in randfernen $\mathbf{x} \in \Omega_h$ hat man $-\tilde{c}_h \leq c_h h^1 \leq \tilde{c}_h$. Da $L_h^{-1} \geq 0$, folgt $-v_h \leq w_h^1 \leq v_h$, d.h.

$$\|w_h^1\|_\infty \leq K = \frac{1}{3} h^3 \|u\|_{C^{2,1}(\bar{\Omega})}. \quad (4.8.12b)$$

(4.8.12a,b) zusammen mit $w_h^1 + w_h^2 = w_h = u_h - R_h u$ beweist den folgenden Satz.

Satz 4.8.6 (Konvergenz des Shortley-Weller-Verfahrens) Ω erfülle die Voraussetzung des Satzes 4.8.4. Das Shortley-Weller-Verfahren konvergiert von zweiter Ordnung, falls $u \in C^{3,1}(\bar{\Omega})$:

$$\|u_h - R_h u_h\|_\infty \leq \frac{1}{3} h^3 \|u\|_{C^{2,1}(\bar{\Omega})} + \|L_h^{-1}\|_\infty \frac{1}{6} h^2 \|u\|_{C^{3,1}(\bar{\Omega})} \leq \left(\frac{1}{3} h^3 + \frac{d^2}{48} h^2 \right) \|u\|_{C^{3,1}(\bar{\Omega})}. \quad (4.8.13)$$

Übungsaufgabe 4.8.7 Man zeige: Verwendet man die Shortley-Weller-Diskretisierung für alle randnahen Gitterpunkte, aber das Mehrstellenverfahren aus Abschnitt 4.6 für alle randfernen Punkte, so ergibt sich ein Verfahren der dritter Konvergenzordnung: $\|u_h - R_h u\|_\infty = \mathcal{O}(h^3)$.

4.8.2 Interpolation in randnahen Punkten

Statt in randnahen Gitterpunkten $\mathbf{x} \in \Omega_h$ die Poisson-Gleichung zu diskretisieren, kann man auch versuchen, $u_h(\mathbf{x})$ durch Interpolation aus den Nachbarpunkten zu bestimmen. Ist z.B. $\mathbf{x} = (x, y) \in \Omega_h$ randnah, $(x - s_\ell h, y) \in \Gamma_h$ und $(x + s_r h, y) \in \Omega_h \cup \Gamma_h$, so ergibt lineare Interpolation den Wert

$$u_h(x, y) = [s_\ell u_h(x + s_r h, y) + s_r u_h(x - s_\ell h, y)] / (s_r + s_\ell).$$

Daher wird im Punkte \mathbf{x} die Gleichung

$$(s_r + s_\ell) u_h(x, y) - s_\ell u_h(x + s_r h, y) - s_r u_h(x - s_\ell h, y) = 0 \quad (4.8.14a)$$

aufgestellt. Da $\xi = (x - s_\ell h, y)$ Randpunkt sein sollte, lässt sich $u_h(\xi)$ durch $\varphi(\xi)$ ersetzen. Ist dagegen $(x, y + s_o h)$ oder $(x, y - s_u h)$ ein Randpunkt, wählen wir die Interpolation in y -Richtung:

$$(s_u + s_o) u_h(x, y) - s_u u_h(x + s_o h, y) - s_o u_h(x - s_u h, y) = 0 \quad (4.8.14b)$$

In allen randfernen $\mathbf{x} \in \Omega_h$ wird die Fünfpunktformel (4.2.3) verwendet:

$$-(\Delta_h u_h)(\mathbf{x}) = f_h(\mathbf{x}) = \tilde{R}_h f(\mathbf{x}) = f(\mathbf{x}) \quad (\mathbf{x} \text{ randfern}). \quad (4.8.14c)$$

Satz 4.8.8 Ω erfülle die Voraussetzungen des Satzes 4.8.4. Die Diskretisierung sei durch (4.8.14a-c) gegeben, wobei die Wahl zwischen (4.8.14a) und (4.8.14b) so getroffen wird, dass stets (mindestens) ein Randpunkt für die Interpolation benutzt wird. Das nach der Elimination der Randwerte $u_h(\xi) = \varphi(\xi)$, $\xi \in \Gamma_h$, entstehende Gleichungssystem sei $L_h u_h = q_h$. L_h ist eine (im Allgemeinen unsymmetrische) M -Matrix, die der Abschätzung (4.8.10) genügt. Die diskreten Lösungen u_h konvergieren von der Ordnung 2, falls $u \in C^{3,1}(\bar{\Omega})$:

$$\|u_h - R_h u_h\|_\infty \leq h^2 \|u\|_{C^{1,1}(\bar{\Omega})} + \|L_h^{-1}\|_\infty \frac{1}{6} h^2 \|u\|_{C^{3,1}(\bar{\Omega})}. \quad (4.8.15)$$

Beweis. 1) Die M -Matrix-Eigenschaft und (4.8.10) beweist man wie in Satz 4.8.4.

2) $\mathbf{x} \in \Omega_h$ sei ein randnaher Punkt, in dem (4.8.14a) verwendet wird. Der Interpolationsfehler beträgt

$$\begin{aligned} c_h^1(x, y) &:= (s_r + s_\ell) R_h u(x, y) - s_\ell R_h u(x + s_r h, y) - s_r R_h u(x - s_\ell h, y), \\ |c_h^1(x, y)| &\leq \frac{1}{2} s_r s_\ell (s_r + s_\ell) h^2 \|u\|_{C^{1,1}(\bar{\Omega})}. \end{aligned}$$

Mit diesem c_h^1 und $K := h^2 \|u\|_{C^{1,1}(\bar{\Omega})}$ lässt sich der Beweis zu Satz 4.8.6 sinngemäß wiederholen. ■

Indem man die Gleichungen (4.8.14a,b) umskaliert und addiert, erhält man

$$\begin{aligned} \frac{1}{h^2} \left\{ \left(\frac{s_r + s_\ell}{s_r s_\ell} + \frac{s_o + s_u}{s_o s_u} \right) u_h(x, y) - \frac{1}{s_\ell} u_h(x - s_\ell h, y) \right. \\ \left. - \frac{1}{s_r} u_h(x + s_r h, y) - \frac{1}{s_o} u_h(x, y + s_o h) - \frac{1}{s_u} u_h(x, y - s_u h) \right\} = 0. \end{aligned} \quad (4.8.16)$$

Hierdurch lässt sich auch bei beliebigem Ω eine symmetrische Matrix L_h erreichen.

Übungsaufgabe 4.8.9 Man zeige: Die Diskretisierung (4.8.14c), (4.8.16) führt zu einer symmetrischen M -Matrix. Die Konvergenzabschätzung lautet

$$\|u_h - R_h u_h\|_\infty \leq 2h^2 \|u\|_{C^{1,1}(\bar{\Omega})} + \frac{1}{6} d^2 h^2 \|u\|_{C^{3,1}(\bar{\Omega})}.$$

In (4.8.14a,b) wurde die lineare Interpolation gewählt, weil hierzu die Werte der Nachbarpunkte ausreichen. Konstante Interpolation durch

$$u(x, y) = u_h(x - s_\ell h, y) = \varphi(x - s_\ell h, y), \quad \text{falls } (x, y) \in \Omega_h, (x - s_\ell h, y) \in \Gamma_h,$$

ist ungünstiger, da sie nur Konvergenz erster Ordnung liefert: $\|u_h - R_h u_h\|_\infty = \mathcal{O}(h)$. Dagegen ist Interpolation höherer Ordnung durchaus anwendbar (vgl. Pereyra-Proskurowski-Widlund [67]). Sie wird aber durch eine Gleichung beschrieben, die auch Punkte in einem Abstand $\geq 2h$ enthält. Höhere Randapproximationen sind insbesondere dann notwendig, wenn man Extrapolationsverfahren anwenden will (vgl. Marchuk-Shaidurov [61, p. 162ff]).

5

Allgemeine Randwertaufgaben

5.1 Dirichlet-Randwertaufgaben für lineare Differentialgleichungen zweiter Ordnung

5.1.1 Problemstellung

In §1.2 wurde bereits die allgemeine lineare Differentialgleichung zweiter Ordnung formuliert:

$$Lu = f \quad \text{in } \Omega \quad \text{mit } L = \sum_{i,j=1}^n a_{ij}(\mathbf{x}) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^n a_i(\mathbf{x}) \frac{\partial}{\partial x_i} + a(\mathbf{x}). \quad (5.1.1a)$$

Es wurde erwähnt, dass ohne Beschränkung der Allgemeinheit

$$a_{ij}(\mathbf{x}) = a_{ji}(\mathbf{x}) \quad (i, j \in \{1, \dots, n\}, \mathbf{x} \in \Omega) \quad (5.1.1b)$$

vorausgesetzt werden kann, sodass die Matrix

$$A(\mathbf{x}) = (a_{ij}(\mathbf{x}))_{i,j=1,\dots,n} \quad (5.1.1c)$$

symmetrisch ist. Ein scheinbar allgemeinerer Differentialoperator als (5.1.1a) ist

$$L = \sum_{i,j=1}^n \left[a_{ij}^I \frac{\partial^2}{\partial x_i \partial x_j} + \frac{\partial}{\partial x_j} a_{ij}^{II} \frac{\partial}{\partial x_i} + \frac{\partial^2}{\partial x_i \partial x_j} a_{ij}^{III} \right] + \sum_{i=1}^n \left[a_i^I \frac{\partial}{\partial x_i} + \frac{\partial}{\partial x_i} a_i^{II} \right] + a(\mathbf{x}). \quad (5.1.2)$$

Da aber z.B. $\frac{\partial}{\partial x_j} (a_{ij}^{II} \frac{\partial}{\partial x_i} u) = a_{ij}^{II} u_{x_i x_j} + (\partial a_{ij}^{II} / \partial x_j) u_{x_i}$ lässt sich der Operator (5.1.2) in der Form (5.1.1a) schreiben, wenn nur die Koeffizienten hinreichend oft differenzierbar sind. Nach Definition 1.2.3 ist Gleichung (5.1.1a) elliptisch in Ω , falls alle Eigenwerte von $A(\mathbf{x})$ das gleiche Vorzeichen besitzen. Ohne Einschränkung der Allgemeinheit kann angenommen werden, dass alle Eigenwerte positiv sind, sodass $A(\mathbf{x})$ positiv definit ist (vgl. Übungsaufgabe 4.3.22a). Damit ist L in Ω elliptisch, wenn

$$\sum_{i,j=1}^n a_{ij}(\mathbf{x}) \xi_i \xi_j > 0 \quad \text{für alle } \mathbf{x} \in \Omega, 0 \neq \xi \in \mathbb{R}^n. \quad (5.1.3a)$$

Für alle $\mathbf{x} \in \Omega$ existiert $c(\mathbf{x}) := \min\{a_{ij}(\mathbf{x}) \xi_i \xi_j : |\xi| = 1\}$ und muss positiv sein ($c(\mathbf{x})$ ist der kleinste Eigenwert von $A(\mathbf{x})$). Somit kann man (5.1.3a) auch in der Form (5.1.3a') schreiben:

$$\sum_{i,j=1}^n a_{ij}(\mathbf{x}) \xi_i \xi_j \geq c(\mathbf{x}) |\xi|^2, \quad c(\mathbf{x}) > 0 \text{ für alle } \mathbf{x} \in \Omega, 0 \neq \xi \in \mathbb{R}^n. \quad (5.1.3a')$$

Definition 5.1.1 Gleichung (5.1.1a) bzw. der Operator L heißen gleichmäßig elliptisch in Ω , wenn

$$\inf\{c(\mathbf{x}) : \mathbf{x} \in \Omega\} > 0 \quad (c(\mathbf{x}) \text{ aus (5.1.3a')}). \quad (5.1.3b)$$

Bemerkung 5.1.2 Wenn L stetige Koeffizienten im Gebiet Ω besitzt und dort elliptisch ist, ist L in jedem Kompaktum $K \subset \Omega$ gleichmäßig elliptisch.

Beweis. a) Da L in K elliptisch ist, gilt $c(\mathbf{x}) > 0$ in K .

b) Nach Voraussetzung ist $c(\cdot)$ stetig und nimmt daher auf K ein Minimum an, das nach a) positiv sein muss. ■

Auf $\Gamma = \partial\Omega$ wird die Dirichlet-Randwertbedingung gestellt:

$$u = \varphi \quad \text{auf } \Gamma. \quad (5.1.4)$$

Übungsaufgabe 5.1.3 Sei $\mathbf{x} \mapsto \Phi(\mathbf{x}) := \mathbf{x}' := \mathbf{x}_0 + \mathbf{S}\mathbf{x}$ (\mathbf{S} : reguläre $n \times n$ -Matrix) eine affine Abbildung von Ω auf Ω' . Man zeige, dass der Operator L' der transformierten Differentialgleichung wieder die gleichen Elliptizitätseigenschaften besitzt und $a'(\Phi(\mathbf{x})) = a(\mathbf{x})$ für den Koeffizienten der nullten Ableitung in $L' = \sum_{i,j=1}^n a'_{ij}(\mathbf{x}') \frac{\partial^2}{\partial x'_i \partial x'_j} + \sum_{i=1}^n a'_i(\mathbf{x}') \frac{\partial}{\partial x'_i} + a'(\mathbf{x}')$ gilt. Hinweis: Übungsaufgabe 1.2.5.

5.1.2 Maximumprinzip

Im Allgemeinen wird für die Gleichung $Lu = f$ weder das Maximumprinzip gelten, noch ist die Lösung der Randwertaufgabe (5.1.3a), (5.1.4) eindeutig bestimmt.

Beispiel 5.1.4 Seien $\Omega = (0, \pi) \times (0, \pi)$, $\varphi = 0$, $f = 0$, $L = \Delta + 2$. Dann sind sowohl $u = 0$ als auch $u(x, y) = \sin(x) \sin(y)$ Lösungen der Randwertaufgabe. Die zweite Lösung nimmt ihr striktes Maximum in dem inneren Punkt $(\pi/2, \pi/2) \in \Omega$ an.

Im vorliegenden Beispiel hat der Koeffizient $a(\mathbf{x}) = 2$ (vgl. (5.1.1b)) das "falsche" Vorzeichen. Sobald $a \leq 0$, ist die folgende Aussage anwendbar, die auf Hopf¹ zurückgeht.

Satz 5.1.5 (Maximum-Minimum-Prinzip) Die Koeffizienten des elliptischen Operators (5.1.1a) seien stetig in Ω . $u \in C^2(\Omega)$ erfülle $Lu = f$. Für jedes Kompaktum $K \subset \bar{\Omega}$ mit $a(\mathbf{x}) \leq 0$ in K und $u \in C^0(K)$ gilt:

- a) Sind $f \leq 0$ und $u \leq 0$ in K , so nimmt u sein Minimum auf ∂K an.
- b) Sind $f \geq 0$ und $u \geq 0$ in K , so nimmt u sein Maximum auf ∂K an.

Beweis. 1) Es reicht, den Teil a) zu beweisen, da b) nach Übergang zu $-u$ mit a) identisch ist.

2) Wir führen den Beweis indirekt. Da K kompakt ist, nimmt u in K ein globales Minimum an. Wenn dieses nicht auf ∂K liegt, besitzt u in $\mathbf{x}^* \in K$ (Inneres von K) ein lokales (und globales) Minimum

$$u(\mathbf{x}^*) < \min_{x \in \partial K} u(\mathbf{x}) \leq 0. \quad (5.1.5a)$$

Folglich verschwinden die ersten Ableitungen: $u_{x_i}(\mathbf{x}^*) = 0$, und die Hesse²-Matrix $B := (u_{x_i x_j}(\mathbf{x}^*))_{i,j=1,\dots,n}$ ist positiv semidefinit (d.h. $B = B^\top$ und $\langle \xi, B\xi \rangle \geq 0$ für alle $\xi \in \mathbb{R}^n$). Auswertung der Differentialgleichung in \mathbf{x}^* liefert

$$0 \geq f(\mathbf{x}^*) = (Lu)(\mathbf{x}^*) = \underbrace{\sum_{i,j=1}^n a_{ij}(\mathbf{x}^*) u_{x_i x_j}(\mathbf{x}^*)}_{\geq 0} + \underbrace{a(\mathbf{x}^*)}_{\leq 0} \underbrace{u(\mathbf{x}^*)}_{< 0} \geq 0. \quad (5.1.5b)$$

Von den unterlegten Ungleichungen ist lediglich $\text{Spur}(A(\mathbf{x}^*)B) = \sum_{i,j=1}^n a_{ij}(\mathbf{x}^*) u_{x_i x_j}(\mathbf{x}^*) \geq 0$ noch zu zeigen. Dies wird Gegenstand der Übungsaufgabe 5.1.6 sein.

Verstärken wir die Annahme $f \leq 0$ zu $f < 0$, liefert (5.1.5b) den Widerspruch $0 > f(\mathbf{x}^*) = \dots \geq 0$.

¹Eberhard Frederick Ferdinand Hopf, geb. am 4. April 1902 in Salzburg, gest. am 24. Juli 1983

²Ludwig Otto Hesse, geb. am 22. April 1811 in Königsberg, gest. am 4. August 1874 in München

3) Es bleibt der Fall $f \leq 0$ zu untersuchen. O.B.d.A. dürfen wir annehmen, dass K in $\{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_2 \geq 1\}$ liegt (vgl. Übungsaufgabe 5.1.3). Die Funktion $q(\mathbf{x}) := \exp(\alpha \|\mathbf{x}\|_2^2)$ ($\alpha \in \mathbb{R}$) hat die Eigenschaften

$$q(\mathbf{x}) > 0, \quad Lq(\mathbf{x}) = \left(4\alpha^2 \sum_{i,j=1}^n a_{ij}(\mathbf{x})x_i x_j + 2\alpha \sum_{i=1}^n a_{ii}(\mathbf{x}) + 2\alpha \sum_{i=1}^n a_i(\mathbf{x})x_i + a(\mathbf{x}) \right) q(\mathbf{x}). \quad (5.1.5c)$$

Wegen $\|\mathbf{x}\|_2 \geq 1$ ($\mathbf{x} \in K$) ist die quadratische Form $\sum_{i,j=1}^n a_{ij}(\mathbf{x})x_i x_j$ positiv, wenn wir voraussetzen, dass L nicht nur elliptisch, sondern auch gleichmäßig elliptisch ist. Für hinreichend großes $|\alpha|$ überwiegt der Anteil $4\alpha^2 \sum_{i,j=1}^n a_{ij}(\mathbf{x})x_i x_j$ und die Klammer in (5.1.5c) ist positiv. Mit dieser Wahl ist $Lq > 0$ in K . Wir setzen

$$v(\mathbf{x}) := u(\mathbf{x}) - cq(\mathbf{x}) \quad \text{mit } c > 0. \quad (5.1.5d)$$

v erfüllt die Differentialgleichung $Lv = g := f - cLq$ und die Bedingung $v \leq 0$. Wegen $Lq > 0$ ist $g < 0$.

Die Ungleichung aus (5.1.5a) gilt auch für v , wenn $c > 0$ hinreichend klein gewählt wird. Damit gelten die Ungleichungen $v \leq 0$ und $Lv = g < 0$ in K , aber das Minimum von v wird nicht auf dem Rand $\partial K_\rho(\mathbf{x}^*)$ angenommen. Dies ist ein Widerspruch zum Resultat aus Beweisteil 2), da das dortige f jetzt das $g < 0$ ist.

4) Wenn L in Ω nicht gleichmäßig elliptisch ist, gilt die Aussage des Satz immer noch für Kompakta $K' \subset \Omega$ (vgl. Bemerkung 5.1.2). Indem wir $K \subset \overset{\circ}{\Omega}$ von innen ausschöpfen:

$$K_1 \subset K_2 \subset \dots \subset \overset{\circ}{K}, \quad K_i = \overline{K_i}, \quad \text{dist}(K_i, \partial K) \rightarrow 0,$$

erhalten wir Minima $u(\mathbf{x}^i) = \min_{\mathbf{x} \in K_i} u(\mathbf{x})$ mit $\mathbf{x}^i \in \partial K_i$. Da $\text{dist}(\mathbf{x}^i, \partial K) \rightarrow 0$, konvergiert eine Teilfolge gegen ein $\mathbf{x}^* \in \partial K$ mit $u(\mathbf{x}^*) = \min_{\mathbf{x} \in K} u(\mathbf{x})$, sodass die Aussage insgesamt bewiesen ist. ■

Übungsaufgabe 5.1.6 Es ist $\text{Spur}(A) := \sum_{i=1}^n a_{ii}$. Man beweise:

a) $a_{ii} \geq 0$ und $\text{Spur}(A) \geq 0$, wenn A positiv semidefinit ist.

b) $\text{Spur}(AB) = \text{Spur}(BA) = \sum_{i,j=1}^n a_{ij}b_{ji}$.

c) $\text{Spur}(AB) \geq 0$, falls A und B positiv semidefinit sind.

Hinweis zu c): $B^{1/2}AB^{1/2}$ ist positiv semidefinit; Übungsaufgabe 4.3.22d.

Das Maximum-Minimum-Prinzip aus Satz 2.3.3 im Falle der Laplace-Gleichung ist weitergehend, da dort ein Maximum bzw. Minimum im Inneren impliziert, dass die Lösung u konstant sein muss. Dies bezeichnet man auch als *starkes Maximum-Minimum-Prinzip*.

Zusatz 5.1.7 Unter den Voraussetzungen von Satz 5.1.5 gilt auch das starke Maximum-Minimum-Prinzip: Werden die Extrema im Inneren von K angenommen, so gilt $u = \text{const}$ in K .

Beweis. Wir beschränken uns auf den Fall eines Minimum im Inneren von K , wobei $Lu = f \leq 0$ und $u \leq 0$ in K vorausgesetzt ist. In $\mathbf{x}_0 \in \overset{\circ}{K}$ werde das globale Minimum angenommen:

$$u(\mathbf{x}_0) = m := \min\{u(\mathbf{x}) : \mathbf{x} \in K\}.$$

Falls die Menge

$$\omega := \{\mathbf{x} \in K : u(\mathbf{x}) > m\}$$

leer ist, folgt die Behauptung $u = m$ in K . Zum indirekten Beweis sei deshalb $\omega \neq \emptyset$ angenommen.

Da ω offen (relativ zu K) ist, kann es keine Teilmenge von ∂K sein. $\partial\omega \subset \partial K$ kann ebenfalls ausgeschlossen werden (da sonst $\overline{\omega} = K$ im Widerspruch zu $\mathbf{x}_0 \in \overset{\circ}{K}$ und $\mathbf{x}_0 \notin \omega$). Also existiert ein $\mathbf{x}' \in \partial\omega \setminus \partial K$ mit $d := \text{dist}(\mathbf{x}', \partial K) > 0$. Sei $\mathbf{y} \in \omega$ ein Punkt in der $d/3$ -Umgebung von \mathbf{x}' . Offenbar gilt

$$\delta := \text{dist}(\mathbf{y}, \partial\omega) \leq d/3 < 2d/3 \leq \text{dist}(\mathbf{y}, \partial K).$$

Die offene Kugel $K_\delta := K_\delta(\mathbf{y})$ erfüllt nach Konstruktion von δ

$$K_\delta \subset \omega, \quad \overline{K_\delta} \subset \overset{\circ}{K} \quad (\text{d.h. } \overline{K_\delta} \cap \partial K = \emptyset), \quad \overline{K_\delta} \cap (\partial\omega \setminus \partial K) \neq \emptyset,$$

d.h. es existiert ein $\mathbf{z} \in \partial K_\delta \cap \partial\omega \setminus \partial K \subset \overset{\circ}{K} \setminus \omega$. Nach Definition der Mengen gilt

$$u(\mathbf{y}) > m, \quad u(\mathbf{z}) = m.$$

O.B.d.A. kann (nach geeigneter affiner Transformation, vgl. Übungsaufgabe 5.1.3) die Situation aus Abbildung 5.1.1 angenommen werden:

$$\mathbf{y} = \mathbf{0}, \quad \mathbf{z} = \delta \mathbf{e} \quad (\mathbf{e}: \text{Einheitsvektor in } x_1\text{-Richtung}).$$

Die Abbildung zeigt weiterhin die Kugel $K_{\delta/2} := K_{\delta/2}(\mathbf{y}) \subset K_\delta$ und das (offene) Ringgebiet $R := K_\delta \setminus \overline{K_{\delta/2}}$. Der Rand ∂R ist die Vereinigung von $\partial K_{\delta/2}$ und ∂K_δ . Auf $\partial K_{\delta/2}$ ist wegen $\partial K_{\delta/2} \subset \omega$

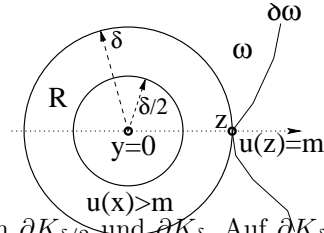


Abb. 5.1.1. Skizze zum Beweis

$$\min\{u(\mathbf{x}) : \mathbf{x} \in \partial K_{\delta/2}\} > m,$$

während $u(\mathbf{x}) \geq m$ für alle $\mathbf{x} \in \partial K_\delta$ mit dem Minimum $u(\mathbf{z}) = m$ in $\mathbf{z} \in \partial K_\delta$.

Die Funktion q aus dem Beweis von Satz 5.1.5 wird geringfügig variiert:

$$p(\mathbf{x}) := \exp(-\alpha \|\mathbf{x}\|_2^2) - \exp(-\alpha \delta^2) \quad \text{mit } \alpha > 0 \text{ für } \delta/2 \leq \|\mathbf{x}\|_2 \leq \delta \text{ (d.h. } \mathbf{x} \in \overline{R})$$

Weiterhin gilt $Lp \geq 0$ und $p \geq 0$ in \overline{R} für hinreichend großes $\alpha > 0$. Damit erfüllt $w := u - cp$ für $c \geq 0$ die Vorzeichenbedingungen $Lw =: g \leq 0$ und $w \leq 0$. Nach Satz 5.1.5 nimmt w als Funktion auf \overline{R} sein Minimum auf $\partial R = \partial K_{\delta/2} \cup \partial K_\delta$ an.

Für hinreichend kleines $c > 0$ ist weiterhin

$$\min\{w(\mathbf{x}) : \mathbf{x} \in \partial K_{\delta/2}\} > m,$$

während $p(\mathbf{x}) = 0$ auf ∂K_δ zu $w|_{\partial K_\delta} = u|_{\partial K_\delta}$ und damit

$$\min\{w(\mathbf{x}) : \mathbf{x} \in \partial K_\delta\} = \min\{u(\mathbf{x}) : \mathbf{x} \in \partial K_\delta\} = u(\mathbf{z}) = w(\mathbf{z}) = m$$

führt. Insbesondere erzwingt das Maximum-Minimum-Prinzip $w(\mathbf{x}) \geq m$ in \overline{R} .

Da w (beschränkt auf \overline{R}) in \mathbf{z} ein Randminimum besitzt, muss (in dem Koordinatensystem der Abbildung) $\frac{\partial}{\partial x_1} w(\mathbf{z}) \leq 0$ gelten. Einsetzen der Definition $w = u - cp$ liefert

$$\frac{\partial u(\mathbf{z})}{\partial x_1} \leq c \frac{\partial p(\mathbf{z})}{\partial x_1} = -2c\delta\alpha e^{-\alpha\delta^2} < 0.$$

Andererseits ist $\mathbf{z} \in \partial K_\delta \subset \overset{\circ}{K}$ ein lokales Minimum von u , sodass $\frac{\partial u(\mathbf{z})}{\partial x_1} = 0$ im Widerspruch zur vorherigen Ungleichung gelten muss. ■

Zusatz 5.1.8 Falls der Koeffizient $a(\mathbf{x})$ von L (vgl. (5.1.1a)) identisch verschwindet, können in Satz 5.1.5 und Zusatz 5.1.7 die Bedingungen $u \leq 0$ (für ein Minimum) bzw. $u \geq 0$ (für ein Maximum) entfallen.

Beweis. Die Vorzeichenbedingungen werden nur benötigt, damit das Produkt au das richtige Vorzeichen besitzt, was durch $a = 0$ ebenfalls gesichert ist. ■

Bemerkung 5.1.9 Die Stetigkeit der Koeffizienten a_{ij}, a_i, a von L kann in Satz 5.1.5 und Zusatz 5.1.7 ersetzt werden durch die Annahme $a < 0$ in Ω oder durch die Voraussetzung: In jeder kompakten Menge $K \subset \Omega$ seien a_{ij}, a_i und a beschränkt und L gleichmäßig elliptisch.

5.1.3 Eindeutigkeit der Lösung und stetige Abhängigkeit

Lemma 5.1.10 Ω sei beschränkt, die Koeffizienten von L seien stetig, es gelte (5.1.3a) und $a \leq 0$ in Ω . $u_1, u_2 \in C^2(\Omega) \cap C^0(\overline{\Omega})$ seien Lösungen der Randwertaufgaben

$$Lu_i = f_i \quad \text{in } \Omega, \quad u_i = \varphi_i \quad \text{auf } \Gamma \quad (i = 1, 2). \quad (5.1.6)$$

Ist $f_1 \geq f_2$ in Ω und $\varphi_1 \leq \varphi_2$ auf Γ , so ist auch $u_1 \leq u_2$ in Ω .

Beweis. $v := u_2 - u_1$ erfüllt $Lv = f_2 - f_1 \leq 0$ in Ω und $v = \varphi_2 - \varphi_1 \geq 0$ auf Γ . Aus Satz 5.1.5 und $v \geq 0$ auf Γ schließt man, dass kein negatives Minimum von v existiert. Also $v \geq 0$, d.h. $u_2 \geq u_1$. ■

Satz 5.1.11 (Eindeutigkeit) *Unter den Voraussetzungen des Lemmas 5.1.10 ist die Lösung der Randwertaufgabe $Lu = f$ in Ω und $u = \varphi$ auf Γ eindeutig bestimmt.*

Beweis. Seien u_1, u_2 zwei Lösungen. Lemma 5.1.10 mit $f_1 = f_2 = f$, $\varphi_1 = \varphi_2 = \varphi$ zeigt $u_1 \leq u_2$ ebenso wie $u_2 \leq u_1$. Also $u_1 = u_2$. ■

Der nächste Satz besagt, dass die Lösung Lipschitz-stetig von f und φ abhängt.

Satz 5.1.12 *L sei gleichmäßig elliptisch in Ω . Unter den Voraussetzungen des Lemmas 5.1.10 gilt*

$$\|u_1 - u_2\|_\infty \leq \|\varphi_1 - \varphi_2\|_\infty + M \|f_1 - f_2\|_\infty \quad (5.1.7)$$

für Lösungen u_1, u_2 von (5.1.6). Dabei hängt die Zahl M nur von $K := \sup\{|a_{ij}(\mathbf{x})|, |a_i(\mathbf{x})|, |a(\mathbf{x})| : \mathbf{x} \in \Omega\}$, von der Elliptizitätskonstante $m := \inf\{c(\mathbf{x}) : \mathbf{x} \in \Omega\} > 0$ (vgl. (5.1.3b)) und vom Durchmesser des Bereichs Ω ab.

Beweis. Sei $\Omega \subset K_R(\mathbf{z})$. Für alle $\mathbf{x} \in \Omega$ gilt $z_1 - R \leq x_1 \leq z_1 + R$. Wir wählen $\alpha \geq 0$ so, dass

$$m\alpha^2 - K(\alpha + 1) \geq 1$$

und definieren

$$w(\mathbf{x}) := \|\varphi_1 - \varphi_2\|_\infty + \left(e^{2R\alpha} - e^{\alpha(x_1 - z_1 + R)}\right) \|f_1 - f_2\|_\infty.$$

Wir vergleichen w mit der Lösung $v := u_1 - u_2$ von $Lv = f_1 - f_2$ in Ω , $v = \varphi_1 - \varphi_2$ auf Γ . Nach Wahl von $K_R(\mathbf{z})$ gilt

$$w(\mathbf{x}) \geq v(\mathbf{x}) \quad \text{für } \mathbf{x} \in \Gamma.$$

Ferner erhält man nach Wahl von α :

$$\begin{aligned} (Lv)(\mathbf{x}) &= \underbrace{a(\mathbf{x}) \|\varphi_1 - \varphi_2\|_\infty}_{\leq 0} + \underbrace{\{ae^{2R\alpha} - e^{\alpha(x_1 - z_1 + R)}\}}_{\geq 1} \underbrace{\{a_{11}(\mathbf{x})\alpha^2 + a_1(\mathbf{x})\alpha + a(\mathbf{x})\}}_{\geq m} \|f_1 - f_2\|_\infty \\ &\leq -\{m\alpha^2 - K(\alpha + 1)\} \|f_1 - f_2\|_\infty \leq -\|f_1 - f_2\|_\infty \leq f_1(\mathbf{x}) - f_2(\mathbf{x}) = (Lv)(\mathbf{x}). \end{aligned}$$

Wendet man Lemma 5.1.10 mit $u_1 = v$, $u_2 = w$ an, ergibt sich $v \leq w$ in Ω , d.h.

$$u_1(\mathbf{x}) - u_2(\mathbf{x}) = v(\mathbf{x}) \leq w(\mathbf{x}) \leq \|w\|_\infty \leq \|\varphi_1 - \varphi_2\|_\infty + M \|f_1 - f_2\|_\infty,$$

wobei $M := e^{2R\alpha}$. Analog beweist man $-w \leq v$, sodass (5.1.7) folgt. ■

Übungsaufgabe 5.1.13 a) Ω sei beschränkt, die Koeffizienten von L seien stetig in Ω . Man zeige: L ist genau dann gleichmäßig elliptisch in Ω , wenn L in $\overline{\Omega}$ elliptisch ist.

b) Satz 5.1.12 gilt für den Spezialfall $f_1 = f_2$ ohne die Voraussetzung gleichmäßiger Elliptizität.

c) Ein Streifen im \mathbb{R}^n sei beschrieben durch

$$S = \{\mathbf{x} \in \mathbb{R}^n : 0 \leq \langle \eta, \mathbf{x} - \mathbf{x}^* \rangle \leq \delta\},$$

wobei $\mathbf{x}^* \in \mathbb{R}^n$ ein Randpunkt von S , $\eta \in \mathbb{R}^n$ mit $|\eta| = 1$ ein Einheitsvektor und δ die Streifenbreite sind. Man zeige: Ungleichung (5.1.7) gilt mit $M := e^{\delta\alpha}$, wenn $\Omega \subset S$ und α wie im Beweis des Satzes 5.1.12.

d) Unter den Voraussetzungen des Satzes 5.1.12 zeige man

$$\|u\|_\infty \leq \|\varphi\|_\infty + M \|f\|_\infty. \quad (5.1.8)$$

Übungsaufgabe 5.1.14 Die Koordinatentransformation $\Phi : \mathbf{x} \in \overline{\Omega} \mapsto \xi \in \overline{\Omega}'$ und die Inverse $\Phi^{-1} : \overline{\Omega}' \rightarrow \overline{\Omega}$ seien stetig differenzierbar. Dem Operator L (in den x -Koordinaten) entspreche L' in den ξ -Koordinaten. Man zeige: Erfüllt L die Voraussetzungen des Lemmas 5.1.7 bzw. des Satzes 5.1.12, so auch L' .

Die rechte Seite f und die Randwerte φ sind nicht die einzigen Parameter, von denen die Lösung u abhängt. Im Folgenden wird untersucht, wie die Lösung von den Koeffizienten a_{ij}, a_i, a des Differentialoperators L abhängt.

Satz 5.1.15 Die Koeffizienten von L^I und L^II seien a_{ij}^I, a_j^I, a^I bzw. $a_{ij}^{II}, a_j^{II}, a^{II}$. u^I und u^{II} seien Lösungen von

$$L^I u^I = L^{II} u^{II} = f \quad \text{in } \Omega, \quad u^I = u^{II} = \varphi \quad \text{auf } \Gamma.$$

L^I erfülle die Voraussetzungen des Satzes 5.1.12 und u^{II} gehöre zu $C^2(\overline{\Omega})$. Mit M aus (5.1.7) gilt dann

$$\|u^I - u^{II}\|_\infty \leq M \left\{ \sum_{i,j=1}^n \|a_{ij}^I - a_{ij}^{II}\|_\infty \|u^{II}\|_{C^2(\overline{\Omega})} + \sum_{i=1}^n \|a_i^I - a_i^{II}\|_\infty \|u^{II}\|_{C^1(\overline{\Omega})} + \|a^I - a^{II}\|_\infty \|u^{II}\|_\infty \right\}. \quad (5.1.9)$$

Falls $a_{ij}^I = a_{ij}^{II}$, genügt die Voraussetzung $u^{II} \in C^1(\overline{\Omega}) \cap C^2(\Omega)$; falls auch $a_i^I = a_i^{II}$, reicht sogar $u^{II} \in C^0(\overline{\Omega}) \cap C^2(\Omega)$.

Beweis. Setze $f' := L^I u^{II}$. $\|f' - f\|_\infty = \|(L^I - L^{II})u^{II}\|_\infty$ ist durch die rechte Seite von (5.1.9) ohne den Faktor M abschätzbar. Satz 5.1.12 angewandt auf $L^I u^{II} = f'$ und $L^I u^I = f$ liefert die Ungleichung $\|u^I - u^{II}\|_\infty \leq M \|f' - f\|_\infty$. ■

5.1.4 Differenzenverfahren für die allgemeine Differentialgleichung zweiter Ordnung

Aus schreibtechnischen Gründen beschränken wir uns auf den zweidimensionalen Fall $n = 2$. Allgemeine Gebiete $\Omega \in \mathbb{R}^2$ erfordern die in Abschnitt 4.8 erläuterten speziellen Diskretisierungen am Rand. Hier sollen nur die Differenzenformeln in randfernen Punkten diskutiert werden. Deshalb genügt es, im Folgenden das Einheitsquadrat

$$\Omega = (0, 1) \times (0, 1)$$

zugrunde zu legen.

L sei durch (5.1.1a) gegeben. Will man ein Differenzenverfahren der Konsistenzordnung 2 erreichen, liegt folgende Wahl nahe, die aus Stabilitätsgründen später modifiziert wird.

$$\begin{aligned} & a_{11}(x, y) \partial_x^+ \partial_x^- + 2a_{12} \partial_x^0 \partial_y^0 + a_{22}(x, y) \partial_y^+ \partial_y^- + a_1(x, y) \partial_x^0 + a_2(x, y) \partial_y^0 + a(x, y) \\ &= h^{-2} \begin{bmatrix} -a_{12}(x, y)/2 & a_{22}(x, y) & a_{12}(x, y)/2 \\ a_{11}(x, y) & -2[a_{11}(x, y) + a_{22}(x, y)] & a_{11}(x, y) \\ a_{12}(x, y)/2 & a_{22}(x, y) & -a_{12}(x, y)/2 \end{bmatrix} \\ &+ \frac{1}{2} h^{-1} \begin{bmatrix} 0 & a_2(x, y) & 0 \\ -a_1(x, y) & 0 & a_1(x, y) \\ 0 & -a_2(x, y) & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & a(x, y) & 0 \\ 0 & 0 & 0 \end{bmatrix}. \end{aligned} \quad (5.1.10)$$

Bemerkung 5.1.16 Das Differenzenverfahren (5.1.10) ist ein Neunpunktschema der Konsistenzordnung 2.

Sei L_h die zu (5.1.10) gehörige Matrix des Systems

$$L_h u_h = q_h \quad (5.1.11)$$

der Differenzengleichungen in Ω_h (vgl. Bemerkung 4.2.5). Die Lösbarkeit von (5.1.11) (für beliebige q_h) ist äquivalent zur Eindeutigkeit. Im kontinuierlichen Fall war die Eindeutigkeit im Wesentlichen durch die Bedingung $a \leq 0$ und die Elliptizität $a_{11}a_{22} > a_{12}^2$ gegeben (vgl. Satz 5.1.11). Diese Bedingungen reichen im Allgemeinen nicht aus, um die Lösbarkeit der Gleichung (5.1.11) zu garantieren. Wir ersetzen (5.1.10) daher durch eine andere Diskretisierung.

Satz 5.1.17 Für die Koeffizienten von L gelte (5.1.3a,b),

$$|a_{12}(x, y)| \leq \min\{a_{11}(x, y), a_{22}(x, y)\} \quad (5.1.12)$$

und $a(x, y) \leq 0$ in Ω . Dann gibt es zu L ein Siebenpunktdifferenzenverfahren der Konsistenzordnung 1, sodass die zugehörige Matrix L_h bis auf das Vorzeichen eine M-Matrix ist. Insbesondere ist die entstehende Gleichung (5.1.11) lösbar.

Man beachte, dass die Bedingung (5.1.12) nur dann aus (5.1.3a) folgt, wenn $a_{11} = a_{22} > 0$.

Beweis. Im Gitterpunkt $(x, y) \in \Omega_h$ (vgl. (4.2.1a)) kürzen wir die Koeffizienten $a_{ij}(x, y)$, $a_i(x, y)$, $a(x, y)$ durch a_{ij} , a_i , a ab. Der Hauptteil $\sum a_{ij} \partial^2 / \partial x_i \partial x_j$ ($x_1 = x$, $x_2 = y$) wird durch die folgenden Differenzensterme diskretisiert:

$$\begin{aligned} \frac{\partial^2}{\partial x_1^2} : \quad & h^{-2} \begin{bmatrix} 0 & & \\ 1 & -2 & 1 \\ & 0 & \end{bmatrix}, \quad \frac{\partial^2}{\partial x_2^2} : \quad h^{-2} \begin{bmatrix} & 1 & \\ & -2 & 0 \\ & & 1 \end{bmatrix}, \\ \frac{\partial^2}{\partial x_1 \partial x_2} : \quad & \begin{cases} \frac{1}{2} h^{-2} \begin{bmatrix} 0 & -1 & 1 \\ -1 & 2 & -1 \\ 1 & -1 & 0 \end{bmatrix}, & \text{falls } a_{12} \geq 0, \\ \frac{1}{2} h^{-2} \begin{bmatrix} -1 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & -1 \end{bmatrix}, & \text{falls } a_{12} < 0. \end{cases} \end{aligned} \quad (5.1.13)$$

Für $a_{12} \geq 0$ erhält man somit

$$\sum_{i,j=1}^2 a_{ij} \frac{\partial^2}{\partial x_i \partial x_j} : \quad h^{-2} \begin{bmatrix} 0 & a_{22} - a_{12} & a_{12} \\ a_{11} - a_{12} & 2(a_{12} - a_{11} - a_{22}) & a_{11} - a_{12} \\ a_{12} & a_{22} - a_{12} & 0 \end{bmatrix}.$$

Führt man $a_{12}^+ := \max\{a_{12}, 0\}$ und $a_{12}^- := \min\{a_{12}, 0\}$ ein, so ergibt sich für $a_{12} \geq 0$ wie auch $a_{12} < 0$ der Siebenpunktstern

$$h^{-2} \begin{bmatrix} -a_{12}^- & a_{22} - |a_{12}| & a_{12}^+ \\ a_{11} - |a_{12}| & 2(|a_{12}| - a_{11} - a_{22}) & a_{11} - |a_{12}| \\ a_{12}^+ & a_{22} - |a_{12}| & -a_{12}^- \end{bmatrix}. \quad (5.1.13')$$

■

Übungsaufgabe 5.1.18 Sei L_h die zur Differenzenformel (5.1.13') gehörende Matrix (man beachte, dass die Koeffizienten in (5.1.13') ortsabhängig sind). Es gelte (5.1.3) und (5.1.12). Man beweise, dass $-L_h$ die Voraussetzungen der Übungsaufgabe 4.3.8a erfüllt.

Die ersten Ableitungsterme $a_i \partial / \partial x_i$ werden durch die Vor- bzw. Rückwärtsdifferenz $a_i \partial_{x_i}^\pm$ ersetzt, wenn $a_i \leq 0$:

$$\sum_{i=1}^2 a_i \frac{\partial}{\partial x_i} + a : \quad h^{-1} \begin{bmatrix} a_2^+ & & \\ a_1^- & -|a_1| - |a_2| & a_1^+ \\ & -a_2^- & \end{bmatrix} + \begin{bmatrix} & & \\ & a & \\ & & \end{bmatrix}, \quad (5.1.14)$$

wobei a_i^\pm wie a_{12}^\pm definiert ist. Wegen $a \leq 0$ ist das Diagonalelement negativ, die anderen nichtnegativ. Addiert man diese Terme zu (5.1.13'), erfüllt die entstehende Matrix $-L_h$ die Voraussetzungen der Übungsaufgabe 4.3.8a. Folglich ist $-L_h$ eine M-Matrix.

Man überzeugt sich leicht, dass die Differenzenformel (5.1.13') von zweiter Konsistenzordnung ist; dagegen enthält (5.1.14) die einseitigen Differenzen, sodass die Diskretisierung insgesamt nur von erster Ordnung ist.

Die Bedingung (5.1.12) lässt sich vermeiden, wenn man größere Differenzensterne in Kauf nimmt, die noch die Werte $u(x_1 \pm \nu h, x_2 \pm \mu h)$ für feste $\nu, \mu \in \mathbb{N}$ (aber im Allgemeinen $|\nu|, |\mu| > 1$) enthalten (vgl. Bramble-Hubbard [15]).

Um ein Verfahren der Konsistenzordnung 2 zu erzielen, hat man $\sum a_i \partial / \partial x_i$ wie in (5.1.10) zu diskretisieren. Der folgende Zusatz zeigt, dass auch dann $-L_h$ eine M-Matrix ist, wenn nur ha_i klein genug ist.

Zusatz 5.1.19 Zusätzlich zu den Voraussetzungen des Satzes 5.1.17 gelte

$$a_{ii} > |a_{12}| + \frac{h}{2} |a_i| \quad (i = 1, 2). \quad (5.1.15)$$

Dann führt die Diskretisierung von $\sum a_i \partial / \partial x_i$ aus (5.1.10) zusammen mit (5.1.13') zu einem Siebenpunkt-differenzenverfahren zweiter Konsistenzordnung, sodass $-L_h$ eine M-Matrix ist.

Beweis. $-L_h$ erfüllt (4.3.1a) und ist irreduzibel diagonaldominant. ■

Übungsaufgabe 5.1.20 Die Bedingung $a_{ii} \geq |a_{12}| + h|a_i|/2$ anstelle von (5.1.15) ist nicht ausreichend. Man konstruiere ein Gegenbeispiel mit $a_{11} = a_{22} = 1$, $a_{12} = 0$, $h = 1/3$, $a_i(x, y)$ variabel, $|a_i| = 6$, sodass L_h singular ist.

Wesentlich schwächere Voraussetzungen für die Nichtsingularität von L_h als in Satz 5.1.17 und Zusatz 5.1.19 werden in §9.2 gebraucht (vgl. Übungsaufgabe 9.2.6, Folgerung 11.3.5).

Im Allgemeinen ist L_h keine symmetrische Matrix. Symmetrie von L_h ist nur dann zu erwarten, wenn auch L symmetrisch ist: $L = L'$. Dabei ist der formal adjungierte Operator L' , der zu L aus (5.1.2) gehört, durch

$$L' = \sum_{i,j=1}^n \left[a_{ij}^{III} \frac{\partial^2}{\partial x_i \partial x_j} + \frac{\partial}{\partial x_i} a_{ij}^{II} \frac{\partial}{\partial x_j} + \frac{\partial^2}{\partial x_i \partial x_j} a_{ij}^I \right] - \sum_{i=1}^n \left[a_i^{II} \frac{\partial}{\partial x_i} + \frac{\partial}{\partial x_i} a_i^I \right] + a \quad (5.1.16)$$

definiert. Man überlegt sich leicht, dass ein *symmetrischer Operator* stets in der Form

$$L = \sum_{i,j=1}^n \frac{\partial}{\partial x_i} a_{ij}(\mathbf{x}) \frac{\partial}{\partial x_j} + a(\mathbf{x}), \quad a_{ij}(\mathbf{x}) = a_{ji}(\mathbf{x}), \quad (5.1.17)$$

geschrieben werden kann. Ein Differenzenverfahren hierfür ist im Falle $n = 2$ und $a_{12} = 0$ durch den Fünfpunktstern

$$h^{-2} \begin{bmatrix} 0 & a_{22}(x, y + \frac{h}{2}) & 0 \\ a_{11}(x - \frac{h}{2}, y) & \begin{Bmatrix} -a_{11}(x - \frac{h}{2}, y) - a_{11}(x + \frac{h}{2}, y) \\ -a_{22}(x, y + \frac{h}{2}) - a_{22}(x, y - \frac{h}{2}) \end{Bmatrix} & a_{11}(x + \frac{h}{2}, y) \\ 0 & a_{22}(x, y - \frac{h}{2}) & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (5.1.18)$$

gegeben.

Satz 5.1.21 Seien $a_{11}, a_{22} \in C^{2,1}(\overline{\Omega})$. Das Differenzenverfahren (5.1.18) ist konsistent von der Ordnung 2. Die zugehörige Matrix L_h ist symmetrisch. Sind $a_{ii} > 0$ (Elliptizität) und $a \leq 0$, so ist $-L_h$ eine positiv definite M-Matrix.

Beweis. a) Zum Konsistenznachweis entwickle man

$$v(x + \frac{h}{2}) := a_{11}(x + \frac{h}{2}, y) [u(x + h, y) + u(x, y)] \quad (y \text{ fest})$$

um $x + h/2$ und danach $v(x + h/2) - v(x - h/2)$ um x .

b) Die Symmetrie ergibt sich aus $L_{\mathbf{x}\xi} = L_{\xi\mathbf{x}}$, $\mathbf{x}, \xi \in \Omega_h$ (vgl. Übungsaufgabe 4.2.2c).

c) L_h ist irreduzibel diagonaldominant, sodass die Kriterien 4.3.10 und 4.3.24 anwendbar sind. ■

Der gemischte Term $\frac{\partial}{\partial x} a_{12}(x, y) \frac{\partial}{\partial y} + \frac{\partial}{\partial y} a_{12}(x, y) \frac{\partial}{\partial x}$ lässt sich durch

$$\frac{1}{2} h^{-2} \begin{bmatrix} -a_{12}(\mathbf{x}_A) - a_{12}(\mathbf{x}_B) & a_{12}(\mathbf{x}_B) + a_{12}(\mathbf{x}_C) & 0 \\ 2a_{12}(\mathbf{x}_A) & 2[a_{12}(\mathbf{x}_A) + a_{12}(\mathbf{x}_C)] & 2a_{12}(\mathbf{x}_C) \\ 0 & a_{12}(\mathbf{x}_A) + a_{12}(\mathbf{x}_D) & -a_{12}(\mathbf{x}_C) - a_{12}(\mathbf{x}_D) \end{bmatrix} \quad (5.1.19)$$

mit

$$\mathbf{x}_A = \left(x - \frac{h}{2}, y\right), \quad \mathbf{x}_B = \left(x - \frac{h}{2}, y + h\right), \quad \mathbf{x}_C = \left(x + \frac{h}{2}, y\right), \quad \mathbf{x}_D = \left(x + \frac{h}{2}, y - h\right)$$

diskretisieren. Eine andere Schreibweise für (5.1.19) ist

$$\frac{1}{2} \left[\partial_y^+ b(\cdot - \frac{h}{2}, \cdot) \partial_x^- + \partial_x^+ b(\cdot - \frac{h}{2}, \cdot) \partial_y^- + \partial_y^- b(\cdot + \frac{h}{2}, \cdot) \partial_x^+ + \partial_x^- b(\cdot + \frac{h}{2}, \cdot) \partial_y^- \right]$$

wobei $b := a_{12}$.

Übungsaufgabe 5.1.22 a) Für die Koeffizienten von L aus (5.1.17) gelte $a_{11} > 0$, $a_{22} > 0$, $a_{12} = a_{21} \leq 0$, $a_{11} + a_{12} > 0$, $2a_{22}(x, y + h/2) + a_{12}(x - h/2, y + h) + a_{12}(x + h/2, y) > 0$, $a \leq 0$. Man zeige, dass das durch (5.1.18) und (5.1.19) beschriebene Differenzenschema die Konsistenzordnung 2 besitzt und die zugehörige Matrix $-L_h$ eine symmetrische, irreduzibel diagonaldominante und positiv definite M-Matrix ist.

b) Wie lautet die für den Fall $a_{12} \geq 0$ geeignete Diskretisierung?

Übungsaufgabe 5.1.23 Die Differenzenformel aus Satz 5.1.17 für den Operator L lautet

$$a_{11}\partial_x^+\partial_x^- + a_{12}(\partial_x^+\partial_y^+ + \partial_x^-\partial_y^-) + a_{22}\partial_y^+\partial_y^- + a_{11}\partial_x^+ + a_{22}\partial_y^+ + a_{11}\partial_x^- + a_{22}\partial_y^- + a_{11} + a_{22},$$

wenn $a_{12} \geq 0$, $a_1 \geq 0$, $a_2 \geq 0$. Die zugehörige Matrix sei L_h . Man beweise: Die transponierte Matrix L_h^\top beschreibt ein Differenzenverfahren für den adjungierten Operator L' und besitzt ebenfalls die Konsistenzordnung 1.

Allgemeiner läßt sich für regelmäßige Differenzenverfahren zeigen, dass L_h^\top eine Diskretisierung von L' ist. Die Rolle der Regelmäßigkeit veranschaulicht das

Beispiel 5.1.24 Sei $Lu := u'' + au'$ in $\Omega = (-1, 1)$ mit $a(x) \leq 0$ für $x \leq 0$ und $a(x) \geq 0$ für $x > 0$. Gemäß (5.1.14) wird au' in $x \leq 0$ durch $a(x)\partial^+u(x)$ und in $x > 0$ durch $a(x)\partial^-u(x)$ diskretisiert. Die zugehörige Matrix sei $L_h = L_{h,2} + L_{h,1}$, wobei $L_{h,2}$ und $L_{h,1}$ jeweils den Termen u'' und au' entsprechen. Nach den vorangehenden Überlegungen sollte $L_{h,1}^\top v_h$ eine Diskretisierung von $-(av)'$ sein. Die Differenzen $L_{h,1}^\top v_h$ in $x = 0$ und $x = h$ lauten aber

$$\begin{aligned} h^{-1} [a(-h)v_h(-h) - a(0)v_h(0) - a(h)v_h(h)] &= -(av)' - h^{-1}a(0)v_h(0) + \mathcal{O}(h), \\ h^{-1} [a(0)v_h(0) - a(h)v_h(h) - a(2h)v_h(2h)] &= -(av)' + h^{-1}a(0)v_h(0) + \mathcal{O}(h), \end{aligned}$$

und sind daher nicht konsistent. Trotzdem ist L_h^\top eine mögliche Diskretisierung von $L'v = v'' - (av)'$, denn es läßt sich zeigen, dass der Fehler $v_h - R_h v$ die Größenordnung $\mathcal{O}(h)$ besitzt.

Zum Nachweis der Stabilität hat man $\|L_h^{-1}\|_\infty \leq \text{const}$ zu beweisen. Offenbar genügt es, diese Ungleichung für hinreichend kleines h nachzuweisen. Im Beweis des Satzes 5.1.12 haben wir davon Gebrauch gemacht, dass

$$-Lw \geq 1 \quad \text{in } \Omega, \quad w \geq 0 \quad \text{auf } \Gamma$$

für $w(\mathbf{x}) := \exp(2R\alpha) - \exp(\alpha(x_1 - z_1 + R))$. Seien $D_h u_h(\mathbf{x})$ die Differenzengleichungen, aus denen $L_h u_h$ nach Elimination der Randwerte entsteht. Wir setzen $w_h := 2R_h w$ (d.h. $w_h(\mathbf{x}) = 2w(\mathbf{x})$ für $\mathbf{x} \in \bar{\Omega}_h$). Es gilt

$$-D_h w_h = -2 \left(D_h R_h - \tilde{R}_h L \right) w - 2 \tilde{R}_h L w \geq 2 - 2 \left(D_h R_h - \tilde{R}_h L \right) w.$$

Jedes konsistente Differenzenverfahren erfüllt $\|(D_h R_h - \tilde{R}_h L)w\|_\infty \rightarrow 0$. Für hinreichend kleines h hat man somit

$$-D_h w_h(\mathbf{x}) \geq 1 \quad (\mathbf{x} \in \Omega_h, h \leq h_0).$$

Diese Ungleichung stimmt für randferne $\mathbf{x} \in \Omega_h$ mit $-L_h w_h(\mathbf{x}) \geq 1$ überein. Für randnahe $\mathbf{x} \in \Omega_h$ enthält $-D_h w_h(\mathbf{x})$ noch die Summe $-\sum_{\xi \in \Gamma_h} L_{\mathbf{x}\xi} w_h(\xi)$, die in $-L_h w_h(\mathbf{x})$ nicht enthalten ist. Für die Diskretisierungen aus Satz 5.1.17, Zusatz 5.1.19, Satz 5.1.21 und Übungsaufgabe 5.1.22 gilt jedoch $L_{\mathbf{x}\xi} \geq 0$ ($\mathbf{x} \in \Omega_h, \xi \in \Gamma_h$), sodass wegen $w_h \geq 0$ auf Γ_h

$$-L_h w_h \geq -D_h w_h \geq 1 \quad (h \leq h_0)$$

für alle Gitterpunkte zutrifft. Satz 4.3.16 beweist den folgenden Satz.

Satz 5.1.25 Die Diskretisierungen aus Satz 5.1.17, Zusatz 5.1.19, Satz 5.1.21 und Übungsaufgabe 5.1.22 sind unter den dort gestellten Voraussetzungen stabil, d.h. $\|L_h^{-1}\|_\infty \leq \text{const}$ für alle Schrittweiten $h \in H = \{1/n : n \in \mathbb{N}\}$. Gemäß Satz 4.5.3 konvergieren die Verfahren. Die Konvergenzordnung stimmt mit der jeweiligen Konsistenzordnung überein.

5.1.5 Greensche Funktion

Das Konzept der Lösungsdarstellung durch die Greensche Funktion läßt sich für die allgemeine Differentialgleichung (5.1.1a) wiederholen. Die Greensche Funktion (erster Art) $g(\xi, \mathbf{x})$ ist in $\xi = \mathbf{x}$ singulär und erfüllt

$$\begin{aligned} L_{\mathbf{x}} g(\xi, \mathbf{x}) &= 0, & L'_\xi g(\xi, \mathbf{x}) &= 0 & \text{für } \mathbf{x}, \xi \in \Omega, \mathbf{x} \neq \xi, \\ g(\xi, \mathbf{x}) &= 0 & & \text{für } \mathbf{x} \in \Gamma \text{ oder } \xi \in \Gamma. \end{aligned}$$

Dabei ist L' der adjungierte Differentialoperator (5.1.16). Sobald $L \neq L'$, ist auch g nicht mehr symmetrisch: $g(\mathbf{x}, \xi) \neq g(\xi, \mathbf{x})$. Unter geeigneten Voraussetzungen ist die Lösung von (5.1.1a), (5.1.4) durch

$$u(\mathbf{x}) = - \int_{\Omega} g(\xi, \mathbf{x}) f(\xi) d\xi - \int_{\Gamma} \varphi(\xi) B_{\xi} g(\xi, \mathbf{x}) d\Gamma_{\xi}$$

darstellbar, wobei $B = B_{\xi} = \sum_{i,j=1}^n n_j \frac{\partial}{\partial \xi_i} a_{ij}$ ein Rand-Differentialoperator (Konormalenableitung) ist (n_j sind die Komponenten des Normalenvektors $\vec{n} = \vec{n}(\xi)$, $\xi \in \Gamma$). Nur wenn der Hauptteil von L mit Δ übereinstimmt, ist B die Normalenableitung.

Im diskreten Fall entspricht die Inverse L_h^{-1} wieder der Greenschen Funktion $g(\cdot, \cdot)$.

5.2 Allgemeine Randbedingungen

5.2.1 Formulierung der Randwertaufgabe

Die Differentialgleichung sei durch (5.1.1a) gegeben. Die Dirichlet-Randbedingung (5.1.4) kann in der Form

$$Bu = \varphi \quad \text{auf } \Gamma \quad (5.2.1a)$$

geschrieben werden, wobei B die Identität (genauer: die Beschränkung auf Γ) darstellt. Allgemeiner kann B ein Differentialoperator - ein sogenannter *Rand-Differentialoperator* - der Ordnung 1 sein:

$$B = \sum_{i=1}^n b_i(\mathbf{x}) \frac{\partial}{\partial x_i} + b_0(\mathbf{x}) \quad (\mathbf{x} \in \Gamma). \quad (5.2.1b)$$

Führt man den Vektor $b(\mathbf{x}) = (b_1(\mathbf{x}), \dots, b_n(\mathbf{x}))^T$ ein, läßt sich Bu in der Form

$$Bu = \langle b(\mathbf{x}), \nabla u(\mathbf{x}) \rangle + b_0(\mathbf{x}) u(\mathbf{x}) \quad \text{oder} \quad B = b^T \nabla + b_0 \quad (5.2.1b')$$

schreiben.

Beispiel 5.2.1 (a) Für $b = 0$ und $b_0(\mathbf{x}) \neq 0$ ergibt sich die Dirichlet-Bedingung $u = \varphi/b_0$ auf Γ , die auch erste Randbedingung genannt wird.

(b) Die Wahl $b = \vec{n}$, $b_0 = 0$ charakterisiert die Neumann-Bedingung, auch zweite Randbedingung genannt.

(c) Gleichung (5.2.1a) mit $\langle b, \vec{n} \rangle \neq 0$, $b_0 \neq 0$ heißt dritte Randbedingung (Robin-Randbedingung).

Bemerkung 5.2.2 Der Fall $\langle b, \vec{n} \rangle = 0$ wird im Allgemeinen ausgeschlossen. Für $\langle b, \vec{n} \rangle = 0$ ist $b^T \nabla$ eine Tangentialableitung. Die Randbedingung $Bu = \varphi$ ist dann einer Dirichlet-Bedingung sehr ähnlich. Die Bedingung “ $u = \varphi$ auf Γ ” impliziert “ $Bu = \tilde{\varphi} := B\varphi$ auf Γ ” (Warum ist $B\varphi$ definiert?).

Die Normalenableitung $B = \partial/\partial n$ (d.h. $b = \vec{n}$) tritt insbesondere im Zusammenhang mit $L = -\Delta$ auf. Für den allgemeinen Operator L aus (5.1.1b) ist die sogenannte *konormale Ableitung* B mit

$$b = A\vec{n} \quad (A = (a_{ij})_{i,j=1}^n \text{ mit } a_{ij} \text{ aus (5.1.1b)})$$

von größerer Bedeutung, wie wir in Abschnitt 7.4 sehen werden. Aussagen über Existenz und Eindeutigkeit der Lösung hängen stets von L und B ab. Wir haben schon gesehen, dass für $L = -\Delta$, $B = I$ (Dirichlet-Bedingung) Eindeutigkeit gewährleistet ist (vgl. Satz 3.1.2), während das zu $L = -\Delta$, $B = \vec{n}^T \nabla = \partial/\partial n$ gehörende Problem im Allgemeinen nicht lösbar ist (vgl. Satz 3.4.1). Die Koeffizienten von B sind ortsabhängig. Selbstverständlich darf $B(\mathbf{x}) = 0$, d.h. $b(\mathbf{x}) = 0$ und $b_0(\mathbf{x}) = 0$, für kein $\mathbf{x} \in \Gamma$ auftreten. Aber es ist möglich, dass $b(\mathbf{x}) = 0$ (und $b_0 \neq 0$) in $\gamma \subset \Gamma$ und $b(\mathbf{x}) \neq 0$ in $\Gamma \setminus \gamma$. Dann liegt eine Dirichlet-Bedingung $u = \varphi/b_0$ im Teilstück γ und eine Randbedingung erster Ordnung auf dem restlichen Randstück $\Gamma \setminus \gamma$ vor. An den Kontaktstellen zwischen γ und $\Gamma \setminus \gamma$ ist die Lösung im Allgemeinen nicht glatt (sie hat Singularitäten in den Ableitungen).

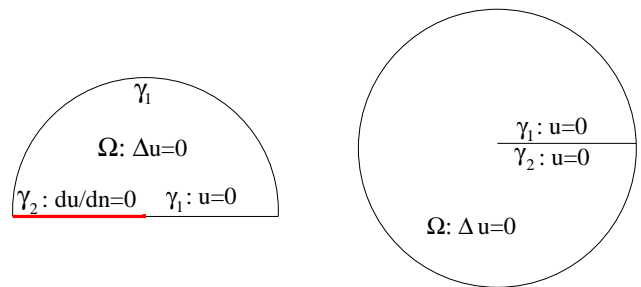


Abb. 5.2.1. a) Randwertproblem mit wechselndem Randbedingungstyp, b) Dirichlet-Problem im geschlitzten Kreis

Beispiel 5.2.3 Ω sei der obere Halbkreis um $x = y = 0$ mit Radius 1. Differentialgleichung und Randbedingungen seien wie in Abbildung 5.2.1a angegeben. Die Randbedingung ändert ihre Ordnung in $x = y = 0$. Die Lösung lautet in den Polarkoordinaten: $u = \sqrt{r} \sin(\varphi/2)$ (vgl. (2.1.3)). Man rechnet nach, dass $u_x = \mathcal{O}(1/\sqrt{r})$ und $u_y = \mathcal{O}(1/\sqrt{r})$.

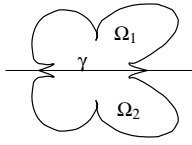


Abb. 5.2.2. Bezüglich γ spiegelsymmetrisches Gebiet

Die gleiche Singularität wie in Beispiel 5.2.3 tritt bei dem in Abbildung 5.2.1b beschriebenen Problem auf; die Lösung ist ebenfalls $r^{1/2} \sin(\varphi/2)$. Diese Dirichlet-Aufgabe und Beispiel 5.2.3 hängen eng miteinander zusammen.

Beispiel 5.2.4 Sei $\Omega = \Omega_1 \cup \Omega_2 \cup \gamma$ ($\Omega_1 \cap \Omega_2 = \emptyset$) wie in Abbildung 5.2.2: Spiegelung von Ω_1 an γ erzeuge Ω_2 . Sucht man eine Lösung von $Lu = f$ in Ω , $Bu = \varphi$ auf $\partial\Omega$, und ist mit u auch die an γ gespiegelte Funktion \bar{u} eine Lösung, so erwartet man $u = \bar{u}$. Diese Lösung erfüllt dann $Lu = f$ in Ω_1 , $Bu = \varphi$ auf $\partial\Omega_1 \setminus \gamma$, $\partial u / \partial n = 0$ auf γ .

Während die Randbedingung $Bu = \varphi$ auf $\partial\Omega$ physikalischen Ursprungs sein mag, zeigt Beispiel 5.2.4, dass eine Neumann-Bedingung auch geometrische Gründe haben kann. Eine andere, geometrisch begründete Randbedingung ist die folgende. Sei Ω wie in Abbildung 5.2.3: γ_1 und γ_2 sind Teile von $\Gamma = \partial\Omega$ mit

$$\gamma = \{(x_i, y) : y_1 \leq y \leq y_2\} \quad (i = 1, 2).$$

Dann lässt sich neben $Bu = \varphi$ auf $\Gamma \setminus (\gamma_1 \cup \gamma_2)$ die periodische Randbedingung

$$u(x_1, y) = u(x_2, y), \quad u_x(x_1, y) = u_x(x_2, y) \quad \text{für } y_1 \leq y \leq y_2 \quad (5.2.2)$$

für γ_1 und γ_2 formulieren. Die Lösung ist in x -Richtung periodisch fortsetzbar (Periode: $x_2 - x_1$). Eine mögliche Herkunft periodischer Randbedingungen wird im folgenden Beispiel diskutiert.

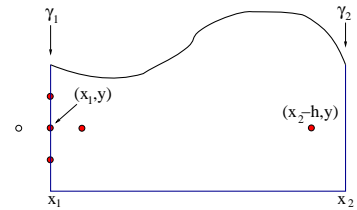


Abb. 5.2.3. Diskretisierung bei periodischen Randbedingungen

Beispiel 5.2.5 a) Sei Ω' ein Ringsektor, der durch die Polarkoordinaten $r \in (r_1, r_2)$, $\varphi \in [0, 2\pi)$ beschrieben wird. Transformation der Differentialgleichung auf Polarkoordinaten ergibt als Bildgebiet das Rechteck $\Omega = (r_1, r_2) \times (0, 2\pi)$. Die ursprünglichen Randbedingungen auf Ω' werden zu Randbedingungen am oberen und unteren Rand, während (5.2.2) die Periodizität der Winkelvariablen $x \in (0, 2\pi)$ beschreibt.

b) Statt auf $\Omega' \subset \mathbb{R}^2$ kann eine Randwertaufgabe auch auf einem Teil der 2-dimensionalen Oberfläche eines 3-dimensionalen Körpers definiert werden. Liegt Ω' z.B. auf der Oberfläche des Zylinders $\{\xi \in \mathbb{R}^3 : \xi_1^2 + \xi_2^2 \leq r^2, \xi_3 \in \mathbb{R}\}$, so ergibt die Auffaltung von Ω' ein Gebiet Ω wie in Abbildung 5.2.3. Auch hier ist (5.2.2) dadurch begründet, dass x die Rolle einer Winkelvariablen spielt.

c) Sucht man Lösungen im unbeschränkten Streifen $\Omega' = \mathbb{R} \times (y_1, y_2)$, kann man stattdessen periodische Lösungen in $\Omega = (0, 2\pi) \times (y_1, y_2)$ mit der Randbedingung (5.2.2) suchen, da diese (nach periodischer Fortsetzung) auch Lösungen der ursprünglichen Aufgabe sind. Lösungen mit periodischer Randbedingung in x - und y -Richtung lassen sich sogar auf $\Omega' = \mathbb{R}^2$ fortsetzen.

5.2.2 Differenzenverfahren bei allgemeinen Randbedingungen

Die geringsten Schwierigkeiten stellen die periodischen Randbedingungen dar. Man definiere Ω_h als die Menge der Gitterpunkte in Ω und auf γ_1 (aber nicht auf γ_2 ; vgl. Abbildung 5.2.3). Die Differenzengleichung in einem Gitterpunkt $(x_1, y) \in \gamma_1$ hat einen linken Nachbarn $(x_1 - h, y)$ außerhalb von $\bar{\Omega}$. $u_h(x_1 - h, y)$ ersetze man in der Differenzengleichung durch $u_h(x_2 - h, y)$. Damit hat man die Periodizität auf die Differenzenlösung übertragen, ohne Gleichung (5.2.2) explizit zu diskretisieren. Selbstverständlich muss die Schrittweite so gewählt sein, dass $x_2 - x_1$ ein Vielfaches von h ist.

Übungsaufgabe 5.2.6 Bei periodischen Randbedingungen ändert sich die Struktur der Matrix L_h . Sei $L = -\Delta$ im Quadrat $\Omega = (0, 1) \times (0, 1)$. Für den oberen und unteren Rand seien Dirichlet-Bedingungen, für die seitlichen Ränder die periodische Bedingung (5.2.2) gegeben. In Analogie zu (4.2.8) gebe man die Gestalt der Matrix L_h bei lexikographischer Anordnung der Gitterpunkte an.

Die Diskretisierung der Randbedingung $B = \vec{n}^\top \nabla = \partial/\partial n$ haben wir in Abschnitt 4.7 für den Fall beschrieben, dass Ω ein Rechteck ist und Γ mit dem Gitternetz übereinstimmt. In der gleichen Situation kann man die allgemeine Randbedingung (5.2.1a,b) wie folgt diskretisieren. Sei $(\bar{x}, \bar{y}) \in \Gamma_h = \bar{\Omega}_h \cap \Gamma$ ein Gitterpunkt auf Γ . Der in Abbildung 5.2.4 eingezeichnete Schnittpunkt $(\bar{x} + h, \tilde{y})$ ist durch $\tilde{y} = \bar{y} + hb_2(\bar{x}, \bar{y})/b_1(\bar{x}, \bar{y})$ gegeben. Man beachte, dass $b_1 \neq 0$, da $b^\top \nabla$ keine tangentielle Ableitung sein soll. Für $u \in C^2(\bar{\Omega})$ wird die Richtungsableitung $b^\top \nabla u = \langle b, \nabla u \rangle$ in (\bar{x}, \bar{y}) durch

$$[u(\bar{x}, \bar{y}) - u(\bar{x} + h, \tilde{y})] / \sqrt{h^2 + (\bar{y} - \tilde{y})^2} = (b^\top \nabla u)(\bar{x}, \bar{y}) + \mathcal{O}(h) \quad (5.2.3a)$$

approximiert. Andererseits ergibt sich $u(\bar{x} + h, \tilde{y})$ unter der gleichen Voraussetzung durch Interpolation bis auf $\mathcal{O}(h^2)$:

$$u(\bar{x} + h, \tilde{y}) = \frac{\tilde{y} - \bar{y}}{h} u(\bar{x} + h, \bar{y} + h) + \frac{\bar{y} + h - \tilde{y}}{h} u(\bar{x} + h, \bar{y}) + \mathcal{O}(h^2). \quad (5.2.3b)$$

Kombination von (5.2.3a) und (5.2.3b) führt zur Differenzenformel

$$(B_h u_h)(\bar{x}, \bar{y}) := \frac{u_h(\bar{x}, \bar{y}) - \frac{\bar{y} + h - \tilde{y}}{h} u_h(\bar{x} + h, \bar{y}) - \frac{\tilde{y} - \bar{y}}{h} u_h(\bar{x} + h, \bar{y} + h)}{\sqrt{h^2 + (\bar{y} - \tilde{y})^2} + b_0(\bar{x}, \bar{y}) u_h(\bar{x}, \bar{y})} = \varphi(\bar{x}, \bar{y}) \quad (5.2.3c)$$

mit $\tilde{y} - \bar{y} = hb_2(\bar{x}, \bar{y})/b_1(\bar{x}, \bar{y})$.

Lemma 5.2.7 R_h sei die Beschränkung auf die Gitterpunkte $\bar{\Omega}_h$. Für $u \in C^2(\bar{\Omega})$ gilt

$$B_h(u_h - R_h u)(\bar{x}, \bar{y}) = \varphi(\bar{x}, \bar{y}) - B_h R_h u(\bar{x}, \bar{y}) = \mathcal{O}(h \|u\|_{C^2(\bar{\Omega})}),$$

d.h. die Randdiskretisierung hat die Konsistenzordnung 1.

Gleichung (5.2.3c) hat die Gestalt

$$(B_h u_h)(\bar{x}, \bar{y}) = c_0 u_h(\bar{x}, \bar{y}) - c_1 u_h(\bar{x} + h, \bar{y}) - c_2 u_h(\bar{x} + h, \bar{y} + h). \quad (5.2.3c')$$

Wir nehmen an, dass die Differenzengleichungen $\sum_{\xi \in \bar{\Omega}_h} L_{\mathbf{x}\xi} u_h(\xi) = f_h(\mathbf{x})$ [vor der Elimination der Werte $u_h(\xi)$, $\xi \in \Gamma_h$] die Ungleichungen

$$L_{\mathbf{x}\xi} \leq 0 \quad (\mathbf{x} \neq \xi), \quad L_{\mathbf{x}\mathbf{x}} \geq - \sum_{\xi \neq \mathbf{x}} L_{\mathbf{x}\xi} \quad (\mathbf{x} \in \Omega_h, \xi \in \bar{\Omega}_h) \quad (5.2.4a)$$

erfüllen. $L_{\mathbf{x}\mathbf{x}} > 0$ und $L_{\mathbf{x}\xi} \leq 0$ entsprechen der Vorzeichenbedingung (4.3.1a). Die zweite Ungleichung in (5.2.4a) stimmt mit (4.3.4b) überein. Die entsprechenden Ungleichungen für (5.2.3c') lauten

$$c_1 \geq 0, \quad c_2 \geq 0, \quad c_0 \geq c_1 + c_2. \quad (5.2.4b)$$

Die Differenzengleichungen in $\mathbf{x} \in \Omega_h$ und die Randgleichungen (5.2.3c') für $(\bar{x}, \bar{y}) \in \Gamma_h$ seien zusammengefasst in $A_h u_h = g_h$. Die Ungleichungen (5.2.4a,b) implizieren

$$a_{\mathbf{x}\xi} \leq 0 \quad (\mathbf{x}, \xi \in \bar{\Omega}_h), \quad a_{\mathbf{x}\mathbf{x}} \geq - \sum_{\xi \neq \mathbf{x}} a_{\mathbf{x}\xi} \quad (\mathbf{x} \in \bar{\Omega}_h).$$

Damit ist A_h eine M-Matrix, wenn A_h irreduzibel ist und für mindestens ein $\mathbf{x} \in \bar{\Omega}_h$ (4.3.4a) gilt.

Die vorangehenden Überlegungen begründen, warum die Randdiskretisierung die Bedingungen (5.2.4b) erfüllen sollte. Für den Fall $b_0 = 0$ gilt (5.2.4b) genau dann, wenn $b_2/b_1 \in [0, 1]$. Falls $b_2/b_1 \in [-1, 0]$, lässt sich (5.2.4b) erfüllen, wenn man zwischen \bar{y} und $\bar{y} - h$ (statt \bar{y} und $\bar{y} + h$) interpoliert. Ist dagegen $|b_2| > |b_1|$, überwiegt also die tangentielle Komponente, könnte man zwischen $(\bar{x} + h, \bar{y})$ und $(\bar{x} + h, \bar{y} \pm kh)$ interpolieren, wobei $|b_2/b_1| \leq k$. Geschickter ist es jedoch, im Falle von $b_2/b_1 \geq 1$ im Punkt $(\bar{x}, \bar{y} + h)$ zwischen $(\bar{x}, \bar{y} + h)$ und $(\bar{x} + h, \bar{y} + h)$ zu interpolieren. Allgemein wähle man als Interpolationspunkt den Schnittpunkt der Geraden mit dem gestrichelten Streckenzug aus Abbildung 5.2.4.

Bei der bisherigen Diskussion sind wir vom Fall $\Omega = (0, 1) \times (0, 1)$ ausgegangen. Ist Ω ein allgemeines Gebiet, bieten sich wieder zwei Diskretisierungstechniken an (vgl. §4.8.1 und §4.8.2).

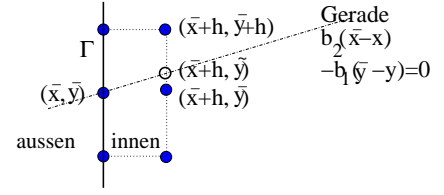


Abb. 5.2.4. Diskretisierung von $Bu = \varphi$

Erste Diskretisierungsmöglichkeit. Ω_h und Γ_h seien wie in §4.8.1 gewählt. Die Differentialgleichung wird in randnahen Punkten durch ein Differenzenschema approximiert, das im Falle $L = -\Delta$ dem Shortley-Weller-Verfahren entspricht. Hierzu benötigt man die Werte von u_h in den Randpunkten $\xi \in \Gamma_h$. Wie in Abbildung 5.2.5 sei $(\bar{x}, \bar{y}) = (\hat{x} - s_t h, \hat{y}) \in \Gamma_h$ ein Randpunkt. Wieder lassen sich zu (5.2.3a-c) analoge Gleichungen aufstellen. In Abbildung 5.2.5 hat (\hat{x}, \hat{y}) die gleiche Position wie $(\bar{x} + h, \bar{y})$ in Abbildung 5.2.4. Im Allgemeinen hat man den Schnittpunkt der Geraden, die auch durch $(\bar{x}, \bar{y}) + t\mathbf{b}$ ($t \in \mathbb{R}$) dargestellt wird, mit dem in Abbildung 5.2.5 gestrichelt gezeichneten Streckenzug zu verwenden.

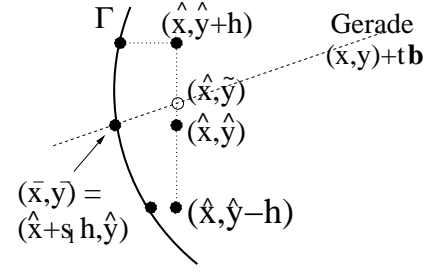


Abb. 5.2.5. erste Randdiskretisierung

Zweite Diskretisierungsmöglichkeit. $\bar{\Omega}_h$ sei das oben verwendete Gitter Ω_h . Ω_h bestehe jetzt aus allen randfernen Punkten von $\bar{\Omega}_h$. Für alle $(x, y) \in \Omega_h$ sind Differenzengleichungen (mit äquidistanten Schrittweiten) erklärt, die $\{u_h(\xi) : \xi \in \bar{\Omega}_h\}$ enthalten. Für jeden Punkt $(\hat{x}, \hat{y}) \in \Gamma_h := \bar{\Omega}_h \setminus \Omega_h$ ist eine Randdiskretisierung zu finden (vgl. Abbildung 5.2.6). Gleichung (5.2.3a) lässt sich mit (\hat{x}, \hat{y}) und $(\hat{x} + h, \hat{y})$ anstelle von (\bar{x}, \bar{y}) und $(\bar{x} + h, \bar{y})$ aufstellen. Die Argumente der Koeffizienten b_1 und b_2 sind (\bar{x}, \bar{y}) . Dieser Punkt ergibt sich implizit aus den beiden Bedingungen

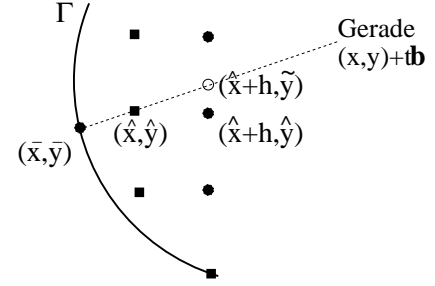


Abb. 5.2.6. zweite Randdiskretisierung

$$b_2(\bar{x}, \bar{y}) (\bar{x} - \hat{x}) = b_1(\bar{x}, \bar{y}) (\bar{y} - \hat{y}), \quad (\bar{x}, \bar{y}) \in \Gamma. \quad (5.2.5)$$

Bemerkung 5.2.8 a) Die erste Diskretisierung erfordert in den randnahen Gitterpunkten ein Differenzenschema für $Lu = f$ mit nichtäquidistanten Schrittweiten. Die zweite Diskretisierung verlangt eine Approximation des nichtlinearen Problems (5.2.5). Programmtechnisch sind beide Verfahren wegen der notwendigen Fallunterscheidungen unbequem.

b) Nähert sich der Vektor \mathbf{b} aus B der Tangentialrichtung, versagen beide Verfahren, da die Gerade nicht mehr den Streckenzug aus den Abbildungen 5.2.5 und 5.2.6 schneidet. Dabei versagt die zweite Diskretisierung früher als die erste.

Eine weitere Möglichkeit, die den geschilderten Schwierigkeiten aus dem Weg geht, besteht darin, zumindest in Randnähe Variationsdifferenzengleichungen zu verwenden (vgl. Bemerkung 8.5.5).

Gelegentlich ist es auch möglich, die Randbedingungen durch Koordinatentransformationen zu vereinfachen. Sei $Bu = \varphi$ auf $\gamma \subset \Gamma$ vorgeschrieben. Findet man eine Transformation $(x, y) \leftrightarrow (\xi, \eta)$ derart, dass $\xi = 0$ und $b_2 x_\xi = b_1 y_\xi$ auf γ erfüllt sind, so gilt für das transformierte Problem $Bu \equiv \sigma \partial u / \partial n + b_0 u = \varphi$ ($\partial / \partial n = -\partial / \partial \xi$) auf einem vertikalen Randstück. Die Diskretisierung kann wie in §4.7 beschrieben durchgeführt werden. Die gleiche Argumentation wie in den vorangegangenen Abschnitten beweist Stabilität und Konvergenz:

Bemerkung 5.2.9 Das Differenzenverfahren $D_h u_h = f_h$ erfülle die Voraussetzungen des Satzes 5.1.25, sodass $-D_h \tilde{w}_h \geq \mathbf{1}$ für $\tilde{w}_h := K\mathbf{1} + w_h$ gilt ($K > 0$: Konstante, w_h : wie im Beweis des Satzes 5.1.25). Die Randdiskretisierung (5.2.3c') erfülle

$$c_0 - c_1 - c_2 = b_0(\mathbf{x}) \geq \varepsilon > 0 \quad \text{für alle } \mathbf{x} \in \Gamma_h.$$

Für hinreichend großes K ist dann $(B_h \tilde{w}_h)(\mathbf{x}) \geq \mathbf{1}$ für $\mathbf{x} \in \Gamma_h$. Damit folgt $A_h \tilde{w}_h \geq \mathbf{1}$, wobei A_h die anschließend an Gleichung (5.2.4b) definierte M -Matrix ist. Da $\|A_h^{-1}\|_\infty \leq \text{const}$ (vgl. Satz 4.3.16), ist die Stabilität bewiesen. Die Konvergenzordnung ist das Minimum von 1 (Konsistenzordnung von B_h ; vgl. Lemma 5.2.7) und der Konsistenzordnung von $D_h u_h = f_h$.

Im Allgemeinen ist es nicht möglich, $\mathbf{b}^\top \nabla$ durch symmetrische Differenzen zu approximieren. Um trotzdem eine Diskretisierung von $Bu = \varphi$ mit Konsistenzordnung 2 zu konstruieren, setze man an:

$$B_h u_h(\bar{x}, \bar{y}) := c_0 u_h(\bar{x}, \bar{y}) - \sum_{i=1}^3 c_i u_h(x_i, y_i) = g(\bar{x}, \bar{y}), \quad (\bar{x}, \bar{y}) \in \Gamma_h,$$

wobei (x_i, y_i) drei Punkte aus Ω_h seien. Benutzt man zum Abgleich in der Taylor-Entwicklung die Differentialgleichung $Lu = f$ in (\bar{x}, \bar{y}) und die Tangentialableitung von $Bu = \varphi$, findet man eine Diskretisierung der Ordnung 2. Für den Spezialfall $L = \Delta$, $B = \partial/\partial n + b_0$ weisen Bramble-Hubbard [16] nach, dass die drei Punkte (x_i, y_i) so gewählt werden können, dass die Ungleichungen $c_i \geq 0$, $c_0 \geq c_1 + c_2 + c_3$ gelten und Stabilität garantieren. Die (x_i, y_i) liegen jedoch im Allgemeinen nicht in der unmittelbaren Nachbarschaft von $(\bar{x}, \bar{y}) \in \Gamma_h$. Die Konstruktion der Diskretisierung scheint allerdings zu kompliziert, als dass sie für praktische Zwecke empfohlen werden könnte.

5.3 Randwertaufgaben höherer Ordnung

5.3.1 Die biharmonische Differentialgleichung

In der Elastomechanik führt die freie Stabschwingung zu (gewöhnlichen) Differentialgleichungen zweiter Ordnung, wenn es sich um Längsschwingungen (= Dehnungsschwingungen) oder um Torsionsschwingungen handelt. Dagegen ergeben Querschwingungen (= Biegungsschwingungen) eine Gleichung der Ordnung 4. Entsprechend führt die Biegungsschwingung einer Platte auf eine partielle Differentialgleichung der Ordnung 4. Es ist die biharmonische Gleichung (Plattengleichung)

$$\Delta^2 u = f \quad \text{in } \Omega \quad (5.3.1)$$

($\Delta^2 = \partial^4/\partial x^4 + 2\partial^4/\partial x^2\partial y^2 + \partial^4/\partial y^4$). u beschreibt dabei die Auslenkung der Platte vertikal zur Oberfläche. Ist die Platte am Rand fest eingespannt, erhält man die Randbedingungen

$$u = \varphi_1 \quad \text{und} \quad \frac{\partial u}{\partial n} = \varphi_2 \quad \text{auf } \Gamma \quad (5.3.2)$$

mit $\varphi_1 = \varphi_2 = 0$. Ein biharmonisches Problem (5.3.1), (5.3.2) entsteht ebenfalls durch Umformung des Stokes-Gleichungen in $\Omega \subset \mathbb{R}^2$ (vgl. Bemerkung 12.1.5). Die Differentialgleichung (5.3.1) lässt sich mit anderen Randwerten als (5.3.2) kombinieren. Ein Beispiel ist

$$u = \varphi_1 \quad \text{und} \quad \Delta u = \varphi_2 \quad \text{auf } \Gamma \quad (5.3.3)$$

(einfach unterstützte Platte), weitere Beispiele findet man in (7.4.12b,e).

Übungsaufgabe 5.3.1 Man zeige: Löst man die Poisson-Gleichung $\Delta v = f$ in Ω , $v = \varphi_2$ auf Γ und anschließend $\Delta u = v$ in Ω und $u = \varphi_1$ auf Γ , so ist u die Lösung der Randwertaufgabe (5.3.1), (5.3.3). Warum lässt sich das Problem (5.3.1), (5.3.2) nicht ebenso behandeln?

Bemerkung 5.3.2 Die Lösungen von $\Delta^2 u = 0$ erfüllen kein Maximum-Minimum-Prinzip (Gegenbeispiel: $u = x^2 + y^2$ in $\Omega = K_R(0)$).

5.3.2 Allgemeine lineare Differentialgleichung der Ordnung $2m$

Die partielle Ableitung D^α ($\alpha \in \mathbb{N}_0^n$: Multiindex) der Ordnung $|\alpha| = \alpha_1 + \dots + \alpha_n$ ist in (3.2.5) definiert. Ein Differentialoperator der Ordnung $2m$ lautet³

$$L = \sum_{|\alpha| \leq 2m} a_\alpha(\mathbf{x}) D^\alpha \quad (\mathbf{x} \in \Omega) \quad (5.3.4a)$$

und definiert die Differentialgleichung der Ordnung $2m$:

$$Lu = f \quad \text{in } \Omega. \quad (5.3.4b)$$

Die Elliptizität wurde bisher nur für Gleichungen zweiter Ordnung erklärt (vgl. Definition 1.2.3).

³Die Notation $\sum_{|\alpha| \leq 2m}$ (oder $\sum_{|\alpha|=2m}$ usw.) bedeutet stets die Summation über alle Multiindizes $\alpha \in \mathbb{N}_0^d$ mit der angegebenen Nebenbedingung.

Definition 5.3.3 Der Differentialoperator L aus (5.3.4a) (mit reellwertigen Koeffizienten a_α) heißt elliptisch (von der Ordnung $2m$) in $\mathbf{x} \in \Omega$, wenn

$$\sum_{|\alpha|=2m} a_\alpha(\mathbf{x}) \xi^\alpha \neq 0 \quad \text{für alle } 0 \neq \xi \in \mathbb{R}^n. \quad (5.3.5a)$$

Dabei ist ζ^α die Abkürzung für das Polynom $\zeta_1^{\alpha_1} \zeta_2^{\alpha_2} \dots \zeta_n^{\alpha_n}$ vom Grad $|\alpha|$. Wir setzen $P(\mathbf{x}, \zeta) := \sum_{|\alpha|=2m} a_\alpha(\mathbf{x}) \zeta^\alpha$. Offenbar ist (5.3.5a) äquivalent zu $P(\mathbf{x}, \zeta) \neq 0$ für alle $\zeta \in \mathbb{R}^n$ mit $|\zeta| = 1$. Aus Stetigkeitsgründen muss entweder $P(\mathbf{x}, \zeta) > 0$ oder $P(\mathbf{x}, \zeta) < 0$ gelten. Ohne Beschränkung der Allgemeinheit können wir $P(\mathbf{x}, \zeta) > 0$ annehmen; sonst skalieren wir mit dem Faktor -1 (Übergang von $Lu(\mathbf{x}) = f(\mathbf{x})$ zu $-Lu(\mathbf{x}) = -f(\mathbf{x})$). Da die Menge $\{\zeta \in \mathbb{R}^n : |\zeta| = 1\}$ kompakt ist, folgt $c(\mathbf{x}) := \min\{P(\mathbf{x}, \zeta) : |\zeta| = 1\} > 0$ und begründet die Formulierung von (5.3.5a) durch

$$\sum_{|\alpha|=2m} a_\alpha(\mathbf{x}) \xi^\alpha \geq c(\mathbf{x}) |\xi|^{2m} \quad \text{für alle } \xi \in \mathbb{R}^n \text{ mit } c(\mathbf{x}) > 0. \quad (5.3.5b)$$

Definition 5.3.4 Der Differentialoperator L heißt gleichmäßig elliptisch in Ω , falls $\inf\{c(\mathbf{x}) : \mathbf{x} \in \Omega\} > 0$ für $c(\mathbf{x})$ aus (5.3.5b).

Übungsaufgabe 5.3.5 (a) Man übersetze L aus (5.1.1b) in die Schreibweise (5.3.4a). Wie lauten die Koeffizienten a_α für $L = \Delta^2$?

(b) Man beweise: Der biharmonische Operator Δ^2 ist gleichmäßig elliptisch.

(c) a_α seien reellwertig. Warum gibt es keine elliptischen Operatoren L ungerader Ordnung?

(d) Sind die Koeffizienten a_α hinreichend glatt, so lässt sich der Operator L aus (5.3.4a) in der Form $L = \sum_{|\alpha| \leq m} \sum_{|\beta| \leq m} (-1)^{|\beta|} D^\beta a_{\alpha\beta}(\mathbf{x}) D^\alpha$ schreiben (vgl. (5.1.1b) und (5.1.2)).

Für $m = 1$ (Gleichung der Ordnung 2) haben wir eine Randbedingung gebraucht; für die biharmonische Gleichung ($m = 2$) treten zwei Randbedingungen auf. Allgemein benötigt man m Randbedingungen

$$(B_j u)(\mathbf{x}) := \sum_{|\alpha| \leq m_j} b_{j\alpha}(\mathbf{x}) D^\alpha u(\mathbf{x}) = \varphi_j(\mathbf{x}) \quad \text{für } \mathbf{x} \in \Gamma. \quad (5.3.6)$$

mit Rand-Differentialoperatoren B_j der Ordnung $0 \leq m_j < 2m$.

Bemerkung 5.3.6 Die Randoperatoren B_j können nicht beliebig gewählt werden, sondern müssen unabhängig voneinander sein (ein sogenanntes "normales System" bilden; vgl. Lions-Magenes [58, p. 113] und Wloka [89, p. 214]). Insbesondere müssen die Ordnungen m_j paarweise verschieden sein.

Eine Dirichlet-Randbedingung liegt vor, wenn die Randoperatoren $B_j = \partial^{j-1}/\partial n^{j-1} = (\partial/\partial n)^{j-1}$ für $j = 1, \dots, m$ lauten (vgl. (5.3.2)). Auf die Darstellung der Lösung mit Hilfe der Greenschen Funktion soll hier nicht eingegangen werden. Bemerkenswert ist jedoch, dass die Greensche Funktion (und insbesondere die Singularitätenfunktion) zu L stetig ist, sobald $2m > n$. Für $L = \Delta^2$ in \mathbb{R}^2 lautet die Singularitätenfunktion beispielsweise $\frac{1}{8\pi} |\mathbf{x} - \mathbf{y}|^2 \log |\mathbf{x} - \mathbf{y}|$ (vgl. Wloka [89, p. 349]).

In Satz 2.4.6 wurde die Frage nach der Stetigkeit der Randwertaufgabe bezüglich Änderungen des Gebietes Ω diskutiert. Das dortige Resultat ist für Differentialgleichungen zweiter Ordnung richtig. Für das hier vorliegende Randwertproblem vierter Ordnung muss die Änderung von Ω glatter sein: Die Metrik, die die Änderung von Ω misst, muss die Krümmung des Randes berücksichtigen. Verwendet man dagegen die übliche Hausdorff-Norm, stellt das sogenannte Babuška-Paradoxon ein schönes Gegenbeispiel dar. Hierzu betrachtet man die biharmonische Gleichung (5.3.1) mit der Randbedingung $u = \Delta u = 0$ auf dem gleichmäßigen n -Eck Ω_n , das dem Einheitskreis Ω eingeschrieben ist. Offenbar konvergiert Ω_n in der Hausdorff-Norm gegen Ω . Aber die Lösungen u_n auf Ω_n konvergieren nicht gegen die Lösung u auf Ω (vgl. Babuška [4] und Rannacher [69]).

5.3.3 Diskretisierung der biharmonischen Differentialgleichung

Die einfachste Differenzenformel für $L = \Delta^2$ ist das 13-Punkt-Differenzenverfahren

$$D_h = \Delta_h^2 = h^{-4} \begin{bmatrix} & & 1 & & \\ & 2 & -8 & 2 & \\ 1 & -8 & 20 & -8 & 1 \\ & 2 & -8 & 2 & \\ & & 1 & & \end{bmatrix}, \quad (5.3.7)$$

das sich als Quadrat des Fünfpunktsterns Δ_h aus (4.2.3) darstellen lässt. Das Gitter $\Omega_h \subset \Omega$ und die Ränder $\Gamma_h \subset \Gamma$, $\gamma_h \subset \mathbb{R}^2 \setminus \overline{\Omega}$ seien wie in Abbildung 5.3.1 definiert (\bullet für innere und \circ für äußere Gitterpunkte, \times für Randpunkte). Die Differenzengleichung

$$D_h u_h = f_h \quad \text{in } \Omega_h \quad (5.3.8a)$$

benötigt die Werte von u_h in $\overline{\Omega}_h := \Omega_h \cup \Gamma_h \cup \gamma_h$. Diese sind durch die Randbedingungen

$$u_h(\mathbf{x}) = \varphi_1(\mathbf{x}) \quad \text{für } \mathbf{x} \in \Gamma_h, \quad (5.3.8b)$$

$$\partial_n^0 u_h(\mathbf{x}) := \frac{1}{2h} [u_h(\mathbf{x} + h\vec{n}) - u_h(\mathbf{x} - h\vec{n})] = \varphi_2(\mathbf{x}) \quad \text{für } \mathbf{x} \in \Gamma_h, \mathbf{x} + h\vec{n} \in \gamma_h \quad (5.3.8c)$$

bestimmt (vgl. (5.3.2)). Dabei werden wie in Gleichung (4.7.14b) den Eckpunkten je zwei Normalenrichtungen zugeordnet. Dementsprechend gehören zu jedem Eckpunkt zwei Gleichungen (5.3.8c). Man beachte, dass zwei ($m = 2$) Randschichten Γ_h und γ_h auftreten und zwei Randbedingungen gegeben sind.

Bemerkung 5.3.7 *Eliminiert man mit Hilfe von (5.3.8b,c) in Gleichung (5.3.8a) die Unbekannten $\{u_h(\xi) : \xi \in \Gamma_h \cup \gamma_h\}$, so ergibt sich ein Gleichungssystem $L_h u_h = q_h$ für $\{u_h(\xi) : \xi \in \Omega_h\}$. L_h ist keine M-Matrix, da die Vorzeichenbedingung (4.3.1a) verletzt ist. Aber auch L_h^{-1} ist im Allgemeinen nicht ≥ 0 (vgl. (4.3.1b)).*

Damit lassen sich die Beweismethoden des Abschnittes 4 nicht wiederholen. Stattdessen werden wir das Analogon des Satzes 4.4.12 beweisen.

Satz 5.3.8 *u_h sei die durch (5.3.8a-c) definierte diskrete Lösung der biharmonischen Gleichung im Quadrat $\Omega = (0,1) \times (0,1)$. Die in Bemerkung 5.3.7 erklärte Matrix L_h ist symmetrisch und positiv definit. Sie erfüllt*

$$\|L_h\|_2 \leq 64h^{-4}, \quad \|L_h^{-1}\|_2 \leq 1/256. \quad (5.3.9a)$$

Wenn $\varphi_1 = 0$ auf Γ_h (vgl. (5.3.8b)), so lässt sich u_h durch

$$\|u_h\|_{\Omega_h} \leq \frac{1}{256} \|f_h\|_{\Omega_h} + \frac{h^{-1/2}}{16\sqrt{2}} |\varphi_2|_{\Gamma_h} \quad (5.3.9b)$$

abschätzen, wobei die Normen wie folgt definiert seien:

$$\|u_h\|_{\Omega_h} := \sqrt{h^2 \sum_{\mathbf{x} \in \Omega_h} |u_h(\mathbf{x})|^2}, \quad |\varphi|_{\Gamma_h} := \sqrt{h \sum_{\mathbf{x} \in \Gamma_h} |\varphi(\mathbf{x})|^2}. \quad (5.3.9c)$$

Die Summe $\sum_{\mathbf{x} \in \Gamma_h}$ in $|\varphi_2|_{\Gamma_h}$ soll die Eckpunkte zweimal enthalten, wobei z.B. in $\mathbf{x} = (0,0)$ der Wert $\varphi(\mathbf{x})$ einmal als $\varphi_2(0+,0)$, einmal als $\varphi_2(0,0+)$ zu interpretieren ist (vgl. Kommentar zu (5.3.8c)).

Beweis. Wir beschränken uns auf den wichtigsten Teil, die Ungleichung (5.3.9b). Im unbeschränkten Gitter $Q_h := \{(x,y) \in \mathbb{R}^2 : x/h, y/h \in \mathbb{N}\}$ setzen wir

$$\bar{u}_h(\mathbf{x}) := u_h(\mathbf{x}) \quad \text{für } \mathbf{x} \in \Omega_h, \quad \bar{u}_h(\mathbf{x}) := 0 \quad \text{für } \mathbf{x} \in Q_h \setminus \Omega_h.$$

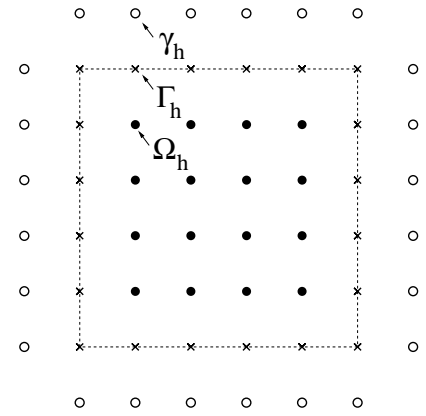


Abb. 5.3.1. Gitter für die biharmonische Gleichung in $\Omega = (0,1) \times (0,1)$

Man erhält

$$\sum_{\mathbf{x} \in \Omega_h} u_h(\mathbf{x}) f_h(\mathbf{x}) = \sum_{\mathbf{x} \in Q_h} \bar{u}_h(\mathbf{x}) \Delta_h^2 u_h(\mathbf{x}) = \sum_{\mathbf{x} \in Q_h} \Delta_h \bar{u}_h(\mathbf{x}) \Delta_h u_h(\mathbf{x})$$

nach partieller Summation (man überlege sich zunächst die Identitäten $\sum v(x) \partial^+ \partial^- u(x) = -\sum (\partial^- v)(x) (\partial^- u)(x) = \sum (\partial^+ \partial^- v)(x) u(x)$ im eindimensionalen Fall). Nach Definition von \bar{u}_h ist $\Delta \bar{u}_h = 0$ auf $Q_h \setminus (\Omega_h \cup \Gamma_h)$. Auf Γ_h hat man $u_h = \bar{u}_h = 0$ und

$$\bar{u}_h(\mathbf{x} - h\vec{n}) = u_h(\mathbf{x} - h\vec{n}), \quad \bar{u}_h(\mathbf{x} + h\vec{n}) = 0, \quad u_h(\mathbf{x} + h\vec{n}) = u_h(\mathbf{x} - h\vec{n}) + 2h\varphi_2(\mathbf{x})$$

für $\mathbf{x} \in \Gamma_h$. Setzt man $\partial_n^- u_h(\mathbf{x}) := [u_h(\mathbf{x}) - u_h(\mathbf{x} - h\vec{n})]/h = -u_h(\mathbf{x} - h\vec{n})/h$ auf $\mathbf{x} \in \Gamma_h$, so zeigt

$$\Delta_h \bar{u}_h(\mathbf{x}) = -h^{-1} \partial_n^- u_h(\mathbf{x}), \quad \Delta_h u_h(\mathbf{x}) = -2h^{-1} \partial_n^- u_h(\mathbf{x}) + 2h^{-1} \varphi_2(\mathbf{x}) \quad \text{in } \mathbf{x} \in \Gamma_h,$$

dass $(u_h, L_h u_h)_{\Omega_h} = (u_h, f_h)_{\Omega_h} = \|\Delta_h u_h\|_{\Omega_h}^2 + 2h^{-1} |\partial_n^- u_h|_{\Gamma_h}^2 - 2h^{-1} (\partial_n^- u_h, \varphi_2)_{\Gamma_h}$, wobei $(\cdot, \cdot)_{\Gamma_h}$ und $(\cdot, \cdot)_{\Omega_h}$ die zu den Normen (5.3.9c) gehörenden Skalarprodukte sind. Gemäß

$$ab \leq a^2 + b^2/4 \quad (a, b \in \mathbb{R}) \quad (5.3.10)$$

schätzt man ab:

$$(\partial_n^- u_h, \varphi_2)_{\Gamma_h} \leq |\partial_n^- u_h|_{\Gamma_h} |\varphi_2|_{\Gamma_h} \leq |\partial_n^- u_h|_{\Gamma_h}^2 + |\varphi_2|_{\Gamma_h}^2 / 4.$$

Somit erhält man

$$\begin{aligned} \|\Delta_h u_h\|_{\Omega_h}^2 + 2h^{-1} |\partial_n^- u_h|_{\Gamma_h}^2 &\leq (u_h, f_h)_{\Omega_h} + 2h^{-1} (\partial_n^- u_h, \varphi_2)_{\Gamma_h} \\ &\leq \|u_h\|_{\Omega_h} \|f_h\|_{\Omega_h} + 2h^{-1} |\partial_n^- u_h|_{\Gamma_h}^2 + h^{-1} |\varphi_2|_{\Gamma_h}^2 / 2. \end{aligned} \quad (5.3.11)$$

Satz 4.4.12 beweist $\|u_h\|_{\Omega_h} \leq \|\Delta_h u_h\|_{\Omega_h} / 16$. Für $\varphi_2 = 0$ liefert (5.3.11) daher $\|u_h\|_{\Omega_h}^2 \leq 16^{-2} \|\Delta_h u_h\|_{\Omega_h}^2 \leq 16^{-2} \|u_h\|_{\Omega_h} \|f_h\|_{\Omega_h}$, also

$$\|u_h\|_{\Omega_h} \leq 16^{-2} \|f_h\|_{\Omega_h}. \quad (5.3.12a)$$

Für $f_h = 0$ liefert (5.3.11) die Ungleichung $\|\Delta_h u_h\|_{\Omega_h}^2 \leq (2h)^{-1} |\varphi_2|_{\Gamma_h}^2$, also

$$\|u_h\|_{\Omega_h} \leq \frac{1}{16} \|\Delta_h u_h\|_{\Omega_h} \leq \frac{1}{16} (2h)^{-1/2} |\varphi_2|_{\Gamma_h}. \quad (5.3.12b)$$

Im allgemeinen Fall schreiben wir u_h als $u_h^I + u_h^{II}$, wobei $\varphi_2^I = 0$, $\varphi_2^{II} = \varphi_2$, $f_h^I = f_h$ und $f_h^{II} = 0$ die Summanden definieren. $\|u_h\|_{\Omega_h} \leq \|u_h^I\|_{\Omega_h} + \|u_h^{II}\|_{\Omega_h}$ zusammen mit (5.3.12a,b) ergibt (5.3.9b). ■

Satz 5.3.9 (Konvergenz) Die Lösung der biharmonischen Randwertaufgabe (5.3.1) und (5.3.2) in $\Omega = (0, 1) \times (0, 1)$ erfülle $u \in C^{5,1}(\bar{\Omega})$. Die diskrete Lösung der Gleichungen (5.3.8a-c) konvergiert dann (mindestens) von der Ordnung $\frac{2}{3}$:

$$\|u_h - R_h u\|_{\Omega_h} \leq C_1 h^2 \|u\|_{C^{5,1}(\bar{\Omega})} + C_2 h^{3/2} \|u\|_{C^{2,1}(\bar{\Omega})} \quad (5.3.13)$$

(R_h : Beschränkung auf $\Omega_h \cup \Gamma_h$).

Beweis. Taylor-Entwicklung zeigt

$$\Delta_h R_h u = \Delta u + \frac{h^2}{4!} (u_{xxxx} + u_{yyyy}) + h^4 R, \quad |R| \leq \frac{1}{360} \|u\|_{C^{5,1}(\bar{\Omega})} \quad \text{in } \Omega_h. \quad (5.3.14)$$

Außerhalb von Ω_h kann man $R_h u$ nach Belieben erklären, da diese Werte nicht in (5.3.13) erscheinen. In γ_h setzen wir

$$R_h u(\mathbf{x} + h\vec{n}) := u(\mathbf{x} - h\vec{n}) + 2h\varphi_2(\mathbf{x}) + \frac{2h^3}{3!} \frac{\partial^3 u(\mathbf{x})}{\partial n^3} - \frac{2h^4}{4!} \frac{\partial^4 u(\mathbf{x})}{\partial n^4} + \frac{2h^5}{5!} \frac{\partial^5 u(\mathbf{x})}{\partial n^5} \quad (\mathbf{x} \in \Gamma_h).$$

Die Wahl ist so getroffen, dass (5.3.14) auch in $\mathbf{x} \in \Gamma_h$ gilt. Anwendung von Δ_h auf $\Delta_h R_h u$ liefert

$$\Delta_h^2 R_h u = D_h R_h u = \tilde{R}_h \Delta^2 u + \mathcal{O}(h^2 \|u\|_{C^{5,1}(\bar{\Omega})}) \quad \text{in } \Omega_h.$$

(\tilde{R}_h : Beschränkung auf Ω_h). Die Differenz $w_h := u_h - R_h u$ (in Ω_h definiert) erfüllt

$$D_h w_h = D_h u_h - D_h R_h u = \tilde{R}_h (f - \Delta^2 u) + \mathcal{O}(h^2 \|u\|_{C^{5,1}(\overline{\Omega})}) =: g_h, \quad (5.3.15a)$$

$$w_h(\mathbf{x}) = u_h(\mathbf{x}) - (R_h u)(\mathbf{x}) = \varphi_1(\mathbf{x}) - u(\mathbf{x}) = 0 \quad \text{für } \mathbf{x} \in \Gamma_h, \quad (5.3.15b)$$

$$\partial_n^0 w_h(\mathbf{x}) = \partial_n^0 u_h(\mathbf{x}) - \partial_n^0 R_h u(\mathbf{x}) = \varphi_2(\mathbf{x}) - \left[\varphi_2(\mathbf{x}) + \mathcal{O}(h^2 \|u\|_{C^{2,1}(\overline{\Omega})}) \right] =: \psi_h \quad \text{für } \mathbf{x} \in \Gamma_h. \quad (5.3.15c)$$

Setzt man $g_h = \mathcal{O}(h^2 \|u\|_{C^{5,1}(\overline{\Omega})})$ und $\psi_h = \mathcal{O}(h^2 \|u\|_{C^{2,1}(\overline{\Omega})})$ anstelle von f_h und φ_2 in Satz 5.3.8 ein (man beachte $\varphi_1 = 0!$), so liefert (5.3.9b) die Behauptung (5.3.13). ■

Mit komplizierteren Mitteln lässt sich der Faktor $h^{-1/2}$ in (5.3.13) durch $C(\varepsilon)h^{-\varepsilon}$ ($\varepsilon > 0$ beliebig) ersetzen, sodass die Konvergenzordnung $\mathcal{O}(h^{2-\varepsilon})$ - also fast 2 - ist.

Bemerkung 5.3.10 Ungleichung (5.3.9a) zeigt $\text{cond}_2(L_h) := \|L_h\|_2 \|L_h^{-1}\|_2 = \mathcal{O}(h^{-4})$. Im Allgemeinen führt ein Differenzenverfahren für eine Differentialgleichung der Ordnung $2m$ auf die Kondition

$$\text{cond}_2(L_h) = \mathcal{O}(h^{-2m}) \quad (5.3.16)$$

Dies weist auf eine größere Rundungsfehlerempfindlichkeit bei höheren Ordnungen $2m$ hin (vgl. [77, §4.4]).

Differenzenverfahren für Randwertaufgaben vierter Ordnung mit variablen Koeffizienten und allgemeinen Gebieten Ω werden in einem Aufsatz von Zlámal [95] behandelt. Auch dort wird Konvergenz der Ordnung $\mathcal{O}(h^{3/2})$ gezeigt. $\mathcal{O}(h^2)$ -Konvergenz wird dagegen von Bramble [14] für den 13-Punkt-Stern (5.3.7) in einem allgemeinen Gebiet Ω bewiesen, wenn man die Randbedingungen geeignet diskretisiert.

6

Exkurs über Funktionalanalysis

Damit auch Leser ohne eingehende Vorkenntnisse aus der Funktionalanalysis den nächsten Kapiteln folgen können, werden hier diejenigen Definitionen und Resultate zusammengestellt, die später benötigt werden.

6.1 Banach-Räume und Hilbert-Räume

6.1.1 Normierte Räume

X sei ein linearer Raum (andere Bezeichnung: Vektorraum) über \mathbb{K} , wobei $\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$. Im Folgenden wird stets der Normalfall $K = \mathbb{R}$ zugrundegelegt. Nur im Zusammenhang mit der Fourier-Transformation tritt $\mathbb{K} = \mathbb{C}$ auf.

Die Norm $\|\cdot\| : X \rightarrow [0, \infty)$ ist in Definition 4.3.12 erklärt. Der mit einer Norm versehene lineare Raum X heißt *normierter Raum* und wird durch das Paar $(X, \|\cdot\|)$ bezeichnet. Sobald zweifelsfrei feststeht, welche Norm zu X gehört, wird diese Norm $\|\cdot\|_X$ genannt und nur X anstelle von $(X, \|\cdot\|_X)$ geschrieben.

Beispiel 6.1.1 a) Die Euklidische Norm $|\cdot|$ aus (4.3.13) wie auch die Maximumnorm (4.3.3) sind Normen des \mathbb{R}^n .

b) Die stetigen Funktionen bilden den (unendlich-dimensionalen) Raum $C^0(\overline{\Omega})$. Wenn Ω beschränkt ist, sind alle $u \in C^0(\overline{\Omega})$ beschränkt, sodass die Supremumsnorm (2.4.1) erklärt ist und die Normaxiome erfüllt. Ist Ω unbeschränkt, bilden die beschränkten, stetigen Funktionen eine echte Teilmenge $(C^0(\overline{\Omega}), \|\cdot\|_{C^0(\overline{\Omega})})$ von $C^0(\overline{\Omega})$. Anstelle von $\|\cdot\|_{C^0(\overline{\Omega})}$ benutzen wir die traditionelle Schreibweise $\|\cdot\|_\infty$.

c) Die in Definition 3.2.8 eingeführten Hölder-stetigen Funktionen bilden den normierten Raum $(C^s(\overline{\Omega}), \|\cdot\|_{C^s(\overline{\Omega})})$.

Mit der Norm ist auf X eine Topologie definiert: $A \subset X$ ist offen, falls für alle $x \in A$ ein $\varepsilon > 0$ existiert, sodass die "Kugel" $K_\varepsilon(x) = \{y \in X : \|x - y\| < \varepsilon\}$ in A enthalten ist. Wir schreiben

$$x_n \rightarrow x \quad \text{oder} \quad x = \lim_{n \rightarrow \infty} x_n, \quad \text{falls} \quad \|x_n - x\| \rightarrow 0.$$

Beispiel 6.1.2 Seien $f_n, f \in C^0(\overline{\Omega})$. Der Grenzprozess $f_n \rightarrow f$ (bezüglich $\|\cdot\|_\infty$) bedeutet die aus der Analysis bekannte gleichmäßige Konvergenz.

Übungsaufgabe 6.1.3 Die Norm $\|\cdot\| : X \rightarrow [0, \infty)$ ist stetig; insbesondere gilt die umgekehrte Dreiecksungleichung

$$\left| \|x\| - \|y\| \right| \leq \|x - y\| \quad \text{für alle } x, y \in X. \quad (6.1.1)$$

Wie Beispiel 6.1.1a zeigt, können auf X mehrere Normen definiert werden. Zwei Normen $\|\cdot\|$ und $\|\cdot\|$ auf X heißen *äquivalent*, falls eine Zahl $0 < C < \infty$ existiert, sodass

$$\frac{1}{C} \|x\| \leq \|x\| \leq C \|x\| \quad \text{für alle } x \in X. \quad (6.1.2)$$

Übungsaufgabe 6.1.4 Äquivalente Normen führen zu der gleichen Topologie auf X .

6.1.2 Operatoren

X und Y seien normierte Räume mit den Normen $\|\cdot\|_X$ bzw. $\|\cdot\|_Y$. Eine lineare Abbildung $T : X \rightarrow Y$ nennt man *Operator*. Wenn die *Operatornorm*

$$\|T\|_{Y \leftarrow X} := \sup \{ \|Tx\|_Y / \|x\|_X : 0 \neq x \in X \} \quad (6.1.3)$$

endlich ist, heißt T beschränkt (vgl. (4.3.9)).

Übungsaufgabe 6.1.5 a) T ist genau dann beschränkt, wenn es stetig ist. b) $\|\cdot\|_{Y \leftarrow X}$ ist Norm.

Mit der Addition $(T_1 + T_2)x = T_1x + T_2x$ und der skalaren Multiplikation bilden die beschränkten Operatoren einen linearen Raum, den man mit $L(X, Y)$ bezeichnet. $(L(X, Y), \|\cdot\|_{Y \leftarrow X})$ ist ein normierter Raum. Mit der zusätzlichen Multiplikation $(T_1 T_2)x = T_1(T_2x)$ bildet $L(X, X)$ sogar eine Algebra mit Eins-Element I , da $\|I\|_{X \leftarrow X} = 1$. Mit I wird stets die Identität bezeichnet: $Ix = x$.

Übungsaufgabe 6.1.6 Man beweise:

$$\|Tx\|_Y \leq \|T\|_{Y \leftarrow X} \|x\|_X \quad \text{für alle } x \in X, T \in L(X, Y).$$

Für $T_1 \in L(Y, Z)$ und $T_2 \in L(X, Y)$ gilt $T_1 T_2 \in L(X, Z)$ mit

$$\|T_1 T_2\|_{Z \leftarrow X} \leq \|T_1\|_{Z \leftarrow Y} \|T_2\|_{Y \leftarrow X}.$$

Wir schreiben $T_n \rightarrow T$ für $T, T_n \in L(X, Y)$, falls $\|T - T_n\|_{Y \leftarrow X} \rightarrow 0$ (Konvergenz in der Operatornorm).

6.1.3 Banach-Räume

Eine Folge $\{x_n \in X : n \geq 1\}$ heißt *Cauchy-Folge*, wenn $\sup\{\|x_n - x_m\|_X : n, m \geq k\} \rightarrow 0$ für $k \rightarrow \infty$. Ein Raum X heißt *vollständig*, wenn jede Cauchy-Folge gegen ein $x \in X$ konvergiert. Ein *Banach-Raum*¹ ist ein normierter und vollständiger Raum.

Beispiel 6.1.7 a) \mathbb{R}^n ist vollständig, \mathbb{Q}^n dagegen nicht (\mathbb{Q} : rationale Zahlen). b) Sei $D \subset \mathbb{R}^n$. $(C^0(D), \|\cdot\|_\infty)$ ist Banach-Raum.

Beweis. Ist $\{u_n\}$ Cauchy-Folge, so existiert $u^*(x) := \lim u_n(x)$ für alle $x \in D$. Da u_n gleichmäßig gegen u^* konvergiert, muss u^* stetig sein, d.h. $u^* \in C^0(D)$. c) $(C^k(D), \|\cdot\|_{C^k(D)})$ für $k \in \mathbb{N}$ und die Hölder-Räume $(C^s(D), \|\cdot\|_{C^s(D)})$, $s > 0$, sowie $(C^{k,1}(D), \|\cdot\|_{C^{k,1}(D)})$ sind vollständig, also Banach-Räume. ■

Mit $L^\infty(D)$ bezeichnet man die Menge der auf D beschränkten und lokal integrierbaren Funktionen. Dabei wird nicht zwischen Funktionen unterschieden, die fast überall übereinstimmen. Die Supremumsnorm wird in diesem Fall durch

$$\|u\|_{L^\infty(D)} := \inf \{ \sup \{|u(x)| : x \in D \setminus A\} : A \text{ Menge vom Maß null} \}$$

erklärt. $(L^\infty(D), \|\cdot\|_{L^\infty(D)})$ ist Banach-Raum.

Übungsaufgabe 6.1.8 X sei normierter Raum und Y Banach-Raum. Man zeige: $L(X, Y)$ ist Banach-Raum.

¹Stefan Banach, geb. 30. März 1892 in Krakau, gest. 31. Aug. 1945 in Lemberg (Lwow)

Übungsaufgabe 6.1.9 X sei Banach-Raum und $Z \subset X$ ein abgeschlossener Unterraum. Der Quotientenraum X/Z hat als Elemente die Klassen $\tilde{x} = \{x + z : z \in Z\}$. Man zeige: X/Z mit der Norm $\|\tilde{x}\| = \inf\{\|x + z\|_X : z \in Z\}$ ist Banach-Raum.

Eine Menge A heißt *dicht* in $(X, \|\cdot\|_X)$, falls $A \subset X$ und $\bar{A} = X$, d.h. jedes $x \in X$ ist Grenzwert einer Folge $a_n \in A$. Ist $(X, \|\cdot\|_X)$ ein normierter, aber nicht vollständiger Raum, so bezeichnet man $(\tilde{X}, \|\cdot\|_{\tilde{X}})$ als Vervollständigung von X , falls X dicht in \tilde{X} und $\|x\|_{\tilde{X}} = \|x\|_X$ für alle $x \in X$. Die Vervollständigung ist bis auf Isomorphie eindeutig bestimmt und kann ebenso konstruiert werden, wie \mathbb{Q} zu \mathbb{R} vervollständigt wird.

Lemma 6.1.10 $(X, \|\cdot\|_X)$ sei normierter linearer Raum und Unterraum eines Banach-Raumes $(Y, \|\cdot\|_Y)$ mit $\|x\|_Y \leq C \|x\|_X$ für alle $x \in X \subset Y$. Jede Cauchy-Folge $\{x_n\}$ in X mit $\|x_n\|_Y \rightarrow 0$ erfülle auch $\|x_n\|_X \rightarrow 0$. Dann existiert eine Vervollständigung \tilde{X} von X in Y , d.h. $X \subset \tilde{X} \subset Y$.

Häufig ist es nicht leicht, das Bild Tx eines Operators $T \in L(X, Y)$ für alle Elemente x des Banach-Raumes X zu beschreiben. Der folgende Satz erlaubt eine erhebliche Vereinfachung: Es genügt T auf einer dichten Menge $X_0 \subset X$ zu untersuchen.

Satz 6.1.11 X_0 sei dichter Unterraum (oder dichte Teilmenge) des normierten Raumes X . Y sei Banach-Raum. a) Ein auf X_0 definierter Operator $T_0 \in L(X_0, Y)$ mit $\|T_0\|_{Y \leftarrow X_0} = \sup\{\|T_0x\|_Y / \|x\|_X : 0 \neq x \in X_0\}$ hat eine eindeutige Fortsetzung $T \in L(X, Y)$ mit $Tx = T_0x$ für alle $x \in X_0$.

b) Für $x_n \rightarrow x$ ($x_n \in X_0, x \in X$) gilt $Tx = \lim_{n \rightarrow \infty} T_0x_n$.

c) Es ist $\|T\|_{Y \leftarrow X} = \|T_0\|_{Y \leftarrow X_0}$.

Beweis. Beweis. Für $x \in X_0$ ist T durch $Tx = T_0x$ definiert, während für $x \in X \setminus X_0$ eine Folge $X_0 \ni x_n \rightarrow x$ existiert und $Tx = \lim_{n \rightarrow \infty} Tx_n = \lim_{n \rightarrow \infty} T_0x_n$ festlegt. Es bleibt zu zeigen, dass $\lim T_0x_n$ existiert und unabhängig von der Auswahl der Folge $x_n \in X_0$ ist. Da $\|T_0x_n - T_0x_m\|_Y \leq \|T_0\|_{Y \leftarrow X} \|x_n - x_m\|_X$, ist T_0x_n eine Cauchy-Folge. Wegen der Vollständigkeit von Y existiert $y \in Y$ mit $T_0x_n \rightarrow y$. Für eine zweite Folge $x'_n \in X_0$ mit $x'_n \rightarrow x$ existiert aus dem gleichen Grunde ein $y' \in Y$, sodass $T_0x'_n \rightarrow y'$. Da $y' - y = \lim(T_0x'_n - T_0x_n)$ und $\|T_0x'_n - T_0x_n\|_Y \leq \|T_0\|_{Y \leftarrow X} \|x'_n - x_n\|_X \leq \|T_0\|_{Y \leftarrow X_0} \|x'_n - x_n\|_X \rightarrow 0$, ist Tx durch b) wohldefiniert. Teil (c) folgt aus $\|Tx\|_Y / \|x\|_X = \lim \|T_0x_n\|_Y / \|x_n\|_X$ für $x_n \rightarrow x, x_n \in X_0$. ■

Übungsaufgabe 6.1.12 X_0 sei dicht in $(X, \|\cdot\|)$. Eine zweite, auf X_0 definierte Norm $\|\cdot\|$ sei mit $\|\cdot\|$ auf X_0 äquivalent. Man zeige: Die Vervollständigung von X_0 bezüglich $\|\cdot\|$ ergibt $(X, \|\cdot\|)$, und $\|\cdot\|$ und $\|\cdot\|$ sind auch auf X äquivalent.

Eine Folgerung des ‘‘Satzes von der offenen Abbildung’’ (vgl. Yosida [91, §II.5]) liefert das folgende, wichtige Resultat:

Satz 6.1.13 X, Y seien Banach-Räume. $T \in L(X, Y)$ sei bijektiv. Dann gilt auch $T^{-1} \in L(Y, X)$.

Übungsaufgabe 6.1.14 X, Y seien Banach-Räume, $T \in L(X, Y)$ sei injektiv. Man zeige:

a) $Y_0 := \text{Bild}(T) := \{Tx : x \in X\}$ mit $\|\cdot\|_Y$ als Norm ist ein Banach-Unterraum von Y .

b) T^{-1} existiert auf Y_0 und ist beschränkt: $T^{-1} \in L(Y_0, X)$.

X und Y seien Banach-Räume mit $X \subset Y$. Offenbar ist die durch I bezeichnete Inklusion $I : x \in X \mapsto x \in Y$ eine lineare Abbildung. Wenn sie beschränkt ist:

$$I \in L(X, Y), \quad \text{d.h. } \|x\|_Y \leq C \|x\|_X \quad \text{für alle } x \in X, \quad (6.1.4)$$

nennt man X *stetig eingebettet* in Y . Ist X außerdem dicht in $(Y, \|\cdot\|_Y)$, so heißt X *dicht und stetig eingebettet* in Y .

Übungsaufgabe 6.1.15 $X \subset Y \subset Z$ seien Banach-Räume. Man zeige: Ist X [dicht und] stetig in Y und ist Y [dicht und] stetig in Z eingebettet, so ist auch X [dicht und] stetig in Z eingebettet.

6.1.4 Hilbert-Räume

Die Abbildung $(\cdot, \cdot) : X \times X \rightarrow \mathbb{K}$ ($\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$) heißt *Skalarprodukt* auf X , falls

$$\begin{aligned} (x, x) &> 0 && \text{für alle } 0 \neq x \in X, \\ (\lambda x + y, z) &= \lambda(x, z) + (y, z) && \text{für alle } \lambda \in \mathbb{K}, x, y, z \in X, \\ (x, y) &= \overline{(y, x)} && \text{für alle } x, y \in X, \end{aligned} \quad (6.1.5)$$

wobei $\bar{\lambda}$ die komplex konjugierte Zahl zu $\lambda \in \mathbb{C}$ ist.

Übungsaufgabe 6.1.16 Man zeige: a) $\|x\| := (x, x)^{1/2}$ ist Norm auf X . Zum Beweis verwende man:
b) Es gilt die Schwarzsche² Ungleichung

$$|(x, y)| \leq \|x\| \|y\| \quad \text{für alle } x, y \in X. \quad (6.1.6)$$

Hinweis: Betrachte $\|x - \lambda y\| \geq 0$ für $\lambda = \|x\|^2 / (x, y)$, falls $(x, y) \neq 0$.

c) $(\cdot, \cdot) : X \times X \rightarrow \mathbb{K}$ ist stetig.

Ein Banach-Raum X heißt *Hilbert³-Raum*, wenn ein Skalarprodukt $(\cdot, \cdot)_X$ auf X existiert, sodass $\|x\|_X = (x, x)_X^{1/2}$ für alle $x \in X$.

$x \in X$ und $y \in X$ heißen *orthogonal* (in Zeichen: $x \perp y$), falls $(x, y)_X = 0$. Ist $A \subset X$ eine Teilmenge des Hilbert-Raumes X , so definiert der *Orthogonalraum* $A^\perp := \{x \in X : (x, a)_X = 0 \text{ für alle } a \in A\}$ einen abgeschlossenen Unterraum von X .

Lemma 6.1.17 U sei ein abgeschlossener Unterraum eines Hilbert-Raumes X . Dann lässt sich X in die direkte Summe $X = U \oplus U^\perp$ zerlegen, d.h. alle $x \in X$ haben eine eindeutige Zerlegung $x = u + v$ mit $u \in U$ und $v \in U^\perp$. Ferner gilt $\|x\|_X^2 = \|u\|_X^2 + \|v\|_X^2$.

Übungsaufgabe 6.1.18 Sei A eine Teilmenge des Hilbert-Raumes X . Man beweise die Äquivalenz der folgenden Aussagen: (a) A liegt dicht in X ; (b) $A^\perp = \{0\}$; (c) Aus $x \in X$ und $(a, x)_X = 0$ für alle $a \in A$ folgt $x = 0$; (d) zu jedem $0 \neq x \in X$ existiert ein $a \in A$ mit $(a, x)_X \neq 0$.

6.2 Sobolev-Räume

Im Folgenden ist Ω stets eine offene Teilmenge des \mathbb{R}^n .

6.2.1 Der Raum $L^2(\Omega)$

Zu $L^2(\Omega)$ gehören alle Lebesgue⁴-messbaren Funktionen, deren Quadrate auf Ω Lebesgue-integrierbar sind. Zwei Funktionen $u, v \in L^2(\Omega)$ werden als gleich betrachtet ($u = v$), wenn $u(x) = v(x)$ für fast alle $x \in \Omega$, d.h. für alle $x \in \Omega \setminus A$, wobei die Ausnahmemenge A das Lebesgue-Maß $\mu(A) = 0$ besitzt.

Satz 6.2.1 $L^2(\Omega)$ bildet einen Hilbert-Raum mit dem Skalarprodukt

$$(u, v)_0 := \int_{\Omega} u(x) \overline{v(x)} \, dx \quad \text{für alle } u, v \in L^2(\Omega) \quad (6.2.1)$$

und der Norm

$$|u|_0 := \|u\|_{L^2(\Omega)} := \sqrt{\int_{\Omega} |u(x)|^2 \, dx}. \quad (6.2.2)$$

²Hermann Amandus Schwarz, geb. am 25. Jan. 1843 in Hermsdorf, gest. am 30. Nov. 1921 in Berlin

³David Hilbert, am 23 Januar 1862 in Königsberg (Preußen) geboren und am 14 Februar 1943 in Göttingen gestorben

⁴Henri Léon Lebesgue, geb. 28. Juni 1875 in Beauvais, Oise, Picardie, Frankreich, gest. 26. Juli 1941 in Paris

Lemma 6.2.2 Die Räume $C^\infty(\Omega) \cap L^2(\Omega)$ und $C_0^\infty(\Omega)$ liegen dicht in $L^2(\Omega)$. Dabei ist

$$C_0^\infty(\Omega) := \{u \in C^\infty(\Omega) : \text{Tr}(u) \text{ kompakt, } \text{Tr}(u) \subset\subset \Omega\}. \quad (6.2.3)$$

$\text{Tr}(u) := \overline{\{x \in \Omega : u(x) \neq 0\}}$ bezeichnet den Träger von u . Die doppelte Inklusion $\Omega' \subset\subset \Omega$ bedeutet, dass Ω' im Inneren von Ω liegt (d.h. $\Omega' \subset \Omega$ und $\Omega' \cap \partial\Omega = \emptyset$).

D^α sei der partielle Ableitungsoperator (3.2.5b). Im Folgenden benötigen wir sogenannte *schwache Ableitungen* $D^\alpha u$, die in nichtklassischer Weise definiert werden.

Definition 6.2.3 $u \in L^2(\Omega)$ besitzt eine (schwache) Ableitung $v := D^\alpha u \in L^2(\Omega)$, falls für dieses $v \in L^2(\Omega)$ gilt:

$$(w, v)_0 = (-1)^{|\alpha|} (D^\alpha w, u)_0 \quad \text{für alle } w \in C_0^\infty(\Omega). \quad (6.2.4)$$

Übungsaufgabe 6.2.4 Man zeige: a) u besitze eine schwache Ableitung $D^\alpha u \in L^2(\Omega)$. Existiert in $\Omega' \subset \Omega$ die klassische Ableitung $D^\alpha u$, so stimmt sie dort (fast überall) mit der schwachen überein.

b) Besitzt u die schwache Ableitung $v_\alpha = D^\alpha u \in L^2(\Omega)$, und hat v_α die schwache Ableitung $v_{\alpha+\beta} = D^\beta v_\alpha \in L^2(\Omega)$, so ist $v_{\alpha+\beta}$ auch die schwache $D^{\alpha+\beta}$ -Ableitung von u , d.h. $D^{\alpha+\beta} u = D^\beta(D^\alpha u)$.

c) $\Omega \subset \mathbb{R}^n$ sei beschränkt und $0 \in \Omega$. $u(x) := |x|^\sigma$ hat erste schwache Ableitungen in $L^2(\Omega)$, falls $\sigma = 0$ oder $2\sigma + n > 2$.

d) Für $u_\nu \in C^\infty(\Omega)$ gelte $u_\nu \rightarrow u \in L^2(\Omega)$ und $D^\alpha u_\nu \rightarrow v \in L^2(\Omega)$ bezüglich der $L^2(\Omega)$ -Norm. Dann ist $v = D^\alpha u$.

Für spätere Anwendungen betrachten wir das

Beispiel 6.2.5 Sei $\bar{\Omega} = \bigcup_{i=1}^N \bar{\Omega}_i$, wobei die beschränkten Teilgebiete Ω_i disjunkt seien und einen stückweise glatten Rand besitzen. Sei $k \in \mathbb{N}$. Eine Funktion $u \in C^{k-1}(\bar{\Omega})$ mit Beschränkungen $u_{\Omega_i} \in C^k(\bar{\Omega}_i)$ ($1 \leq i \leq N$) besitzt eine (schwache) k -fache Ableitung $v_\alpha = D^\alpha u \in L^2(\Omega)$, $|\alpha| \leq k$, die in $\bigcup \Omega_i$ mit der klassischen übereinstimmt.

Beweis. Die $(k-1)$ -fachen Ableitungen existieren als klassische Ableitungen, sodass die Behauptung nur noch für $k=1$ zu zeigen ist. Sei $D^\alpha = \partial/\partial x_j$, $w \in C_0^\infty(\Omega)$. Die Behauptung folgt durch partielle Integration:

$$-(D^\alpha w, u)_0 = -(w_{x_j}, u)_0 = - \int_\Omega w_{x_j} u dx = - \sum_{i=1}^N \int_{\Omega_i} w_{x_j} u dx = \sum_{i=1}^N \left[\int_{\Omega_i} w u_{x_j} dx - \int_{\partial\Omega_i} w u n_j^{(i)} d\Gamma \right],$$

wobei $n_j^{(i)}$ die j -te Komponente des Normalenvektors $\vec{n}^{(i)}$ bezüglich Ω_i ist. Für Punkte $x \in \partial\Omega_i \cap \partial\Omega_k$ ($i \neq k$) sind die Vorzeichen entgegengesetzt: $\vec{n}^{(i)} = -\vec{n}^{(k)}$. Daher fallen in $\sum_{i=1}^N \int_{\partial\Omega_i} w u n_j^{(i)} d\Gamma$ die Beiträge der inneren Seiten heraus. Randbeiträge von $\partial\Omega$ verschwinden ebenfalls, da $w = 0$ auf $\partial\Omega$. Damit folgt

$$-(D^\alpha w, u)_0 = \sum_{i=1}^N \int_{\Omega_i} w u_{x_j} dx = \sum_{i=1}^N \int_{\Omega_i} w v_\alpha dx = \int_\Omega w v_\alpha dx = (w, v_\alpha)_0,$$

wobei $v_\alpha \in L^2(\Omega)$ auf allen Teilbereichen Ω_i mit u_{x_j} übereinstimmt. ■

Die Ungleichung $|(u, v)_0| \leq |u|_0 |v|_0$ (vgl. (6.1.6)) lautet explizit

$$\left| \int_\Omega u(x) v(x) dx \right| \leq \sqrt{\int_\Omega |u(x)|^2 dx} \sqrt{\int_\Omega |v(x)|^2 dx}. \quad (6.2.5a)$$

Für $v = 1$ ergibt sich

$$\left| \int_\Omega u(x) dx \right| \leq \sqrt{\mu(\Omega)} \sqrt{\int_\Omega |u(x)|^2 dx} \quad (\mu(\Omega) : \text{Maß von } \Omega). \quad (6.2.5b)$$

Für $a \in L^\infty(\Omega)$ und $u, v \in L^2(\Omega)$ hat man

$$\left| \int_\Omega a(x) u(x) v(x) dx \right| \leq \|a\|_{L^\infty(\Omega)} \|u\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)} = \|a\|_\infty |u|_0 |v|_0. \quad (6.2.5c)$$

6.2.2 Die Räume $H^k(\Omega)$ und $H_0^k(\Omega)$

Sei $k \in \mathbb{N}_0$. $H^k(\Omega) \subset L^2(\Omega)$ sei die Menge aller Funktionen, die schwache Ableitungen $D^\alpha u \in L^2(\Omega)$ für $|\alpha| \leq k$ besitzen:

$$H^k(\Omega) = \{u \in L^2(\Omega) : D^\alpha u \in L^2(\Omega) \text{ für } |\alpha| \leq k\}.$$

Der hier mit $H^k(\Omega)$ bezeichnete Sobolev⁵-Raum wird auch mit $W_2^k(\Omega)$ oder $W^{k,2}(\Omega)$ bezeichnet.

Beispiel 6.2.6 Sei $\Omega = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$ der Einheitskreis. Man zeige, dass die Funktion $u(x, y) = \log(\log \frac{2}{\sqrt{x^2 + y^2}})$ zu $H^1(\Omega)$ gehört, aber unstetig ist.

Satz 6.2.7 $H^k(\Omega)$ bildet einen Hilbert-Raum mit dem Skalarprodukt

$$(u, v)_k := (u, v)_{H^k(\Omega)} := \sum_{|\alpha| \leq k} (D^\alpha u, D^\alpha v)_{L^2(\Omega)} \quad (6.2.6)$$

und der (Sobolev-)Norm

$$|u|_k := \|u\|_{H^k(\Omega)} := \sqrt{\sum_{|\alpha| \leq k} \|D^\alpha u\|_{L^2(\Omega)}^2}. \quad (6.2.7)$$

Lemma 6.2.8 $C^\infty(\Omega) \cap H^k(\Omega)$ liegt dicht in $H^k(\Omega)$.

Lemma 6.2.8, dessen Beweis beispielsweise bei Wloka [89, p. 74] zu finden ist, gestattet eine zweite Definition des Sobolev-Raumes $H^k(\Omega)$:

Bemerkung 6.2.9 Sei $X_0 := \{u \in C^\infty(\Omega) : |u|_k < \infty\}$. Die Vervollständigung von X_0 in $L^2(\Omega)$ bezüglich der Norm (6.2.7) ergibt $H^k(\Omega)$.

Definition 6.2.10 Die Vervollständigung von $C_0^\infty(\Omega)$ in $L^2(\Omega)$ bezüglich der Norm (6.2.7) wird mit $H_0^k(\Omega)$ bezeichnet.

Satz 6.2.11 Der Hilbert-Raum $H_0^k(\Omega)$ ist ein Unterraum von $H^k(\Omega)$ mit dem gleichen Skalarprodukt (6.2.6) und gleicher Norm (6.2.7). $C_0^\infty(\Omega)$ liegt dicht in $H_0^k(\Omega)$. Für $k = 0$ gilt

$$H_0^0(\Omega) = H^0(\Omega) = L^2(\Omega). \quad (6.2.8)$$

Beweis. 1) $C_0^\infty(\Omega) \subset X_0$ (vgl. Bemerkung 6.2.9) impliziert $H_0^k(\Omega) \subset H^k(\Omega) \subset L^2(\Omega)$.

2) Definitionsgemäß ist $C_0^\infty(\Omega)$ dicht in $H_0^k(\Omega)$.

3) Für $k = 0$ stimmen die Normen (6.2.2) und (6.2.7) überein. Lemma 6.2.2 beweist $H_0^0(\Omega) = L^2(\Omega)$. Wegen $H_0^k(\Omega) \subset H^k(\Omega) \subset L^2(\Omega)$ ($k \in \mathbb{N}_0$ beliebig) folgt (6.2.8). ■

Lemma 6.2.12 Für beschränktes Ω sind $\|\cdot\|_{H^k(\Omega)}$ und

$$|u|_{k,0} := \sqrt{\sum_{|\alpha|=k} \|D^\alpha u\|_{L^2(\Omega)}^2} \quad (6.2.9)$$

äquivalente Normen in $H_0^k(\Omega)$.

Beweis. 1) Offenbar ist $|\cdot|_{k,0} \leq \|\cdot\|_{H^k(\Omega)}$.

2) Sei $u \in C_0^\infty(\Omega)$. Für ein α mit $|\alpha| = k - 1$ setze man $v := D^\alpha u \in C_0^\infty(\Omega)$. Es gibt ein R mit $\Omega \subset K_R(0)$. Für jedes $x \in \Omega$ ist $x_1 \in [-R, R]$, sodass

$$|v(x)|^2 = \left| \int_{-R}^{x_1} v_{x_1}(\xi, x_2, \dots, x_n) d\xi \right|^2 \stackrel{(6.2.5b)}{\leq} (x_1 + R) \int_{-R}^{x_1} |v_{x_1}(\xi, x_2, \dots, x_n)|^2 d\xi \leq 2R \int_{-R}^R |v_{x_1}(\xi, x_2, \dots, x_n)|^2 d\xi.$$

⁵Sergei Lwowitsch Sobolew, geb. am 6. Oktober 1908 in St. Petersburg, gest. am 3. Januar 1989 in Leningrad

Integration über $x \in \Omega$ liefert $|v|_0^2 \leq 4R^2 |v_{x_1}|_0^2 \leq 4R^2 |u|_{k,0}^2$, da v_{x_1} k -fache Ableitung von u ist. Die Summation über alle α mit $|\alpha| = k-1$ ergibt $|u|_{k-1,0}^2 \leq C_{k-1} |u|_{k,0}^2$. Ebenso folgt $|u|_{j-1,0}^2 \leq C_{j-1} |u|_{j,0}^2$ für alle $1 \leq j \leq k$ und somit $|u|_{j,0}^2 \leq C_j C_{j+1} \cdots C_{k-1} |u|_{k,0}^2$. Da $|u|_k^2 \leq \sum_{j=0}^k |u|_{j,0}^2 \leq C |u|_{k,0}^2$ für alle $u \in C_0^\infty(\Omega)$, folgt die Behauptung mit Übungsaufgabe 6.1.12. ■

Die Konstante c_Ω in der sogenannten Poincaré⁶-Friedrichs⁷-Ungleichung $\|u\|_{L^2(\Omega)} \leq c_\Omega |u|_{k,0}$, die im Beweis als $\sqrt{C_j C_{j+1} \cdots C_{k-1}}$ gewählt ist, wird entsprechend Poincaré-Friedrichs-Konstante genannt.

Übungsaufgabe 6.2.13 Man zeige: a) Für beschränktes Ω und $k \geq 1$ sind $H^k(\Omega)$ und $H_0^k(\Omega)$ verschieden. Hinweis: Man betrachte die konstante Funktion $u(x) = 1$ und verwende Lemma 6.2.12.

b) Lemma 6.2.12 gilt auch, wenn Ω in einer Richtung beschränkt ist, d.h. wenn Ω in einem Streifen $\{x \in \mathbb{R}^n : |x_k| < R\}$ für ein $k \in \{1, \dots, n\}$ liegt.

Satz 6.2.14 Sei $m \geq 1$. Es gibt Konstanten $C = C(m)$ und $\eta(\varepsilon) = \eta(\varepsilon, m)$, sodass

$$|u|_k \leq C |u|_m^{k/m} |u|_0^{(m-k)/m} \quad \text{für alle } 0 \leq k \leq m, u \in H_0^m(\Omega), \quad (6.2.10a)$$

$$|u|_k \leq \varepsilon |u|_m + \eta(\varepsilon) |u|_0 \quad \text{für alle } \varepsilon > 0, 0 \leq k < m, u \in H_0^m(\Omega). \quad (6.2.10b)$$

Beweis. 1) Für die n Multiindizes α mit $|\alpha| = 1$ zeigt partielle Integration, dass $|D^\alpha u|_0^2 = (D^\alpha u, D^\alpha u)_0 = -(D^{2\alpha} u, u)_0 \leq |D^{2\alpha} u|_0 |u|_0 \leq |u|_2 |u|_0$. Da auch $|u|_0^2 \leq |u|_2 |u|_0$, folgt $|u|_1^2 \leq (n+1) |u|_2 |u|_0$. Ersetzt man u durch $D^\beta u$ mit $|\beta| = 1$, ergibt sich ebenso

$$|u|_\ell^2 \leq \tilde{C} |u|_{\ell+1} |u|_{\ell-1} \quad \text{für alle } 0 \leq \ell < m, u \in H_0^m(\Omega). \quad (6.2.10c)$$

2) Sei $u \neq 0$ fest. Setze $v_\ell := \log |u|_\ell$, $w_\ell := [\ell v_m + (m-\ell)v_0]/m$ und $z_\ell := v_\ell - w_\ell$. (6.2.10c) führt auf $2z_\ell - z_{\ell-1} - z_{\ell+1} \leq \tilde{c} := \log \tilde{C}$, wobei $z_0 = z_m = 0$. Für $z = (z_1, \dots, z_{m-1})^\top$ ergibt sich $Az \leq \tilde{c} \mathbf{1}$. Dabei ist A die M-Matrix (4.1.8b) für $h = 1$. $A^{-1} \geq 0$ beweist $z \leq c := \tilde{c} A^{-1} \mathbf{1}$. $v_\ell = z_\ell + w_\ell \leq c_\ell + w_\ell$ ($1 \leq \ell < m$) impliziert $|u|_\ell = \exp(v_\ell) \leq \exp(c_\ell + w_\ell)$, also (6.2.10a) mit $C := \exp(\tilde{c} \|A^{-1} \mathbf{1}\|_\infty)$.

3) Elementare Rechnung zeigt: Für jedes $\varepsilon > 0$, $0 \leq \Theta \leq \Theta_0 < 1$ gibt es ein $\eta(\varepsilon) = \eta(\varepsilon, \Theta_0)$, sodass

$$a^\Theta b^{1-\Theta} \leq \varepsilon a + \eta(\varepsilon) b \quad \text{für alle } a, b \geq 0. \quad (6.2.10d)$$

(6.2.10b) ist eine Folgerung aus (6.2.10a,d). ■

Mit ähnlichen Mitteln (zusammen mit (5.3.10)) beweist man

$$(D^\alpha u, D^\beta u)_0 \leq \varepsilon |u|_m^2 + \frac{1}{4\varepsilon} |u|_k^2 \quad \text{für } \varepsilon > 0, |\alpha| \leq m, |\beta| \leq k \leq m, u \in H^m(\Omega), \quad (6.2.10e)$$

$$(D^\alpha u, D^\beta u)_0 \leq \varepsilon |u|_m^2 + \eta(\varepsilon) |u|_0^2 \quad \text{für } \varepsilon > 0, |\alpha| \leq m, |\beta| < m, u \in H_0^m(\Omega), \quad (6.2.10f)$$

Bemerkung 6.2.15 Die Menge $\{u \in C^\infty(\Omega) : \text{Tr}(u) \text{ kompakt}, |u|_k < \infty\}$ liegt dicht in $H^k(\Omega)$.

Beweis. Seien $u \in H^k(\Omega)$ und $\varepsilon > 0$. Gemäß Lemma 6.2.8 gibt es ein $u_\varepsilon \in C^\infty(\Omega)$ mit $|u - u_\varepsilon|_k \leq \varepsilon/2$. Es gibt $a \in C^\infty(\mathbb{R}^n)$ mit $a(x) = 1$ für $|x| \leq 1$ und $a(x) = 0$ für $|x| \geq 2$. Für hinreichend großes R hat man $|u_\varepsilon(x) - a(x/R)u_\varepsilon(x)|_k \leq \varepsilon/2$. Damit existiert $v(x) = a(x/R)u_\varepsilon(x) \in C_0^\infty(\Omega)$ mit $|u - v|_k \leq \varepsilon$. ■

Da “ $\text{Tr}(u)$ kompakt” schon “ $\text{Tr}(u) \subset \subset \mathbb{R}^n$ ” impliziert, folgt der

Zusatz 6.2.16 $H_0^k(\mathbb{R}^n) = H^k(\mathbb{R}^n)$ für alle $k \geq 0$.

Die Leibnizsche⁸ Ableitungsregel für Produkte beweist

Satz 6.2.17 $\|au\|_{H^k(\Omega)} \leq C_k \|a\|_{C^k(\overline{\Omega})} \|u\|_{H^k(\Omega)}$ für alle $a \in C^k(\overline{\Omega})$ und $u \in H^k(\Omega)$.

Satz 6.2.17 zusammen mit der Substitutionsregel für Volumenintegrale beweist den folgenden Satz.

⁶Jules Henri Poincaré, geb. 29. April 1854 in Nancy, Lorraine, gest. 17. Juli 1912 in Paris

⁷Kurt Otto Friedrichs, geb. 28. Sept. 1901 in Kiel, gest. 31. Dez. 1982 in New Rochelle (N.Y., USA)

⁸Gottfried Wilhelm von Leibniz, geb. am 1. Juli 1646 in Leipzig, gest. am 14. Nov. 1716 in Hannover

Satz 6.2.18 (Transformationssatz) Sei $T \in C^{\max(k,1)}(\overline{\Omega}) : \Omega \rightarrow \Omega'$ bijektiv mit $|\det dT/dx| \geq \delta > 0$ in Ω . Wir schreiben $v = u \circ T$ für $v(x) = u(T(x))$. Dann impliziert $u \in H^k(\Omega')$ [$u \in H_0^k(\Omega')$] auch $u \circ T \in H^k(\Omega)$ [$u \circ T \in H_0^k(\Omega)$] und

$$\|u \circ T\|_{H^k(\Omega)} \leq C_k \|T\|_{C^k(\overline{\Omega})} \|u\|_{H^k(\Omega)} / \sqrt{\delta}. \quad (6.2.11)$$

6.2.3 Fourier-Transformation und $H^k(\mathbb{R}^n)$

Für $u \in C_0^\infty(\mathbb{R}^n)$ definiert man die Fourier-transformierte Funktion \hat{u} mittels

$$\hat{u}(\xi) := (\mathcal{F}u)(\xi) := (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{-i\langle \xi, x \rangle} u(x) dx \quad (\xi \in \mathbb{R}^n). \quad (6.2.12)$$

Man beachte, dass \hat{u} durch ein eigentliches Integral beschrieben wird, da der Träger von u beschränkt ist.

Lemma 6.2.19 Sei $u \in C_0^\infty(\mathbb{R}^n)$. Für $R \rightarrow \infty$ konvergiert

$$I_R(u; y) := (2\pi)^{-n} \int_{|\xi|_\infty \leq R} \left[\int_{\mathbb{R}^n} e^{-i\langle \xi, x-y \rangle} u(x) dx \right] d\xi$$

gleichmäßig auf $Tr(u)$ gegen $u(y)$. Genauer gilt $I_R(u; y) = u(y) + \mathcal{O}(1/R)$.

Beweis. Es genügt, den Fall $n = 1$ zu behandeln (Satz von Fubini). Integration bezüglich ξ ergibt

$$I_R(u; y) = \frac{1}{\pi} \int_{\mathbb{R}} (x-y)^{-1} \sin(R(x-y)) u(x) dx.$$

Es ist $I_R(1; y) = (1/\pi) \int_{\mathbb{R}} t^{-1} \sin(t) dt = 1$ für alle $R > 0$. Da $u \in C^\infty(\mathbb{R}^n)$, gehört auch die Funktion $w(x, y) := [u(x) - u(y)]/(x-y)$ zu $C^\infty(\mathbb{R}^{2n})$. Die Abschätzungen $w(x, y) = \mathcal{O}(1/[1 + |x-y|])$ und $w_x(x, y) = \mathcal{O}(1/[1 + |x-y|]^2)$ gelten gleichmäßig auf $y \in Tr(u)$. Partielle Integration liefert

$$\begin{aligned} I_R(u(\cdot) - u(y); y) &= I_R((\cdot - y)w(\cdot, y); y) = \frac{1}{\pi} \int_{\mathbb{R}} \sin(R(x-y)) w(x, y) dx \\ &= -\frac{1}{\pi R} \int_{\mathbb{R}} \cos(R(x-y)) w_x(x, y) dx = \mathcal{O}(1/R). \end{aligned}$$

Die Behauptung folgt aus $I_R(u; y) = u(y)I_R(1; y) + I_R(u(\cdot) - u(y); y) = u(y) + \mathcal{O}(1/R)$. ■

Lemma 6.2.20 $\hat{u} \in L^2(\mathbb{R}^n)$ und $|\hat{u}|_0 = |u|_0$ für alle $u \in C_0^\infty(\mathbb{R}^n)$.

Beweis. $\int_{|\xi|_\infty \leq R} |\hat{u}(\xi)|^2 d\xi = (2\pi)^{-n} \int_{|\xi|_\infty \leq R} \left[\int_{\mathbb{R}^n} e^{-i\langle \xi, x \rangle} u(x) dx \right] \left[\int_{\mathbb{R}^n} e^{i\langle \xi, y \rangle} \overline{u(y)} dy \right] d\xi = \int_{\mathbb{R}^n} I_R(u; y) \overline{u(y)} dy \rightarrow \int_{\mathbb{R}^n} |u(y)|^2 dy$ gemäß Lemma 6.2.19. ■

Lemma 6.2.21 Die inverse Fourier-Transformation $\mathcal{F}^{-1}\hat{u} = u$ ist für $u \in C_0^\infty(\mathbb{R}^n)$ definiert durch

$$(\mathcal{F}^{-1}\hat{u})(x) := (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{i\langle \xi, x \rangle} \hat{u}(\xi) d\xi := (2\pi)^{-n/2} \lim_{R \rightarrow \infty} \int_{|\xi|_\infty \leq R} e^{i\langle \xi, x \rangle} \hat{u}(\xi) d\xi \quad (6.2.13)$$

Beweis. $(2\pi)^{-n/2} \int_{|\xi|_\infty \leq R} e^{i\langle \xi, x \rangle} \hat{u}(\xi) d\xi = (2\pi)^{-n} \int_{|\xi|_\infty \leq R} \int_{\mathbb{R}^n} e^{i\langle \xi, x-y \rangle} u(y) dy d\xi \stackrel{\text{Lemma 6.2.19}}{=} u(x) + \mathcal{O}(1/R)$. ■

Satz 6.2.22 a) $\mathcal{F}, \mathcal{F}^{-1} \in L(L^2(\mathbb{R}^n), L^2(\mathbb{R}^n))$ mit $\|\mathcal{F}\|_{L^2(\mathbb{R}^n) \leftarrow L^2(\mathbb{R}^n)} = \|\mathcal{F}^{-1}\|_{L^2(\mathbb{R}^n) \leftarrow L^2(\mathbb{R}^n)} = 1$.

b) Das Skalarprodukt erfüllt $(u, v)_0 = (\hat{u}, \hat{v})_0$ für alle $u, v \in L^2(\mathbb{R}^n)$.

Beweis. a) Da $C_0^\infty(\mathbb{R}^n)$ dicht in $L^2(\mathbb{R}^n)$ liegt (vgl. Lemma 6.2.2), lässt sich \mathcal{F} zu $\mathcal{F} : L^2(\mathbb{R}^n) \rightarrow L^2(\mathbb{R}^n)$ fortsetzen (vgl. Satz 6.1.11). Die Normabschätzung folgt aus Lemma 6.2.20. Die Rollen von \mathcal{F} und \mathcal{F}^{-1} lassen sich vertauschen (vgl. (6.2.12) und (6.2.13)); also gilt auch $\mathcal{F}^{-1} \in L(L^2(\mathbb{R}^n), L^2(\mathbb{R}^n))$.

b) $(u, v)_0 = \frac{1}{2} (|u + v|_0^2 - |u|_0^2 - |v|_0^2) = \frac{1}{2} (|\hat{u} + \hat{v}|_0^2 - |\hat{u}|_0^2 - |\hat{v}|_0^2) = (\hat{u}, \hat{v})_0$ dank Lemma 6.2.20. ■

Übungsaufgabe 6.2.23 Man beweise: a) Mit $\xi^\alpha = \xi_1^{\alpha_1} \dots \xi_n^{\alpha_n}$ gilt

$$\mathcal{F}(D^\alpha u)(\xi) = i^{|\alpha|} \xi^\alpha \hat{u}(\xi) \quad \text{für } u \in C_0^\infty(\mathbb{R}^n). \quad (6.2.14)$$

b) Es gibt $C = C(k)$, sodass $\frac{1}{C}(1 + |\xi|^2)^k \leq \sum_{|\alpha| \leq k} |\xi^\alpha|^2 \leq C(1 + |\xi|^2)^k$ für alle $\xi \in \mathbb{R}^n$.

Lemma 6.2.24 a) Es gilt $|u|_k = \left| \sqrt{\sum_{|\alpha| \leq k} |\xi^\alpha|^2} \hat{u}(\xi) \right|_0$ für alle $u \in H^k(\mathbb{R}^n)$.

b) Eine zu $|\cdot|_k$ äquivalente Norm auf $H^k(\mathbb{R}^n)$ ist

$$|\hat{u}|_k := \left| (1 + |\xi|^2)^k \hat{u}(\xi) \right|_0. \quad (6.2.15)$$

Beweis. a) Es genügt, die Behauptung für $u \in C_0^\infty(\mathbb{R}^n)$ zu zeigen (vgl. Zusatz 6.2.16, Satz 6.1.11):

$$|u|_k^2 = \sum_{|\alpha| \leq k} |D^\alpha u|_0^2 = \sum_{|\alpha| \leq k} |\mathcal{F} D^\alpha u|_0^2 = \sum_{|\alpha| \leq k} |\xi^\alpha \hat{u}(\xi)|_0^2 = \left| \sqrt{\sum_{|\alpha| \leq k} |\xi^\alpha|^2} \hat{u}(\xi) \right|_0^2.$$

b) Die zweite Behauptung folgt aus Übungsaufgabe 6.2.23b. ■

Lemma 6.2.25 $\partial_{h,j}$ sei der Differenzenoperator

$$\partial_{h,j} u(x) := [u(x + \frac{h}{2} \mathbf{e}_j) - u(x - \frac{h}{2} \mathbf{e}_j)]/h, \quad h > 0, \quad \mathbf{e}_j : j\text{-ter Einheitsvektor}.$$

Wenn $u \in H^k(\mathbb{R}^n)$ und $|\partial_{h,j} u|_k \leq C$ für alle $h > 0$, $1 \leq j \leq n$, so gilt $u \in H^{k+1}(\mathbb{R}^n)$. Umgekehrt gilt $|\partial_{h,j} u|_k \leq |u|_{k+1}$ für alle $u \in H^{k+1}(\mathbb{R}^n)$.

Beweis. a) Aus $u(\widehat{\cdot + \delta \mathbf{e}_j}) = \exp(-i\delta \xi_j) \hat{u}(\xi)$ folgt $\widehat{\partial_{h,j} u}(\xi) = \frac{2i}{h} \sin(\xi_j h/2) \hat{u}(\xi)$. Da $4h^{-2} \sin^2(\xi_j h/2) \geq \xi_j^2$ für $h \leq 1/|\xi|$, folgt $|\widehat{\partial_{h,j} u}|^2 \geq \xi_j^2 |\hat{u}|^2$ für $|\xi| \leq 1/h$. Summation über j und Integration über ξ liefern wegen $(1 + |\xi|^2)^{k+1} = (1 + |\xi|^2)^k + (1 + |\xi|^2)^k |\xi|^2$

$$\begin{aligned} \left(|\hat{u}|_{k+1} \right)^2 &= \int_{\mathbb{R}^n} (1 + |\xi|^2)^{k+1} |\hat{u}(\xi)|^2 d\xi = \int_{\mathbb{R}^n} (1 + |\xi|^2)^k |\hat{u}(\xi)|^2 d\xi + \int_{\mathbb{R}^n} (1 + |\xi|^2)^k |\xi|^2 |\hat{u}(\xi)|^2 d\xi \\ &\leq \left(|\hat{u}|_k \right)^2 + \int_{|\xi| \geq 1/h} (1 + |\xi|^2)^k |\xi|^2 |\hat{u}(\xi)|^2 d\xi + \int_{|\xi| \leq 1/h} (1 + |\xi|^2)^k \left| \widehat{\partial_{h,j} u}(\xi) \right|^2 d\xi \\ &= \left(|\hat{u}|_k \right)^2 + \left(|\partial_{h,j} u|_k \right)^2 + \int_{|\xi| \geq 1/h} \dots \end{aligned}$$

Das Integral über $|\xi| \geq 1/h$ verschwindet für $h \rightarrow 0$, sodass die Behauptung folgt.

b) Für die Umkehrung verwende man $(\partial_{h,j} u)(x) = \int_{-1/2}^{1/2} u_{x_j}(x + t\mathbf{e}_j) dt$. ■

6.2.4 $H^s(\Omega)$ für reelles $s \geq 0$

Sei $s \geq 0$. Für $\Omega = \mathbb{R}^n$ lassen sich das folgende Skalarprodukt (6.2.16a) und die Sobolev-Norm (6.2.16b) für alle $u \in C_0^\infty(\mathbb{R}^n)$ definieren:

$$(u, v)_s := \int_{\mathbb{R}^n} (1 + |\xi|^2)^s \hat{u}(\xi) \overline{\hat{v}(\xi)} d\xi, \quad (6.2.16a)$$

$$|\hat{u}|_s := \|(1 + |\xi|^2)^{s/2} \hat{u}(\xi)\|_{L^2(\mathbb{R}^n)}. \quad (6.2.16b)$$

Die Vervollständigung in $L^2(\mathbb{R}^n)$ definiert den Sobolev-Raum $H^s(\mathbb{R}^n)$ auch für eine nichtganzzahlige Ordnung s . Auf Grund von Lemma 6.2.24b und Übungsaufgabe 6.1.12 stimmen die neu definierten $H^s(\mathbb{R}^n)$ für $s \in \mathbb{N}_0$ mit den bisherigen Sobolev-Räumen überein.

Sei $\Omega \subset \mathbb{R}^n$ ein Teilgebiet. $s \geq 0$ wird zerlegt in $s = k + \lambda$ mit $k \in \mathbb{N}_0$ und $0 < \lambda < 1$. Wir definieren

$$(u, v)_s := \sum_{|\alpha| \leq k} \left[\int_{\Omega} D^{\alpha} u(x) D^{\alpha} v(x) dx + \iint_{\Omega \times \Omega} \frac{[D^{\alpha} u(x) - D^{\alpha} u(y)][D^{\alpha} v(x) - D^{\alpha} v(y)]}{|x - y|^{n+2\lambda}} dx dy \right], \quad (6.2.17a)$$

$$\|u\|_s := \|u\|_{H^s(\Omega)} := \sqrt{(u, u)_s} \quad (s = k + \lambda, 0 < \lambda < 1). \quad (6.2.17b)$$

$|\cdot|_s$ heißt die *Sobolev-Slobodeckij-Norm*. In der gleichen Weise wie im Fall $s = k \in \mathbb{N}$ erklärt man die Hilbert-Räume $H^s(\Omega)$ und $H_0^s(\Omega)$. Die Eigenschaften dieser Räume sind im folgenden Satz zusammengefasst (vgl. Adams [1], Wloka [89]):

Satz 6.2.26 Sei $s \geq 0$. a) Für $\Omega = \mathbb{R}^n$ sind die Normen (6.2.16b) und (6.2.17b) äquivalent, d.h. beide Normen definieren den gleichen Raum $H^s(\mathbb{R}^n) = H_0^s(\mathbb{R}^n)$.

b) $\{u \in C^\infty(\Omega) : \text{Tr}(u) \text{ kompakt, } |u|_s < \infty\}$ liegt dicht in $H^s(\Omega)$.

c) $C_0^\infty(\Omega)$ liegt dicht in $H_0^s(\Omega)$.

d) $a D^{\alpha}(bu) \in H^{s-|\alpha|}(\Omega)$, falls $|\alpha| \leq s$, $u \in H^s(\Omega)$, $a \in C^{t-|\alpha|}(\overline{\Omega})$, $b \in C^t(\overline{\Omega})$, wobei $t = s \in \mathbb{N}_0$ oder $t > s$.

e) $H^s(\Omega) \subset H^t(\Omega)$, $H_0^s(\Omega) \subset H_0^t(\Omega)$ für $s \geq t$.

f) In (6.2.10a,b) können k und m reell sein.

g) In (6.2.11) ist $\|T\|_{C^k(\overline{\Omega})}$ durch $\|T\|_{C^t(\overline{\Omega})}$ mit $t > k$ zu ersetzen, falls $k \notin \mathbb{N}$.

Übungsaufgabe 6.2.27 Man prüfe in beiden Normen (6.2.16b) und (6.2.17b) nach, dass die charakteristische Funktion $u(x) = 1$ in $[-1, 1]$, $u(x) = 0$ für $|x| > 1$ genau dann zu $H^s(\mathbb{R})$ gehört, wenn $0 \leq s < 1/2$.

6.2.5 Spur- und Fortsetzungssätze

Die Natur der Randwertaufgaben verlangt, dass man sinnvoll Randwerte $u|_{\Gamma}$ (Beschränkung oder sogenannte *Spur* von u auf $\Gamma = \partial\Omega$) bilden kann. Eine Hölder-stetige Funktion $u \in C^s(\overline{\Omega})$ hat, wie man leicht einsieht, eine Beschränkung $u|_{\Gamma} \in C^s(\Gamma)$, wenn nur Γ hinreichend glatt ist. Aus $u \in H^s(\Gamma)$ folgt aber keineswegs $u|_{\Gamma} \in H^s(\Gamma)$. Da die Gleichheit $u = v$ in $H^s(\Omega)$ nur bedeutet, dass $u(x) = v(x)$ fast überall in Ω , aber Γ eine Menge vom Maß null ist, kann $u(x) \neq v(x)$ überall auf Γ gelten. Der Randwert $u(x)$ ($x \in \Gamma$) kann auch nicht durch stetige Fortsetzung definiert werden, da z.B. $u \in H^1(\Omega)$ nicht stetig zu sein braucht (vgl. Übungsaufgabe 6.2.4c).

Das inverse Problem zur Definition von $u|_{\Gamma}$ ist die Fortsetzung: Existiert zu einem gegebenen Randwert φ auf Γ eine Funktion $u \in H^s(\Omega)$, sodass $\varphi = u|_{\Gamma}$? Im negativen Falle kann es auch keine Lösung $u \in H^s(\Omega)$ des Dirichlet-Randwertproblems geben. Wir werden diese Fragestellungen zunächst an dem Halbraum

$$\Omega = \mathbb{R}_+^n := \{(x_1, \dots, x_n) \in \mathbb{R}^n : x_n > 0\} \quad \text{mit } \Gamma = \partial\Omega = \mathbb{R}^{n-1} \times \{0\} \quad (6.2.18)$$

studieren. $u \in H^s(\mathbb{R}_+^n)$ wird zunächst zu $\bar{u} \in H^s(\mathbb{R}^n)$ fortgesetzt, bevor \bar{u} auf Γ beschränkt wird.

Satz 6.2.28 Sei $s \geq 0$. Es gibt einen Fortsetzungsoperator $\phi_s \in L(H^s(\mathbb{R}_+^n), H^s(\mathbb{R}^n))$, sodass für alle $u \in H^s(\mathbb{R}_+^n)$ die Fortsetzung $\bar{u} = \phi_s u$ mit u auf \mathbb{R}_+^n übereinstimmt: $\bar{u}|_{\mathbb{R}_+^n} = u$.

Beweis. Für $s = 0$ setze man $\bar{u} = u$ auf \mathbb{R}_+^n und $\bar{u} = 0$ sonst. Da $\|\bar{u}\|_{L^2(\mathbb{R}^n)} = \|u\|_{L^2(\mathbb{R}_+^n)}$, ist die durch $\phi_0 u := \bar{u}$ definierte Abbildung beschränkt: $\|\phi_0\|_{L^2(\mathbb{R}^n) \leftarrow L^2(\mathbb{R}_+^n)} = 1$. Für $s \leq 1$ erkläre man $\phi_s u = \bar{u}$ durch

$$\bar{u} = u \text{ auf } \mathbb{R}_+^n, \quad \bar{u}(x_1, \dots, x_{n-1}, -x_n) = \bar{u}(x_1, \dots, x_{n-1}, x_n) \text{ für } x_n > 0$$

(d.h. durch Spiegelung an Γ). Für $s = 1$ ergibt sich z.B. $|\bar{u}|_1 = \sqrt{2}|u|_1$, d.h. $\phi_s \in L(H^s(\mathbb{R}_+^n), H^s(\mathbb{R}^n))$. Für größere s benutzt man für $\bar{u}(\dots, -x_n)$ höhere Interpolationsformeln (vgl. Übungsaufgabe 9.1.17, vgl. Wloka [89, p. 105]). ■

Die Beschränkung $u|_{\Gamma}$ wird im Folgenden als γu notiert. Zunächst wird γ nur auf $C_0^\infty(\mathbb{R}^n)$ definiert:

$$\gamma : C_0^\infty(\mathbb{R}^n) \rightarrow C_0^\infty(\Gamma) \subset L^2(\mathbb{R}^{n-1}), \quad (\gamma u)(x) := u(x) \quad \text{für alle } x \in \Gamma. \quad (6.2.19)$$

Wir schreiben $x = (x', x_n)$ mit $x' = (x_1, \dots, x_{n-1}) \in \mathbb{R}^{n-1}$. Der Rand Γ wird mit \mathbb{R}^{n-1} identifiziert, d.h. $H^s(\Gamma) = H^s(\mathbb{R}^{n-1})$.

Satz 6.2.29 Sei $s > 1/2$. γ aus (6.2.19) ist fortsetzbar zu $\gamma \in L(H^s(\mathbb{R}^n), H^{s-1/2}(\mathbb{R}^{n-1}))$. Insbesondere gilt $|\gamma u|_{s-1/2} \leq C_s |u|_s$ für $u \in H^s(\mathbb{R}^n)$. Im Falle $n = 1$ (d.h. $\gamma u = u(0)$) gilt $|\gamma u| \leq C_s |u|_s$.

Beweis. Es genügt $|\gamma u|_{s-1/2} \leq C'_s |u|_s$ für $u \in C_0^\infty(\mathbb{R}^n)$ zu zeigen (vgl. Satz 6.1.11 und Satz 6.2.26a). Die Fourier-Transformierten zu $u \in C_0^\infty(\mathbb{R}^n)$ und $w := \gamma u \in C_0^\infty(\mathbb{R}^{n-1})$ seien $\hat{u} = \mathcal{F}_n u$ und $\hat{w} = \mathcal{F}_{n-1} w$ (\mathcal{F}_k : k -dimensionale Fourier-Transformation). \mathcal{F}_n kann als Produkt $\mathcal{F}_1 \circ \mathcal{F}_{n-1}$ geschrieben werden, wobei \mathcal{F}_{n-1} auf $x' \in \mathbb{R}^{n-1}$ und \mathcal{F}_1 auf x_n wirkt. Daher hat $\hat{W}(\cdot, x_n) := \mathcal{F}_{n-1} u(\cdot, x_n)$ die Eigenschaften $\hat{u}(\xi', \cdot) = \mathcal{F}_1 \hat{W}(\xi', \cdot)$ und $\hat{w} = \hat{W}(\cdot, 0)$. Laut Lemma 6.2.21 hat $\hat{w}(\xi') = \hat{W}(\xi', 0) = [\mathcal{F}_1^{-1} \hat{u}(\xi', \cdot)]|_{x_n=0}$ die Darstellung

$$\hat{w}(\xi') = (2\pi)^{-1/2} \int_{\mathbb{R}^n} \hat{u}(\xi', \xi_n) d\xi_n \quad \text{für alle } \xi' \in \mathbb{R}^{n-1}. \quad (6.2.20)$$

Da $u \in H^s(\mathbb{R}^n)$, liegt $\hat{U}(\xi', \xi_n) := (1 + |\xi'|^2 + \xi_n^2)^{s/2} \hat{u}(\xi', \xi_n)$ in $L^2(\mathbb{R}^n)$ (vgl. Lemma 6.2.24b). Ungleichung (6.2.5a) liefert

$$2\pi |\hat{w}(\xi')|^2 = \left| \int_{\mathbb{R}^n} \hat{u}(\xi', \xi_n) d\xi_n \right|^2 \leq \int_{\mathbb{R}^n} (1 + |\xi'|^2 + \xi_n^2)^s d\xi_n \int_{\mathbb{R}^n} |\hat{U}(\xi', \xi_n)|^2 d\xi_n.$$

Der erste Faktor errechnet sich zu $K_s(1 + |\xi'|^2)^{1/2-s}$ mit $K_s = \int_{\mathbb{R}} (1 + x^2)^{-s} dx < \infty$, da $s > 1/2$. Der zweite Faktor lautet $V(\xi') := \|\hat{U}(\xi', \cdot)\|_{L^2(\mathbb{R})}^2 \in L^2(\mathbb{R}^{n-1})$, weil $\|V\|_{L^2(\mathbb{R}^{n-1})} = \|\hat{U}\|_{L^2(\mathbb{R}^n)}$ (Satz von Fubini). Zusammen:

$$(1 + |\xi'|^2)^{s-1/2} |\hat{w}(\xi')|^2 \leq \frac{K_s}{2\pi} V(\xi')^2 \quad \text{für } \xi' \in \mathbb{R}^{n-1}.$$

Integration über $\xi' \in \mathbb{R}^{n-1}$ ergibt

$$\int_{\mathbb{R}^{n-1}} (1 + |\xi'|^2)^{s-1/2} |\hat{w}(\xi')|^2 d\xi' \leq \frac{K_s}{2\pi} |V|_0^2 = \frac{K_s}{2\pi} |\hat{U}|_0^2 = \frac{K_s}{2\pi} \int_{\mathbb{R}^n} (1 + |\xi|^2)^s |\hat{u}(\xi)|^2 d\xi.$$

Somit ist $|w|_{s-1/2} \leq C'_s |u|_s$ mit $C'_s = (K_s/(2\pi))^{1/2}$ bewiesen (vgl. (6.2.16b)). Im Falle $n = 1$ stellt $\hat{w}(\xi')$ schon $\gamma u = u(0)$ dar, und die Integration über ξ' entfällt. ■

Satz 6.2.29 beschreibt die Beschränkung $u(\cdot, 0) = \gamma u$ auf $x_n = 0$. Selbstverständlich gilt ebenso $|u(\cdot, x_n)|_{s-1/2} \leq C_s |u|_s$ für jedes andere $x_n \in \mathbb{R}$ mit der gleichen Konstanten C_s . Die Abbildung $x_n \mapsto u(\cdot, x_n)$ ist im folgenden Sinne stetig bzw. Hölder-stetig.

Satz 6.2.30 Für $s > 1/2$ gilt

$$\lim_{y_n \rightarrow x_n} \|u(\cdot, x_n) - u(\cdot, y_n)\|_{H^{s-1/2}(\mathbb{R}^{n-1})} = 0 \quad \text{für alle } x_n \in \mathbb{R}, u \in H^s(\mathbb{R}^n), \quad (6.2.21a)$$

$$\|u(\cdot, x_n) - u(\cdot, y_n)\|_{H^{s-1/2}(\mathbb{R}^{n-1})} \leq K_{s,\lambda} |x_n - y_n|^\lambda \|u\|_{H^s(\mathbb{R}^n)} \quad (6.2.21b)$$

für $u \in H^s(\mathbb{R}^n)$, $0 \leq \lambda < 1$ (und für $\lambda = 1$, falls $s > 3/2$).

Beweis. 1) Sei $u_\nu \in C_0^\infty(\mathbb{R}^n)$ eine Folge mit $u_\nu \rightarrow u \in H^s(\mathbb{R}^n)$ und setze $\varphi_\nu(x) := \|u_\nu(\cdot, x)\|_{H^{s-1/2}(\mathbb{R}^{n-1})}$. Die Funktion φ_ν ist stetig in \mathbb{R} und konvergiert gleichmäßig gegen $\|u(\cdot, x)\|_{H^{s-1/2}(\mathbb{R}^{n-1})}$, da $|u_\nu(\cdot, x) - u(\cdot, x)|_{s-1/2} \leq C_s |u_\nu - u|_s$ für alle $x \in \mathbb{R}$. Also folgt (6.2.21a).

2) $u_\varepsilon(\cdot, x_n) := u(\cdot, x_n + \varepsilon) - u(\cdot, x_n)$ hat die Fourier-Transformierte $\hat{u}_\varepsilon(\xi) = \hat{u}_\varepsilon(\xi', \xi_n) = [\exp(i\xi_n \varepsilon) - 1] \hat{u}(\xi)$, sodass $|\hat{u}_\varepsilon(\xi)|^2 = 4 \sin^2(\xi_n \varepsilon/2) |\hat{u}(\xi)|^2$. Wie im Beweis zu Satz 6.2.29 setze man $\hat{W}(\cdot, x_n) := \mathcal{F}_{n-1} u_\varepsilon(\cdot, x_n)$ und $\hat{w} = \hat{W}(\cdot, 0)$. Das erste Integral in der Abschätzung von $2\pi |\hat{w}(\xi')|^2$ lautet jetzt

$$\int_{\mathbb{R}^n} (1 + |\xi'|^2 + \xi_n^2)^{-s} \sin^2(\xi_n \varepsilon/2) d\xi_n = (1 + |\xi'|^2 + \xi_n^2)^{1/2-s} \int_{\mathbb{R}^n} (1 + t^2) \sin^2(\eta t) dt$$

mit $\eta = \varepsilon/2(1 + |\xi'|^2)^{1/2}$. Die Zerlegung des letzten Integrals in die Teilintegrale über $|t| \leq 1/\eta$ und $|t| \geq 1/\eta$ zeigt $\int_{\mathbb{R}} (1 + t^2)^{-s} \sin^2(\eta t) dt \leq C_{s,\lambda} \eta^{2\lambda}$. Die weitere Argumentation verläuft wie im Beweis zu Satz 6.2.29. ■

Bisher wurde $H^s(\mathbb{R}^n)$ durch Vervollständigung von $C_0^\infty(\mathbb{R}^n)$ in $L^2(\mathbb{R}^n)$ gewonnen. Der nächste Satz zeigt, dass man für hinreichend großes s auch in $C^0(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ vervollständigen kann, sodass $H^s(\mathbb{R}^n)$ nur klassische Funktionen (d.h. stetige, Hölder-stetige, [Hölder]-stetig differenzierbare Funktionen) enthält.

Satz 6.2.31 (Sobolevsche Einbettung) Es gilt $H^s(\mathbb{R}^n) \subset C^k(\mathbb{R}^n)$ für $k \in \mathbb{N}_0$, $s > k + n/2$ und $H^s(\mathbb{R}^n) \subset C^t(\mathbb{R}^n)$ für $0 < t \notin \mathbb{N}$, $s \geq t + n/2$.

Beweis. 1) Sei $s \geq t + n/2$, $0 < t < 1$, $u \in H^s(\mathbb{R}^n)$. Für vorgegebene $x, y \in \mathbb{R}^n$ soll $|u(x) - u(y)| \leq C|x - y|^t$ gezeigt werden (C unabhängig von x, y). Die Koordinaten des \mathbb{R}^n können so gedreht werden, dass $x = (x_1, 0, \dots, 0)$, $y = (y_1, 0, \dots, 0)$. $(n-1)$ -fache Anwendung des Satzes 6.2.29 auf $u(\cdot)$, $u(\cdot, 0)$, $u(\cdot, 0, 0)$ usw. ergibt $u(\cdot, 0, 0, \dots, 0) \in H^{s-(n-1)/2}(\mathbb{R})$. Satz 6.2.30 liefert die gewünschte Abschätzung mit $C = K_{s-(n-1)/2, t} |u(\cdot, 0, \dots, 0)|_{s-(n-1)/2} \leq C' |u|_s$, also $u \in C^t(\mathbb{R}^n)$ und $\|\cdot\|_{C^t(\mathbb{R}^n)} \leq C \|\cdot\|_{H^s(\mathbb{R}^n)}$. Die Behauptung $u \in C^0(\mathbb{R}^n)$ folgt aus $C^t(\mathbb{R}^n) \subset C^0(\mathbb{R}^n)$.

2) Sei $s \geq t + n/2$, $1 < t < 2$ und $u \in H^s(\mathbb{R}^n)$. Auf $D^\alpha u \in H^{s-1}(\mathbb{R}^n)$ (vgl. Satz 6.2.26d,e) mit $|\alpha| \leq 1$ ist Teil 1) anwendbar: $D^\alpha u \in C^{t-1}(\mathbb{R}^n)$ für $|\alpha| \leq 1$. Also $u \in C^t(\mathbb{R}^n)$. Usw. ■

Wir kehren zu den Aussagen der Sätze 6.2.28, 6.2.29 zurück. Für alle $u \in C_0^\infty(\mathbb{R}_+^n)$ stimmt die Beschränkung $\gamma u = u(\cdot, 0)$ mit $\gamma \phi_s u$ überein. Vervollständigung in $H^s(\mathbb{R}_+^n)$ liefert

$$\|\gamma\|_{H^{s-1/2}(\mathbb{R}^{n-1}) \leftarrow H^s(\mathbb{R}_+^n)} \leq \|\gamma\|_{H^{s-1/2}(\mathbb{R}^{n-1}) \leftarrow H^s(\mathbb{R}^n)} \|\phi_s\|_{H^s(\mathbb{R}^n) \leftarrow H^s(\mathbb{R}_+^n)}.$$

Dies beweist das

Zusatz 6.2.32 Sei $s > 1/2$. Für $\gamma u := u(\cdot, 0)$ gilt $\gamma \in L(H^s(\mathbb{R}_+^n), H^{s-1/2}(\mathbb{R}^{n-1}))$.

Bei der Beschränkung auf $x_n = 0$ verliert man offenbar eine halbe Differentiationsordnung. Umgekehrt gewinnt man eine halbe Ordnung, wenn man $w \in H^{s-1/2}(\mathbb{R}^{n-1})$ geeignet in \mathbb{R}^n fortsetzt.

Satz 6.2.33 Seien $s > 1/2$ und $w \in H^{s-1/2}(\mathbb{R}^{n-1})$. Es gibt ein $u \in H^s(\mathbb{R}^n)$ [$u \in H^s(\mathbb{R}_+^n)$], sodass $|u|_s \leq C_s |w|_{s-1/2}$ und $\gamma u = w$, d.h. $w = u(\cdot, 0)$.

Beweis. Seien $\hat{u} = \mathcal{F}_n u$ und $\hat{w} = \mathcal{F}_{n-1} w$ die Fourier-Transformierten. $\gamma u = w$ ist äquivalent zu (6.2.20). Für

$$\hat{u}(\xi) = \hat{u}(\xi', \xi_n) := \hat{w}(\xi') (1 + |\xi'|^2)^{s-1/2} (1 + |\xi'|^2 + \xi_n^2)^{-s} / K_s, \quad K_s = \int_{\mathbb{R}^n} (1 + t^2)^{-s} dt,$$

prüft man nach, dass (6.2.20) und $|u|_s^\wedge = K_s^{-1/2} |w|_{s-1/2}^\wedge$ gelten. Beschränkung von $u \in H^s(\mathbb{R}^n)$ auf \mathbb{R}_+^n beweist den Klammerzusatz. ■

Ersetzt man \mathbb{R}_+^n durch ein allgemeines Gebiet $\Omega \subset \mathbb{R}^n$, wird $\mathbb{R}^{n-1} \cong \mathbb{R}^{n-1} \times \{0\} = \partial \mathbb{R}_+^n$ zu $\Gamma = \partial \Omega$, und es ergibt sich die Notwendigkeit, den Sobolev-Raum $H^s(\Gamma)$ zu definieren. Wir beginnen mit der

Definition 6.2.34 Sei $0 < t \in \mathbb{R} \cup \{\infty\}$ [bzw. $k \in \mathbb{N}_0$]. Wir schreiben $\Omega \in C^t$ [bzw. $\Omega \in C^{k,1}$], falls zu jedem $x \in \Gamma := \partial \Omega$ eine Umgebung $U \subset \mathbb{R}^n$ existiert, sodass es eine bijektive Abbildung $\phi : U \rightarrow K_1(0) = \{\xi \in \mathbb{R}^n : |\xi| < 1\}$ gibt mit

$$\phi \in C^t(\bar{U}), \quad \phi^{-1} \in C^t(\overline{K_1(0)}) \quad [\phi \in C^{k,1}(\bar{U}), \quad \phi^{-1} \in C^{k,1}(\overline{K_1(0)})], \quad (6.2.22a)$$

$$\phi(U \cap \Gamma) = \{\xi \in K_1(0) : \xi_n = 0\}, \quad (6.2.22b)$$

$$\phi(U \cap \Omega) = \{\xi \in K_1(0) : \xi_n > 0\}, \quad (6.2.22c)$$

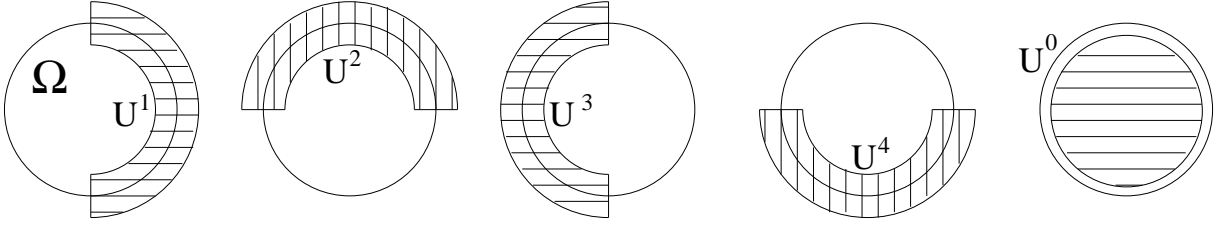
$$\phi(U \cap (\mathbb{R}^n \setminus \Omega)) = \{\xi \in K_1(0) : \xi_n < 0\}. \quad (6.2.22d)$$

$K_1(0)$ ist ein Kreis (Kugel), wenn $|\cdot|$ die Euklidische Norm ist. Für die Maximumnorm $|\cdot|_\infty$ ist $K_1(0)$ ein Quadrat (Quader). Ebenso gut kann $K_1(0)$ durch jeden anderen Kreis $K_R(z)$ oder jeden anderen Quader $(x'_1, x''_2) \times \dots \times (x'_n, x''_n)$ ersetzt werden.

Beispiel 6.2.35 Ω sei der Kreis $K_1(0) \subset \mathbb{R}^2$. Eine Umgebung von $x^* = (1, 0)$ ist U^1 aus Abbildung 6.2.1. Die Abbildung $x \in U^1 \mapsto \phi(x) := \xi \in (-1, 1) \times (-1, 1)$ mit $x_1 = (1 - \xi_2/2) \cos(\pi \xi_1/2)$, $x_2 = (1 - \xi_2/2) \sin(\pi \xi_1/2)$ ist bijektiv und erfüllt (6.2.22a-d) mit $t = \infty$. Entsprechendes gilt für jedes $x \in \Gamma$. Damit ist $\Omega \in C^\infty$.

Übungsaufgabe 6.2.36 a) Das Rechteck $\Omega = (x'_1, x''_1) \times (x'_2, x''_2)$ sowie das L -Gebiet aus Beispiel 2.1.4 sind Gebiete $\Omega \in C^{0,1}$.

b) Der geschlitzte Kreis aus Abbildung 5.2.1b gehört nicht zu $C^{0,1}$.

Abb. 6.2.1. Überdeckende Umgebungen von Γ und Ω

Lemma 6.2.37 $\Omega \in C^t$ [$\Omega \in C^{k,1}$] sei ein beschränktes Gebiet. Dann gibt es $N \in \mathbb{N}$, U^i ($0 \leq i \leq N$), U_i , α_i ($1 \leq i \leq N$) mit

$$U^i \text{ offen, beschränkt } (0 \leq i \leq N), \quad \bigcup_{i=0}^N U^i \supset \bar{\Omega}, \quad U^0 \subset\subset \Omega, \quad (6.2.23a)$$

$$U_i := U^i \cap \Gamma \quad (1 \leq i \leq N), \quad \bigcup_{i=0}^N U_i = \Gamma, \quad (6.2.23b)$$

$$\alpha_i : U_i \rightarrow \alpha_i(U_i) \subset \mathbb{R}^{n-1} \text{ bijektiv für alle } i = 1, \dots, N, \quad (6.2.23c)$$

$$\alpha_i \circ \alpha_j^{-1} \in C^t(\overline{\alpha_j(U_i \cap U_j)}) \quad \text{bzw.} \quad \alpha_i \circ \alpha_j^{-1} \in C^{k,1}(\overline{\alpha_j(U_i \cap U_j)}). \quad (6.2.23d)$$

Auf U^i ($1 \leq i \leq N$) sind Abbildungen ϕ_i mit den Eigenschaften (6.2.22a-d) definiert.

Beweis. Zu jedem $x \in \Gamma$ existieren $U = U(x)$ und $\phi = \phi_x$ gemäß Definition 6.2.34. α_x sei die Beschränkung von ϕ_x auf $U(x) \cap \Gamma$. V^i ($i \in \mathbb{N}$) seien die offenen Mengen $\{x \in \Omega : \text{dist}(x, \Gamma) > 1/i\} \subset\subset \Omega$. $\bigcup_{x \in \Gamma} U(x) \cup \bigcup_i V^i$ ist eine offene Überdeckung der kompakten Menge $\bar{\Omega}$. Daher existiert eine endliche Überdeckung durch $U^i := U(x_i)$ ($1 \leq i \leq N$) und höchstens ein V^j , das mit U^0 bezeichnet wird. Setzt man $U_i := U^i \cap \Gamma$, $\alpha_i = \alpha_{x_i}$, $\phi_i = \phi_{x_i}$, folgen die Aussagen aus (6.2.23a-d). ■

Eine Paarmenge $\{(U_i, \alpha_i) : 1 \leq i \leq N\}$ mit (6.2.23b-d) heißt C^t - bzw. $C^{k,1}$ -Koordinatensystem von Γ .

In Beispiel 6.2.35 hat man $N = 4$. α_1^{-1} ist durch $\alpha_1^{-1}(\xi_1) = (\cos(\pi\xi_1/2), \sin(\pi\xi_1/2)) \in U_1$ beschrieben; ebenso $\alpha_4^{-1}(\xi_1) = (\cos((3+\xi_1)\pi/2), \sin((3+\xi_1)\pi/2)) \in U_4$, wobei jeweils $-1 < \xi_1 < 1$. Auf $\alpha_1(U_1 \cap U_4)$ ergibt sich $\alpha_4(\alpha_1^{-1}(\xi_1)) = \xi_1 + 1$.

Lemma 6.2.38 (Partition der Eins) $\{U^i : 0 \leq i \leq N\}$ erfülle (6.2.23a). Es existieren Funktionen $\sigma_i \in C_0^\infty(\mathbb{R}^n)$, $0 \leq i \leq N$, mit

$$\text{Tr}(\sigma_i) \subset U^i, \quad \sum_{i=0}^N \sigma_i^2(x) = 1 \quad \text{für alle } x \in \bar{\Omega}. \quad (6.2.24)$$

Die allgemeine Konstruktion der σ_i findet sich z.B. bei Wloka [89, p. 14]. Im speziellen Fall von Abbildung 6.2.1 kann man wie folgt vorgehen. Sei $\sigma(t) := 0$ für $|t| \geq 1$ und $\sigma(t) := \exp(1/(t^2 - 1))$ für $t \in (-1, 1)$. Es gilt $\sigma \in C_0^\infty(\mathbb{R})$ und $\text{Tr}(\sigma) = [-1, 1]$. In U^i aus Abbildung 6.2.1 definiere man z.B.

$$\psi_0(x) := \sigma(9|x|^2/4), \quad \psi_1(x) := \sigma(2r-2)\sigma(2\varphi/\pi) \quad \text{für } x = r \begin{pmatrix} \cos \varphi \\ \sin \varphi \end{pmatrix}, \quad \text{etc.}$$

$\sigma_i(x) := \psi_i(x)/\sqrt{\sum \psi_i^2(x)}$ erfüllt (6.2.24).

Eine Funktion u auf Γ kann in der Form $\sum \sigma_i^2 u$ geschrieben werden. Jeder Summand $\sigma_i^2 u$ ist über $\alpha_i(U_i) \subset \mathbb{R}^{n-1}$ parametrisierbar: $(\sigma_i^2 u) \circ \alpha_i^{-1} : \alpha_i(U_i) \subset\subset \mathbb{R}^{n-1} \rightarrow \mathbb{R}$. Dies ermöglicht die

Definition 6.2.39 Sei $\Omega \in C^t$ [$\in C^{k,1}$]. (U_i, α_i) und σ_i erfüllen (6.2.23b-d) und (6.2.24). Sei $s \leq t \in \mathbb{N}$ [$s \leq k+1$] oder $s < t \notin \mathbb{N}$, $t > 1$. Der Sobolev-Raum $H^s(\Gamma)$ sei die Menge aller Funktionen $u : \Gamma \rightarrow \mathbb{R}$, sodass $(\sigma_i u) \circ \alpha_i^{-1} \in H_0^s(\mathbb{R}^{n-1})$ ($1 \leq i \leq N$).

Satz 6.2.40 a) $H^s(\Gamma)$ ist ein Hilbert-Raum mit dem Skalarprodukt

$$(u, v)_s := (u, v)_{H^s(\Gamma)} := \sum_{i=1}^N ((\sigma_i u) \circ \alpha_i^{-1}, (\sigma_i v) \circ \alpha_i^{-1})_{H^s(\mathbb{R}^{n-1})}.$$

b) Ist $\{(\tilde{U}_i, \tilde{\alpha}_i) : 1 \leq i \leq N\}$ ein anderes C^t - $[C^{k,1}]$ -Koordinatensystem von Γ und $\{\tilde{\sigma}_i\}$ eine andere Zerlegung der Eins, so ist der hierdurch definierte Raum $\tilde{H}^s(\Gamma)$ mengenmäßig mit $H^s(\Gamma)$ gleich. Die Normen von $H^s(\Gamma)$ und $\tilde{H}^s(\Gamma)$ sind äquivalent.

Beweis zu b). Man verwende Transformationssatz 6.2.18 bzw. Satz 6.2.26g. Für $\Omega \in C^{0,1}$ vgl. Wloka [89, p. 90]. ■

Die Spur- und Fortsetzungssätze (Zusatz 6.2.32 und Satz 6.2.33) lassen sich auf ein hinreichend glattberandetes Gebiet übertragen. γ bezeichnet die Beschränkung auf $\Gamma = \partial\Omega : \gamma u = u|_\Gamma$.

Satz 6.2.41 Sei $\Omega \in C^t$ mit $1/2 < s < t \in \mathbb{N}$ oder $1/2 < s < t$ [bzw. $\Omega \in C^{k,1}$, $1/2 < s = k + 1 \in \mathbb{N}$].

a) Die Beschränkung γu von $u \in H^s(\Omega)$ gehört zu $H^{s-1/2}(\Gamma)$: $\gamma \in L(H^s(\Omega), H^{s-1/2}(\Gamma))$.

b) Zu jedem $w \in H^{s-1/2}(\Gamma)$ existiert ein $u \in H^s(\Omega)$ mit $w = \gamma u$, $|u|_s \leq C_s |w|_{s-1/2}$.

c) Zu jedem $w \in H^s(\Omega)$ existiert eine Fortsetzung $Ew \in H^s(\mathbb{R}^n)$ mit $E \in L(H^s(\Omega), H^s(\mathbb{R}^n))$.

Die Beweise folgen jeweils dem gleichen Muster. U^i, U_i und ϕ_i, α_i seien wie in Lemma 6.2.37. Der Summand $u_i = \sigma_i^2 u$ aus $u = \sum \sigma_i^2 u$ (bzw. $w_i = \sigma_i^2 w$) hat seinen Träger in U^i (bzw. U_i) und kann mittels ϕ_i (bzw. α_i) auf \mathbb{R}_+^n (bzw. $\mathbb{R}^{n-1} \cong \mathbb{R}^{n-1} \times \{0\} = \partial\mathbb{R}^n$) abgebildet werden. Dort gelten Zusatz 6.2.32 und Satz 6.2.33. Die Beschränkung auf \mathbb{R}^{n-1} (bzw. Fortsetzung auf \mathbb{R}_+^n oder \mathbb{R}^n) kann zurück abgebildet werden. Im Detail wird die erste Aussage des Satzes bewiesen.

Seien $u_i := \sigma_i^2 u$ und $\tilde{u}_i := u_i \circ \phi_i^{-1}$. Auf Grund der Sätze 6.2.18, 6.2.26g gehört \tilde{u}_i zu $H^s(\mathbb{R}_+^n)$. Damit liegt die Beschränkung $\gamma_+ \tilde{u}_i := u_i(\cdot, 0)$ in $H^{s-1/2}(\mathbb{R}^{n-1})$ (vgl. Zusatz 6.2.32) und hat $\alpha_i(U_i)$ als Träger. Setze $w_i := (\gamma_+ \tilde{u}_i) \circ \alpha_i$ auf U_i , $w_i := 0$ auf $\Gamma \setminus U_i$. Gemäß Definition 6.2.39 gehört $w := \sum w_i$ zu $H^{s-1/2}(\Gamma)$. Da α_i die Beschränkung von ϕ_i auf U_i darstellt, findet man für alle $u \in C^t(\overline{\Omega})$ ($C^t(\overline{\Omega})$ ist dicht in $H^s(\Omega)$):

$$w(x) = \sum_i w_i(x) = \sum_i (\gamma_+ \tilde{u}_i(\alpha_i(x))) = \sum_i \tilde{u}_i(\phi_i(x)) = \sum_i (\sigma_i^2 u)(x) = u(x) \quad \text{für } x \in \Gamma.$$

Da alle Teilabbildungen beschränkt sind, findet man $|\gamma u|_{s-1/2} = |w|_{s-1/2} \leq C_s |u|_s$.

Bemerkung 6.2.42 a) Unter der Voraussetzung des Satzes 6.2.41 und der zusätzlichen Bedingung $s > |\alpha| + 1/2$ existiert die Beschränkung $\gamma D^\alpha u \in H^{s-|\alpha|-1/2}(\Gamma)$ der Ableitung von $u \in H^s(\Omega)$.

b) Für jedes $u \in H_0^s(\Omega)$ mit $s < \ell + 1/2$ gilt $\gamma D^\alpha u = 0$, wenn $|\alpha| \leq \ell$.

Satz 6.2.43 Für $\Omega \in C^{0,1}$ gilt $H_0^1(\Omega) = \{u \in H^1(\Omega) : u|_\Gamma = 0\}$.

Beweis. a) $u \in H_0^1(\Omega)$ ist Limes von $u_\nu \in C_0^\infty(\Omega)$, die sämtlich $u_\nu|_\Gamma = 0$ erfüllen. Da $u_\nu|_\Gamma \rightarrow u|_\Gamma$ bezüglich $H^{1/2}(\Gamma)$, folgt $u \in H^1(\Omega)$ und $u|_\Gamma = 0$.

b) Sei umgekehrt $u \in H^1(\Omega)$ und $u|_\Gamma = 0$. Der Beweis von $u \in H_0^1(\Omega)$ lässt sich wie folgt gliedern :

ba) Indem man von der Partition der Eins über $\{U^i : 0 \leq i \leq N\}$ (vgl. Lemma 6.2.38) Gebrauch macht, reduziert sich die Behauptung auf den Fall $\Omega = \mathbb{R}_+^n$.

bb) Ohne Beschränkung der Allgemeinheit lässt sich $n = 1$ annehmen: $\Omega = \mathbb{R}_+$.

bc) Nach Bemerkung 6.2.15 gibt es zu jedem $\varepsilon > 0$ ein $\tilde{u}_\varepsilon \in C^\infty(\mathbb{R}_+)$ mit endlichem Träger, sodass $|u - \tilde{u}_\varepsilon|_1 \leq \eta/2$. Zu jedem $\eta > 0$ existiert $\varphi_\eta \in C^\infty(\mathbb{R}_+)$ mit

$$\varphi_\eta(x) = 0 \quad \text{für } x \leq \eta/2, \quad \varphi_\eta(x) = 1 \quad \text{für } x \geq \eta, \quad |\varphi'_\eta(x)| \leq 3/\eta, \quad 0 \leq \varphi_\eta(x) \leq 1.$$

$u_\varepsilon := \tilde{u}_\varepsilon - (1 - \varphi_1)\tilde{u}_\varepsilon(0)$ erfüllt 1) $u_\varepsilon \in C^\infty(\mathbb{R}_+)$, 2) $u_\varepsilon(0) = 0$, 3) $Tr(u_\varepsilon)$ endlich und 4) $|u - u_\varepsilon|_1 \leq C_\varepsilon$. Damit liegt $X := \{v \in C^\infty(\mathbb{R}_+) : v(0) = 0, Tr(v) \text{ endlich}\}$ dicht in $\{u \in H^1(\mathbb{R}_+) : u(0) = 0\}$.

bd) Die Behauptung wäre bewiesen, wenn auch $C_0^\infty(\mathbb{R}_+)$ bezüglich $|\cdot|_1$ dicht in X liegt. Sei $v \in X$. Offenbar gilt $v_\eta := \varphi_\eta v \in C_0^\infty(\mathbb{R}_+)$ für alle $\eta > 0$. Man setze $\Theta(\eta) := \|v'\|_{L^2(0,\eta)}$ und beachte $\Theta(\eta) \rightarrow 0$ für $\eta \rightarrow 0$.

Da $v(x) = v_\eta(x)$ für $x \geq \eta$, bleiben die Größen $\|v - v_\eta\|_{L^2(0,\eta)}$ und $\|v' - v'_\eta\|_{L^2(0,\eta)}$ abzuschätzen. Wegen $v'_\eta - v' = \varphi'_\eta v + (\varphi_\eta - 1)v'$ ergibt sich

$$\|v'_\eta - v'\|_{L^2(0,\eta)} \leq \|\varphi'_\eta\|_{L^\infty(0,\eta)} \|v\|_{L^2(0,\eta)} + \|\varphi_\eta - 1\|_{L^\infty(0,\eta)} \|v'\|_{L^2(0,\eta)} \leq \frac{3}{\eta} \|v\|_{L^2(0,\eta)} + \Theta(\eta).$$

Da $|v(x)| \leq \int_0^x v'(\zeta) d\zeta$, liefert (6.2.5b) die Abschätzung $|v(x)| \leq \sqrt{\eta} \Theta(\eta)$ für $0 \leq x \leq \eta$ und damit $\|v\|_{L^2(0,\eta)} \leq \eta \Theta(\eta)$. Über $|v - v_\eta|_1^2 \leq \|v - v_\eta\|_{L^2(0,\eta)}^2 + \|v' - v'_\eta\|_{L^2(0,\eta)}^2 \leq C \Theta^2(\eta) \rightarrow 0$ folgt die Behauptung. ■

Wenn $\Omega \in C^1$, existiert in allen Randpunkten die Normalenrichtung \vec{n} . In Analogie zu Satz 6.2.43 beweist man

Zusatz 6.2.44 Für $\Omega \in C^1$ und $k \in \mathbb{N}$ gilt

$$\begin{aligned} H_0^k(\Omega) &= \{u \in H^k(\Omega) : \partial^\ell u / \partial n^\ell|_\Gamma = 0 \text{ für alle } 0 \leq \ell \leq k-1\} \\ &= \{u \in H^k(\Omega) : D^\alpha u|_\Gamma = 0 \text{ für alle } 0 \leq |\alpha| \leq k-1\}. \end{aligned}$$

6.3 Dualräume

6.3.1 Dualraum eines normierten Raumes

X sei ein normierter, linearer Raum über \mathbb{R} . Als *Dualraum* X' bezeichnet den Raum aller beschränkten, linearen Abbildungen von X auf \mathbb{R} :

$$X' = L(X, \mathbb{R}).$$

Gemäß Übungsaufgabe 6.1.8 ist X' ein Banach-Raum mit der Norm (*Dualnorm*)

$$\|x'\|_{X'} := \|x'\|_{\mathbb{R} \leftarrow X} = \sup \{|x'(x)| / \|x\|_X : 0 \neq x \in X\}. \quad (6.3.1)$$

Die Elemente $x' \in X'$ nennt man *lineare Funktionale* auf X . Statt $x'(x)$ (Anwendung von x' auf x) schreibt man auch $\langle x, x' \rangle_{X \times X'}$ oder $\langle x', x \rangle_{X' \times X}$ und nennt $\langle \cdot, \cdot \rangle_{X \times X'}$ die *Dualform* auf $X \times X'$:

$$\langle x, x' \rangle_{X \times X'} = \langle x', x \rangle_{X' \times X} = x'(x).$$

Lemma 6.3.1 Der Banach-Raum X sei dicht und stetig im Banach-Raum Y eingebettet. Dann ist Y' stetig in X' eingebettet.

Beweis. 1) Sei $y' \in Y'$. Wegen $X \subset Y$ ist y' auf X definiert.

2) Da X ein dichter Teilraum von Y ist, gilt nach Satz 6.1.11 und (6.1.4) für jedes $y' \in Y'$:

$$\|y'\|_{Y'} = \sup_{0 \neq x \in X} |y'(x)| / \|x\|_Y \geq \frac{1}{C} \sup_x |y'(x)| / \|x\|_X = \frac{1}{C} \|y'\|_{X'},$$

d.h. Y' ist stetig in X' eingebettet. ■

Der transponierten Matrix im endlich-dimensionalen Fall entspricht die *Dualabbildung* (oder der *Dualoperator*).

Lemma 6.3.2 Seien X und Y normiert und $T \in L(X, Y)$. Für jedes $y' \in Y'$ definiert die Gleichung

$$\langle Tx, y' \rangle_{Y \times Y'} = \langle x, x' \rangle_{X \times X'} \quad \text{auf alle } x \in X \quad (6.3.2)$$

ein eindeutiges $x' \in X'$. Die lineare Abbildung $y' \mapsto x'$ definiert den Dualoperator $T' : Y' \rightarrow X'$ mit $T'y' = x'$. Es gilt $T' \in L(Y', X')$ und

$$\|T'\|_{X' \leftarrow Y'} = \|T\|_{Y \leftarrow X}. \quad (6.3.3)$$

Beweis. Schreibt man (6.3.2) in der Form $y'(Tx) = x'(x)$, so sieht man, dass $x' = y' \circ T$. Aus

$$\begin{aligned} \|T'\|_{X' \leftarrow Y'} &= \sup_{y' \neq 0} \|T'y'\|_{X'} / \|y'\|_{Y'} = \sup_{x \neq 0, y' \neq 0} \underbrace{\langle x, T'y' \rangle}_{=x'}_{X \times X'} / [\|x\|_X \|y'\|_{Y'}] \\ &= \sup_{x \neq 0, y' \neq 0} \langle Tx, y' \rangle_{Y \times Y'} / [\|x\|_X \|y'\|_{Y'}] = \sup_{x' \neq 0} \|Tx\|_Y / \|x\|_X = \|T\|_{Y \leftarrow X} \end{aligned}$$

folgt (6.3.3) nach Definition der Normen. ■

Beispiel 6.3.3 Seien $\Omega = (0, 1)$, $X = (C^0(\overline{\Omega}), \|\cdot\|_\infty)$ und sei $x \in \Omega$ fest gewählter Punkt. Die Abbildung $\delta_x : u \in C^0(\overline{\Omega}) \mapsto u(x) \in \mathbb{R}$ ist ein Funktional: $\delta_x \in C^0(\overline{\Omega})'$ (die sogenannte Diracsche Deltafunktion). Der Laplace-Operator Δ gehört zu $L(C^2(\overline{\Omega}), C^0(\overline{\Omega}))$. Die Dualabbildung $\Delta' \in L(C^0(\overline{\Omega})', C^2(\overline{\Omega})')$ ist auf δ_x anwendbar: $\Delta'\delta_x$ ist durch $(\Delta'\delta_x)u = \Delta u(x)$ für alle $u \in C^2(\overline{\Omega})$ charakterisiert.

Übungsaufgabe 6.3.4 Seien $S \in L(X, Y)$ und $T \in L(Y, Z)$. Man zeige: $(TS)' = S'T'$.

Übungsaufgabe 6.3.5 Man zeige: Ist $T \in L(X, Y)$ surjektiv, so ist T' injektiv.

6.3.2 Adjungierte Operatoren

X sei ein Hilbert-Raum (über \mathbb{R}). Jedes $y \in X$ definiert durch

$$f_y(x) := (x, y)_X$$

ein lineares Funktional $f_y \in X'$ mit $\|f_y\|_{X'} = \|y\|_X$. Die Umkehrung gilt ebenfalls (vgl. Yosida [91, §III.6]):

Satz 6.3.6 (Darstellungssatz von Riesz) X sei Hilbert-Raum und $f \in X'$ ein Funktional. Dann existiert genau ein $y_f \in X$, sodass

$$f(x) = (x, y_f)_X \quad \text{für alle } x \in X \quad \text{und} \quad \|f\|_{X'} = \|y_f\|_X.$$

Folgerung 6.3.7 (Riesz-Isomorphismus) X sei Hilbert-Raum. a) Es existiert eine eindeutige Zuordnung (der sogenannte Riesz⁹-Isomorphismus) $J_X \in L(X, X')$ mit $J_X y = f_y$, $J_X^{-1} f = y_f$, die die Norm erhält: $\|J_X\|_{X' \leftarrow X} = \|J_X^{-1}\|_{X \leftarrow X'} = 1$.

b) X' ist ein Hilbert-Raum mit dem Skalarprodukt $(x', y')_{X'} = (J_X^{-1} x', J_X^{-1} y')_X$. Die Dualnorm $\|x'\|_{X'}$ aus (6.3.1) stimmt mit der durch $(x', x')_X^{1/2}$ induzierten Norm überein.

c) Man identifiziert stets X mit X'' vermöge $x(x') := x'(x)$. Hieraus folgt

$$J_{X'} = J_X^{-1}, \quad J_X = J_{X'}' \quad \text{und} \quad T'' = T \quad \text{für } T \in L(X, Y),$$

wenn auch $Y = Y''$ Hilbert-Raum ist.

d) Man kann X und X' identifizieren: $X = X'$, $J_X = I$.

Seien X, Y Hilbert-Räume und $T \in L(X, Y)$. Die durch $T^* := J_X^{-1} T' J_Y \in L(Y, X)$ definierte Abbildung heißt der zu T adjungierte Operator und erfüllt

$$(Tx, y)_Y = (x, T^* y)_X \quad \text{für alle } x \in X, y \in Y, \quad \|T\|_{Y \leftarrow X} = \|T^*\|_{X \leftarrow Y}. \quad (6.3.4)$$

Der adjungierte und der duale Operator fallen nur dann zusammen (d.h. $T^* = T'$), wenn jeweils die Dualräume X', Y' mit X, Y identifiziert werden. $T \in L(X, X)$ heißt selbstadjungiert (oder symmetrisch), wenn $T = T^*$. Der Operator $T \in L(X, X)$ heißt Projektion, wenn $T^2 = T$. Eine orthogonale Projektion liegt vor, wenn T außerdem selbstadjungiert ist.

Bemerkung 6.3.8 Sei X_0 ein abgeschlossener Unterraum des Hilbert-Raumes X . Durch $Tx := y \in X_0$ mit

$$\|x - y\|_X := \inf \{\|x - \eta\|_X : \eta \in X_0\} \quad (6.3.5)$$

wird eine orthogonale Projektion beschrieben. Ist umgekehrt $T \in L(X, X)$ eine orthogonale Projektion mit dem Bild $X_0 := \{Tx : x \in X\}$, so gilt (6.3.5) für $y = Tx$. Eine orthogonale Projektion hat stets die Norm $\|T\|_{X \leftarrow X} \leq 1$.

⁹Frigyes Riesz, geb. am 22. Januar 1880 in Győr, gest. am 28. Februar 1956 in Budapest (nicht zu verwechseln mit seinem Bruder Marcel, der ebenfalls Mathematiker war)

Beweis. 1) x lässt sich eindeutig in $x = y + z$ ($y \in X_0$, $z \in X_0^\perp$) zerlegen (vgl. Lemma 6.1.17). y ist eindeutige Lösung von (6.3.5). $x \in X_0$ impliziert $y = x$, also $T^2 = T$. Die analoge Zerlegung $x' = y' + z'$ zeigt $(x, T^*x') = (Tx, x') = (y, x') = (y, y' + z') = (y, y') = (y + z, y') = (x, Tx')$, somit $T = T^*$.

2) Sei nun T eine orthogonale Projektion mit dem Bild X_0 . $x = y + z$ sei wie oben zerlegt. $T^2 = T$ zeigt $Ty = y$. Für jedes $y' \in X_0$ gilt $(Tz, y') = (z, T^*y') = (z, Ty') = (z, y') = 0$, also $Tz \in X_0^\perp$. Zusammen mit $Tz \in X_0$ folgt $Tz = 0$, sodass $Tx = Ty + Tz = y$.

3) $Tx = y$ und $\|x\|_X^2 = \|y\|_X^2 + \|z\|_X^2 \geq \|y\|_X^2$ beweisen $\|T\|_{X \leftarrow X} \leq 1$. ■

6.3.3 Skalen von Hilbert-Räumen

Wir setzen voraus:

$$V \subset U \quad \text{seien zwei Hilbert-Räume mit stetiger und dichter Einbettung.} \quad (6.3.6)$$

Lemma 6.3.9 *Unter der Voraussetzung (6.3.6) ist U' stetig und dicht in V' eingebettet.*

Beweis. Die Stetigkeit der Einbettung $U' \subset V'$ wurde in Lemma 6.3.1 festgestellt. Zum Beweis, dass U' dicht in V' liegt, verwenden wir Übungsaufgabe 6.1.18d ($A := U'$, $X := V'$). Seien $0 \neq v' \in V'$ beliebig und $u := J_V^{-1}v' \in V \subset U$. Definitionsgemäß ist $u' := J_U u \in U' \subset V'$ durch $u'(x) = (x, u)_U$ für alle $x \in U$ charakterisiert. Für $x := u = J_V^{-1}v' \in V$ folgt

$$(v', u')_{V'} = (J_V^{-1}v', J_V^{-1}u')_V = (u, J_V^{-1}u')_V = u'(u) = (u, u)_U > 0$$

und damit insbesondere $(v', u')_{V'} \neq 0$. Somit trifft die Charakterisierung aus Übungsaufgabe 6.1.18d zu. ■

Gemäß Folgerung 6.3.7d lassen sich U und U' identifizieren. Damit erhält man den *Gelfand*¹⁰-Dreier

$$V \subset U \subset V' \quad (V \subset U \text{ stetig und dicht eingebettet}). \quad (6.3.7)$$

Folgerung 6.3.10 *In einem Gelfand-Dreier (6.3.7) sind auch V und U stetig und dicht in V' eingebettet.*

Beweis. Zu $U \subset V'$ vergleiche man Lemma 6.3.9, $V \subset V'$ folgt mit Übungsaufgabe 6.1.15. ■

Achtung. Man könnte ebenso gut V mit V' identifizieren und erhielte $U' \subset V' = V \subset U$. Es ist aber nicht möglich, gleichzeitig U mit U' und V mit V' zu identifizieren. Im ersten Falle interpretiert man $x(y) = \langle y, x \rangle_{U \times U'}$ für $x, y \in U$ (insbesondere für $x, y \in V \subset U$) als $\langle y, x \rangle_U$, im zweiten Falle als $\langle y, x \rangle_{V'}$.

Übungsaufgabe 6.3.11 *Es gelte (6.3.7). Man setze $W := \{J_V^{-1}u : u \in U\}$ und definiere $(x, y)_W := (J_V X, J_V y)_U$ als Skalarprodukt auf W . Man zeige: a) W ist Hilbert-Raum,*

b) $W \subset V$ ist stetige und dichte Einbettung,

c) $(v, w)_V = (v, J_V w)_U$ für alle $v \in V, w \in W$,

d) $|(x, y)_V| \leq \|x\|_U \|y\|_W$ ($x, y \in W$).

Das Skalarprodukt $(x, y)_U$ kann wegen $U = U'$ auch in der Form $y(x) = \langle x, y \rangle_{U \times U'}$ geschrieben werden. Ist $x \in V$, so gilt auch $y(x) = \langle x, y \rangle_{V \times V'}$. Das heißt, dass $(x, y)_U = \langle x, y \rangle_{V \times V'}$ für alle $x \in V, y \in U \subset V'$. Ebenso erhält man $(x, y)_U = \langle x, y \rangle_{V' \times V}$ für alle $x \in U$ und $y \in V$. Die dichte und stetige Einbettung $V \subset V'$ beweist die

Bemerkung 6.3.12 *$V \subset U \subset V'$ sei ein Gelfand-Dreier (6.3.7). Die stetige Fortsetzung des Skalarproduktes $(\cdot, \cdot)_U$ auf $V \times V'$ [$V' \times V$] ergibt die Dualform $\langle \cdot, \cdot \rangle_{V \times V'}$ [$\langle \cdot, \cdot \rangle_{V' \times V}$]. Deshalb ist folgende Schreibweise sinnvoll:*

$$\langle x, y \rangle_{V \times V'} = (x, y)_U \quad \text{für } x \in V, y \in V', \quad \langle x, y \rangle_{V' \times V} = (x, y)_U \quad \text{für } x \in V', y \in V.$$

Im Zusammenhang mit Sobolev-Räumen wird stets $U := L^2(\Omega)$ gewählt, sodass die Einbettungen wie folgt lauten:

$$H_0^s(\Omega) \subset L^2(\Omega) \subset (H_0^s(\Omega))' \quad (s \geq 0), \quad (6.3.8)$$

$$H^s(\Omega) \subset L^2(\Omega) \subset (H^s(\Omega))' \quad (s \geq 0), \quad (6.3.9)$$

¹⁰Israil Moiseevic Gelfand, am 2 Sept. 1913 in Krasnye Okny, Odessa, Ukraine geboren.

Übungsaufgabe 6.3.13 Man zeige: (6.3.8) und (6.3.9) sind Gelfand-Dreier.

Den Dualraum von $H_0^s(\Omega)$ bezeichnet man auch mit $H^{-s}(\Omega)$ oder $H_0^{-s}(\Omega)$:

$$H_0^{-s}(\Omega) := H^{-s}(\Omega) := (H_0^s(\Omega))' \quad (s \geq 0).$$

Die Norm von $H^{-s}(\Omega)$ lautet gemäß (6.3.1):

$$|u|_{-s} := \sup\{|(u, v)_{L^2(\Omega)}| : |v|_s : 0 \neq v \in H_0^s(\Omega)\},$$

wobei $(u, v)_{L^2(\Omega)}$ die Dualform auf $H_0^s(\Omega) \times H^{-s}(\Omega)$ bezeichnet (vgl. Bemerkung 6.3.12).

Bemerkung 6.3.14 a) Sei $\Omega = \mathbb{R}^n$. Die zu $|\cdot|_s$ gehörige Dualnorm $|u|_{-s} := \sup_{0 \neq v \in H^s(\mathbb{R}^n)} |(u, v)_0| / |v|_s$ ist äquivalent zu $|\cdot|_{-s}$ und hat die Darstellung (6.2.16b) (mit $-s$ statt s).

b) Die Fourier-Transformation zeigt $D^\alpha \in L(H^s(\mathbb{R}^n), H^{s-|\alpha|}(\mathbb{R}^n))$ für alle $s \in \mathbb{R}$.

c) Es ist $au \in H^s(\Omega)$, falls $u \in H^s(\Omega)$, $a \in C^t(\overline{\Omega})$, wobei entweder $t = |s| \in \mathbb{N}_0$ oder $t > |s|$.

6.4 Kompakte Operatoren

Definition 6.4.1 Eine Teilmenge K eines Banach-Raumes heißt präkompakt [kompakt], falls jede Folge $x_i \in K$ ($i \in \mathbb{N}$) eine konvergente Teilfolge x_{i_k} enthält [und $\lim_{k \rightarrow \infty} x_{i_k} \in K$].

Eine andere Definition der Kompaktheit lautet: Jede offene Überdeckung von K enthält bereits eine endliche Überdeckung von K . Beide Definitionen sind in metrischen Räumen äquivalent (vgl. Pflaumann-Unger [68, p. 66]).

Bemerkung 6.4.2 a) $K \subset \mathbb{R}^n$ ist genau dann präkompakt [kompakt], wenn K beschränkt [und abgeschlossen] ist.

b) X sei Banach-Raum. Die Einheitskugel $\{x \in X : \|x\| \leq 1\}$ ist genau dann kompakt, wenn $\dim(X) < \infty$.

Definition 6.4.3 X und Y seien Banach-Räume. Eine Abbildung $T \in L(X, Y)$ heißt kompakt, falls $\{Tx : x \in X, \|x\|_X \leq 1\}$ (Bild der Einheitskugel in X) präkompakt in Y ist.

Übungsaufgabe 6.4.4 Wann ist die Identität $I \in L(X, X)$ kompakt?

Lemma 6.4.5 X, Y, Z seien Banach-Räume. a) Eine der Abbildungen $T_1 \in L(X, Y), T_2 \in L(Y, Z)$ sei kompakt. Dann ist auch $T_2 T_1 \in L(X, Z)$ kompakt.

b) $T \in L(X, Y)$ ist genau dann kompakt, wenn $T' \in L(Y', X')$ kompakt ist.

Beweis. a) Sei $K_1 := \{x \in X : \|x\|_X \leq 1\}$. Ist T_1 kompakt, d.h. $T_1(K_1)$ präkompakt, so ist auch $T_2(T_1(K_1))$ präkompakt und damit $T_2 T_1$ kompakt. Ist dagegen T_2 kompakt, beweist man die Behauptung wie folgt. Da eine Skalierung die Kompaktheit nicht ändert, kann ohne Beschränkung der Allgemeinheit $\|T_1\|_{Y \leftarrow X} \leq 1$ angenommen werden. Damit ist $T_1(K_1)$ eine Teilmenge der Einheitskugel in Y und somit $T_2(T_1(K_1))$ präkompakt.

b) Man vergleiche Yosida [91, §X]. ■

Eine spezielle kompakte Abbildung ist die kompakte Einbettung:

Definition 6.4.6 $X \subset Y$ sei eine stetige Einbettung. X heißt kompakt eingebettet in Y , falls die Inklusion $I \in L(X, Y), Ix = x$, kompakt ist.

Zusammen mit den Definitionen 6.4.1 und 6.4.3 erhält man: $X \subset Y$ ist kompakt eingebettet, falls jede Folge $x_i \in X$ mit $\|x_i\|_X \leq 1$ eine in Y konvergente Teilfolge enthält.

Beispiel 6.4.7 Ω sei beschränkt. $C^s(\overline{\Omega}) \subset C^0(\overline{\Omega})$ ist für $s > 0$ eine kompakte Einbettung.

Beweis. Funktionen $u_i \in C^s(\overline{\Omega})$ mit $\|u_i\|_{C^s(\overline{\Omega})} \leq 1$ sind gleichgradig stetig und gleichmäßig beschränkt. Die Behauptung folgt nach dem Satz von Arzelà-Ascoli (vgl. Yosida [91, §III.3]). ■

Analoge Resultate erhält man für Sobolev-Räume (vgl. Adams [1, p. 144], Wloka [89, p. 117ff]):

Satz 6.4.8 $\Omega \subset \mathbb{R}^n$ sei offen und beschränkt. a) Die Einbettungen $H_0^s(\Omega) \subset H_0^t(\Omega)$ ($s, t \in \mathbb{R}$, $s > t$) sind kompakt.

b) Sei zudem $\Omega \in C^{0,1}$. Die Einbettungen $H^k(\Omega) \subset H^l(\Omega)$ ($k, l \in \mathbb{N}_0$, $k > l$) sind kompakt.

c) Sei $0 < t < s$ und $\Omega \in C^r$ ($r > t$, $r > 1$) oder $\Omega \in C^{k,1}$ ($k+1 > t$). Dann ist die Einbettung $H^s(\Omega) \subset H^t(\Omega)$ kompakt.

Bemerkung 6.4.9 In Satz 6.4.8b kann $\Omega \in C^{0,1}$ durch die “gleichmäßige Kegelbedingung” ersetzt werden (vgl. Wloka [89, p. 45]). Hinreichend für $\Omega \in C^{0,1}$ ist, dass der Rand $\partial\Omega$ stückweise glatt ist und der Innenwinkel der möglichen Ecken kleiner als 2π ausfällt.

Einspringende Ecken (vgl. Abbildung 2.1.1) sind damit zugelassen, ein geschlitztes Gebiet (vgl. Abbildung 5.2.1b) ist dagegen ausgeschlossen.

In Abschnitt 6.5 wird die folgende Situation entstehen:

$$V \subset U \subset V', \quad T \in L(V', V). \quad (6.4.1)$$

Wegen der stetigen Einbettungen gehört T auch zu $L(V', V')$, $L(U, U)$, $L(V, V)$ und $L(U, V)$.

Satz 6.4.10 Es gelte (6.4.1). $V \subset U$ sei eine kompakte Einbettung. Dann sind $T \in L(V', V')$, $T \in L(U, U)$, $T \in L(V, V)$, $T \in L(V', U)$ und $T \in L(U, V)$ kompakt.

Beweis. Als Beispiel sei $T \in L(U, V)$ vorgeführt. Mit der Inklusion $I \in L(V, U)$ ist auch $I \in L(U, V)$ kompakt (vgl. Lemma 6.4.5b). $T \in L(U, V)$ als Produkt der kompakten Abbildung $I \in L(U, V')$ mit $T \in L(V', V)$ ist kompakt (vgl. Lemma 6.4.5a). ■

Übungsaufgabe 6.4.11 Seien $\dim X < \infty$ oder $\dim Y < \infty$. Man zeige: $T \in L(X, Y)$ ist kompakt.

Die Bedeutung kompakter Operatoren $T \in L(X, X)$ liegt darin, dass die Gleichung $Tx - \lambda x = y$ ($x, y \in X$, y gegeben, x gesucht) analoge Eigenschaften wie im endlich-dimensionalen Falle hat.

Satz 6.4.12 (Riesz-Schauder-Theorie) X sei ein Banach-Raum; $T \in L(X, X)$ sei kompakt. a) Für jedes $\lambda \in \mathbb{C} \setminus \{0\}$ gilt eine der folgenden Alternativen:

$$i) (T - \lambda I)^{-1} \in L(X, X) \quad \text{oder} \quad ii) \lambda \text{ ist Eigenwert.}$$

Im Falle i) hat die Gleichung $Tx - \lambda x = y$ für alle $y \in X$ eine eindeutige Lösung $x \in X$. Im Falle ii) gibt es einen endlich-dimensionalen Eigenraum $E(\lambda, T) := \text{Kern}(T - \lambda I) \neq \{0\}$. Alle $x \in E(\lambda, T)$ lösen das Eigenwertproblem $Tx = \lambda x$.

b) Das Spektrum $\sigma(T)$ von T besteht definitionsgemäß aus allen Eigenwerten und, falls nicht $T^{-1} \in L(X, X)$, $\lambda = 0$. Es gibt höchstens abzählbar viele Eigenwerte, die sich nur in null häufen können. Es gilt $\lambda \in \sigma(T)$ genau dann, wenn $\bar{\lambda} \in \sigma(T')$. Ferner ist $\dim(E(\lambda, T)) = \dim(E(\bar{\lambda}, T')) < \infty$.

c) Für $\lambda \in \sigma(T) \setminus \{0\}$ hat $Tx - \lambda x = y$ genau dann mindestens eine Lösung $x \in X$, wenn $\langle y, x' \rangle_{X \times X'} = 0$ für alle $x' \in E(\bar{\lambda}, T')$.

In Lemma 6.5.18 werden wir das folgende Lemma benötigen.

Lemma 6.4.13 $X \subset Y \subset Z$ seien stetig eingebettete Banach-Räume, wobei $X \subset Y$ kompakt eingebettet sei. Dann gibt es zu jedem $\varepsilon > 0$ ein C_ε , sodass

$$\|x\|_Y \leq \varepsilon \|x\|_X + C_\varepsilon \|x\|_Z \quad \text{für alle } x \in X. \quad (6.4.2)$$

Beweis. Sei $\varepsilon > 0$ fest. Die Negation von (6.4.2) lautet: Es gibt $x_i \in X$ mit $(\|x_i\|_Y - \varepsilon \|x_i\|_X) / \|x_i\|_Z \rightarrow \infty$. Für $y_i := (\varepsilon \|x_i\|_X)^{-1} x_i \in X$ gilt somit $(\|y_i\|_Y - 1) / \|y_i\|_Z \rightarrow \infty$. Hieraus folgert man $\|y_i\|_Z \rightarrow 0$ und $\|y_i\|_Y > 1$ für hinreichend große i . Da $\|y_i\|_X < 1/\varepsilon$ und $X \subset Y$ kompakt, konvergiert eine Teilfolge y_{i_k} gegen $y^* \in Y$. $\|y_i\|_Y > 1$ impliziert $\|y^*\|_Y \geq 1$, d.h. $y^* \neq 0$. Andererseits konvergiert y_{i_k} auch in Z gegen y^* , da $Y \subset Z$ stetig eingebettet ist. $\|y_i\|_Z \rightarrow 0$ liefert den Widerspruch: $y^* = 0$. ■

6.5 Bilinearformen

Im Folgenden sei angenommen, dass V ein Hilbert-Raum ist. Die Abbildung $a(\cdot, \cdot) : V \times V \rightarrow \mathbb{R}$ heißt *Bilinearform*, falls

$$a(x + \lambda y, z) = a(x, z) + \lambda a(y, z), \quad a(x, y + \lambda z) = a(x, y) + \lambda a(x, z) \quad \text{für alle } \lambda \in \mathbb{R}, x, y, z \in V$$

(im komplexen Falle, wobei $a(x, \lambda y) = \bar{\lambda} a(x, y)$, spricht man von *Sesquilinearformen*).

$a(\cdot, \cdot)$ heißt *stetig* (oder *beschränkt*), falls ein C_S existiert, sodass

$$|a(x, y)| \leq C_S \|x\|_V \|y\|_V \quad \text{für alle } x, y \in V. \quad (6.5.1)$$

Lemma 6.5.1 a) Einer stetigen Bilinearform kann man eineindeutig einen Operator $A \in L(V, V')$ zuordnen, sodass

$$a(x, y) = \langle Ax, y \rangle_{V' \times V} \quad \text{für alle } x, y \in V, \quad \|A\|_{V' \leftarrow V} \leq C_S. \quad (6.5.2)$$

Umgekehrt ist $C_S := \|A\|_{V' \leftarrow V}$ die beste Schranke in (6.5.1).

b) V_1 und V_2 seien dicht in V . $a(\cdot, \cdot)$ sei auf $V_1 \times V_2$ definiert und erfülle (6.5.1) mit “ $x \in V_1, y \in V_2$ ” statt “ $x, y \in V$ ”. Dann ist $a(\cdot, \cdot)$ eindeutig auf $V \times V$ stetig fortsetzbar, sodass (6.5.1) mit gleichem C_S für alle $x, y \in V$ gilt.

Beweis. a) $x \in V$ sei fest. $\varphi_x(y) := a(x, y)$ definiert ein Funktional $\varphi_x \in V'$ mit $\|\varphi_x\|_{V'} \leq C_S \|x\|_V$. Da $x \mapsto \varphi_x$ eine lineare Abbildung ist, lässt sich $Ax := \varphi_x$ für $x \in V$ definieren. $\|Ax\|_{V'} \leq C_S \|x\|_V$ beweist $\|A\|_{V' \leftarrow V} \leq C_S$. Die Definitionen zeigen $\langle Ax, y \rangle_{V' \times V} = \langle \varphi_x, y \rangle_{V' \times V} = \varphi_x(y) = a(x, y)$. Umgekehrt ist für jedes $A \in L(V, V')$ auch $a(x, y) := \langle Ax, y \rangle_{V' \times V}$ eine Bilinearform mit $\langle Ax, y \rangle_{V' \times V} \leq \|Ax\|_{V'} \|y\|_V \leq \|A\|_{V' \leftarrow V} \|x\|_V \|y\|_V$.

b) Gemäß Satz 6.1.11 lässt sich A auch dann eindeutig festlegen, wenn $a(\cdot, \cdot)$ nur auf $V_1 \times V_2$ erklärt ist. $\langle Ax, y \rangle_{V' \times V}$ stellt dann die stetige Fortsetzung dar. ■

Der Beweis zeigt

$$\|A\|_{V' \leftarrow V} = \sup \{ |a(x, y)| : x, y \in V, \|x\|_V = \|y\|_V = 1 \}. \quad (6.5.3)$$

A heißt der $a(\cdot, \cdot)$ zugeordnete oder zu $a(\cdot, \cdot)$ gehörende Operator.

Die zu $a(\cdot, \cdot)$ adjungierte Bilinearform $a^*(\cdot, \cdot)$ ist durch $a^*(x, y) := a(y, x)$ ($x, y \in V$) gegeben. Die Bilinearform heißt *symmetrisch*, falls $a(\cdot, \cdot) = a^*(\cdot, \cdot)$.

Übungsaufgabe 6.5.2 a) Gehört A zu $a(\cdot, \cdot)$, so A' zu $a^*(\cdot, \cdot)$. b) Ist $a(\cdot, \cdot)$ symmetrisch, so gilt $A = A'$.

Lemma 6.5.3 (inf-sup-Bedingungen) $A \in L(V, V')$ sei der zu einer stetigen Bilinearform $a(\cdot, \cdot)$ gehörende Operator. Dann sind die folgenden Aussagen (i), (ii), (iii) äquivalent:

- (i) $A^{-1} \in L(V', V)$ existiert;
- (ii) Es existieren $\varepsilon, \varepsilon' > 0$, sodass

$$\inf_{x \in V, \|x\|_V = 1} \sup_{y \in V, \|y\|_V = 1} |a(x, y)| = \varepsilon > 0, \quad (6.5.4a)$$

$$\inf_{y \in V, \|y\|_V = 1} \sup_{x \in V, \|x\|_V = 1} |a(x, y)| = \varepsilon' > 0. \quad (6.5.4b)$$

- (iii) Es gelten die Ungleichungen (6.5.4a) und (6.5.4c):

$$\sup_{x \in V, \|x\|_V = 1} |a(x, y)| > 0 \quad \text{für alle } y \in V. \quad (6.5.4c)$$

Gilt eine der Aussagen (i)-(iii), so ist

$$\varepsilon = \varepsilon' = 1 / \|A^{-1}\|_{V' \leftarrow V}, \quad (\varepsilon, \varepsilon' \text{ aus (6.5.4a, b)}). \quad (6.5.4d)$$

Aus der Gleichung (6.5.4a) folgt trivialerweise die Ungleichung

$$\inf_{x \in V, \|x\|_V=1} \sup_{y \in V, \|y\|_V=1} |a(x, y)| \geq \varepsilon > 0. \quad (6.5.4e)$$

Umgekehrt folgt (6.5.4a) aus (6.5.4e) mit einem eventuell größeren $\varepsilon > 0$. (6.5.4e) nennt man auch die *Babuška-Bedingung* (oder auch die *inf-sup-Bedingung*). (6.5.4e) ist äquivalent zu

$$\sup_{y \in V, \|y\|_V=1} |a(x, y)| \geq \varepsilon \|x\|_V \quad \text{für alle } x \in V, \quad (6.5.4e')$$

denn (6.5.4e) ist gleichbedeutend mit (6.5.4e') für alle $x \in V, \|x\|_V = 1$. Die Normierungsbedingung $\|x\|_V = 1$ kann man offenbar fallen lassen. Die linke Seite in (6.5.4e') stimmt mit der Definition der Dualnorm von Ax überein, sodass (6.5.4e) und (6.5.4e') auch zu (6.5.4e'') äquivalent sind:

$$\|Ax\|_{V'} \geq \varepsilon \|x\|_V \quad \text{für alle } x \in V. \quad (6.5.4e'')$$

Beweis des Lemmas 6.5.3. a) “(i) \Rightarrow (ii)”: Die Inverse $A^{-1} \in L(V', V)$ existiere. (6.5.4a) folgt aus

$$\begin{aligned} \inf_{\substack{x \in V \\ \|x\|_V=1}} \sup_{\substack{y \in V \\ \|y\|_V=1}} |a(x, y)| &= \inf_{\substack{x \in V \\ x \neq 0}} \sup_{\substack{y \in V \\ y \neq 0}} \frac{|a(x, y)|}{\|x\|_V \|y\|_V} = \inf_{\substack{x \in V \\ x \neq 0}} \sup_{\substack{y \in V \\ y \neq 0}} \frac{|\langle Ax, y \rangle_{V' \times V}|}{\|x\|_V \|y\|_V} \\ &= \inf_{\substack{x' \in V' \\ x' \neq 0}} \sup_{\substack{y \in V \\ y \neq 0}} \frac{|\langle AA^{-1}x', y \rangle_{V' \times V}|}{\|A^{-1}x'\|_V \|y\|_V} = \inf_{\substack{x' \in V' \\ x' \neq 0}} \left(\frac{1}{\|A^{-1}x'\|_V} \sup_{\substack{y \in V \\ y \neq 0}} \frac{|\langle x', y \rangle_{V' \times V}|}{\|y\|_V} \right) \\ &= \inf_{\substack{x' \in V' \\ x' \neq 0}} \frac{1}{\|A^{-1}x'\|_V} \|x'\|_{V'} = 1 / \sup_{\substack{x' \in V' \\ x' \neq 0}} \frac{\|A^{-1}x'\|_V}{\|x'\|_{V'}} = 1 / \|A^{-1}\|_{V \leftarrow V'} =: \varepsilon \end{aligned}$$

und zeigt überdies die Charakterisierung von ε aus (6.5.4d). Ebenso zeigt man (6.5.4b) mit $\varepsilon' = 1 / \|A'^{-1}\|_{V \leftarrow V'}$. Wegen $A'^{-1} = (A^{-1})'$, (6.3.3) und $V'' = V$ folgt $\varepsilon = \varepsilon'$.

b) “(ii) \Rightarrow (iii)”: (6.5.4c) ist eine Abschwächung von (6.5.4b).

c) “(iii) \Rightarrow (i)”: $\varepsilon > 0$ in (6.5.4a) beweist, dass A injektiv ist. Um auch Surjektivität nachzuweisen, wollen wir zunächst zeigen, dass das Bild $W := \{Ax : x \in V\} \subset V'$ abgeschlossen ist. Für eine Folge $\{w_\nu\}$ mit $\|w^* - w_\nu\|_{V'} \rightarrow 0$ ist daher $w^* \in W$ zu zeigen. Nach Definition von W gibt es $x_\nu \in V$ mit $Ax_\nu = w_\nu$. Von (6.5.4a) schließt man über (6.5.4e) und (6.5.4e'') (mit $x := x_\nu - x_\mu$) auf $\|x_\nu - x_\mu\|_V \leq \|w_\nu - w_\mu\|_{V'} / \varepsilon$. Da $\{w_\nu\}$ eine Cauchy-Folge ist, überträgt sich diese Eigenschaft auf $\{x_\nu\}$. Es existiert ein $x^* \in V$ mit $x_\nu \rightarrow x^*$ in V . Die Stetigkeit von $A \in L(V, V')$ beweist $w_\nu = Ax_\nu \rightarrow Ax^*$, sodass $w^* = Ax^* \in W$. Gemäß Lemma 6.1.17 lässt sich V' in $W \oplus W^\perp$ zerlegen. Wäre A nicht surjektiv (also $W \neq V'$), gäbe es ein $w \in W^\perp$ mit $w \neq 0$. $y := J_{V'} w = J_{V'}^{-1} w \in V$ würde $y \neq 0$ erfüllen (vgl. Satz 6.3.6, Folgerung 6.3.7). Da $a(x, y) = \langle Ax, y \rangle_{V' \times V} = \langle Ax, w \rangle_{V'} = 0$ für alle $x \in V$, ergäbe sich ein Widerspruch zu (6.5.4c). Also ist A auch surjektiv, und Satz 6.1.13 beweist $A^{-1} \in L(V', V)$.

d) (6.5.4d) ergab sich bereits aus Teil a) des Beweises. ■

Es wird sich herausstellen, dass für die interessanten Fälle die Bedingungen (6.5.4a) und (6.5.4b) äquivalent sind (vgl. Lemma 6.5.17). Ein besonders einfacher Fall liegt vor in

Übungsaufgabe 6.5.4 Man zeige: Ist $\dim V < \infty$, so impliziert (6.5.4a) die Aussage (6.5.4b) mit $\varepsilon' = \varepsilon$ und umgekehrt.

Definition 6.5.5 (V-Elliptizität) Eine Bilinearform heißt V-elliptisch, falls sie auf $V \times V$ stetig ist und eine Zahl $C_E > 0$ existiert, sodass

$$a(x, x) \geq C_E \|x\|_V^2 \quad \text{für alle } x \in V. \quad (6.5.5)$$

In der Bezeichnung “V-elliptisch” ist V als eine Textvariable für einen Hilbert-Raum anzusehen. Im Falle von $V = H_0^1(\Omega)$ heißt die Bilinearform beispielsweise $H_0^1(\Omega)$ -elliptisch.

Übungsaufgabe 6.5.6 Man zeige: a) Ist $W \subset V$ ein Hilbert-Unterraum mit gleicher (oder äquivalenter) Norm wie in V , so ist eine V -elliptische Bilinearform auch W -elliptisch.

b) $a(\cdot, \cdot) : V \times V \rightarrow \mathbb{R}$ sei stetig. Gilt $a(x, x) \geq C_E \|x\|_V^2$ für alle $x \in V_0$, wobei V_0 dicht in V , so folgt (6.5.5) mit gleichem C_E .

c) $a(\cdot, \cdot) : V \times V \rightarrow \mathbb{R}$ sei stetig, symmetrisch, nichtnegativ ($a(v, v) \geq 0$ für alle $v \in V$) und erfülle (6.5.4a,c). Dann ist $a(\cdot, \cdot)$ V -elliptisch mit $C_E \geq \varepsilon^2/C_S$ (ε aus (6.5.4a,b), C_S aus (6.5.1)). Hinweis: Man beweise zunächst $|a(u, v)| \leq [a(u, u)a(v, v)]^{1/2}$ (vgl. Übungsaufgabe 6.1.16).

Lemma 6.5.7 V -Elliptizität (6.5.5) impliziert (6.5.1) und (6.5.4a,b) mit $\varepsilon = \varepsilon' \geq C_E$.

Beweis. Sei $x \in V$, $\|x\|_V = 1$. $\sup\{|a(x, y)| : y \in V, \|y\|_V = 1\} \geq |a(x, x)| \geq C_E$ beweist (6.5.4a) mit $\varepsilon \geq C_E$. Analog folgt (6.5.4b). ■

Die Zusammenfassung der Lemmata 6.5.1, 6.5.3, 6.5.7 zusammen mit $\|A'^{-1}\|_{V \leftarrow V'} = \|A^{-1}\|_{V \leftarrow V'}$ (vgl. Lemma 6.3.2) beweist den

Satz 6.5.8 Die Bilinearform sei V -elliptisch [oder erfülle (6.5.1), (6.5.4a,c)]. Dann genügt der zugehörige Operator A den Bedingungen

$$\begin{aligned} A &\in L(V, V'), & \|A'\|_{V' \leftarrow V} &= \|A\|_{V' \leftarrow V} \leq C_S, \\ A^{-1} &\in L(V', V), & \|A'^{-1}\|_{V \leftarrow V'} &= \|A^{-1}\|_{V \leftarrow V'} \leq C' \end{aligned} \quad (6.5.6)$$

mit C_S aus (6.5.1) und $C' = 1/C_E$ [bzw. $C' = 1/\varepsilon = 1/\varepsilon'$].

Mit Hilfe der Bilinearform $a(\cdot, \cdot)$ und eines Funktionals $f \in V'$ lässt sich das folgende *Variationsproblem* formulieren:

$$\text{suche } x \in V \text{ mit } a(x, y) = f(y) \quad \text{für alle } y \in V. \quad (6.5.7)$$

Gemäß Lemma 6.5.1 kann man (6.5.7) in der Form $\langle Ax - f, y \rangle_{V' \times V} = 0$ für alle $y \in V$ schreiben, d.h. $Ax = f$ in V' . Die Gleichung $Ax = f$ ist genau dann für alle $f \in V'$ lösbar, wenn $A^{-1} \in L(V', V)$. Somit erhält man den

Satz 6.5.9 Die Bilinearform sei stetig (vgl. (6.5.1)) und erfülle die Stabilitätsbedingung (6.5.4a,c) [hinreichend ist, dass $a(\cdot, \cdot)$ V -elliptisch ist]. Dann hat das Problem (6.5.7) genau eine Lösung $x := A^{-1}f$. Diese erfüllt $\|x\|_V \leq C \|f\|_{V'}$ mit $C = 1/\varepsilon = 1/\varepsilon'$ [bzw. $C = 1/C_E$].

Zusatz 6.5.10 (adjungierte Aufgabe) Unter den Voraussetzungen des Satzes 6.5.9 gilt die analoge Aussage mit der gleichen Abschätzung für das adjungierte Variationsproblem

$$\text{suche } x^* \in V \text{ mit } a^*(x^*, y) = f(y) \quad \text{für alle } y \in V. \quad (6.5.8)$$

Beweis. $\|A^{-1}\|_{V \leftarrow V'} = \|A'^{-1}\|_{V \leftarrow V'}$ (vgl. Übungsaufgabe 6.5.2a und Satz 6.5.8). ■

Übungsaufgabe 6.5.11 $a(\cdot, \cdot) : V \times V \rightarrow \mathbb{R}$ sei stetig. V_0 sei dicht in V . Man zeige: Die Lösung $x \in V$ von (6.5.7) ist schon durch " $a(x, y) = f(y)$ für alle $y \in V_0$ " eindeutig bestimmt. Gleiches gilt für (6.5.8).

Das Problem (6.5.7) kann einem *Minimierungsproblem* äquivalent sein:

Satz 6.5.12 $a(\cdot, \cdot)$ sei V -elliptisch und symmetrisch, ferner sei $f \in V'$. Dann nimmt

$$J(x) := a(x, x) - 2f(x) \quad (x \in V) \quad (6.5.9)$$

sein eindeutiges Minimum für die Lösung x der Aufgabe (6.5.7) an.

Beweis. Sei x die Lösung von (6.5.7). Für ein beliebiges $z \in V$ setze man $y := z - x$. Aus

$$\begin{aligned} J(z) &= J(x + y) = a(x + y, x + y) - 2f(x + y) = a(x, x) + a(x, y) + a(y, x) + a(y, y) - 2f(x) - 2f(y) \\ &\stackrel{\text{Symmetrie}}{=} J(x) + a(y, y) + 2\underbrace{[a(x, y) - f(y)]}_{=0, \text{ da } x \text{ Lösung}} = J(x) + a(y, y) \geq J(x) + C_E \|y\|_V^2 = J(x) + C_E \|z - x\|_V^2 \end{aligned}$$

liest man $J(z) > J(x)$ für alle $z \neq x$ ab. ■

Die Bezeichnung “V-elliptisch” scheint anzudeuten, dass den elliptischen Randwertproblemen V-elliptische Bilinearformen entsprechen. Dies ist im Allgemeinen nicht der Fall. Den elliptischen Randwertaufgaben werden V-koerzive Formen zugeordnet sein. Ihre Definition erfordert die Einführung eines Gelfand-Dreiers (vgl. (6.3.7): $V \subset U \subset V'$, $U = U'$, $V \subset U$ stetig und dicht eingebettet).

Definition 6.5.13 (V-Koerzivität) $V \subset U \subset V'$ sei ein Gelfand-Dreier. Eine Bilinearform $a(\cdot, \cdot)$ heißt V-koerziv, wenn sie stetig ist und wenn es Konstanten $C_K \in \mathbb{R}$ und $C_E > 0$ gibt, sodass

$$a(x, x) \geq C_E \|x\|_V^2 - C_K \|x\|_U^2 \quad \text{für alle } x \in V. \quad (6.5.10)$$

Offenbar unterscheidet sich die V-Koerzivität von der V-Elliptizität durch den Term $C_K \|x\|_U^2$, wobei an C_K keine Vorzeichenbedingungen gestellt werden. Entscheidend ist, dass die U-Norm wegen $V \subset U$ schwächer als die V-Norm ist.

Übungsaufgabe 6.5.14 Setze $\tilde{a}(x, y) := a(x, y) + C_K(x, y)_U$ mit C_K aus (6.5.10). Man zeige:

- Die Koerzivitätsbedingung (6.5.10) ist äquivalent zur V-Elliptizität von \tilde{a} .
- $I : V \rightarrow V'$ sei die Inklusion. Gehört $A \in L(V, V')$ zu $a(\cdot, \cdot)$, so $\tilde{A} := A + C_K I$ zu $\tilde{a}(\cdot, \cdot)$. Warum gilt $\tilde{A} \in L(V, V')$?

Die Ergebnisse der Riesz-Schauder-Theorie (Satz 6.4.12) übertragen sich auf A , sobald die Einbettung $V \subset U$ nicht nur stetig, sondern auch kompakt ist.

Satz 6.5.15 $V \subset U \subset V'$ sei ein Gelfand-Dreier mit kompakter Einbettung $V \subset U$. Die Bilinearform $a(\cdot, \cdot)$ sei V-koerziv mit zugehörigem Operator $A : V \rightarrow V'$ sei die Inklusion. a) Für jedes $\lambda \in \mathbb{C}$ gilt eine der Alternativen i), ii):

$$i) (A - \lambda I)^{-1} \in L(V', V) \text{ und } (A' - \bar{\lambda} I)^{-1} \in L(V', V) \quad \text{oder} \quad ii) \lambda \text{ ist Eigenwert.}$$

Im Falle i) ist sowohl $Ax - \lambda x = f$ als auch $A'x^* - \lambda x^* = f$ für alle $f \in V'$ eindeutig lösbar (d.h. $a(x, y) - \lambda(x, y)_U = f(y)$ und $a^*(x^*, y) - \bar{\lambda}(x^*, y)_U = f(y)$ für alle $y \in V$). Im Falle ii) gibt es endlich-dimensionale Eigenräume $\{0\} \neq E(\lambda) := \text{Kern}(A - \lambda I)$ und $\{0\} \neq E'(\lambda) := \text{Kern}(A' - \bar{\lambda} I)$, sodass

$$\begin{aligned} Ax &= \lambda x \quad \text{für } x \in E(\lambda), & \text{d.h. } a(x, y) &= \lambda(x, y)_U \quad \text{für alle } y \in V, \\ A'x^* &= \lambda x^* \quad \text{für } x^* \in E'(\lambda), & \text{d.h. } a^*(x^*, y) &= \bar{\lambda}(x^*, y)_U \quad \text{für alle } y \in V. \end{aligned} \quad (6.5.11)$$

b) Das Spektrum $\sigma(A)$ von A besteht aus höchstens abzählbar vielen Eigenwerten, die sich in \mathbb{C} nicht häufen können. Es ist $\lambda \in \sigma(A)$ genau dann, wenn $\bar{\lambda} \in \sigma(A')$. Ferner ist $\dim E(\lambda) = \dim E'(\bar{\lambda}) < \infty$.

c) Für $\lambda \in \sigma(A)$ hat $Ax - \lambda x = f \in V'$ genau dann mindestens eine Lösung $x \in V$, wenn $f \perp E'(\lambda)$, d.h. wenn $\langle f, x^* \rangle_{V' \times V} = (f, x^*)_U = 0$ für alle $x^* \in E'(\lambda)$.

Beweis. Mit $V \subset U$ ist auch $V \subset V'$ eine kompakte Einbettung, d.h. die Inklusion $I : V \rightarrow V'$ ist kompakt. $A + C_K I$ mit C_K aus (6.5.10) erfüllt $A + C_K I \in L(V, V')$, $(A + C_K I)^{-1} \in L(V', V)$ (vgl. Übungsaufgabe 6.5.14a). Lemma 6.4.5 zeigt, dass $K := (A + C_K I)^{-1} I : V \rightarrow V$ kompakt ist. Somit ist die Riesz-Schauder-Theorie auf $K - \mu I$ anwendbar. Da

$$K - \mu I = -\mu(I - \frac{1}{\mu}K) = -\mu(A + C_K I)^{-1}\{A + C_K I - \frac{1}{\mu}I\} = -\mu(A + C_K I)^{-1}(A - \lambda I) \quad \text{mit } \lambda = -C_K + \frac{1}{\mu},$$

übertragen sich die Aussagen des Satzes 6.4.12 über $K - \mu I$ auf $A - \lambda I = -\mu^{-1}(A + C_K I)(K - \mu I)$. ■

Bemerkung 6.5.16 Das Spektrum $\subset (A)$ hat das Maß null, sodass unter den Voraussetzungen des Satzes 6.5.15 die Lösbarkeit von $Ax - \lambda x = f$ für fast alle λ garantiert wird. Das Problem (6.5.7) ist lösbar, wenn nicht “zufällig” $0 \in \sigma(A)$.

Lemma 6.5.17 Unter den Voraussetzungen des Satzes 6.5.15 sind die Ungleichungen (6.5.4a,b) äquivalent.

Beweis. (6.5.4a) beweist, dass A injektiv ist, d.h. $0 \notin \sigma(A)$. Satz 6.5.15a zeigt $A^{-1} \in L(V', V)$, sodass die zweite Bedingung (6.5.4b) aus Lemma 6.5.3 folgt. Analog folgt (6.5.4a) aus (6.5.4b). ■

Offenbar bleibt $a(\cdot, \cdot)$ V-koerziv, wenn man $\text{const} \cdot (\cdot, \cdot)_U$ addiert. Generell gilt das

Lemma 6.5.18 $a(\cdot, \cdot)$ sei V -koerziv, wobei $V \subset U \subset V'$. Dann ist $a(\cdot, \cdot) + b(\cdot, \cdot)$ ebenfalls V -koerziv, wenn die Bilinearform $b(\cdot, \cdot)$ eine der folgenden Bedingungen a-c erfüllt: a) Zu jedem $\varepsilon > 0$ existiert C_ε , sodass

$$|b(x, x)| \leq \varepsilon \|x\|_V^2 + C_\varepsilon \|x\|_U^2 \quad \text{für alle } x \in V. \quad (6.5.12a)$$

b) Die Einbettungen $V \subset X$ und $V \subset Y$ seien stetig, und mindestens eine von ihnen sei kompakt. Ferner gelte

$$|b(x, x)| \leq C_B \|x\|_X \|x\|_Y \quad \text{für alle } x \in V. \quad (6.5.12b)$$

c) Die Einbettungen $V \subset X$, $V \subset Y$ seien stetig. Es gelte (6.5.12b). Für $\|\cdot\|_X$ oder $\|\cdot\|_Y$ sei vorausgesetzt: Zu jedem $\varepsilon > 0$ existiere ein C'_ε , sodass

$$\|x\|_X \leq \varepsilon \|x\|_V + C'_\varepsilon \|x\|_U \quad \text{oder} \quad \|x\|_Y \leq \varepsilon \|x\|_V + C'_\varepsilon \|x\|_U \quad (x \in V). \quad (6.5.12c)$$

Beweis. a) Wähle $\varepsilon = C_E/2$ in (6.5.12a) mit C_E aus (6.5.10). Dann erfüllt $a(\cdot, \cdot) + b(\cdot, \cdot)$ die V -Koerzitivitätsbedingung mit $C_E/2 > 0$ und $C_K + C_\varepsilon$ anstelle von C_E und C_K .

b) Lemma 6.4.13 beweist (6.5.12c), sodass nur noch Teil c) zu beweisen ist.

c) Es gelte z.B. die erste Ungleichung aus (6.5.12c). Da $V \subset Y$ stetig, existiert C_Y mit $\|x\|_Y \leq C_Y \|x\|_V$. Wähle $\varepsilon' = \varepsilon/(2C_B C_Y)$ in (6.5.12c):

$$|b(x, x)| \stackrel{(6.5.12b,c)}{\leq} C_B (\varepsilon' \|x\|_V + C'_{\varepsilon'} \|x\|_U) C_Y \|x\|_V \stackrel{(5.3.10)}{\leq} \frac{\varepsilon}{2} \|x\|_V^2 + \frac{1}{2\varepsilon} K^2 \|x\|_U^2 \quad \text{mit } K := C_B C_Y C'_{\varepsilon'}.$$

Damit ist (6.5.12a) gezeigt. ■

7

Variationsformulierung

7.1 Historische Bemerkungen

In den vorangegangenen Kapiteln konnte noch nicht einmal für das Dirichlet-Problem der Potentialgleichung (2.1.1a,b) geklärt werden, ob beziehungsweise unter welcher Bedingung eine klassische Lösung $u \in C^2(\Omega) \cap C^0(\overline{\Omega})$ existiert. Green [34] war der Ansicht, dass seine 1828 beschriebene Greensche Funktion stets existiert und die Lösung explizit liefert. Dies trifft nicht zu. Lebesgue bewies 1913, dass für gewisse Gebiete die Greensche Funktion nicht existiert.

Von Thomson (1847), Kelvin (1847) und Dirichlet stammt eine andere Argumentation. Das *Dirichletsche Integral*

$$J(u) := \int_{\Omega} |\nabla u|^2 dx = \int_{\Omega} \sum_{i=1}^n u_{x_i}^2(x) dx \quad (7.1.1)$$

beschreibt in der Physik die Energie. Bei vorgegebenen Randwerten $u = \varphi$ auf Γ sucht man $J(u)$ zu minimieren. Dieses Variationsproblem ist äquivalent zu

$$I(u, v) := \int_{\Omega} \langle \nabla u, \nabla v \rangle dx = \int_{\Omega} \sum_{i=1}^n u_{x_i}(x) v_{x_i}(x) dx \quad \text{für alle } v \text{ mit } v = 0 \text{ auf } \Gamma. \quad (7.1.2)$$

Der Beweis der Äquivalenz ergibt sich aus $J(u + v) = J(u) + 2I(u, v) + J(v)$ und $I(v) \geq 0$ für alle v (vgl. Satz 6.5.12). Die Greensche Formel (2.2.5a) liefert $I(u, v) = \int_{\Omega} v \Delta u dx = 0$ für alle v mit $v = 0$ auf Γ , sodass $\Delta u = 0$ folgt. Somit ist das Variationsproblem $J(u) = \min$ wie auch (7.1.2) äquivalent zum Dirichlet-Problem $\Delta u = 0$ in Ω , $u = \varphi$ auf Γ .

Das sogenannte Dirichletsche Prinzip besagt, dass $J(u)$, da es nach unten durch $I(u) \geq 0$ beschränkt ist, für ein u minimal werden muss. Nach den vorangehenden Überlegungen würde damit die Existenz einer Lösung des Dirichlet-Problems gesichert.

Weierstraß¹ wandte 1870 gegen diese Argumentation ein, dass zwar das Infimum von $J(u)$ über $\{u \in C^2(\Omega) \cap C^0(\overline{\Omega}) : u = \varphi \text{ auf } \Gamma\}$ existieren mag, aber nicht notwendigerweise in dieser Menge angenommen zu werden braucht. Beispielsweise nimmt das Integral $J(u) := \int_0^1 u^2(x) dx$ in $\{u \in C^0([0, 1]) : u(0) = 0, u(1) = 1\}$ nie den Wert $\inf J(u) = 0$ an.

Das folgende Beispiel von Hadamard² zeigt zudem, dass kein endliches Infimum des Dirichletschen Integrals zu existieren braucht. r und φ seien die Polarkoordinaten im Kreis $\Omega = K_1(0)$. Die Funktion $u(r, \varphi) = \sum_{n=1}^{\infty} r^{n!} n^{-2} \sin(n! \varphi)$ ist in Ω harmonisch, aber das Integral $J(u)$ existiert nicht.

¹Karl Theodor Wilhelm Weierstraß, geb. 31. Okt. 1815 in Ostenfelde, gest. 19. Febr. 1897 in Berlin

²Jacques Salomon Hadamard, am 8 Dez. 1865 in Versailles geboren und am 17 Okt. 1963 in Paris gestorben.

Die geschilderten Schwierigkeiten verschwinden, wenn man die Lösungen in den dazu geeigneten Sobolev-Räumen statt in $C^2(\Omega) \cap C^0(\overline{\Omega})$ sucht.

7.2 Gleichungen mit homogenen Dirichlet-Randbedingungen

Wir untersuchen im Folgenden die elliptische Gleichung

$$Lu = g \quad \text{in } \Omega, \quad (7.2.1a)$$

$$L = \sum_{|\alpha| \leq m} \sum_{|\beta| \leq m} (-1)^{|\beta|} D^\beta a_{\alpha\beta}(x) D^\alpha \quad (7.2.1b)$$

der Ordnung $2m$ (vgl. Abschnitt 5.3.2, Übungsaufgabe 5.3.5d). Der Hauptteil von L ist

$$L = (-1)^m \sum_{|\alpha|=|\beta|=m} D^\beta a_{\alpha\beta}(x) D^\alpha. \quad (7.2.2)$$

Gemäß Definition 5.3.4 ist L gleichmäßig elliptisch in $\overline{\Omega}$, falls $\varepsilon > 0$ existiert, sodass

$$\sum_{|\alpha|=|\beta|=m} a_{\alpha\beta}(x) \xi^{\alpha+\beta} \geq \varepsilon |\xi|^{2m} \quad \text{für alle } x \in \Omega \text{ und } \xi \in \mathbb{R}^n. \quad (7.2.3)$$

Sobald nur $a_{\alpha\beta} \in L^\infty(\overline{\Omega})$ vorausgesetzt wird, ist “für alle $x \in \Omega$ ” durch “für fast alle $x \in \Omega$ ” zu ersetzen.

Als Randbedingungen werden die *homogenen Dirichlet-Randwerte*

$$u = 0, \quad \frac{\partial u}{\partial n} = 0, \quad \left(\frac{\partial}{\partial n} \right)^2 u = 0, \quad \dots, \quad \left(\frac{\partial}{\partial n} \right)^{m-1} u = 0 \quad \text{auf } \Gamma \quad (7.2.4)$$

verlangt, die allerdings nur sinnvoll sind, falls $\Gamma = \partial\Omega$ hinreichend glatt ist (andernfalls ist die Normalenrichtung \vec{n} nicht definiert). Man beachte, dass für den Standardfall $m = 1$ (Gleichung zweiter Ordnung) Bedingung (7.2.4) zu $u = 0$ wird.

Da mit $u = 0$ auf Γ auch die Tangentialableitungen verschwinden, sind nicht nur die k -fachen Normalenableitungen ($k \leq m - 1$), sondern sämtliche Ableitungen der Ordnung $\leq m - 1$ gleich null:

$$D^\alpha u = 0 \quad \text{in } x \in \Gamma \text{ für } |\alpha| \leq m - 1. \quad (7.2.4')$$

Bedingung (7.2.4') verlangt nicht mehr die Existenz einer Normalenrichtung. Gemäß Zusatz 6.2.44 lässt sich (7.2.4') auch durch

$$u \in H_0^m(\Omega) \quad (7.2.4'')$$

formulieren.

Sei $u \in C^{2m}(\Omega) \cap H_0^m(\Omega)$ eine klassische Lösung von (7.2.1a) und (7.2.4). Zur Herleitung der Variationsformulierung bilden wir für ein beliebiges $v \in C_0^\infty(\Omega)$ das Skalarprodukt

$$(Lu, v)_0 = \sum_{|\alpha| \leq m} \sum_{|\beta| \leq m} (-1)^{|\beta|} \int_{\Omega} v(x) D^\beta a_{\alpha\beta}(x) D^\alpha u(x) dx.$$

Da $v \in C_0^\infty(\Omega)$, verschwindet der Integrand in der Nähe von Γ , sodass man partiell integrieren kann, ohne dass ein Randintegral auftritt:

$$(-1)^{|\beta|} \int_{\Omega} v(x) D^\beta a_{\alpha\beta}(x) D^\alpha u(x) dx = \int_{\Omega} a_{\alpha\beta}(x) [D^\alpha u(x)] [D^\beta v(x)] dx.$$

Damit erreicht man die Variationsformulierung

$$\sum_{|\alpha| \leq m} \sum_{|\beta| \leq m} \int_{\Omega} a_{\alpha\beta}(x) [D^\alpha u(x)] [D^\beta v(x)] dx = \int_{\Omega} g(x) v(x) dx \quad \text{für alle } v \in C_0^\infty(\Omega), \quad (7.2.5)$$

da $Lu = g$.

Erfüllt umgekehrt $u \in C^{2m}(\Omega)$ mit Randbedingung (7.2.4) die Forderung (7.2.5), so lässt sich die partielle Integration rückgängig machen, und $\int_{\Omega} (g - Lu)v dx = 0$ für alle $v \in C_0^{\infty}(\Omega)$ beweist $Lu = g$. Das heißt, eine klassische Lösung des Variationsproblems (7.2.5) mit Randbedingung (7.2.4) ist auch Lösung der ursprünglichen Randwertaufgabe. Somit sind die Differentialgleichung (7.2.1a,b) und die Variationsformulierung (7.2.5) bezüglich klassischer Lösungen äquivalent. Wir führen die Bilinearform

$$a(u, v) := \sum_{|\alpha| \leq m} \sum_{|\beta| \leq m} \int_{\Omega} a_{\alpha\beta}(x) [D^{\alpha}u(x)] [D^{\beta}v(x)] dx \quad (7.2.6)$$

und das Funktional

$$f(v) := \int_{\Omega} g(x)v(x) dx \quad (7.2.7)$$

ein. Wie oben bemerkt, bedeutet die Randbedingung (7.2.4) für klassische Lösungen $u \in C^{2m}(\Omega) \cap H^m(\Omega)$, dass $u \in H_0^m(\Omega)$. Damit lautet die "Variationsformulierung" oder "schwache Formulierung" der Randwertaufgabe (7.2.1), (7.2.4) wie folgt:

$$\text{Suche } u \in H_0^m(\Omega) \text{ mit } a(u, v) = f(v) \text{ für alle } v \in C_0^{\infty}(\Omega). \quad (7.2.8)$$

Eine Lösung der Aufgabe (7.2.8), die definitionsgemäß in $H_0^m(\Omega)$, aber nicht notwendigerweise in $C^{2m}(\Omega)$ liegt, heißt *schwache Lösung*.

Übungsaufgabe 7.2.1 a) Ω sei beschränkt. Man zeige: Eine klassische Lösung $u \in C^{2m}(\Omega) \cap C^m(\overline{\Omega})$ ist auch eine schwache Lösung.

b) Anhand des Beispiels 2.4.2 zeige man, dass diese Aussage für unbeschränkte Gebiete falsch wird.

Satz 7.2.2 Seien $a_{\alpha\beta} \in L^{\infty}(\Omega)$. Die durch (7.2.6) definierte Bilinearform ist auf $H_0^m(\Omega) \times H_0^m(\Omega)$ beschränkt.

Beweis. Seien $u, v \in C_0^{\infty}(\Omega)$. Die Ungleichung (6.2.5c) liefert

$$|a(u, v)| \leq \sum_{|\alpha| \leq m} \sum_{|\beta| \leq m} \|a_{\alpha\beta}\|_{L^{\infty}(\Omega)} |D^{\alpha}u|_0 |D^{\beta}v|_0 \leq \text{const } |u|_m |v|_m.$$

Da $C_0^{\infty}(\Omega)$ dicht in $H_0^m(\Omega)$ (vgl. Satz 6.2.11), ist $a(\cdot, \cdot)$ auf $H_0^m(\Omega) \times H_0^m(\Omega)$ fortsetzbar und durch die gleiche Konstante beschränkt (vgl. Lemma 6.5.1b). ■

Auch $f(v)$ ist für $v \in H_0^m(\Omega)$ definiert und beschränkt, wenn z.B. $g \in L^2(\Omega)$. Gemäß Übungsaufgabe 6.5.11 ist die Variationsformulierung (7.2.8) äquivalent zu der folgenden Variation über $H_0^m(\Omega)$:

$$\text{Suche } u \in H_0^m(\Omega) \text{ mit } a(u, v) = f(v) \text{ für alle } v \in H_0^m(\Omega). \quad (7.2.9)$$

Die Form $Lu = f$ gewinnt man zurück, wenn man Lemma 6.5.1 anwendet. $L \in L(H_0^m(\Omega), H^{-m}(\Omega))$ und $f \in H^{-m}(\Omega) = (H_0^m(\Omega))'$ seien durch $a(u, v) = \langle Lu, v \rangle_{H^{-m}(\Omega) \times H_0^m(\Omega)}$ und $f(v) = \langle f, v \rangle_{H^{-m}(\Omega) \times H_0^m(\Omega)}$ für alle $v \in H_0^m(\Omega)$ definiert. Gleichung (7.2.9) besagt, dass

$$Lu = f. \quad (7.2.9')$$

Während (7.2.1a) eine Gleichung $Lu = g$ in $C^0(\Omega)$ (d.h. für eine klassische Lösung) darstellt, ist (7.2.9') eine Gleichung in $H^{-m}(\Omega)$.

Satz 6.5.9 garantiert die eindeutige Lösbarkeit der Gleichung (7.2.9), falls $a(\cdot, \cdot)$ $H_0^m(\Omega)$ -elliptisch ist. Wir untersuchen zunächst den Standardfall $m = 1$ (Gleichungen der Ordnung $2m = 2$).

Satz 7.2.3 Seien Ω beschränkt, $m = 1$ und $a_{\alpha\beta} \in L^{\infty}(\Omega)$. L erfülle (7.2.3) (gleichmäßige Elliptizität) und stimme mit dem Hauptteil L_0 überein, d.h. $a_{\alpha\beta} = 0$ für $|\alpha| + |\beta| \leq 1$. Dann ist die Bilinearform $a(\cdot, \cdot)$ $H_0^1(\Omega)$ -elliptisch:

$$a(u, u) \geq \varepsilon' |u|_1^2 \quad \text{für alle } u \in H_0^1(\Omega), \text{ wobei } \varepsilon' > 0. \quad (7.2.10)$$

Beweis. Da $|\alpha| = |\beta| = 1$, lassen sich α und β gemäß $D^\alpha = \partial/\partial x_i$, $D^\beta = \partial/\partial x_j$ mit Indizes $i, j \in \{1, \dots, n\}$ identifizieren. Für festes $x \in \Omega$ wende man (7.2.3) mit $\xi = \nabla u(x)$ an:

$$\sum_{|\alpha|=|\beta|=1} a_{\alpha\beta}(x) \xi^{\alpha+\beta} = \sum_{i,j=1}^n a_{ij}(x) \xi_i \xi_j \stackrel{(7.2.3)}{\geq} \varepsilon |\xi|^2 = \varepsilon |\nabla u(x)|^2.$$

Integration über Ω liefert $a(u, u) \geq \varepsilon \int_\Omega |\nabla u|^2 dx$. Da $\int_\Omega |\nabla u|^2 dx \geq C_\Omega |u|_1^2$ (vgl. Lemma 6.2.12), folgt (7.2.10) mit $\varepsilon' = \varepsilon C_\Omega$. ■

Zusatz 7.2.4 Die Bedingung “ Ω beschränkt” kann entfallen, wenn für $\alpha = \beta = 0$ Positivität von $a_{00}(x) \geq \eta > 0$ (anstelle von $a_{00} = 0$ in Satz 7.2.3) vorausgesetzt wird.

Beispiel 7.2.5 Die Helmholtz³-Gleichung $-\Delta u + u = f$ in Ω führt auf die Bilinearform

$$a(u, v) := \int_\Omega \sum_{i=1}^n [u_{x_i}(x) v_{x_i}(x) + u(x) v(x)] dx = \int_\Omega \sum_{i=1}^n [\langle \nabla u, \nabla v \rangle + uv] dx.$$

$a(u, v)$ ist das Skalarprodukt in $H_0^1(\Omega)$ (und $H^1(\Omega)$). $a(u, u) = |u|_1^2$ beweist die $H_0^1(\Omega)$ -Elliptizität.

Übungsaufgabe 7.2.6 Die Voraussetzungen des Satzes 7.2.3 oder des Zusatzes 7.2.4 seien erfüllt, angenommen, dass die Koeffizienten $a_{\alpha 0}$ und $a_{0\beta}$ ($|\alpha| = |\beta| = 1$) der ersten Ableitungen beliebige Konstanten seien. Man zeige, dass Ungleichung (7.2.10) unverändert gilt.

Satz 7.2.3 lässt sich nicht ohne Weiteres auf $m > 1$ ausdehnen.

Satz 7.2.7 Die Hauptteilkoeffizienten seien konstant: $a_{\alpha\beta} = \text{const}$ für $|\alpha| = |\beta| = m$. Außerdem gelte $a_{\alpha\beta} = 0$ für $0 < |\alpha| + |\beta| \leq 2m - 1$ und $a_{00}(x) \geq 0$ für $\alpha = \beta = 0$. L sei gleichmäßig elliptisch (vgl. (7.2.3)). Ferner gelte entweder, dass Ω beschränkt ist oder dass $a_{00} \geq \eta > 0$. Dann ist $a(\cdot, \cdot)$ $H_0^m(\Omega)$ -elliptisch.

Beweis. Wir setzen $u \in H_0^m(\Omega)$ durch $u = 0$ auf \mathbb{R}^n fort. Satz 6.2.22b und Übungsaufgabe 6.2.23 zeigen

$$\begin{aligned} a(u, u) - \int_\Omega a_{00} u^2 dx &= \sum_{|\alpha|=|\beta|=m} \int_\Omega a_{\alpha\beta} D^\alpha u D^\beta v dx = \sum_{|\alpha|=|\beta|=m} a_{\alpha\beta} \int_{\mathbb{R}^n} D^\alpha u(x) D^\beta v(x) dx \\ &= \sum_{|\alpha|=|\beta|=m} a_{\alpha\beta} \cdot (D^\alpha u, D^\beta v)_0 = \sum_{|\alpha|=|\beta|=m} a_{\alpha\beta} \cdot (\widehat{D^\alpha u}, \widehat{D^\beta v})_0 = \sum_{|\alpha|=|\beta|=m} a_{\alpha\beta} \int_{\mathbb{R}^n} [(i\xi)^\alpha \hat{u}(\xi)] \overline{(i\xi)^\beta \hat{v}(\xi)} d\xi \\ &= \int_{\mathbb{R}^n} \left[\sum_{|\alpha|=|\beta|=m} a_{\alpha\beta} \xi^{\alpha+\beta} \right] |\hat{u}(\xi)|^2 d\xi \stackrel{(7.2.3)}{\geq} \varepsilon \int_{\mathbb{R}^n} |\xi|^{2m} |\hat{u}(\xi)|^2 d\xi. \end{aligned}$$

Sei $a_{00} \geq \eta > 0$ angenommen. Es gibt ein $\varepsilon' > 0$, sodass $\varepsilon |\xi|^{2m} \geq \varepsilon' \sum_{|\alpha| \leq m} |\xi^\alpha|^2 - \eta$ für alle $\xi \in \mathbb{R}^n$. Hieraus folgt $\varepsilon \int |\xi|^{2m} |\hat{u}(\xi)|^2 d\xi \geq \varepsilon' |u|_m^2 - \eta |u|_0^2$ (vgl. Lemma 6.2.24a) und $a(u, u) \geq \varepsilon' |u|_m^2$. Falls Ω beschränkt ist, benutze man Lemma 6.2.12. ■

Auf $H_0^m(\Omega)$ -elliptische Formen $a(\cdot, \cdot)$ ist Satz 6.5.9 anwendbar:

Satz 7.2.8 (Existenz und Eindeutigkeit schwacher Lösungen) Ist die Bilinearform $a(\cdot, \cdot)$ $H_0^m(\Omega)$ -elliptisch, so existiert genau eine Lösung $u \in H_0^m(\Omega)$ der Aufgabe (7.2.9). Sie erfüllt

$$|u|_m \leq \frac{1}{C_E} |f|_{-m} \quad (C_E \text{ aus (6.5.5)}). \quad (7.2.11)$$

Da (7.2.11) für alle $f \in H^{-m}(\Omega)$ und $u = L^{-1}f$ gilt (vgl. (7.2.9')), ist Ungleichung (7.2.11) äquivalent zu

$$\|L^{-1}\|_{H_0^m(\Omega) \leftarrow H^{-m}(\Omega)} \leq C := 1/C_E. \quad (7.2.11')$$

Die Variationsaufgabe (7.2.9) kann in eine Minimierungsaufgabe übergehen:

³Hermann Ludwig Ferdinand von Helmholtz, geb. am 31. August 1821 in Potsdam, gest. am 8. Sept. 1894 in Berlin

Satz 7.2.9 $a(\cdot, \cdot)$ sei eine $H_0^m(\Omega)$ -elliptische und symmetrische Bilinearform. Dann ist (7.2.9) äquivalent zum Minimierungsproblem

$$\text{Suche } u \in H_0^m(\Omega), \text{ sodass } J(u) \leq J(v) \text{ für alle } v \in H_0^m(\Omega), \quad (7.2.12a)$$

$$\text{wobei } J(v) := \frac{1}{2}a(v, v) - f(v). \quad (7.2.12b)$$

Achtung. Falls $a(\cdot, \cdot)$ nicht $H_0^m(\Omega)$ -elliptisch oder unsymmetrisch ist, bleibt das Problem (7.2.9) sinnvoll, obwohl seine Lösung das Funktional $J(u)$ nicht minimiert.

Beispiel 7.2.10 Die Poisson-Gleichung $-\Delta u = f$ in Ω , $u = 0$ auf Γ , führt auf die Bilinearform

$$a(u, v) = \int_{\Omega} \langle \nabla u(x), \nabla v(x) \rangle dx.$$

Für ein beschränktes Gebiet Ω ist $a(\cdot, \cdot)$ $H_0^1(\Omega)$ -elliptisch (vgl. Satz 7.2.3), sodass für jedes $f \in H^{-1}(\Omega)$ genau eine (schwache) Lösung $u \in H_0^1(\Omega)$ der Poisson-Gleichung existiert. Diese ist auch Lösung des Variationsproblems $\frac{1}{2} \int_{\Omega} |\nabla u|^2 dx - f(u) = \min$.

Eine schwächere Bedingung als die $H_0^m(\Omega)$ -Elliptizität ist die $H_0^m(\Omega)$ -Koerzivität: $a(u, u) \geq \varepsilon |u|_m^2 - C |u|_0^2$.

Satz 7.2.11 Sei $m = 1$. Die Koeffizienten $a_{\alpha\beta} \in L^\infty(\Omega)$ mögen die Bedingung (7.2.3) der gleichmäßigen Elliptizität erfüllen. Dann ist $a(\cdot, \cdot)$ $H_0^1(\Omega)$ -koerziv.

Beweis. Wir schreiben L als $L = L_I + L_{II}$, wobei L_I die Annahmen des Satzes 7.2.3 bzw. des Zusatzes 7.2.4 erfülle und L_{II} nur Ableitungen der Ordnung ≤ 1 enthalte. Dann lässt sich Lemma 7.2.12 anwenden. ■

Lemma 7.2.12 Sei $a(\cdot, \cdot) = a'(\cdot, \cdot) + a''(\cdot, \cdot)$ so zerlegt, dass $a'(\cdot, \cdot)$ $H_0^m(\Omega)$ -elliptisch oder auch nur $H_0^m(\Omega)$ -koerziv ist, während

$$a''(u, v) = \sum_{\substack{|\alpha| \leq m, |\beta| \leq m \\ |\alpha| + |\beta| < 2m}} \int_{\Omega} a_{\alpha\beta}(x) [D^\alpha u(x)] [D^\beta v(x)] dx$$

mit $a_{\alpha\beta} \in L^\infty(\Omega)$ nur Ableitungen der Ordnung $\leq 2m - 1$ enthält. Dann ist auch $a(\cdot, \cdot)$ $H_0^m(\Omega)$ -koerziv.

Beweis. (6.5.12c) folgt aus (6.2.10b) mit $X = H_0^{|\alpha|}(\Omega)$, $Y = H_0^{|\beta|}(\Omega)$, $V = H_0^m(\Omega)$ und $U = L^2(\Omega)$, sodass Lemma 6.5.18c die Behauptung beweist. ■

Die Verallgemeinerung des Satzes 7.2.11 auf beliebiges $m \geq 1$ verlangt stärkere Voraussetzungen an die Koeffizienten des Hauptteils.

Satz 7.2.13 (Gårding) Es gelte (7.2.3) und $a_{\alpha\beta} \in L^\infty(\Omega)$. Außerdem seien die Koeffizienten $a_{\alpha\beta}$ mit $|\alpha| = |\beta| = m$ gleichmäßig stetig in Ω . Dann ist $a(\cdot, \cdot)$ $H_0^m(\Omega)$ -koerziv. Gilt umgekehrt $a_{\alpha\beta} \in C(\Omega)$ für $|\alpha| = |\beta| = m$ und $a_{\alpha\beta} \in L^\infty(\Omega)$ sonst, so folgt aus der $H_0^m(\Omega)$ -Koerzivität die gleichmäßige Elliptizität (7.2.3).

Der Beweis kann z.B. bei Wloka [89, p. 282] nachgelesen werden. Für den ersten Teil des Satzes verwendet man die Zerlegung der Eins.

Die Bedeutung der Koerzivität liegt in der folgenden Aussage.

Satz 7.2.14 Sei Ω beschränkt und $a(\cdot, \cdot)$ $H_0^m(\Omega)$ -koerziv. Dann gilt eine der folgenden Alternativen:

- (i) Das Problem (7.2.9) hat genau eine (schwache) Lösung $u \in H_0^m(\Omega)$.
- (ii) Die Kerne $E = \text{Kern}(L)$ und $E^* = \text{Kern}(L')$ sind k -dimensional für ein $k \in \mathbb{N}$, d.h.

$$a(e, v) = 0, \quad a(v, e^*) \quad \text{für alle } e \in E, e^* \in E^*, v \in H_0^m(\Omega).$$

Ferner hat die Eigenwertaufgabe

$$a(e, v) = \lambda(e, v)_{L^2(\Omega)} \quad \text{für alle } v \in H_0^m(\Omega) \quad (7.2.13)$$

abzählbar viele Eigenwerte, die sich nicht in \mathbb{C} häufen.

Beweis. Da für beschränktes Ω die Einbettung $V := H_0^m(\Omega) \subset U := L^2(\Omega)$ kompakt ist, ist Satz 6.5.15 anwendbar. ■

7.3 Inhomogene Dirichlet-Randbedingung

Als Nächstes betrachten wir die Randwertaufgabe

$$Lu = g \quad \text{in } \Omega, \quad u = \varphi \quad \text{auf } \Gamma, \quad (7.3.1)$$

wobei L ein Differentialoperator zweiter Ordnung sei (d.h. $m = 1$). Die Variationsformulierung der Randwertaufgabe (7.3.1) lautet:

$$\text{Suche } u \in H^1(\Omega) \text{ mit } u = \varphi \text{ auf } \Gamma, \quad \text{sodass } a(u, v) = f(v) \quad \text{für alle } v \in H_0^1(\Omega). \quad (7.3.2)$$

Gemäß Abschnitt 6.2.5 ist die Beschränkung $u|_\Gamma$ von $u \in H^1(\Omega)$ auf Γ aber wohldefiniert als Funktion des $H^{1/2}(\Gamma)$. Daher ist “ $u = \varphi$ auf Γ ” als die Gleichheit $u|_\Gamma = \varphi$ in $H^{1/2}(\Gamma)$ zu verstehen. Im Gegensatz zum vorhergehenden Abschnitt wird $a(\cdot, \cdot)$ in (7.3.2) als Bilinearform auf $H^1(\Omega) \times H_0^1(\Omega)$ benutzt. Man überzeugt sich leicht, dass $a(\cdot, \cdot)$ auf diesem Produkt wohldefiniert und beschränkt ist.

Bemerkung 7.3.1 *Notwendig für die Lösbarkeit der Aufgabe (7.3.2) ist:*

$$\text{es existiert ein } u_0 \in H^1(\Omega) \quad \text{mit} \quad u_0|_\Gamma = \varphi. \quad (7.3.3)$$

Ist eine Funktion u_0 mit der Eigenschaft (7.3.3) bekannt, ergibt sich eine zweite Charakterisierung der schwachen Lösung:

$$u_0 \text{ erfülle (7.3.3),} \quad \text{suche } w \in H_0^1(\Omega), \text{ sodass} \quad (7.3.4a)$$

$$a(w, v) = f'(v) := f(v) - a(u_0, v) \quad \text{für alle } v \in H_0^1(\Omega). \quad (7.3.4b)$$

Bemerkung 7.3.2 *Die Variationsaufgaben (7.3.2) und (7.3.4a,b) sind äquivalent. Sind u_0 und w die Lösung von (7.3.4a,b), so erfüllt $u = u_0 + w$ die Aufgabe (7.3.2). Ist u eine Lösung von (7.3.2), erfüllen z.B. $u_0 = u$ und $w = 0$ die Aufgabe (7.3.4a,b).*

Übungsaufgabe 7.3.3 *Man zeige $f' \in H^{-1}(\Omega)$ für f' aus (7.3.4b) und*

$$|f'|_{-1} \leq |f|_{-1} + C_S |u_0|_1 \quad (7.3.5)$$

mit C_S aus $|a(u, v)| \leq C_S |u|_1 |v|_1$ (vgl. (6.5.1)).

Bemerkung 7.3.4 *Die Aufgabe (7.3.1) und die Variationsformulierung (7.3.2) haben die gleichen klassischen Lösungen, falls solche existieren.*

Beweis. Es genügt $v \in C_0^\infty(\Omega)$ in (7.3.2) anzunehmen. Die partielle Integration lässt sich wie in Abschnitt 7.2 durchführen und beweist die Behauptung. ■

Satz 7.3.5 (Existenz und Eindeutigkeit) *Die Aufgabe (7.2.9) (mit homogenen Randwerten) sei für alle $f \in H^{-1}(\Omega)$ eindeutig lösbar. Dann ist Bedingung (7.3.3) hinreichend und notwendig für die eindeutige Lösbarkeit der Aufgabe (7.3.2).*

Beweis. Existiert eine Lösung $u \in H^1(\Omega)$ von (7.3.2), so ist (7.3.3) erfüllt. Gilt dagegen (7.3.3), erhält man mittels (7.3.4a,b) eine eindeutige Lösung, da (7.3.4b) mit (7.2.9) übereinstimmt. ■

Bemerkung 7.3.6 *Unter der Bedingung $\Omega \in C^{0,1}$ ist (7.3.3) äquivalent zu*

$$\varphi \in H^{1/2}(\Gamma). \quad (7.3.6)$$

Beweis. Ist $\varphi \in H^{1/2}(\Gamma)$, so garantiert der Satz 6.2.41 eine Fortsetzung $u_0 \in H^1(\Omega)$ auf Ω mit $u_0|_\Gamma = \varphi$ und

$$|u|_1 \leq C |\varphi|_{1/2}. \quad (7.3.7)$$

Erfüllt umgekehrt u_0 die Bedingung (7.3.3), so zeigt gleiche Satz, dass $\varphi \in H^{1/2}(\Gamma)$. ■

Es gelte Ungleichung (7.2.11') im Falle homogener Randwerte. Gleichung (7.3.5) zeigt $|u|_1 \leq |u_0|_1 + |w|_1 \leq |u_0|_1 + (|f|_{-1} + C' |u_0|_1)/\varepsilon$ für die Lösung der Aufgabe (7.3.4a,b). Abschätzung (7.3.7) beweist den

Satz 7.3.7 Sei $\Omega \in C^{0,1}$. Die Bilinearform sei auf $H^1(\Omega) \times H_0^1(\Omega)$ beschränkt und erfülle (7.2.11'). Dann existiert zu jedem $f \in H^{-1}(\Omega)$ und $\varphi \in H^{1/2}(\Gamma)$ genau eine Lösung $u \in H^1(\Omega)$ der Aufgabe (7.3.2) mit

$$|u|_1 \leq C \left[|f|_{-1} + |\varphi|_{1/2} \right]. \quad (7.3.8)$$

Übungsaufgabe 7.3.8 $a(\cdot, \cdot)$ sei symmetrisch und $H_0^1(\Omega)$ -elliptisch. Man zeige: Aufgabe (7.3.2) mit $f=0$ ist äquivalent zum Variationsproblem: Suche $u \in H^1(\Omega)$ mit $u|_\Gamma = \varphi$, sodass $a(u, u)$ minimal ist (vgl. (7.1.1)).

7.4 Natürliche Randbedingungen

Die in (7.2.6) definierte Bilinearform $a(\cdot, \cdot)$ ist auch auf $H^m(\Omega) \times H^m(\Omega)$ wohldefiniert. In Analogie zu Satz 7.2.2 gilt der

Satz 7.4.1 Seien $a_{\alpha\beta} \in L^\infty(\Omega)$. Die durch (7.2.6) definierte Bilinearform ist auf $H^m(\Omega) \times H^m(\Omega)$ beschränkt: $|a(u, v)| \leq \sum_{\alpha, \beta} \|a_{\alpha\beta}\|_{L^\infty(\Omega)} |u|_m |v|_m$ für alle $u, v \in H^m(\Omega)$.

f sei jetzt ein Funktional aus $(H^m(\Omega))'$. (7.2.7) mit $g \in L^2(\Omega)$ beschreibt z.B. ein solches Funktional, aber (7.2.7) ist nur ein Spezialfall des nachfolgend in (7.4.1a) definierten Funktionals f , das wir im Weiteren zugrundelegen wollen.

Übungsaufgabe 7.4.2 Γ sei hinreichend glatt, und es gelte $g \in L^2(\Omega)$, $\varphi \in L^2(\Gamma)$. Man zeige, dass durch

$$f(v) := \int_{\Omega} g(x)v(x)dx + \int_{\Gamma} \varphi(x)v(x)d\Gamma \quad (v \in H^1(\Omega)) \quad (7.4.1a)$$

ein Funktional aus $(H^1(\Omega))'$ mit $\|f\|_{(H^1(\Omega))'} \leq C(\|g\|_{L^2(\Omega)} + \|\varphi\|_{L^2(\Gamma)})$ gegeben ist. Dies impliziert $f \in (H^m(\Omega))'$ für alle $m > 1$. Genauer gilt sogar

$$\|f\|_{(H^1(\Omega))'} \leq C \left[\|g\|_{(H^1(\Omega))'} + \|\varphi\|_{H^{-1/2}(\Gamma)} \right]. \quad (7.4.1b)$$

Häufig haben Variationsprobleme zu Differentialgleichung der Ordnung $2m$ die einfache Form:

$$\text{suche } u \in H^m(\Omega), \text{ sodass } a(u, v) = f(v) \quad \text{für alle } v \in H^m(\Omega). \quad (7.4.2)$$

Im Gegensatz zur Bedingung $u \in H_0^m(\Omega)$ aus Abschnitt 7.2 enthält $u \in H^m(\Omega)$ keine Randbedingung. Trotzdem hat die Aufgabe (7.4.2) eine eindeutige Lösung, wenn $a(\cdot, \cdot)$ $H^m(\Omega)$ -elliptisch ist. Diese Bedingung ist leicht zu erfüllen.

Satz 7.4.3 Unter den Voraussetzungen des Zusatzes 7.2.4 ist $a(\cdot, \cdot)$ $H^1(\Omega)$ -elliptisch: $a(u, u) \geq \varepsilon |u|_1^2$ für alle $u \in H^1(\Omega)$. Aufgabe (7.4.2) (mit $m = 1$) hat genau eine Lösung, die der Abschätzung (7.4.3) genügt:

$$|u|_1 \leq \frac{1}{\varepsilon} \|f\|_{(H^1(\Omega))'}. \quad (7.4.3)$$

Beweis. Ebenso wie für Zusatz 7.2.4 und Satz 7.2.8. ■

Zusatz 7.4.4 a) Eine eindeutige Lösung, die die Abschätzung (7.4.3) erfüllt, existiert auch, wenn man anstelle der $H^1(\Omega)$ -Elliptizität voraussetzt: $a(\cdot, \cdot)$ ist $H^1(\Omega)$ -koerziv, $\Omega \in C^{0,1}$ ist beschränkt und $\lambda = 0$ ist kein Eigenwert (d.h. $a(u, v) = 0$ für alle $v \in H^1(\Omega)$ impliziert $u = 0$).

b) Die Kombination der Ungleichungen (7.4.3) und (7.4.1b) ergibt

$$|u|_1 \leq C \left[\|g\|_{(H^1(\Omega))'} + \|\varphi\|_{H^{-1/2}(\Gamma)} \right]$$

für die Lösung von (7.4.2), wenn f durch (7.4.1a) mit $g \in (H^1(\Omega))'$ und $\varphi \in H^{-1/2}(\Gamma)$ definiert ist.

Beweis. Gemäß Satz 6.4.8b ist $H^1(\Omega)$ kompakt in $L^2(\Omega)$ eingebettet, sodass sich die Aussage des Satzes 7.2.14 übertragen lässt. Ist $\lambda = 0$ kein Eigenwert, so gilt $L^{-1} \in L((H^1(\Omega))', H^1(\Omega))$ (vgl. Satz 6.4.12). ■

Um herauszufinden, welche klassische Randwertaufgabe der Variationsformulierung (7.4.2) entspricht, nehmen wir an, dass (7.4.2) eine klassische Lösung $u \in H^m(\Omega) \cap C^{2m}(\overline{\Omega})$ besitzt. Weiterhin kann $v \in C^\infty(\Omega)$ angenommen werden (vgl. Lemma 6.5.1b). Der Einfachheit halber beschränken wir uns auf den Fall $m = 1$. Unter der Voraussetzung $a_{\alpha\beta} \in C^1(\Omega)$ und unter geeigneten Bedingungen an Ω ist die folgende allgemeine Greensche Formel anwendbar:

$$\begin{aligned} a(u, v) &= \int_{\Omega} \left[\sum_{i,j=1}^n a_{ij} u_{x_i} v_{x_j} + \sum_{i=1}^n a_{0i} u v_{x_i} + \sum_{i=1}^n a_{i0} u_{x_i} v + a_{00} u v \right] dx \\ &= \int_{\Omega} \left[- \sum_{i,j=1}^n (a_{ij} u_{x_i})_{x_j} - \sum_{i=1}^n (a_{0i} u)_{x_i} + \sum_{i=1}^n a_{i0} u_{x_i} + a_{00} u \right] v dx + \int_{\Gamma} \left[\sum_{i,j=1}^n n_j a_{ij} u_{x_i} + \sum_{i=1}^n n_i a_{0i} u \right] v d\Gamma. \end{aligned} \quad (7.4.4)$$

Dabei sind n_i die Komponenten der Normalenrichtung $\vec{n} = \vec{n}(x)$, $x \in \Gamma$. Wir definieren den *Rand-Differentialoperator*

$$B := \sum_{i,j=1}^n n_j a_{ij} \frac{\partial}{\partial x_i} + \sum_{i=1}^n n_i a_{0i}, \quad (7.4.5)$$

während L durch (7.2.1b) beschrieben sei. Gleichung (7.4.4) wird zu $a(u, v) = \int_{\Omega} v L u dx + \int_{\Gamma} v B u d\Gamma$. Nach Aufgabenstellung stimmt $a(u, v)$ mit $f(v)$ aus (7.4.1a) überein. Wählen wir zunächst $v \in H_0^1(\Omega) \subset H^1(\Omega)$, so entfallen die Randintegrale, und wir erhalten $Lu = g$ als notwendige Bedingung wie in Abschnitt 7.2. Damit reduziert sich die Identität $a(u, v) = f(v)$ zu $\int_{\Gamma} v B u d\Gamma = \int_{\Gamma} \varphi v d\Gamma$ für alle $v \in H^1(\Omega)$. Nach Satz 6.2.41b durchläuft $v|_{\Gamma}$ die Menge $H^{1/2}(\Gamma)$, wenn v den Raum $H^1(\Omega)$ durchläuft, sodass $\int_{\Gamma} \psi (Bu - \varphi) d\Gamma = 0$ für alle $\psi \in H^{1/2}(\Gamma)$ gilt; also $Bu = \varphi$. Dies beweist den

Satz 7.4.5 Γ sei hinreichend glatt. Eine klassische Lösung der Aufgabe (7.4.2) mit f aus (7.4.1a) ist auch klassische Lösung der Randwertaufgabe

$$Lu = g \quad \text{in } \Omega, \quad Bu = \varphi \quad \text{auf } \Gamma \quad (7.4.6)$$

und umgekehrt.

Die Bedingung $Bu = 0$ nennt man die *natürliche Randbedingung*. Sie ergibt sich aus der Tatsache, dass in (7.4.2) (anders als in (7.2.9)) die Funktion v beliebige Randwerte annehmen kann. Im Gegensatz dazu nennt man die Dirichlet-Bedingung $u|_{\Gamma} = 0$ auch gelegentlich eine Zwangsrandbedingung, da sie durch die Wahl $u \in H_0^1(\Omega)$ aufgezwungen wird. Man beachte, dass die Bilinearform sowohl L als auch B festlegt.

Übungsaufgabe 7.4.6 Man zeige: Die Bilinearform aus Beispiel 7.2.10 für $-\Delta u = g$ hat als natürliche Randbedingung die Neumann-Bedingung $-\partial u / \partial n = 0$.

Satz 7.4.7 Seien $\Omega \in C^{0,1}$ beschränkt und $a(\cdot, \cdot)$ $H^m(\Omega)$ -koerziv. Dann gelten die Aussagen des Satzes 7.2.14 mit $H^m(\Omega)$ anstelle von $H_0^m(\Omega)$.

Beispiel 7.4.8 $\Omega \in C^{0,1}$ sei ein beschränktes Gebiet. Die zur Helmholtz-Gleichung

$$-\Delta u + cu = f \quad \text{in } \Omega \text{ mit } c > 0, \quad \partial u / \partial n = \varphi \quad \text{auf } \Gamma$$

gehörende Bilinearform $a(u, v) = \int_{\Omega} [\langle \nabla u, \nabla v \rangle + cuv] dx$ ist $H^1(\Omega)$ -elliptisch, da $a(u, u) \geq \min(1, c) |u|_1^2$. Für $c = 0$ ist $a(\cdot, \cdot)$ dagegen nur $H^1(\Omega)$ -koerziv. Wie aus Satz 3.4.1 bekannt, ist die Neumann-Randwertaufgabe für die Poisson-Gleichung (d.h. für $c = 0$) nicht eindeutig lösbar. Gemäß Alternative (ii) aus 7.2.14 gibt es einen nichttrivialen Kern $E = \text{Kern}(L)$. $u \in E$ erfüllt $a(u, u) = \int_{\Omega} \langle \nabla u, \nabla u \rangle dx = 0$, also $\nabla u = 0$. Da Ω zusammenhängend ist, folgt $u(x) = \text{const}$ und somit $\dim E = 1$. Da $a(\cdot, \cdot)$ symmetrisch ist, stimmt $E^* := \text{Kern}(L')$ mit E überein. Nach Satz 7.4.7 ist die Neumann-Randwertaufgabe $a(u, v) = f(v)$ ($v \in H^1(\Omega)$) genau dann lösbar, wenn $f \perp E$, d.h. $f(1) = 0$. Falls $f(v) := \int_{\Omega} g(x)v(x)dx$, liest sich $f(1) = 0$ als $\int_{\Omega} g dx = 0$. Ist f dagegen durch (7.4.1a) gegeben, lautet die Integrabilitätsbedingung $f(1) = \int_{\Omega} g dx + \int_{\Gamma} \varphi d\Gamma = 0$ (das ist Gleichung (3.4.2), in der f durch $-g$ zu ersetzen ist).

Bemerkung 7.4.9 Während die klassische Formulierung einer Randbedingung wie $\partial u / \partial n = 0$ Voraussetzungen an den Rand Γ erfordert, kann die Aufgabe (7.4.2) für beliebige, messbare Ω formuliert werden.

Im Folgenden wollen wir den umgekehrten Weg gehen und fragen: Gibt es zu einem klassisch formulierten Randwertproblem $Lu = g$ in Ω , $Bu = \varphi$ auf Γ mit vorgegebenen L und B eine Bilinearform $a(\cdot, \cdot)$, sodass (7.4.2) die zugehörige Variationsformulierung ist? Dies würde bedeuten, dass der frei vorgegebene Randoperator B die natürliche Randbedingung darstellt.

Für $m = 1$ lautet die allgemeine Form eines Randoperators

$$B = \sum_{i=1}^n b_i(x) \frac{\partial}{\partial x_i} + b_0(x) \quad (x \in \Gamma). \quad (7.4.7)$$

Mit $\vec{b}^\top = (b_1, \dots, b_n)$ lässt sich auch $B = \vec{b}^\top \nabla + b_0$ schreiben (vgl. (5.2.1b, b')). Wir müssen ausschließen, dass $\vec{b}^\top \nabla$ eine Tangentialableitung darstellt (vgl. Bemerkung 5.2.2):

$$\langle \vec{b}(x), \vec{n}(x) \rangle \neq 0 \quad \text{für alle } x \in \Gamma. \quad (7.4.8)$$

Bemerkung 7.4.10 Sei $m = 1$. Es gelte (7.4.8). $A(x)$ sei die Matrix $A = (a_{ij})$ (vgl. (5.1.1c)). Durch den Übergang von $Bu = \varphi$ zur äquivalenten, skalierten Gleichung $\sigma Bu = \sigma \varphi$ mit $\sigma(x) = \langle \vec{n}(x), A(x) \vec{n}(x) \rangle / \langle \vec{b}(x), \vec{n}(x) \rangle$ lässt sich $\langle \vec{n}, \sigma \vec{b} \rangle = \langle \vec{n}, A \vec{n} \rangle$ erreichen. Daher wird im Weiteren stets angenommen, dass \vec{b} bereits $\langle \vec{n}, \vec{b} \rangle = \langle \vec{n}, A \vec{n} \rangle$ erfüllt.

Beweis. Wegen (7.4.8) ist σ wohldefiniert. Damit $Bu = \varphi$ und $\sigma Bu = \sigma \varphi$ äquivalent sind, muss $\sigma \neq 0$ gelten. Dies wird durch die gleichmäßige Elliptizität garantiert: $\langle n, An \rangle \geq \varepsilon |n|^2 = \varepsilon$. ■

Satz 7.4.11 (Konstruktion der Bilinearform) Sei $m = 1$. L und B seien durch (7.2.1b) und (7.4.7) vorgegeben, wobei \vec{b} Bedingung (7.4.8) erfülle. Dann gibt es eine Bilinearform $a(\cdot, \cdot)$ auf $H^1(\Omega) \times H^1(\Omega)$, sodass der Variationsaufgabe (7.4.2) die klassische Formulierung $Lu = g$ in Ω , $Bu = \varphi$ auf Γ entspricht.

Beweis. Die gesuchte Bilinearform ist keineswegs eindeutig festgelegt. Wir werden zwei verschiedene Konstruktionsmöglichkeiten angeben. Zunächst wird der absolute Term in (7.4.7) diskutiert. Auf Grund der Bemerkung 7.4.10 wird $\langle \vec{n}, \vec{b} \rangle = \langle \vec{n}, A \vec{n} \rangle$ angenommen.

1) Die Vektorfunktion $\beta(x) := (\beta_1(x), \dots, \beta_n(x)) \in C^1(\overline{\Omega})$ sei beliebig. Der Differentialoperator

$$L_1 := - \sum_{i=1}^n \frac{\partial}{\partial x_i} \beta_i + \sum_{i=1}^n \beta_i \frac{\partial}{\partial x_i} + \sum_{i=1}^n (\beta_i)_{x_i}$$

bildet jedes $u \in C^1(\Omega)$ in null ab: $L_1 u = 0$ (Produktregel der Differentiation). Daher darf L durch $L + L_1$ ersetzt werden, ohne die Randwertaufgabe zu ändern. Sei $a(\cdot, \cdot)$ gemäß (7.2.6) aus den Koeffizienten von $L + L_1$ konstruiert. Gleichung (7.4.5) zeigt, dass der zu $a(\cdot, \cdot)$ gehörende Randoperator den absoluten Term

$$\sum_{i=1}^n n_i (a_{0i} + \beta_i) \quad (7.4.9)$$

besitzt. Falls $b_0 = 0$, führt die Wahl $\beta_i = -a_{0i}$ zum Ziel. Andernfalls stehen zwei Möglichkeiten offen.

1a) Man wähle β_i so, dass auf Γ $\beta_i(x) = b_0(x)n_i(x) - a_{0i}(x)$ gilt. Da $|n| = 1$, stimmt dann der Term (7.4.9) mit $b_0(x)$ überein. Die praktische Schwierigkeit dieser Methode besteht darin, eine glatte Fortsetzung der Randwerte $\beta_i(x)$, $x \in \Gamma$, auf Ω zu konstruieren.

1b) Man setze $\beta_i = -a_{0i}$ und füge ein geeignetes Randintegral hinzu:

$$a(u, v) := \int_{\Omega} \left\{ \sum_{i,j=1}^n a_{ij} u_{x_i} v_{x_j} + \sum_{i=1}^n (a_{i0} - a_{0i}) u_{x_i} v + \left[a_{00} + \sum_{i=1}^n (a_{0i})_{x_i} \right] uv \right\} dx + \int_{\Gamma} b_0 u v d\Gamma. \quad (7.4.10)$$

Die oben beschriebene partielle Integration zeigt, dass der zu (7.4.10) gehörende Randoperator

$$\tilde{B} = \sum_{i,j=1}^n n_j a_{ij} \frac{\partial}{\partial x_i} + b_0 \quad (7.4.11)$$

lautet.

2) Der Operator (7.4.11) lässt sich in der Form $\tilde{B} = \tilde{b}^\top \nabla + b_0$ mit $\tilde{b} = A\tilde{n}$ schreiben. Da schon $\langle \tilde{n}, \tilde{b} \rangle = \langle \tilde{n}, A\tilde{n} \rangle = \langle \tilde{n}, \tilde{b} \rangle$ vorausgesetzt war, ist $d := \tilde{b} - \tilde{b}$ orthogonal zu \tilde{n} . Um \tilde{b} in \tilde{b} zu ändern, gibt es wieder zwei Konstruktionsmöglichkeiten.

2a) Definiere die $n \times n$ -Matrix A^s auf Γ durch $A^s = dn^\top - nd^\top$, d.h. $a_{ij}^s = d_i n_j - n_i d_j$. $A^s(x)$ ist schiefssymmetrisch: $A^{s\top} = -A^s$. Man setze die Matrix $A^s(x)$, die zunächst nur auf Γ definiert ist, zu einer schiefssymmetrischen Matrix $A^s \in C^1(\Omega)$ fort. [Hier liegt die gleiche praktische Schwierigkeit wie im Schritt 1a.)] Die Koeffizienten von A^s definieren

$$L_2 := \sum_{i,j=1}^n \left[-\frac{\partial}{\partial x_j} a_{ij}^s \frac{\partial}{\partial x_i} + (a_{ij}^s) \frac{\partial}{\partial x_i} \right].$$

Wieder gilt $L_2 u = 0$ für alle $u \in C^2(\Omega)$, da

$$-(a_{ij}^s u_{x_i})_{x_j} - (a_{ji}^s u_{x_j})_{x_i} + (a_{ij}^s)_{x_j} u_{x_i} + (a_{ji}^s)_{x_i} u_{x_j} = -a_{ij}^s u_{x_i x_j} - a_{ji}^s u_{x_j x_i} = 0.$$

Somit kann L durch $L + L_2$ ersetzt werden, ohne die Randwertaufgabe zu ändern. Die zu $L + L_2$ gehörenden Koeffizienten ergeben einen Randoperator B , dessen Ableitungsterme $\sum_{i,j} n_j (a_{ij} + a_{ij}^s) \partial / \partial x_i = [(A + A^s)\tilde{n}]^\top \nabla$ lauten. Nach Konstruktion von A^s gilt $A^s \tilde{n} = (d\tilde{n}^\top - \tilde{n}d^\top)\tilde{n} = d = \tilde{b} - \tilde{b}$, da $\langle \tilde{n}, \tilde{n} \rangle = 1$ und $\langle d, \tilde{n} \rangle = 0$. Zusammen mit $A\tilde{n} = \tilde{b}$ ergibt sich der Ableitungsterm in B wie gewünscht zu $\tilde{b}^\top \nabla$. Der Übergang $L \rightarrow L + L_2$ ändert den absoluten Term in B nicht, sodass aus (7.4.11) $B = \tilde{b}^\top \nabla + b_0$ folgt.

2b) Gegeben sei $B = \tilde{b}^\top \nabla + b_0$ (vgl. (7.4.11)), sodass $d = \tilde{b} - \tilde{b}$ orthogonal zu \tilde{n} ist. Damit ist der Randoperator $T := d^\top \nabla$ eine Richtungsableitung in Tangentialrichtung (falls $n = 2$) bzw. in der Tangentialhyperebene (falls $n \geq 3$). Ist f hinreichend glatt, so ist die Beschränkung $v|_\Gamma$ von $v \in H^1(\Omega)$ ein Element des $H^{1/2}(\Gamma)$. Auf Grund von Bemerkung 6.3.14a lässt sich zeigen, dass $T \in L(H^{1/2}(\Gamma), H^{-1/2}(\Gamma))$. Da $T(u|_\Gamma) \in H^{-1/2}(\Gamma)$, ist $\int_\Gamma \psi T(u|_\Gamma) d\Gamma$ für $\psi \in H^{1/2}(\Gamma)$ wohldefiniert, insbesondere für $\psi = v|_\Gamma$, wenn $v \in H^1(\Omega)$. Somit ist

$$b(u, v) := \int_\Gamma (v|_\Gamma) T(u|_\Gamma) d\Gamma$$

eine auf $H^1(\Omega) \times H^1(\Omega)$ beschränkte Bilinearform. Wir addieren $b(\cdot, \cdot)$ zu $a(\cdot, \cdot)$ aus (7.4.10). Partielle Integration liefert den Randoperator $\tilde{B} + T = \tilde{b}^\top \nabla + b_0 + (\tilde{b} - \tilde{b})^\top \nabla = \tilde{b}^\top \nabla + b_0 = B$. ■

Satz 7.4.12 Die Bilinearform (7.2.6) sei $H^1(\Omega)$ -koerziv (vgl. Satz 7.4.3). Ihre Koeffizienten ebenso wie der Rand Γ seien hinreichend glatt ($\in C^1$). Dann ergeben die Konstruktionen des vorhergehenden Beweises wieder eine $H^1(\Omega)$ -koerzive Form.

Beweis. 1) Im Schritt 1a) werden nur Terme niedriger Ordnung addiert, sodass Lemma 7.2.12 anwendbar ist. Bezüglich 1b) vergleiche 2b).

2a) Im Teil 2a) wird $b(u, v) := \int_\Omega \sum_{i,j} [a_{ij}^s u_{x_i} v_{x_j} + (a_{ij}^s)_{x_j} v u_{x_i}] dx$ addiert. Auch hier ist Lemma 7.2.12 anwendbar, denn die Schiefssymmetrie von A^s ergibt $b(u, u) = \int_\Omega \sum (a_{ij}^s)_{x_j} u u_{x_i} dx$.

2b) Im Falle der Konstruktion im Teil 2b) ist Lemma 7.2.12 sinngemäß anwendbar. Man macht sich das am einfachsten im Falle von $n = 2$ klar. Γ sei durch $\{(x_1(s), x_2(s)) : 0 \leq s \leq 1\}$ beschrieben. Sind $x_1, x_2 \in C^1([0, 1])$ und $x'_i(0) = x'_i(1)$, und gilt $d \in C^1(\Gamma)$ für d aus $T = d^\top \nabla$, so hat $b(\cdot, \cdot)$ die Darstellung $b(u, v) = \int_0^1 \bar{v}(s) \tau(s) \bar{u}'(s) ds$, wobei $\bar{u}(s) = u(x_1(s), x_2(s))$, $\bar{v}(s) = v(x_1(s), x_2(s))$, $\tau \in C^1([0, 1])$. Auf Grund der Periodizität ergibt partielle Integration $b(u, v) = -\int_0^1 (\tau \bar{v})' \bar{u} ds$ ohne Randterme, sodass $b(u, u) = \frac{1}{2} [\int_0^1 \bar{u} \tau \bar{u}' ds - \int_0^1 (\tau \bar{u})' \bar{u} ds] = -\frac{1}{2} \int_0^1 \tau' \bar{u}^2 ds$ folgt. Das impliziert

$$|b(u, u)| \leq \frac{1}{2} \|\tau\|_{C^1([0,1])} \|u|_\Gamma\|_{L^2(\Omega)}^2 \leq C \|\tau\|_{C^1([0,1])} \|u|_\Gamma\|_{H^s(\Gamma)}^2$$

für jedes $s \in (1/2, 1)$ (vgl. Satz 6.2.41a). Da $|u|_s^2 \leq \varepsilon |u|_1^2 - C |u|_0^2$, lässt sich Lemma 6.5.18c anwenden. ■

Der Fall $m \geq 2$ wurde (abgesehen von Satz 7.4.1) in diesem Abschnitt ausgeklammert. Randwertaufgaben der Ordnung $2m$ brauchen m Randbedingungen $B_j u = \varphi_j$ auf Γ ($j = 1, \dots, m$) (vgl. Abschnitt 5.3). Für $m \geq 2$ wird der Nachweis der $H^m(\Omega)$ -Koerzivität komplizierter. Um Satz 7.2.13 zu übertragen, braucht man zusätzlich die sogenannte Bedingung von Agmon (vgl. Wloka [89, p. 286], Lions-Magenes [58, p. 210]).

Die entstehenden Komplikationen lassen sich anhand der biharmonischen Gleichung veranschaulichen.

Beispiel 7.4.13 a) Dem Variationsproblem: suche $u \in H^2(\Omega)$ mit

$$a(u, v) := \int_{\Omega} \Delta u \Delta v dx = f(v) := \int_{\Omega} g(x)v(x) dx + \int_{\Gamma} \left(\varphi_1(x) \frac{\partial v(x)}{\partial n} - \varphi_2(x)v(x) \right) d\Gamma \quad \text{für alle } v \in H^2(\Omega) \quad (7.4.12a)$$

entspricht die klassische Formulierung

$$\Delta^2 u = g \quad \text{in } \Omega, \quad \Delta u = \varphi_1 \quad \text{und} \quad \frac{\partial}{\partial n} \Delta u = \varphi_2 \quad \text{auf } \Gamma. \quad (7.4.12b)$$

Aber die Bilinearform $a(\cdot, \cdot)$ ist nicht $H^2(\Omega)$ -koerziv.

b) Dem Variationsproblem: suche $u \in H^2(\Omega) \cap H_0^1(\Omega)$ mit

$$a(u, v) = f(v) := \int_{\Omega} g(x)v(x) dx + \int_{\Gamma} \varphi \frac{\partial v}{\partial n} d\Gamma \quad \text{für alle } v \in H^2(\Omega) \cap H_0^1(\Omega) \quad (7.4.12c)$$

($a(\cdot, \cdot)$ wie in (7.4.12a)) entspricht die klassische Formulierung

$$\Delta^2 u = g \quad \text{in } \Omega, \quad \Delta u = 0 \quad \text{und} \quad \Delta u = \varphi \quad \text{auf } \Gamma. \quad (7.4.12d)$$

Die Bilinearform ist $H^2(\Omega) \cap H_0^1(\Omega)$ -koerziv.

c) Die Randbedingungen in

$$\Delta^2 u = g \quad \text{in } \Omega, \quad u = 0 \quad \text{und} \quad \frac{\partial}{\partial n} \Delta u = \varphi \quad \text{auf } \Gamma. \quad (7.4.12e)$$

sind zulässig. Trotzdem kann diese Randwertaufgabe nicht in der bisherigen Form als Variationsproblem geschrieben werden. Eine Variationsformulierung für (7.4.12e) lautet:

$$\text{suche } u \in H^2(\Omega) \cap H_0^1(\Omega), \text{ sodass } a(u, v) = f(v) := \int_{\Omega} g v dx + \int_{\Gamma} \varphi v d\Gamma \quad \text{für alle } v \in H^2(\Omega) \text{ mit } \frac{\partial v}{\partial n} = 0 \quad (7.4.12f)$$

($a(\cdot, \cdot)$ wie in (7.4.12a)). Diese entspricht aber nicht dem bisherigen Konzept, da u und v aus verschiedenen Räumen stammen.

Beweis. Die Äquivalenz der Variations- und klassischen Formulierung zeigt man durch partielle Integration:

$$a(u, v) = \int_{\Omega} v \Delta^2 u dx + \int_{\Gamma} \left[\Delta u \frac{\partial v}{\partial n} - \frac{\partial \Delta u}{\partial n} v \right] d\Gamma.$$

Die Nichtkoerzivität im Teil a) ergibt sich wie folgt. $\Omega \subset \mathbb{R}^n$ sei beschränkt. Für alle $\alpha \in \mathbb{R}$ liegt $u_{\alpha}(x_1, \dots, x_n) = \sin(\alpha x_1) \exp(\alpha x_2)$ in $H^2(\Omega)$ und erfüllt $\Delta u_{\alpha} = 0$, also auch $a(u_{\alpha}, u_{\alpha}) = 0$. Wäre $a(\cdot, \cdot)$ koerziv, gäbe es ein C mit $0 = a(u_{\alpha}, u_{\alpha}) \geq \varepsilon |u_{\alpha}|_2 - C |u_{\alpha}|_0$ für alle α , d.h. $|u_{\alpha}|_2 \leq (C/\varepsilon) |u_{\alpha}|_0$. Der Widerspruch ergibt sich aus $|u_{\alpha}|_2 \geq |\partial^2 u_{\alpha} / \partial x_1^2|_0 = \alpha^2 |u_{\alpha}|_0$ für hinreichend großes α . ■

Die natürlichen und Zwangsrandbedingungen können zusammen auftreten. In Beispiel 7.4.13b ist $u = 0$ eine Zwangsbedingung und $\partial \Delta u / \partial n = \varphi$ die natürliche Randbedingung. Aber auch im Falle $m = 1$ können beide Arten der Randbedingung auftreten.

Beispiel 7.4.14 Sei γ eine nichtleere, echte Teilmenge von Γ . Die Randwertaufgabe

$$-\Delta u = g \quad \text{in } \Omega, \quad u = 0 \quad \text{auf } \gamma, \quad \frac{\partial}{\partial n} u = \varphi \quad \text{auf } \Gamma \setminus \gamma. \quad (7.4.13)$$

lautet in der Variationsformulierung wie folgt: Suche $u \in H_{\gamma}^1(\Omega)$, sodass

$$a(u, v) := \int_{\Omega} \langle \nabla u, \nabla v \rangle dx = f(v) := \int_{\Omega} g v dx + \int_{\Gamma \setminus \gamma} \varphi v d\Gamma \quad \text{für alle } v \in H_{\gamma}^1(\Omega),$$

wobei $H_{\gamma}^1(\Omega) := \{u \in H^1(\Omega) : u = 0 \text{ auf } \gamma\}$. Die Randbedingung in (7.4.13) wird eine gemischte Bedingung genannt.

Übungsaufgabe 7.4.15 Die Bilinearform $a(u, v) := \int_{\Omega} [\langle \nabla u, \nabla v \rangle + auv] dx$ mit $a > 0$ sei auf $V \times V$ mit $V := \{u \in H^1(\Omega) : u \text{ konstant auf } \Gamma\}$ definiert. Man zeige:

a) $a(\cdot, \cdot)$ ist V -elliptisch.

b) Die schwache Formulierung $u \in V$, $a(u, v) = \int_{\Omega} g v dx + \int_{\Gamma} \varphi v d\Gamma$ für alle $v \in V$, entspricht dem Problem

$$-\Delta u + au = g \quad \text{in } \Omega, \quad u \text{ konstant auf } \Gamma \quad \text{mit} \quad \int_{\Gamma} \frac{\partial u}{\partial n} d\Gamma = \int_{\Gamma} \varphi d\Gamma, \quad (7.4.14)$$

das auch Adler-Problem genannt wird.

Abschließend soll auf die Schwierigkeiten, eine schwache Lösung klassisch zu interpretieren, hingewiesen werden. In der Variationsformulierung (7.4.2) werden die rechten Seiten g und φ der Differentialgleichung und der Randbedingung zu dem Funktional f zusammengezogen. In der Variationsformulierung sind die Bestandteile g und φ nicht unterscheidbar! $u \in H^1(\Omega)$ besitzt erste Ableitungen in $L^2(\Omega)$, deren Beschränkung auf Γ nicht sinnvoll erklärbar zu sein braucht. Deshalb ist Bu im Allgemeinen nicht definierbar; $Bu = \varphi$ lässt sich nicht als Gleichheit im Raum $H^{-1/2}(\Gamma)$ verstehen, obwohl $\varphi \in H^{-1/2}(\Gamma)$ (vgl. Zusatz 7.4.4b).

Aber auch wenn eine klassische Lösung vorliegt, gibt es folgendes Paradoxon. Sei u eine klassische Lösung von $Lu = 0$ in Ω , $Bu = \varphi$ auf Γ . Gemäß (7.4.1a) sei $f_{\varphi} \in (H^1(\Omega))'$ durch $f_{\varphi}(v) := \int_{\Gamma} \varphi v d\Gamma$ definiert. Man kann u ebenso als Lösung von $Lu = f_{\varphi}$ in Ω , $Bu = 0$ auf Γ auffassen. Diese Gleichungen lassen sich sogar in folgender Weise klassisch interpretieren. Es gibt $f_{\nu} \in C^{\infty}(\Omega)$ ($\nu \in \mathbb{N}$) mit $f_{\nu} \rightarrow f_{\varphi}$ in $(H^1(\Omega))'$. u_{ν} sei die klassische Lösung von $Lu_{\nu} = f_{\nu}$, $Bu_{\nu} = 0$. Dann strebt u_{ν} in $H^1(\Omega)$ gegen die oben erwähnte klassische Lösung u .

Die Verlagerung der Randwerte $Bu = \varphi$ in die Differentialgleichung $Lu = f_{\varphi}$ entspricht einer Umformung des diskretisierten Problems, die ständig in Kapitel 4 angewandt wurde. Die Differenzengleichungen $D_h u_h = f_h$ in Ω_h (innere Gitterpunkte) und die Randbedingungen $u_h = \varphi$ auf Γ_h ergaben das Gleichungssystem $L_h u_h = q_h := f_h + \varphi_h$ (vgl. (4.2.6b)). Definiert man \bar{u}_h durch $\bar{u}_h = u_h$ in Ω_h , $\bar{u}_h = 0$ auf Γ_h , so erfüllt \bar{u}_h die Gleichungen $D_h \bar{u}_h = q_h$ in Ω_h , $\bar{u}_h = 0$ auf Γ_h . So wie sich das Funktional f nicht eindeutig in g und φ trennen lässt, lassen sich f_h und φ_h nicht aus q_h rekonstruieren. Anders als im diskreten Fall ist die Trennung von f in g und φ jedoch möglich, sobald stärkere Bedingungen an g gestellt werden als $g \in (H^1(\Omega))'$ (z.B. $g \in L^2(\Omega)$).

8

Die Methode der finiten Elemente

In Kapitel 7 wurde die Variationsformulierung zunächst nur für den Zweck eingeführt, die Existenz einer (schwachen) Lösung zu beweisen. Es wird sich jetzt herausstellen, dass die Variationsformulierung eine neue Diskretisierungsmethode begründet.

8.1 Das Ritz-Galerkin-Verfahren

8.1.1 Grundlagen

Das Randwertproblem sei in der Variationsformulierung gegeben:

$$\text{suche } u \in V, \quad \text{sodass} \quad a(u, v) = f(v) \quad \text{für alle } v \in V, \quad (8.1.1)$$

wobei wir insbesondere an $V = H_0^m(\Omega)$ und $V = H^1(\Omega)$ denken (vgl. Abschnitte 7.2 und 7.4). Selbstverständlich ist vorausgesetzt, dass $a(\cdot, \cdot)$ eine auf $V \times V$ definierte und beschränkte Bilinearform ist und dass $f \in V'$:

$$|a(u, v)| \leq C_S \|u\|_V \|v\|_V \quad \text{für } u, v \in V, \quad f \in V'. \quad (8.1.2)$$

Die Differenzenverfahren entstanden durch Diskretisierung des Differentialoperators. Jetzt wird der in $a(\cdot, \cdot)$ verborgene Differentialoperator unverändert gelassen. Die *Ritz¹-Galerkin²-Diskretisierung* besteht darin, den unendlich-dimensionalen Raum V durch einen endlich-dimensionalen Unterraum V_N zu ersetzen:

$$V_N \subset V \quad \text{mit} \quad \dim V_N = N < \infty. \quad (8.1.3)$$

V_N versehen mit der Norm $\|\cdot\|_V$ ist wieder ein Banach-Raum. Da $V_N \subset V$, sind sowohl $a(u, v)$ als auch $f(v)$ für $u, v \in V_N$ definiert. Damit lässt sich die Aufgabe (8.1.4) stellen, die aus (8.1.1) entsteht, indem man V durch V_N ersetzt:

$$\text{suche } u^N \in V_N, \quad \text{sodass} \quad a(u^N, v) = f(v) \quad \text{für alle } v \in V_N. \quad (8.1.4)$$

Der Übergang von (8.1.1) zu (8.1.4) heißt *Ritz-Galerkin-* oder kurz *Galerkin-Diskretisierung*. Genauer handelt es sich um eine *konforme* Galerkin-Diskretisierung, da die Unterraumbedingung $V_N \subset V$ auf Grund von (8.1.3) gegeben ist.

¹Das Verfahren geht auf Ritz [70] (1908) zurück, wobei allerdings dort anstelle des allgemeinen Variationsproblems (8.1.1) ein Minimierungsproblem (7.2.12a,b) zugrundegelegt wurde. In diesem Fall spricht man auch vom *Ritz-* oder *Rayleigh-Ritz-*Verfahren. Die Verallgemeinerung auf (8.1.1) stammt von Galerkin.

²Boris Grigorievich Galerkin, am 4 März 1871 in Polotsk (Weißrussland) geboren, am 12 Juni 1945 in Moskau gestorben.

Die Lösung von (8.1.4), falls sie existiert, heißt die (zu V_N gehörende) *Ritz-Galerkin-Lösung* der Randwertaufgabe (8.1.1).

Zur konkreten Berechnung der Lösung u_N benötigt man eine Basis von V_N . Sei $\{b_1, \dots, b_N\}$ eine solche Basis, d.h.

$$V_N = \text{span}\{b_1, \dots, b_N\}. \quad (8.1.5)$$

Für jeden *Koeffizientenvektor* $\mathbf{v} = (v_1, \dots, v_N)^\top \in \mathbb{R}^N$ (\mathbb{C}^N ebenso möglich) definieren wir die Abbildung³

$$P: \mathbb{R}^N \rightarrow V_N \subset V, \quad P\mathbf{v} = \sum_{i=1}^N v_i b_i. \quad (8.1.6)$$

Bemerkung 8.1.1 P ist ein Isomorphismus zwischen \mathbb{R}^N und V_N . Die inverse Abbildung $P^{-1}: V_N \rightarrow \mathbb{R}^N$ ist daher auf V_N wohldefiniert.

Lemma 8.1.2 Unter der Voraussetzung (8.1.5) ist das Problem (8.1.4) äquivalent zu

$$\text{suche } u^N \in V_N, \quad \text{sodass } a(u^N, b_i) = f(b_i) \quad \text{für alle } i = 1, \dots, N. \quad (8.1.7)$$

Beweis. a) “(8.1.7) \Rightarrow (8.1.4)” Sei $v = \sum_{i=1}^N v_i b_i \in V_N$ beliebig. Linearität von $a(u^N, \cdot)$ und f ergeben (8.1.4):

$$a(u^N, v) - f(v) = a\left(u^N, \sum_{i=1}^N v_i b_i\right) - f\left(\sum_{i=1}^N v_i b_i\right) = \sum_{i=1}^N v_i [a(u^N, b_i) - f(b_i)] = 0.$$

b) “(8.1.4) \Rightarrow (8.1.7)” $v = b_i$ in (8.1.4) ergibt die Bedingung (8.1.7). ■

Wir suchen nun den Vektor $\mathbf{u} \in \mathbb{R}^N$, der über $u^N = P\mathbf{u}$ die Ritz-Galerkin-Lösung produziert. Der folgende Satz überführt das Problem (8.1.4) bzw. (8.1.7) in ein lineares Gleichungssystem.

Satz 8.1.3 Es gelte (8.1.5). Die $N \times N$ -Matrix $\mathbf{L} = (L_{ij})_{i,j=1,\dots,N}$ und der N -Vektor $\mathbf{f} = (f_1, \dots, f_N)^\top$ seien definiert durch

$$L_{ij} := a(b_j, b_i) \quad (i, j = 1, \dots, N), \quad (8.1.8a)$$

$$f_i := f(b_i) \quad (i = 1, \dots, N). \quad (8.1.8b)$$

Dann sind das Problem (8.1.4) und das Gleichungssystem

$$\mathbf{L}\mathbf{u} = \mathbf{f} \quad (8.1.9)$$

äquivalent. Ist \mathbf{u} eine Lösung von (8.1.9), so löst $u^N := P\mathbf{u}$ das Problem (8.1.4). Ist umgekehrt u^N Lösung von (8.1.4), so ist $\mathbf{u} := P^{-1}u^N$ (vgl. Bemerkung 8.1.1) Lösung von (8.1.9).

Beweis. (8.1.4) ist äquivalent zu (8.1.7). In (8.1.7) machen wir den Ansatz $u^N = P\mathbf{u} = \sum u_j b_j$:

$$a(u^N, b_i) = a\left(\sum_{j=1}^N u_j b_j, b_i\right) = \sum_j u_j a(b_j, b_i) = \sum_j L_{ij} u_j = f(b_i) = f_i,$$

und erhalten $\mathbf{L}\mathbf{u} = \mathbf{f}$. Umgekehrt führt die Lösung von $\mathbf{L}\mathbf{u} = \mathbf{f}$ zur Lösung $a(u^N, b_i) = f(b_i)$. ■

In ingenieurtechnischen Anwendungen, wo die Randwertaufgaben aus der Kontinuumsmechanik stammen, nennt man \mathbf{L} die *Steifigkeitsmatrix*. Im Folgenden nennen wir diese Matrix neutraler als *Systemmatrix* (“FEM-Matrix” wäre eine Alternative). Der Zusammenhang zwischen \mathbf{L} und $a(\cdot, \cdot)$ einerseits und \mathbf{f} und $f(\cdot)$ andererseits wird deutlich in

Bemerkung 8.1.4 $\langle \mathbf{u}, \mathbf{v} \rangle := \sum_i u_i v_i$ sei das übliche Skalarprodukt im \mathbb{R}^N . Den Vektoren $\mathbf{u}, \mathbf{v} \in \mathbb{R}^N$ seien die Funktionen $u = P\mathbf{u}$, $v = P\mathbf{v}$ zugeordnet. Dann gilt $a(u, v) = \langle \mathbf{L}\mathbf{u}, \mathbf{v} \rangle$, $f(v) = \langle \mathbf{f}, \mathbf{v} \rangle$.

Eine triviale Folgerung aus Satz 8.1.3 lautet:

Zusatz 8.1.5 Die Ritz-Galerkin-Diskretisierung (8.1.4) hat genau dann für jedes $f \in V'$ eine eindeutige Lösung u^N , wenn die Matrix L aus (8.1.8a) nichtsingulär ist.

³Der Buchstabe P steht für “Prolongation”, da bei den späteren Anwendungen die Werte v_i die Knotenwerte an bestimmten Punkten x_i sein werden, die mittels $P\mathbf{v}$ zu einer Funktion aus V_N fortgesetzt werden.

8.1.2 Diskussion der diskreten Gleichung

Definition 8.1.6 (Diskretisierung durch Projektion) Seien $\Pi : V' \rightarrow V'_N = V_N$ eine beliebige Projektion auf V_N und L der zu $a(\cdot, \cdot)$ gehörige Operator. Die zur Aufgabe (8.1.1) äquivalente Gleichung $Lu = f$ wird ersetzt durch die projizierte Gleichung $\Pi Lu^N = \Pi f$ für ein zu bestimmendes $u^N \in V_N$.

Die diskrete Gleichung $\Pi Lu^N = \Pi f$ kann mit Hilfe der Bilinearform als $a(u^N, \Pi^* v) = f(\Pi^* v)$ für alle $v \in V$ geschrieben werden. Letzteres ist äquivalent zu $a(u^N, w) = f(w)$ für alle $w \in \text{Bild}(\Pi^*)$. Man beachte, dass im Allgemeinen $\text{Bild}(\Pi^*) \neq \text{Bild}(\Pi) = V_N$. Das Galerkin-Verfahren (8.1.4) ergibt sich als spezielles Projektionsverfahren für die Wahl von Π als *orthogonale Projektion* auf V_N (d.h. $\Pi^* = \Pi$).

Im Folgenden wird das diskrete Problem $\mathbf{L}u = \mathbf{f}$ mit Hilfe des Operators L_N formuliert, der zur Bilinearform $a(\cdot, \cdot)$ beschränkt auf $V_N \times V_N$ gehört. Ferner wird die orthogonale Projektion auf V_N explizit konstruiert.

Dem Gelfand-Dreier $V \subset U \subset V'$ entspricht im endlich-dimensionalen Falle $V_N \subset U_N \subset V'_N$, wobei diese Räume mengenmäßig übereinstimmen:

$$V_N = U_N = V'_N, \quad (8.1.10a)$$

jedoch drei unterschiedliche Normen tragen:

$$V_N = (V_N, \|\cdot\|_V), \quad U_N = (U_N, \|\cdot\|_U), \quad V'_N = (V'_N, \|\cdot\|_{V'_N}). \quad (8.1.10b)$$

V_N hat $\|\cdot\|_V$ und U^N hat $\|\cdot\|_U$ als Norm, während die Dualnorm definitionsgemäß

$$\|v\|_{V'_N} := \sup \{ |(v, u)_U| / \|u\|_V : 0 \neq u \in V_N \} \quad \text{für } v \in V'_N = V_N \quad (8.1.10c)$$

lautet. $\|\cdot\|_{V'_N}$ ist auch auf V' definiert (da $u \in V_N \subset V$), stellt dort aber keine Norm dar, da $\|v\|_{V'_N} = 0$ für jedes zu V_N orthogonale v . Ein Zusammenhang zwischen $\|\cdot\|_{V'_N}$ und $\|\cdot\|_{V'}$ ist gegeben durch

Bemerkung 8.1.7 Es gilt $\|f\|_{V'_N} \leq \|f\|_{V'}$ für alle $f \in V'$.

Beweis. Definitionsgemäß ist $\|f\|_{V'_N}$ das Supremum von $|(f, u)_U| / \|u\|_V$ für alle $u \in V_N \setminus \{0\}$. Vergrößert man $V_N \setminus \{0\}$ zu $V \setminus \{0\}$, vergrößert sich auch das Supremum, das jetzt $\|f\|_{V'}$ definiert. ■

Gemäß Lemma 6.5.1 gibt es zu $a(\cdot, \cdot) : V_N \times V_N \rightarrow \mathbb{R}$ einen zugehörigen Operator⁴

$$L_N : V_N \rightarrow V'_N, \quad \text{sodass} \quad a(u, v) = (L_N u, v)_U \quad \text{für alle } u, v \in V_N. \quad (8.1.11a)$$

Hierbei schreiben wir $(\cdot, \cdot)_U$ für $\langle \cdot, \cdot \rangle_{V'_N \times V_N}$ (vgl. Bemerkung 6.3.12).

$P : \mathbb{R}^N \rightarrow V_N$ kann wegen $V_N \subset V$ auch als Abbildung $P : \mathbb{R}^N \rightarrow V$ angesehen werden. Die zu P adjungierte Abbildung $P^* \in L(V', \mathbb{R}^N)$ ist durch

$$\langle P^* u, \mathbf{v} \rangle = (u, P \mathbf{v})_U \quad \text{für alle } u \in V', \mathbf{v} \in \mathbb{R}^N. \quad (8.1.11b)$$

definiert (vgl. (6.3.4)).

Übungsaufgabe 8.1.8 a) Man zeige: Der Kern von P^* ist $V_N^\perp \subset U$ (Orthogonalraum von $V_N \subset U$ bezüglich des Skalarproduktes $(\cdot, \cdot)_U$). $P^*|_{V_N} : V_N \rightarrow \mathbb{R}^N$ ist ein Isomorphismus. Der Vektor $P^* u \in \mathbb{R}^N$ hat die Komponenten $(P^* u)_j = (u, b_j)_U$ mit dem Basiselement b_j aus (8.1.5).

b) Da $P^{-1} : V_N \rightarrow \mathbb{R}^N$ existiert (vgl. Bemerkung 8.1.1), können wir

$$C_P := \|P^{-1}\|_{\mathbb{R}^N \leftarrow U_N} = \max \{ \|P^{-1} u\|_{\mathbb{R}^N} / \|u\|_U : 0 \neq u \in V_N \} \quad (8.1.12)$$

definieren, wobei $\|\cdot\| = \|\cdot\|_{\mathbb{R}^N}$ die Euklidische Norm des \mathbb{R}^N ist. Man zeige: Die Matrix $P^* P : \mathbb{R}^N \rightarrow \mathbb{R}^N$ hat eine Inverse mit der Spektralnorm $\|(P^* P)^{-1}\| \leq C_P^2$.

c) Man zeige:

$$Q_N := P (P^* P)^{-1} P^* : U \rightarrow U \quad (8.1.13)$$

ist die (bezüglich $\|\cdot\|_U$) orthogonale Projektion auf $U_N = V_N$. Weiterhin gilt $Q_N \in L(V', V)$.

⁴Schreibt man $a_N(\cdot, \cdot)$ für die Beschränkung von $a(\cdot, \cdot)$ auf $V_N \times V_N$, so wird augenfälliger, dass a und a_N zu verschiedenen Operatoren L und L_N führen. Zur Vereinfachung verwenden wir jedoch in beiden Fällen das Symbol $a(\cdot, \cdot)$.

Lemma 8.1.9 Seien $L_N \in L(V_N, V'_N)$ der zu $a(\cdot, \cdot) : V_N \times V_N \rightarrow \mathbb{R}$ und $L \in L(V, V')$ der zu $a(\cdot, \cdot) : V \times V \rightarrow \mathbb{R}$ gehörige Operator. Zwischen den Operatoren L_N , L und der Systemmatrix \mathbf{L} gelten die folgenden Beziehungen:

$$\mathbf{L} = P^*LP = P^*L_NP : V \rightarrow V', \quad (8.1.14a)$$

$$L_N = P^{*-1}\mathbf{L}P^{-1} : V_N \rightarrow V'_N, \quad (8.1.14b)$$

$$L_N \text{ ist die Beschränkung von } Q_NL \text{ und } Q_NLQ_N \text{ auf } V_N. \quad (8.1.14c)$$

Die ersten beiden Gleichungen entsprechen dem folgenden kommutativen Diagramm:

$$\begin{array}{ccc} V & \xrightarrow{L} & V' & (\text{kontinuierliche Ebene}) \\ P \uparrow & & \downarrow P^* & \\ \mathbb{R}^N & \xrightarrow{\mathbf{L}} & \mathbb{R}^N & (\text{Abbildung als Matrix, Diskretisierungsebene}) \\ P \downarrow \uparrow P^{-1} & & P^* \downarrow \uparrow P^{*-1} & \\ V_N & \xrightarrow{L_N} & V'_N & (\text{abstrakte Abbildung, Diskretisierungsebene}) \end{array}$$

Beweis. a) Für alle $\mathbf{u}, \mathbf{v} \in \mathbb{R}^N$ gilt

$$\langle \mathbf{L}\mathbf{u}, \mathbf{v} \rangle \stackrel{\text{Bemerkung 8.1.4}}{=} a(P\mathbf{u}, P\mathbf{v}) = \begin{cases} (L_N P\mathbf{u}, P\mathbf{v})_U & = \langle P^*L_N P\mathbf{u}, \mathbf{v} \rangle, \\ (LP\mathbf{u}, P\mathbf{v})_U & = \langle P^*LP\mathbf{u}, \mathbf{v} \rangle, \end{cases}$$

was (8.1.14a) beweist.

b) Sei $u \in V_N$. Da das Bild von Q_NLu in $V'_N = U_N$ liegt, ist $(L_Nu, v)_U = (Q_NLu, v)_U$ für alle $v \in U_N$ nachzuweisen, was aus $(L_Nu, v)_U = a(u, v)$ und $(Q_NLu, v)_U \stackrel{Q_N=Q_N^*}{=} (Lu, Q_Nv)_U \stackrel{v \in U_N}{=} (Lu, v)_U = a(u, v)$ folgt. ■

Gemäß Bemerkung 6.3.12 schreiben wir $(f, v)_U$ für $f(v)$. Wenn $v = P\mathbf{v}$, folgt $f(v) = (f, P\mathbf{v})_U = \langle P^*f, \mathbf{v} \rangle$ aus (8.1.11b). Mit Bemerkung 8.1.4 ergibt sich somit

$$\mathbf{f} = P^*f \quad (8.1.15)$$

für die rechte Seite \mathbf{f} in Gleichung (8.1.9) (vgl. Übungsaufgabe 8.1.8a).

Im Falle des kontinuierlichen Variationsproblems garantierte die V -Elliptizität die eindeutige Lösbarkeit. Die gleiche Bedingung reicht auch im diskreten Fall:

Satz 8.1.10 Es gelte (8.1.3). Die Bilinearform sei V -elliptisch: $a(u, u) \geq C_E \|u\|_V^2$ für alle $u \in V$ mit $C_E > 0$. Dann ist die Systemmatrix \mathbf{L} aus Gleichung (8.1.9) nichtsingulär und die Ritz-Galerkin-Lösung $u^N \in V_N$ erfüllt

$$\|u^N\|_V \leq \frac{1}{C_E} \|f\|_{V'_N} \leq \frac{1}{C_E} \|f\|_{V'}. \quad (8.1.16)$$

Beweis. \mathbf{L} ist nichtsingulär, da für jedes $\mathbf{u} \neq 0$ auch $P\mathbf{u} \neq 0$, also

$$\langle \mathbf{L}\mathbf{u}, \mathbf{u} \rangle = a(P\mathbf{u}, P\mathbf{u}) \geq C_E \|P\mathbf{u}\|_V^2 > 0$$

und damit insbesondere $\mathbf{L}\mathbf{u} \neq 0$. Nach Übungsaufgabe 6.5.6a ist $a(\cdot, \cdot)$ auch V_N -elliptisch mit der gleichen Konstanten C_E . Gemäß Satz 6.5.8 gilt $\|L_N^{-1}\|_{V'_N \leftarrow V_N} \leq 1/C_E$, d.h. $\|u^N\|_V = \|L_N^{-1}f\|_V \leq C_E^{-1} \|f\|_{V'_N}$. Die zweite Ungleichung in (8.1.16) ergibt sich aus Bemerkung 8.1.7. ■

Übungsaufgabe 8.1.11 Man zeige: a) Ist $a(\cdot, \cdot)$ symmetrisch, so ist auch die Matrix \mathbf{L} symmetrisch.

b) Ist $a(\cdot, \cdot)$ symmetrisch und V -elliptisch, so ist \mathbf{L} positiv definit. Unter den gleichen Voraussetzungen ist die Ritz-Galerkin-Lösung u^N die Lösung des folgenden Minimierungsproblems (vgl. Satz 6.5.12):

$$J(u^N) \leq J(u) := a(u, u) - 2f(u) \quad \text{für alle } u \in V_N.$$

8.1.3 Beispiele

Die folgenden, sehr niederdimensionalen Beispiele sollen das Prinzip der Galerkin-Diskretisierung illustrieren.

Beispiel 8.1.12 Die Randwertaufgabe laute

$$-\Delta u = 1 \quad \text{in } \Omega = (0, 1) \times (0, 1), \quad u = 0 \quad \text{auf } \Gamma.$$

Die schwache Formulierung ist durch (8.1.1) mit

$$V = H_0^1(\Omega), \quad a(u, v) = \int_{\Omega} \langle \nabla u, \nabla v \rangle \, d\mathbf{x} = \int_0^1 \int_0^1 (u_x v_x + u_y v_y) \, dx dy, \quad f(v) = \int_0^1 \int_0^1 v \, dx dy$$

gegeben. Die Funktionen

$$\begin{aligned} b_1(x, y) &= \sin(\pi x) \sin(\pi y), & b_2(x, y) &= \sin(3\pi x) \sin(\pi y), \\ b_3(x, y) &= \sin(\pi x) \sin(3\pi y), & b_4(x, y) &= \sin(3\pi x) \sin(3\pi y) \end{aligned}$$

erfüllen die Randbedingung $u = 0$ auf Γ und gehören somit zu $V = H_0^1(\Omega)$. Sie bilden die Basis von $V_N = V_4 := \text{span}\{b_1, \dots, b_4\} \subset V$. Die Matrixelemente $L_{ii} = a(b_i, b_i)$ errechnet man zu

$$L_{11} = \pi^2/2, \quad L_{22} = L_{33} = 5\pi^2/2, \quad L_{44} = 9\pi^2/2.$$

Im Übrigen erweist sich die gewählte Basis als $a(\cdot, \cdot)$ -orthogonal:

$$L_{ij} = 0 \quad \text{für } i \neq j,$$

sodass die Systemmatrix \mathbf{L} diagonal ist. Weiter findet man für $f_i = f(b_i) = \int_{\Omega} b_i(x, y) \, dx dy$ die Werte

$$f_1 = 4/\pi^2, \quad f_2 = f_3 = 4/(3\pi^2), \quad f_4 = 4/(9\pi^2).$$

$\mathbf{u} = \mathbf{L}^{-1}\mathbf{f}$ ergibt sich folglich zu

$$u_1 = 8/\pi^4, \quad u_2 = u_3 = 8/(15\pi^4), \quad u_4 = 8/(81\pi^4),$$

und die Ritz-Galerkin-Lösung lautet

$$u^N(x, y) = \frac{8}{\pi^4} \left[\sin(\pi x) \sin(\pi y) + \frac{1}{15} \left(\sin(3\pi x) \sin(\pi y) + \sin(\pi x) \sin(3\pi y) \right) + \frac{1}{81} \sin(3\pi x) \sin(3\pi y) \right].$$

Die Ritz-Galerkin-Lösung und die exakte Lösung ausgewertet in $x = y = 1/2$ ergeben

$$\begin{aligned} u^N\left(\frac{1}{2}, \frac{1}{2}\right) &= \frac{2848}{405} \pi^{-4} = 0.072\,191\,40\dots, \\ u\left(\frac{1}{2}, \frac{1}{2}\right) &= \sum_{\nu, \mu=0}^{\infty} \frac{16}{\pi^4} \frac{(-1)^{\nu+\mu}}{(1+2\nu)(1+2\mu) \left[(1+2\nu)^2 + (1+2\mu)^2 \right]} = 0.073\,671\,3\dots \end{aligned}$$

Beispiel 8.1.13 (natürliche Randbedingungen) Die Randwertaufgabe sei

$$-\Delta u = \pi^2 \cos(\pi x) \quad \text{in } \Omega = (0, 1) \times (0, 1), \quad \partial u / \partial n = 0 \quad \text{auf } \Gamma.$$

Die Lösung lautet $u = \cos \pi x + \text{const.}$ Die schwache Formulierung ist durch (8.1.1) mit

$$V = H^1(\Omega), \quad a(u, v) = \int_0^1 \int_0^1 (u_x v_x + u_y v_y) \, dx dy, \quad f(v) = \pi^2 \int_0^1 \int_0^1 v(x, y) \cos(\pi x) \, dx dy$$

gegeben. Die Randwertaufgabe hat eine eindeutige Lösung in $W := \{v \in V : \int_{\Omega} v \, dx dy = 0\}$. Die Basisfunktionen

$$b_1(x, y) := x - \frac{1}{2}, \quad b_2(x, y) := \left(x - \frac{1}{2}\right)^3$$

liegen in W . Die Systemmatrix \mathbf{L} und der Vektor \mathbf{f} lauten

$$\mathbf{L} = \begin{bmatrix} 1 & 1/4 \\ 1/4 & 9/80 \end{bmatrix}, \quad \mathbf{f} = \begin{bmatrix} -2 \\ -\frac{3}{2} + \frac{12}{\pi^2} \end{bmatrix}, \quad \text{sodass } \mathbf{u} = \mathbf{L}^{-1}\mathbf{f} = \begin{bmatrix} 3 - 60/\pi^2 \\ -20 + 240/\pi^2 \end{bmatrix}.$$

Die Ritz-Galerkin-Lösung ist $u^N(x, y) = (3 - 60/\pi^2)(x - 1/2) - (20 - 240/\pi^2)(x - 1/2)^3$. Die Lösung erfüllt die Randbedingung $\partial u / \partial n = 0$ ebenso wie die Differentialgleichung nur näherungsweise:

$$\partial u^N / \partial n = 12 - 120/\pi^2 = -0.158\,54\dots$$

In $x = 1/4$ lautet die Näherung: $u^N(1/4, y) = -7/16 + 45/(4\pi^2) = 0,702\,36\dots$, während $u(1/4, y) = \cos \pi/4 = 0,707\,1\dots$ die exakte Lösung ist.

8.1.4 Diskrete inf-sup-Bedingung

Im Folgenden untersuchen wir den Fall, dass $a(\cdot, \cdot)$ zwar nicht mehr V -elliptisch, aber noch V -koerziv ist. Die V -Koerzivitt von $a(\cdot, \cdot)$ garantierte, dass entweder das Problem (8.1.1) lsbar oder $\lambda = 0$ ein Eigenwert ist. Setzt man die V -Koerzivitt und die Lsbarkeit der Aufgabe (8.1.1) voraus, so kann man daraus noch *nicht* die Lsbarkeit des diskreten Problems (8.1.4) ableiten.

Beispiel 8.1.14 Sei $\Omega = (0, 1)$. $a(u, v) := \int_0^1 (u'v' - 10uv)dx$ ist $H_0^1(0, 1)$ -koerziv und das Variationsproblem $a(u, v) = f(v) := \int_0^1 gvdv$ ($v \in H_0^1(0, 1)$) ist eindeutig lsbar. V_N sei durch $b_1(x) = x(1 - x) \in V = H_0^1(0, 1)$ aufgespannt (d.h. $N = 1$). Das diskrete Problem (8.1.4) ist wegen $\mathbf{L} = 0$ nicht lsbar.

Ersetzt man in (6.5.4a) aus Lemma 6.5.3 den Raum V durch V_N , so erhlt man die Bedingung

$$\inf_{u \in V_N, \|u\|_V=1} \sup_{v \in V_N, \|v\|_V=1} |a(u, v)| = \varepsilon_N > 0. \quad (8.1.17)$$

Lemma 6.5.3 zusammen mit bungsaufgabe 6.5.4 beweist den

Satz 8.1.15 Das Problem (8.1.4) ist genau dann fr alle $f \in V'$ lsbar, wenn (8.1.17) fr ein $\varepsilon_N > 0$ gilt. Aus (8.1.17) folgt die Abschtzung (8.1.16) mit ε_N statt C_E :

$$\|u^N\|_V \leq \frac{1}{\varepsilon_N} \|f\|_{V'_N} \leq \frac{1}{\varepsilon_N} \|f\|_{V'}. \quad (8.1.18)$$

Da (8.1.4) quivalent zum Gleichungssystem (8.1.9) ist, hat man den

Zusatz 8.1.16 Die Matrix \mathbf{L} ist genau dann nicht singular, wenn (8.1.17) fr ein $\varepsilon_N > 0$ gilt.

bungsaufgabe 8.1.17 Bedingung (8.1.17) ist quivalent zu (8.1.17') wie auch zu (8.1.17''):

$$\|u\|_V \leq \frac{1}{\varepsilon_N} \sup_{v \in V_N, \|v\|_V=1} |a(u, v)| \quad \text{fr alle } u \in V_N, \quad (8.1.17')$$

$$\|L_N^{-1}\|_{V_N \leftarrow V'_N} = \frac{1}{\varepsilon_N} \quad (L_N \text{ aus (8.1.11a)}). \quad (8.1.17'')$$

Achtung. Die Bedingung (8.1.17) fr V_N folgt *nicht* aus der analogen Bedingung (6.5.4a) fr V . Man beachte aber die asymptotische Aussage aus Satz 8.2.8.

Die Abschtzung (8.1.17'') garantiert zwar die Existenz von \mathbf{L}^{-1} , sagt aber nichts ber die Matrixkondition $\text{cond}(\mathbf{L}) = \|\mathbf{L}\| \|\mathbf{L}^{-1}\|$ aus, die fr die Empfindlichkeit des Gleichungssystems $\mathbf{L}\mathbf{u} = \mathbf{f}$ entscheidend ist. Whlt man z.B. fr $a(u, v) := \int_0^1 u'v'dx$ die Monombasis $b_i = x^i$ ($i = 1, \dots, N$), so erhlt man die sehr schlecht konditionierte Matrix $L_{ij} = ij/(i + j - 1)$. Die Kondition von \mathbf{L} wrde optimal, wenn man die Basis $a(\cdot, \cdot)$ -orthogonal whlt: $a(b_i, b_j) = \delta_{ij}$, wie es in Beispiel 8.1.12 bis auf einen Skalierungsfaktor der Fall ist.

8.2 Fehlerabschtzungen

8.2.1 Quasioptimalitt

Im Falle der Differenzenverfahren sind die Lsung u und die Gitterfunktion u_h auf verschiedenen Mengen definiert. Die Ritz-Galerkin-Lsung u^N ist dagegen direkt mit u vergleichbar. Man kann den Diskretisierungsfehler durch $\|u - u^N\|_V$ oder auch $\|u - u^N\|_U$ messen. Wir werden zunchst nur die V -Norm verwenden. Fehlerabschtzungen mit $\|\cdot\|_U$ werden in §8.4.4 folgen.

Sei u die Lsung von (8.1.1): $a(u, v) = f(v)$ fr $v \in V$. Gehrt (zufllig oder auf Grund geschickter Wahl von V_N) u auch zu V_N , so erfllt $u^N := u$ auch (8.1.4). Das heit: Der Diskretisierungsfehler ist null, wenn $u \in V_N$. Wir werden nun zeigen: Je "nher" u an V_N , desto kleiner ist der Diskretisierungsfehler.

Satz 8.2.1 (Cea) Es gelte (8.1.2), (8.1.3), (8.1.17). $u \in V$ sei eine Lsung der Aufgabe (8.1.1), whrend $u^N \in V_N$ die Ritz-Galerkin-Lsung von (8.1.4) sei. Dann gilt die Abschtzung

$$\|u - u^N\|_V \leq \left(1 + \frac{C_S}{\varepsilon_N}\right) \inf_{w \in V_N} \|u - w\|_V \quad (C_S \text{ aus (8.1.2), } \varepsilon_N \text{ aus (8.1.17)}). \quad (8.2.1)$$

Die Aussage (8.2.1) bedeutet die *Quasioptimalität* der diskreten Lösung u^N , da ihr Fehler bis auf einen Faktor mit dem bestmöglichen übereinstimmt. Dies ist einerseits eine sehr starke Aussage (weil der Fehler höchstens noch um einen Faktor verbesserbar wäre), andererseits ist die Aussage schwach, da der *Best-approximationsfehler* $\inf_{w \in V_N} \|u - w\|_V$ unbekannt bleibt. $\inf_{w \in V_N} \|u - w\|_V$ ist der Abstand der Funktion u von V_N und wird im Folgenden abgekürzt durch

$$d(u, V_N) := \inf_{w \in V_N} \|u - w\|_V. \quad (8.2.2)$$

Beweis von Satz 8.2.1. u erfüllt $a(u, v) = f(v)$ für alle $v \in V$, also insbesondere auch für $v \in V_N \subset V$. Da auch $a(u^N, v) = f(v)$ für $v \in V_N$, folgt

$$a(u - u^N, v) = 0 \quad \text{für alle } v \in V_N. \quad (8.2.3)$$

Für beliebige $v, w \in V_N$ mit $\|v\|_V = 1$ schließt man deshalb auf

$$\begin{aligned} a(u^N - w, v) &= a([u^N - u] + [u - w], v) = a(u - w, v) \\ \text{und} \quad |a(u^N - w, v)| &\leq C_S \|u - w\|_V \|v\|_V = C_S \|u - w\|_V. \end{aligned} \quad (*)$$

(8.1.17') zeigt

$$\|u^N - w\|_V \leq \frac{1}{\varepsilon_N} \sup_{v \in V_N, \|v\|_V=1} |a(u^N - w, v)| \stackrel{(*)}{\leq} \frac{C_S}{\varepsilon_N} \|u - w\|_V.$$

Die Dreiecksungleichung liefert

$$\|u - u^N\|_V \leq \|u - w\|_V + \|w - u^N\|_V \leq \left(1 + \frac{C_S}{\varepsilon_N}\right) \|u - w\|_V.$$

Da $w \in V_N$ beliebig ist, folgt Behauptung (8.2.1). ■

Man beachte, dass in Satz 8.2.1 nicht die eindeutige Lösbarkeit der Aufgabe (8.1.1), sondern nur die Existenz mindestens einer Lösung vorausgesetzt wurde. In Satz 8.2.2 wird die (eindeutige) Existenz einer Lösung von $Lu = f$ auf den Voraussetzungen (8.1.17) an die diskreten Aufgaben folgen.

8.2.2 Konvergenz der Ritz-Galerkin-Lösungen

Wenn der Diskretisierungsfehler gegen null streben soll, braucht man eine Folge von Unterräumen $V_{N_i} \subset V$, die im folgenden Sinne gegen V streben:

Satz 8.2.2 $V^{(i)} := V_{N_i} \subset V$ ($i \in \mathbb{N}$) sei eine Folge von Unterräumen mit

$$\lim_{i \rightarrow \infty} d(u, V^{(i)}) = 0 \quad \text{für alle } u \in V. \quad (8.2.4a)$$

Bedingung (8.1.17) gelte mit $\varepsilon_{N_i} \geq \tilde{\varepsilon} > 0$ für alle $i \in \mathbb{N}$. Ferner sei (8.1.2) (Stetigkeit von $a(\cdot, \cdot)$) vorausgesetzt. Dann existiert eine eindeutige Lösung u der Aufgabe (8.1.1), und die Ritz-Galerkin-Lösung $u^{(i)} := u^{N_i}$ konvergiert gegen u :

$$\|u - u^{(i)}\|_V \rightarrow 0 \quad \text{für } i \rightarrow \infty.$$

Hinreichend für (8.2.4a) ist

$$V^{(1)} \subset V^{(2)} \subset \dots \subset V^{(i-1)} \subset V^{(i)} \subset \dots \subset V, \quad \bigcup_{i=1}^{\infty} V^{(i)} \text{ dicht in } V. \quad (8.2.4b)$$

Beweis. 1) Es sei zunächst die Existenz einer Lösung u angenommen.

1a) Abschätzung (8.2.1) beweist die Konvergenz: $\|u - u^{(i)}\|_V \leq (1 + \frac{C_S}{\varepsilon_{N_i}}) d(u, V^{(i)}) \leq (1 + \frac{C_S}{\tilde{\varepsilon}}) d(u, V^{(i)}) \rightarrow 0$.

1b) Wir wollen nun zeigen, dass (8.2.4a) aus (8.2.4b) folgt. Die Inklusion $V^{(i-1)} \subset V^{(i)}$ impliziert $d(u, V^{(i)}) \leq d(u, V^{(i-1)})$. $d(u, V^{(i)})$ wäre daher Nullfolge, wenn zu jedem $\varepsilon > 0$ ein i mit $d(u, V^{(i)}) \leq \varepsilon$

existierte. Nach Voraussetzung (8.2.4b) gibt es zu jedem $u \in V$ und $\varepsilon > 0$ ein $w \in \bigcup_i V^{(i)}$ mit $\|u - w\|_V \leq \varepsilon$. Damit gilt $w \in V^{(i)}$ für ein $i \in \mathbb{N}$, und $d(u, V^{(i)}) \leq \|u - w\|_V \leq \varepsilon$ beweist (8.2.4a). Die Konvergenz $u^{(i)} \rightarrow u$ beweist die Eindeutigkeit der Lösung u .

2) Als Nächstes ist zu zeigen, dass das Bild $W := \{Lv : v \in V\} \subset V'$ des zu $a(\cdot, \cdot)$ gehörenden Operators $L : V \rightarrow V'$ abgeschlossen ist. Für jedes $f \in W$ gibt es ein $u \in V$ mit $Lu = f$, sodass Teil 1) des Beweises die Konvergenz $u^{(i)} \rightarrow u$ beweist. Da $\|u^{(i)}\|_V \leq \|f\|_{V'}/\tilde{\varepsilon}$, folgt $\|u\|_V = \lim \|u^{(i)}\|_V \leq \|f\|_{V'}/\tilde{\varepsilon}$. Sei $f_\nu \in W$ eine Folge mit $f_\nu \rightarrow f^*$ in V' und $f_\nu = Lu_\nu$. Die Cauchy-Konvergenz $\|f_\nu - f_\mu\|_{V'} \rightarrow 0$ beweist $\|u_\nu - u_\mu\|_V \leq \|f_\nu - f_\mu\|_{V'}/\tilde{\varepsilon} \rightarrow 0$, sodass $u^* = \lim u_\nu \in V$ existiert. Die Stetigkeit von $L \in L(V, V')$ zeigt $f^* = \lim f_\nu = \lim Lu_\nu = Lu^*$ und somit $f^* \in W$. Also ist W abgeschlossen.

3) Um die Existenz einer Lösung u der Aufgabe (8.1.1) zu beweisen, ist die Surjektivität von $L : V \rightarrow V'$ zu zeigen. Wäre L nicht surjektiv (d.h. $W \neq V'$), gäbe es ein $f \in W^\perp$ mit $\|f\|_{V'} = 1$. $J_V : V \rightarrow V'$ sei der Riesz-Isomorphismus (vgl. Folgerung 6.3.7). Man setze $v := -J_V^{-1}f \in V$. Es gilt

$$f(v) = \langle f, v \rangle_{V' \times V} = -\langle f, f \rangle_{V'} = -1, \quad a(u, v) = \langle Lu, v \rangle_{V' \times V} = -\langle Lu, f \rangle_{V'} =_{f \in W^\perp} 0 \quad \text{für alle } u \in V.$$

$u^{(i)} \in V^{(i)}$ seien die Ritz-Galerkin-Lösungen. Auch für sie müsste $a(u^{(i)}, v) = 0$ gelten, d.h.

$$a(u^{(i)}, v) - f(v) = 1.$$

Wir zerlegen v in $v^{(i)} + w^{(i)}$, wobei $v^{(i)} \in V^{(i)}$ und $w^{(i)} \perp V^{(i)}$. Gemäß (8.2.4a) (mit v anstelle von u) kann $\|w^{(i)}\|_V \rightarrow 0$ garantiert werden. Dies zeigt

$$1 = a(u^{(i)}, v) - f(v) = a(u^{(i)}, v^{(i)}) - f(v^{(i)}) + a(u^{(i)}, w^{(i)}) - f(w^{(i)}) = a(u^{(i)}, w^{(i)}) - f(w^{(i)})$$

und $1 = |a(u^{(i)}, v) - f(v)| \leq [C_S \|u^{(i)}\|_V + \|f\|_{V'}] \|w^{(i)}\|_V$.

Da $\|u^{(i)}\|_V \leq \|f\|_{V'}/\tilde{\varepsilon}$ gleichmäßig beschränkt ist und $\|w^{(i)}\|_V \rightarrow 0$, ergibt sich ein Widerspruch. Also muss L surjektiv sein, sodass zu jedem $f \in V'$ eine Lösung u von $Lu = f$, d.h. der Aufgabe (8.1.1), existiert. ■

Zusatz 8.2.3 Die Bedingung (8.1.17) mit $\varepsilon_{N_i} \geq \tilde{\varepsilon} > 0$ aus Satz 8.2.2 ist mit $\tilde{\varepsilon} := C_E$ erfüllt, falls $a(\cdot, \cdot)$ V -elliptisch mit der Konstanten $C_E > 0$ ist: $a(u, u) \geq C_E \|u\|_V^2$.

Übungsaufgabe 8.2.4 Man zeige: (8.2.4a) impliziert, dass $\bigcup_{i=1}^\infty V^{(i)}$ und $\bigcup_{i=1}^\infty V_i$ dicht in V liegen.

8.2.3 Ritz-Projektion

Q_N sei die orthogonale Projektion auf V_N (vgl. Übungsaufgabe 8.1.8c). Die Faktoren $L : V \rightarrow V'$, $Q_N : V' \rightarrow V'_N = V_N$ und $L_N^{-1} : V'_N = V_N \rightarrow V_N \subset V$ ergeben die Komposition

$$S_N := L_N^{-1} Q_N L : V \rightarrow V_N \subset V. \quad (8.2.5)$$

Übungsaufgabe 8.2.5 Man beweise die Darstellung

$$S_N = PL^{-1}P^*L. \quad (8.2.5')$$

Lemma 8.2.6 (Ritz-Projektion) S_N ist Projektion auf V_N und wird Ritz-Projektion genannt. Sie bildet die Lösung u der Aufgabe (8.1.1) in die Ritz-Galerkin-Lösung $u^N \in V_N$ ab: $u^N = S_N u$. Unter den Voraussetzungen (8.1.2) und (8.1.17) gilt die Abschätzung

$$\|S_N\|_{V \leftarrow V} \leq C_S/\varepsilon_N \quad (C_S \text{ aus (8.1.2), } \varepsilon_N \text{ aus (8.1.17)}). \quad (8.2.6)$$

Eine zu (8.2.5) äquivalente Definition von S_N lautet

$$S_N u \in V_N \quad \text{und} \quad a(S_N u, v) = a(u, v) \quad \text{für alle } v \in V_N, u \in V. \quad (8.2.7)$$

Beweis. Da $a(u, v) = \langle Lu, v \rangle_{V' \times V} = (Lu, v)_U = (Lu, Q_N v)_U = (Q_N Lu, v)_U$ für alle $v \in V_N$, ist $S_N u$ aus (8.2.7) die Ritz-Galerkin-Lösung zur rechten Seite $f := Q_N Lu \in V'$, d.h. $S_N u = L_N^{-1} Q_N Lu$. Umgekehrt schließt man analog und beweist so die Äquivalenz der Definitionen (8.2.5) und (8.2.7).

2) (8.2.7) zeigt, dass $u \in V_N$ zu $S_N u = u$ führt. Somit ist $S_N^2 = S_N$, d.h. S_N ist eine Projektion. Abschätzung (8.2.6) ergibt sich aus $\|S_N u\|_V \leq \|Q_N Lu\|_{V'_N}/\varepsilon_N$ (vgl. Beweis von (8.1.16)) und $\|Q_N Lu\|_{V'_N} = \|Lu\|_{V'_N} \leq \|Lu\|_{V'} \leq C_S \|u\|_V$ mit C_S aus (8.1.2) (vgl. Bemerkung 8.1.7). ■

Bemerkung 8.2.7 $a(\cdot, \cdot)$ sei V -elliptisch und symmetrisch. $\|v\|_V := \sqrt{a(v, v)}$ ist eine zu $\|\cdot\|_V$ äquivalente Norm. Die Ritz-Projektion S_N ist bezüglich $\|\cdot\|_V$ eine orthogonale Projektion auf V_N . Insbesondere gilt

$$\|S_N\|_{V \leftarrow V} \leq 1. \quad (8.2.8)$$

Beweis. Das zu $\|\cdot\|_V$ gehörige Skalarprodukt ist $a(\cdot, \cdot)$, sodass $a(S_N v, w) = a(v, S_N w)$ zu zeigen ist. Die Definition (8.2.7) beweist $a(S_N v, S_N w) = a(v, S_N w)$, da $S_N w \in V_N$. Symmetrie von $a(\cdot, \cdot)$ und Vertauschung von v und w liefert $a(S_N v, w) = a(w, S_N v) = a(S_N w, S_N v) = a(S_N v, S_N w)$, sodass $a(S_N v, w) = a(v, S_N w)$. (8.2.8) ergibt sich aus Bemerkung 6.3.8. ■

Bemerkung 8.2.7 zeigt noch einmal, dass die Ritz-Galerkin-Lösung $u^N = S_N u$ die im Sinne der Norm $\|\cdot\|_V$ beste Approximation von u in V_N ist. Diese Aussage ist äquivalent mit der Variationsformulierung als Minimierungsaufgabe $J(u^N) \leq J(v)$ für alle $v \in V_N$ (vgl. Übungsaufgabe 8.1.11b). Die Stabilitätsbedingung (8.1.17) mit $\varepsilon_N \geq \varepsilon > 0$ ist außer für V -elliptische Bilinearformen schwer nachweisbar. Der folgende Satz zeigt jedoch, dass diese Bedingung für hinreichend gut approximierende Unterräume gilt.

Satz 8.2.8 Die Bilinearform $a(\cdot, \cdot)$ sei V -koerziv, wobei $V \subset U \subset V'$ stetige, dichte und kompakte Einbettungen sind. Das Problem (8.1.1) sei für alle $f \in V'$ lösbar. Für die Unterräume $V_{N_i} \subset V$ gelte (8.2.4a). Für hinreichend große i ist dann die Stabilitätsbedingung (8.1.17) mit $\varepsilon_{N_i} \geq \varepsilon > 0$ erfüllt. Der Beweis wird im Anschluss an Lemma 11.2.7 nachgeholt werden.

8.2.4 Weitere Stabilitäts- und Fehlerabschätzungen

In der Variationsformulierung (8.1.1) können sowohl die rechte Seite f als auch die Bilinearform a gestört werden:

$$\text{suche } \tilde{u}^N \in V_N, \quad \text{sodass} \quad \tilde{a}(\tilde{u}^N, v^N) = \tilde{f}(v^N) \quad \text{für alle } v^N \in V_N, \quad (8.2.9a)$$

wobei

$$\tilde{a}(z^N, v^N) = a(z^N, v^N) + \delta a(z^N, v^N), \quad \tilde{f}(v^N) = f(v^N) + \delta f(v^N) \quad \text{für alle } z^N, v^N \in V_N. \quad (8.2.9b)$$

Die Störungen δa und δf brauchen nur auf $V_N \times V_N$ bzw. V_N definiert sein.

Eine übliche Quelle der Störung ist numerische Quadratur: Das Integral $\int_{\Omega} f v dx$ wird durch eine Quadraturformel ersetzt (vgl. Übungsaufgabe 8.3.5), die auf Punktauswertungen des Integranden $f v$ beruht. Dies ist offenbar nur möglich, wenn $f v$ stetig ist. Da $v \in V$ für $d \geq 2$ keine Stetigkeit garantiert (vgl. Beispiel 6.2.6), ist die Störung δf in der Norm $\|\cdot\|_{-1}$ unbeschränkt. Ansatzfunktionen $v^N \in V_N$ haben dagegen bessere Glattheitseigenschaften. Allerdings muss die Norm $\|\cdot\|_{-1}$ durch $\|\cdot\|_{V'_N}$ aus (8.1.10c) ersetzt werden. Entsprechendes gilt für δa .

Satz 8.2.9 (Erstes Lemma von Strang) Neben (8.1.2) und (8.1.3) seien $\delta a : V_N \times V_N \rightarrow \mathbb{R}$ und $\delta f : V_N \rightarrow \mathbb{R}$ stetig. Ferner sei $\tilde{a} = a + \delta a$ V_N -elliptisch. u und u^N seien die Lösungen von (8.1.1) und (8.1.4). Dann gilt für die Lösung \tilde{u}^N des gestörten Problems (8.2.9a)

$$\|u - \tilde{u}^N\|_V \leq \text{const} \left[\|\delta f\|_{V'_N} + \inf_{z^N \in V_N} \{ \|u - z^N\|_V + \|\delta a(z^N, \cdot)\|_{V'_N} \} \right]. \quad (8.2.10)$$

Beweis. Für beliebige $z^N, v^N \in V_N$ gilt

$$\begin{aligned} \tilde{a}(\tilde{u}^N - z^N, v^N) &= \tilde{a}(\tilde{u}^N, v^N) - a(z^N, v^N) - \delta a(z^N, v^N) \stackrel{(8.2.9a)}{=} \tilde{f}(v^N) - a(z^N, v^N) - \delta a(z^N, v^N) \\ &= \tilde{f}(v^N) + \underbrace{a(u, v^N) - f(v^N)}_{=0 \text{ nach (8.1.1)}} - a(z^N, v^N) - \delta a(z^N, v^N) \\ &= \delta f(v^N) + a(u - z^N, v^N) - \delta a(z^N, v^N). \end{aligned}$$

Mit $v^N := \tilde{u}^N - z^N$ folgt

$$\begin{aligned} \tilde{C}_E \|\tilde{u}^N - z^N\|_V^2 &\leq \tilde{a}(\tilde{u}^N - z^N, \tilde{u}^N - z^N) = \delta f(\tilde{u}^N - z^N) + a(\tilde{u}^N - z^N, \tilde{u}^N - z^N) - \delta a(z^N, \tilde{u}^N - z^N) \\ &= \|\delta f\|_{V'_N} \|\tilde{u}^N - z^N\|_V + C_S \|u - z^N\|_V \|\tilde{u}^N - z^N\|_V + \|\delta a(z^N, \cdot)\|_{V'_N} \|\tilde{u}^N - z^N\|_V. \end{aligned}$$

Nach Kürzen beider Seiten durch $\|\tilde{u}^N - z^N\|_V$ folgt

$$\tilde{C}_E \|\tilde{u}^N - z^N\|_V \leq \|\delta f\|_{V'_N} + C_S \|u - z^N\|_V + \|\delta a(z^N, \cdot)\|_{V'_N}.$$

Da $z^N \in V_N$ beliebig ist, folgt (8.2.10) aus $\|u - \tilde{u}^N\|_V \leq \|u - z^N\|_V + \|z^N - \tilde{u}^N\|_V$ und der vorherigen Ungleichung. ■

8.3 Finite Elemente

Ein Spezialfall des Ritz-Galerkin-Verfahrens ist die Finite-Elemente-Methode (oft mit FEM abgekürzt). Zum Zweck der Einführung wird zunächst der eindimensionale Fall behandelt.

8.3.1 Einführung: Lineare Elemente für $\Omega = (a, b)$

Sobald die Dimension $N = \dim V_N$ größer wird, zeigt sich ein entscheidender Nachteil des allgemeinen Ritz-Galerkin-Verfahrens. Die Matrix \mathbf{L} ist im Allgemeinen vollbesetzt, d.h. es ist $L_{ij} \neq 0$ für alle $i, j = 1, \dots, N$. Man braucht somit N^2 Integrationen, um $L_{ij} = a(b_j, b_i) = \int_{\Omega} \dots$ exakt oder näherungsweise zu bestimmen. Die abschließende Auflösung des Gleichungssystems $Lu = f$ würde $\mathcal{O}(N^3)$ Rechenoperationen benötigen. Sobald N nicht mehr klein ist, erweist sich das allgemeine Ritz-Galerkin-Verfahren daher als unbrauchbar. Ein Blick auf die Differenzenverfahren zeigt, dass die dort auftretenden Matrizen schwach besetzt sind. Damit liegt es nahe zu fragen, ob es möglich ist, die Basis $\{b_1, \dots, b_N\}$ so zu wählen, dass die Systemmatrix $L_{ij} = a(b_j, b_i)$ ebenfalls schwach besetzt ist. Die günstigste Situation läge vor, wenn b_i bezüglich $a(\cdot, \cdot)$ orthogonal wäre: $a(b_j, b_i) = 0$ für $i \neq j$. Eine solche Basis lässt sich aber nur für spezielle Modellprobleme wie etwa in Beispiel 8.1.12 finden. Stattdessen gründen sich die weiteren Überlegungen auf

Bemerkung 8.3.1 Die Bilinearform $a(\cdot, \cdot)$ sei durch (7.2.6) gegeben. B_i sei das Innere des Trägers $\text{Tr}(b_i)$ der Basisfunktion b_i , d.h. $B_i := \text{Tr}(b_i) \setminus \partial \text{Tr}(b_i)$. Eine hinreichende Bedingung für $L_{ij} = a(b_j, b_i) = 0$ ist $B_i \cap B_j = \emptyset$.

Beweis. Die Integration $a(b_j, b_i) = \int_{\Omega} \dots$ kann auf $B_i \cap B_j$ reduziert werden. ■

Um Bemerkung 8.3.1 anwenden zu können, sollten die Basisfunktionen möglichst kleine Träger haben. Bei der Konstruktion geht man im Allgemeinen umgekehrt vor: Man definiert Zerlegungen von Ω in kleine Teilstücke, die sogenannten finiten Elemente, aus denen sich die Träger von b_i zusammensetzen werden. Zur Einführung untersuchen wir die eindimensionale Randwertaufgabe

$$-u''(x) = g(x) \quad \text{für } a < x < b, \quad u(a) = u(b) = 0. \quad (8.3.1)$$

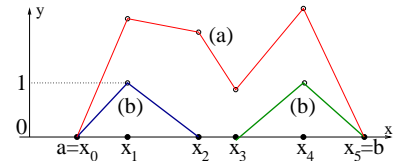


Abb. 8.3.1. a) stückweise lineare Funktion, b) Basisfunktionen zu x_1 und x_4

Eine Zerlegung des Intervalls $[a, b]$ sei durch $a = x_0 < x_1 < \dots < x_{N+1} = b$ gegeben. Die Teilintervalle seien mit $I_i := (x_{i-1}, x_i)$ ($1 \leq i \leq N+1$) bezeichnet. Als Unterraum $V_N \subset H_0^1(a, b)$ wählen wir die *stückweise linearen Funktionen*:

$$V_N = \{u \in C^0([a, b]) : u|_{I_i} \text{ linear für alle } 1 \leq i \leq N+1 \text{ und } u(a) = u(b) = 0\} \quad (8.3.2)$$

(vgl. Abbildung 8.3.1). Die Stetigkeit " $u \in C^0([a, b])$ " ist gleichbedeutend mit der Stetigkeit in den Knotenpunkten x_i ($1 \leq i \leq N$): $u(x_i + 0) = u(x_i - 0)$.

Bemerkung 8.3.2 $u \in V_N$ ist eindeutig durch die Knotenwerte $u(x_i)$ ($1 \leq i \leq N$) gegeben:

$$u(x) = [u(x_i)(x_{i+1} - x) + u(x_{i+1})(x - x_i)] / [x_{i+1} - x_i] \quad \text{für } x \in I_{i+1}, 0 \leq i \leq N, \quad (8.3.3a)$$

wobei $u(x_0) = u(x_{N+1}) = 0$. Aus Satz 6.2.43 ergibt sich $V_N \subset H_0^1(a, b)$. Die (schwache) Ableitung $u' \in L^2(a, b)$ ist stückweise konstant:

$$u'(x) = [u(x_{i+1}) - u(x_i)] / [x_{i+1} - x_i] \quad \text{für } x \in I_{i+1}, 0 \leq i \leq N. \quad (8.3.3b)$$

Die Basisfunktionen können durch

$$b_i(x) = \begin{cases} (x - x_{i-1}) / (x_i - x_{i-1}) & \text{für } x_{i-1} < x \leq x_i \\ (x_{i+1} - x) / (x_{i+1} - x_i) & \text{für } x_i < x < x_{i+1} \\ 0 & \text{sonst} \end{cases} \quad (1 \leq i \leq N) \quad (8.3.4)$$

definiert werden (vgl. Abbildung 8.3.1b). Bemerkung 8.3.2 beweist die

Bemerkung 8.3.3 Es gilt die Darstellung $u = \sum_{i=1}^N u(x_i) b_i$. Die Träger der Basisfunktionen b_i sind $\text{Tr}(b_i) = \bar{I}_{i-1} \cup \bar{I}_i = [x_{i-1}, x_{i+1}]$. B_i aus Bemerkung 8.3.1 ist $B_i = (x_{i-1}, x_{i+1})$.

Die zur Randwertaufgabe (8.3.1) gehörende schwache Formulierung lautet $a(u, v) = f(v)$ mit

$$a(u, v) := \int_a^b u' v' dx, \quad f(v) := \int_a^b g v dx \quad \text{für } u, v \in H_0^1(a, b).$$

Gemäß Bemerkung 8.3.1 gilt für die Matrixelemente $L_{ij} = 0$, sobald $|i - j| \geq 2$, da dann $B_i \cap B_j = \emptyset$. Für $|i - j| \leq 1$ erhält man

$$\begin{aligned} L_{i,i-1} &= a(b_{i-1}, b_i) = \int_{x_{i-1}}^{x_i} \frac{-1}{x_i - x_{i-1}} \frac{1}{x_i - x_{i-1}} dx = \frac{-1}{x_i - x_{i-1}}, \\ L_{i,i} &= a(b_i, b_i) = \int_{x_{i-1}}^{x_i} (x_i - x_{i-1})^{-2} dx + \int_{x_i}^{x_{i+1}} (x_{i+1} - x_i)^{-2} dx = \frac{1}{x_i - x_{i-1}} + \frac{1}{x_{i+1} - x_i}, \\ L_{i,i+1} &= a(b_{i+1}, b_i) = L_{i+1,i} = \frac{-1}{x_{i+1} - x_i}. \end{aligned} \quad (8.3.5a)$$

Die rechte Seite $f = (f_1, \dots, f_N)^\top$ ist durch (8.3.5b) gegeben:

$$f_i = f(b_i) \int_{x_{i-1}}^{x_{i+1}} g b_i dx = \frac{1}{x_i - x_{i-1}} \int_{x_{i-1}}^{x_i} g(x) (x - x_{i-1}) dx + \frac{1}{x_{i+1} - x_i} \int_{x_i}^{x_{i+1}} g(x) (x_{i+1} - x) dx. \quad (8.3.5b)$$

Bemerkung 8.3.4 Das durch (8.3.5a,b) definierte Gleichungssystem $\mathbf{L} \mathbf{u} = \mathbf{f}$ ist tridiagonal, also insbesondere schwach besetzt. Für eine äquidistante Unterteilung $x_i := a + ih$ mit $h := (b - a)/(N + 1)$ lauten die Koeffizienten

$$L_{i,i\pm 1} = -1/h, \quad L_{ii} = 2/h, \quad f_i = h \int_0^1 [g(x_i + th) + g(x_i - th)](1 - t) dt.$$

Falls $g \in C^0(a, b)$, gilt nach dem Zwischenwertsatz $f_i = hg(x_i + \Theta_i h)$, $|\Theta_i| < 1$. Damit ist das Gleichungssystem $\mathbf{L} \mathbf{u} = \mathbf{f}$ bis auf eine Skalierung identisch mit der Differenzengleichung $L_h u_h = f_h$ aus Abschnitt 4.1, wenn man f_h durch $f_h(x_i) := f_i = g(x_i + \Theta_i h)$ statt durch $f_h(x_i) = g(x_i)$ definiert: $\mathbf{L} = h L_h$, $\mathbf{f} = h f_h$.

Dieses Beispiel zeigt, dass es möglich ist, Basisfunktionen mit kleinen Trägern zu finden, sodass \mathbf{L} verhältnismäßig leicht berechenbar ist. Außerdem zeigt sich eine Ähnlichkeit der entstehenden Diskretisierung mit Differenzenverfahren. Abschließend sei auf die numerische Berechnung der Integrale (8.3.5b) hingewiesen.

Übungsaufgabe 8.3.5 Man zeige: Die Gauß-Quadraturformel⁵ für $\int_{x_{i-1}}^{x_{i+1}} g b_i dx$ mit b_i als Gewichtsfunktion und einer Stützstelle lautet:

$$f_i := \frac{x_{i+1} - x_{i-1}}{2} g\left(\frac{x_{i+1} + x_i + x_{i-1}}{3}\right). \quad (8.3.5c)$$

Was ergibt sich für eine äquidistante Intervallunterteilung?

Ersetzt man die Dirichlet-Randbedingung in (8.3.1) durch die Neumann-Bedingung, so ist $V = H^1(a, b)$. Der Unterraum $V_N \subset H^1(a, b)$ ergibt sich aus (8.3.2) nach Streichung der Bedingung $u(a) = u(b) = 0$. Damit $\dim V_N = N$ gilt, muss die Nummerierung geändert werden: Die Zerlegung von (a, b) lautet jetzt $a = x_1 < x_2 < \dots < x_N = b$.

⁵Johann Carl Friedrich Gauß, geb. 30. April 1777 in Braunschweig, gest. 23. Feb. 1855 in Göttingen

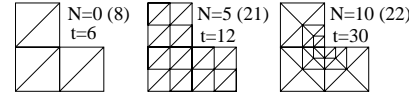
Übungsaufgabe 8.3.6 Sei $\partial u(a)/\partial n = g_a$, $\partial u(b)/\partial n = g_b$ und $a(\cdot, \cdot)$ wie vorher. Die Zerlegung von (a, b) sei äquidistant: $x_i = a + (b - a)i/N$. Wie lautet die Gleichung $\mathbf{L}\mathbf{u} = \mathbf{f}$? Man zeige: $\mathbf{L}\mathbf{u} = \mathbf{f}$ hat mindestens eine Lösung, falls $\int_a^b g(x)dx + g_a + g_b = 0$.

In §8.3.2-8.3.4 diskutieren wir typische Beispiele für Finite-Element-Funktionen in zwei Dimensionen. Weitere findet man bei Braess [12, Tabelle 2 auf Seite 63].

8.3.2 Lineare Elemente für $\Omega \subset \mathbb{R}^2$

Wir nehmen an:

$$\Omega \subset \mathbb{R}^2 \text{ sei Polygon.} \quad (8.3.6)$$



Wie in Abbildung 8.3.2 in drei Versionen angedeutet, zerlegen wir Ω in Dreiecke ("finite Elemente"). Die Zerlegung wird *Triangulation* genannt, wobei der Name auch beibehalten wird, wenn später andere Elemente (Vierecke, Tetraeder usw.) verwendet werden.

Abb. 8.3.2. Triangulation von Ω ; t : Anzahl der Dreiecke, N : Anzahl der inneren Knoten (aller Knoten)

Definition 8.3.7 $\mathcal{T} := \{T_1, \dots, T_t\}$ heißt zulässige Triangulation von Ω , falls die folgenden Bedingungen erfüllt sind:


$$T_i \quad (1 \leq i \leq t) \text{ sind offene Dreiecke,} \quad (8.3.7a)$$

$$T_i \text{ sind disjunkt: } T_i \cap T_j = \emptyset \text{ für } i \neq j, \quad (8.3.7b)$$

$$\bigcup_{i=1, \dots, t} \bar{T}_i = \bar{\Omega}, \quad (8.3.7c)$$

$$\text{für } i \neq j \text{ ist } \bar{T}_i \cap \bar{T}_j \begin{cases} i) & \text{entweder leer oder} \\ ii) & \text{gemeinsame Seite der Elemente } T_i, T_j \text{ oder} \\ iii) & \text{gemeinsame Ecke der Elemente } T_i, T_j. \end{cases} \quad (8.3.7d)$$

Bemerkung 8.3.8 a) Die Bedingungen (8.3.7a) und (8.3.7c) implizieren die Polygongestalt von Ω , d.h. die Voraussetzung (8.3.6).

b) Abbildung 8.3.2 zeigt nur zulässige Triangulationen. Eine unzulässige Triangulation läge z.B. vor, wenn ein Quadrat wie folgt zerlegt wird: , da die halbe Diagonale nicht Seite beider sich berührenden Dreiecke ist.

\mathcal{T} sei eine zulässige Triangulation. \mathbf{x} heißt *Knoten* (von \mathcal{T}), falls \mathbf{x} Eckpunkt eines Dreiecks $T \in \mathcal{T}$ ist. Dabei unterscheidet man innere und Rand-Knoten, je nachdem ob $\mathbf{x} \in \Omega$ oder $\mathbf{x} \in \partial\Omega$. Sei N die Gesamtanzahl der inneren Knoten. Wir definieren V_N als den Unterraum der stückweise linearen Funktionen:

$$V_N = \left\{ u \in C^0(\bar{\Omega}) : u = 0 \text{ auf } \partial\Omega, \text{ auf jedem Dreieck } T \in \mathcal{T} \text{ stimmt } u|_T \text{ mit einer linearen Funktion überein, d.h. } u(x, y) = a_{T,1} + a_{T,2}x + a_{T,3}y \text{ auf } T \right\} \quad (8.3.8)$$

Bemerkung 8.3.2 überträgt sich auf den vorliegenden zweidimensionalen Fall:

Bemerkung 8.3.9 a) Es gilt $V_N \subset H_0^1(\Omega)$. b) Jede Funktion $u \in V_N$ ist durch ihre Knotenwerte $u(\mathbf{x}^i)$ in den inneren Knoten \mathbf{x}^i , $1 \leq i \leq N$, eindeutig festgelegt.

Beweis. a) Beispiel 6.2.5 zeigt $V_N \subset H^1(\Omega)$. Da $u = 0$ auf $\partial\Omega$, beweist Satz 6.2.43, dass $V_N \subset H_0^1(\Omega)$.

b) Seien $\mathbf{x}, \mathbf{x}', \mathbf{x}''$ die drei Eckpunkte von $T \in \mathcal{T}$. Die lineare Funktion $u(x, y) = a_{i1} + a_{i2}x + a_{i3}y$ ist durch die Werte $u(\mathbf{x}), u(\mathbf{x}'), u(\mathbf{x}'')$ in T eindeutig bestimmt. ■

Die Umkehrung der Bemerkung 8.3.9b lautet wie folgt:

Bemerkung 8.3.10 \mathbf{x}^i ($1 \leq i \leq N$) seien die inneren Knoten von \mathcal{T} . Für beliebige u_i ($1 \leq i \leq N$) existiert genau ein $u \in V_N$ mit $u(\mathbf{x}^i) = u_i$. Es ist dargestellt durch $u = \sum_{i=1}^N u_i b_i$, wobei die Basisfunktionen b_i durch

$$b_i(\mathbf{x}^i) = 1, \quad b_i(\mathbf{x}^j) = 0 \quad \text{für alle } i \neq j \quad (1 \leq i, j \leq N) \quad (8.3.9a)$$

charakterisiert sind. Ist $T \in \mathcal{T}$ ein Dreieck mit den Eckpunkten $\mathbf{x}^i = (x_i, y_i)$ [\mathbf{x}^i wie in (8.3.9a)] und $\mathbf{x}' = (x', y')$, $\mathbf{x}'' = (x'', y'')$, so ist

$$b_i(x, y) = \frac{(x - x')(y'' - y') - (y - y')(x'' - x')}{(x_i - x')(y'' - y') - (y_i - y')(x'' - x')} \quad \text{auf } T. \quad (8.3.9b)$$

Auf allen Dreiecken $T \in \mathcal{T}$, die \mathbf{x}^i nicht als Eckpunkt besitzen, gilt $b_i|_T = 0$.

Beweis. Offenbar ist mit (8.3.9b) eine auf allen $T \in \mathcal{T}$ lineare Funktion gefunden, die (8.3.9a) erfüllt. (8.3.9a) erzwingt die Stetigkeit in allen Knotenpunkten. Sind die Ecken \mathbf{x}^j und \mathbf{x}^k durch eine Dreiecksseite direkt verbunden, ergibt (8.3.9b) als Beschränkung auf die Kante die lineare Funktion $b_i(\mathbf{x}^j + s(\mathbf{x}^k - \mathbf{x}^j)) = sb_i(\mathbf{x}^k) + (1 - s)b_i(\mathbf{x}^j)$ mit $s \in [0, 1]$ für beide Dreiecke, die diese Kante gemeinsam haben. Damit ist b_i auch auf den Kantenlinien stetig, sodass $b_i \in C^0(\Omega)$. Die Anwendung dieser Überlegungen auf zwei Randknoten $\mathbf{x}', \mathbf{x}''$ mit $b_i(\mathbf{x}') = b_i(\mathbf{x}'') = 0$ ergibt $b_i = 0$ auf $\partial\Omega$. Folglich gehört b_i zu V_N . ■

Bemerkung 8.3.11 a) Die Dimension des durch (8.3.8) beschriebenen Unterraumes ist die Anzahl der inneren Knotenpunkte von \mathcal{T}

b) Träger der Basisfunktion b_i zum Knoten \mathbf{x}^i ist


$$Tr(b_i) = \bigcup \{ \bar{T} : T \in \mathcal{T} \text{ hat } \mathbf{x}^i \text{ als Ecke} \}.$$

c) B_i sei das Innere von $Tr(b_i)$. $B_i \cap B_j = \emptyset$ gilt genau dann, wenn die Knoten \mathbf{x}^i und \mathbf{x}^j nicht direkt durch eine Kante verbunden sind.

Folgerung 8.3.12 $a(u, v) = \int_{\Omega} \langle \nabla u, \nabla v \rangle dx$ ist die zur Poisson-Gleichung gehörende Bilinearform. Die Integrale $L_{ij} = a(b_j, b_i) = \sum_k \int_{T_k} \langle \nabla b_j, \nabla b_i \rangle dx$ sind über folgende Dreiecke T_k zu erstrecken:

$$\begin{aligned} &\text{alle } T_k \text{ mit } \mathbf{x}^i \text{ als Ecke,} && \text{falls } i = j, \\ &\text{alle } T_k \text{ mit } \mathbf{x}^i \text{ und } \mathbf{x}^j \text{ als Ecke,} && \text{falls } i \neq j. \end{aligned}$$

Es gilt $L_{ij} = 0$, falls \mathbf{x}^i und \mathbf{x}^j nicht direkt durch eine Dreiecksseite verbunden sind.

Eine besonders regelmäßige Triangulation liegt vor, wenn man zunächst in Quadrate der Seitenlänge h zerlegt, die anschließend in zwei Dreiecke unterteilt werden () . Die erste und zweite Triangulation in Abbildung 8.3.2 ist von dieser Art. Wir bezeichnen sie als die "Quadrattgittertriangulation". Die zugehörige Basisfunktion ist in Abbildung 8.3.3 wiedergegeben. Sie hat einen aus sechs Dreiecken bestehenden Träger. Man erwartet deshalb, dass die Matrix \mathbf{L} einer Siebenpunktformel entspricht. Für den Laplace-Operator erhält man trotzdem die aus Abschnitt 4.2 wohlbekannte Fünfpunktformel (4.2.11):

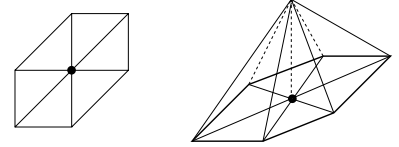


Abb. 8.3.3. Stückweise lineare Basisfunktion; links: Dreiecke des Trägers, rechts: Funktionsskizze

Übungsaufgabe 8.3.13 \mathcal{T} sei eine Quadrattgittertriangulation. Ferner sei $a(u, v) = \int_{\Omega} \langle \nabla u, \nabla v \rangle dx$. Die Basisfunktionen sind durch Abbildung 8.3.3 beschrieben. Man zeige: Für die Systemmatrix \mathbf{L} gilt

$$L_{ii} = 4, \quad L_{ij} = -1, \quad \text{falls } \mathbf{x}^i - \mathbf{x}^j \in \{(0, h), (0, -h), (h, 0), (-h, 0)\}, \quad L_{ij} = 0 \text{ sonst,}$$

d.h. \mathbf{L} stimmt mit der Matrix $h^2 L_h$ aus (4.2.8) überein.

Wenn auch $\mathbf{L} = h^2 L_h$ im Falle der Poisson-Gleichung $-\Delta u = g$ gilt, so stimmen die Finite-Element-Diskretisierung und das Differenzenverfahren doch nicht überein, denn $h^2 f_h$ hat $h^2 g(x^i)$ als Komponenten und weicht somit von $f_i = \int_{\Omega} g b_i dx$ ab.

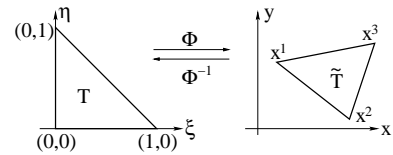
Die Integration $\int_{T_i} \dots dx$ über die (im Allgemeinen verschieden geformten) Dreiecke $T_i \in \mathcal{T}$ scheint zunächst umständlich. Für alle i kann man jedoch $\int_{T_i} \dots dx$ durch ein Integral über das Referenzdreieck T aus Abbildung 8.3.4 ausdrücken. Die Details finden sich in

Übungsaufgabe 8.3.14 $\mathbf{x}^i = (x_i, y_i)$ ($i = 1, 2, 3$) seien die Ecken von $\tilde{T} \in \mathcal{T}$, und T sei das Einheitsdreieck aus Abbildung 8.3.4. Man zeige:

- a) $\Phi : (\xi, \eta) \mapsto \mathbf{x}^1 + \xi(\mathbf{x}^2 - \mathbf{x}^1) + \eta(\mathbf{x}^3 - \mathbf{x}^1)$ bildet T auf \tilde{T} ab.
 b) $\det \Phi'(\xi, \eta) = (x_2 - x_1)(y_3 - y_1) - (y_2 - y_1)(x_3 - x_1)$ für alle $\xi, \eta \in \mathbb{R}$ ist eine Konstante.
 c) Die Substitutionsregel liefert

$$\int_{\tilde{T}} v(x, y) dx dy = [(x_2 - x_1)(y_3 - y_1) - (y_2 - y_1)(x_3 - x_1)] \int_T v(\Phi(\xi, \eta)) d\xi d\eta. \quad (8.3.10)$$

Im Allgemeinen wird man das Integral $\int_T \dots d\xi d\eta$ über dem Einheitsdreieck numerisch auswerten. Beispiele für Integrationsformeln findet man bei Schwarz [75, p. 107] und Ciarlet [22, §4.1]. Die erforderliche Ordnung der Quadraturformel wird an gleicher Stelle von Ciarlet [22] und von Witsch [88] diskutiert. Anders als Differenzenverfahren bietet die Finite-Element-Diskretisierung die Möglichkeit, die Dreiecksgröße lokal zu verändern. Die dritte Triangulation aus Abbildung 8.3.2

Abb. 8.3.4. Referenzdreieck T

enthält Dreiecke, die sich mit zunehmender Nähe zur einspringenden Ecke verkleinern. Diese Flexibilität der Finite-Element-Methode ist ein wesentlicher Vorteil. Andererseits erhält man komplizierter strukturierte Gleichungssysteme $\mathbf{L}\mathbf{u} = \mathbf{f}$, denn (a) \mathbf{u} lässt sich nicht mehr in einem zweidimensionalen Array speichern, (b) \mathbf{L} lässt sich nicht durch einen Stern (4.2.12) charakterisieren.

Bemerkung 8.3.15 Ersetzt man die Dirichlet-Bedingung $u = 0$ auf $\partial\Omega$ durch eine natürliche Randbedingung, so ergeben sich folgende Änderungen:

- a) $N = \dim V_N$ ist die Anzahl aller Knoten (innerer und Randknoten).
 b) $V_N \subset H^1(\Omega)$ ist durch (8.3.8) ohne die Einschränkung “ $u = 0$ auf $\partial\Omega$ ” gegeben.
 c) In den Bemerkungen 8.3.9-11 muss es “Knoten” anstelle von “innere Knoten” heißen.
 d) Die in Übungsaufgabe 8.3.13 berechneten Matrixelemente L_{ij} gelten nur für innere Knoten \mathbf{x}^i . Im Falle eines Neumann-Randwertproblems im Quadrat $\Omega = (0, 1) \times (0, 1)$, stimmt \mathbf{L} mit $h^2 D_h L_h$ aus Übungsaufgabe 4.7.8b überein.
 e) Bei der Berechnung von $f_i = f(b_i)$ hat man noch eventuell auftretende Randintegrale über $\partial\Omega \cap \text{Tr}(b_i)$ zu berücksichtigen, falls $\text{Tr}(b_i) \cap \partial\Omega \neq \emptyset$.

8.3.3 Bilineare Elemente für $\Omega \subset \mathbb{R}^2$

Die Differenzenverfahren im quadratischen Gitter legen eine Zerlegung von Ω in Quadrate der Seitenlänge h nahe (siehe linke Abbildung 8.3.5). Ersetzt man die Quadrate allgemeiner durch Parallelogramme, erhält man Zerlegungen wie in Abbildung 8.3.5, was allerdings nur für spezielle Polygone Ω möglich ist. Eine zulässige Parallelogrammzerlegung ist durch die Bedingungen (8.3.7a-d) beschrieben, wenn in (8.3.7a) die Bezeichnung “Dreieck” durch “Parallelogramm” ersetzt wird.

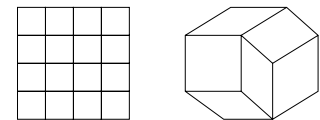
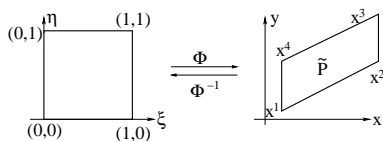
Abb. 8.3.5. Zerlegung von Ω in Parallelogramme

Abb. 8.3.6. Einheitsquadrat als Referenzparallelogramm

Würde man den Unterraum V_N dadurch definieren, dass u in jedem Parallelogramm eine lineare Funktion ist, wären nur noch drei der vier Eckwerte beliebig vorgebar. Für eine Zerlegung wie in Abbildung 8.3.5 (rechts) kann man sich überlegen, dass die einzige stückweise lineare Funktion u mit $u = 0$ auf $\partial\Omega$ die Nullfunktion ist. In jedem Parallelogramm muss u daher eine Funktion sein, die vier freie Parameter enthält. Wir betrachten zunächst den Fall eines *achsenparallelen Rechtecks* $P = (x_1, x_2) \times (y_1, y_4)$ und definieren eine bilineare Funktion auf P mittels

$$u(x, y) := a_{P,1} + a_{P,2}x + a_{P,3}y + a_{P,4}xy \quad \text{in } P. \quad (8.3.11a)$$

u ist in jeder achsenparallelen Richtung – also insbesondere auf den Rechtecksseiten – linear. Für ein beliebiges Parallelogramm \tilde{P} wie in Abbildung 8.3.6 ist die Beschränkung der Funktion (8.3.11a) auf eine Parallelogrammseite jedoch eine im Allgemeinen quadratische Funktion. Daher verallgemeinert man die Definition wie folgt. Sei

$$\Phi : (\xi, \eta) \mapsto \mathbf{x}^1 + \xi(\mathbf{x}^2 - \mathbf{x}^1) + \eta(\mathbf{x}^4 - \mathbf{x}^1) \in \mathbb{R}^2 \quad (8.3.11b)$$

die Abbildung des Einheitsquadrates $(0, 1) \times (0, 1)$ auf das Parallelogramm \tilde{P} (vgl. Abbildung 8.3.6). Eine bilineare Funktion auf \tilde{P} ist durch

$$u(x, y) := v(\Phi^{-1}(x, y)), \quad v(\xi, \eta) := \alpha + \beta\xi + \gamma\eta + \delta\xi\eta \quad (8.3.11c)$$

definiert. Es ist nicht notwendig, $v(\Phi^{-1}(\cdot, \cdot))$ explizit auszurechnen, da sich alle Integrationen über dem Einheitsquadrat $(0, 1) \times (0, 1)$ als Referenzparallelogramm ausführen lassen (vgl. Übungsaufgabe 8.3.14).

Ist $\pi = \{P_1, \dots, P_t\}$ eine zulässige Zerlegung in Parallelogramme, so definiert man $V_N \subset H^1(\Omega)$ bzw. $V_N \subset H_0^1(\Omega)$ durch

$$V_N = \{u \in C^0(\bar{\Omega}) : \text{auf allen } P \in \pi \text{ ist } u \text{ eine bilineare Funktion}\} \quad (8.3.12a)$$

bzw.

$$V_N = \{u \in C^0(\bar{\Omega}) : u = 0 \text{ auf } \Gamma = \partial\Omega, \text{ auf allen } P \in \pi \text{ stimmt } u \text{ mit einer bilinearen Funktion überein}\}. \quad (8.3.12b)$$

Dabei ist $N = \dim V_N$ die Anzahl der aller Knoten (Fall (8.3.12a)) bzw. der inneren Knoten (Fall (8.3.12b)). Eine auf $P \in \pi$ bilineare Funktion ist entlang jeder Seite von P linear. Stetigkeit in den Knotenpunkten impliziert daher schon Stetigkeit in Ω .

Bemerkung 8.3.16 Die Bemerkungen 8.3.8a, 8.3.9-11 gelten sinngemäß.

Übungsaufgabe 8.3.17 Sei $a(u, v) = \int_{\Omega} \langle \nabla u, \nabla v \rangle dx$ die zu $-\Delta$ gehörende Bilinearform. $\Omega = (0, 1) \times (0, 1)$ sei wie in Abbildung 8.3.5 (links) in Quadrate der Seitenlänge h zerlegt. Wie lauten die durch (8.3.9a) charakterisierten Basisfunktionen? Man zeige: \mathbf{L} stimmt mit dem Differenzenstern $\frac{1}{3} \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$ überein.

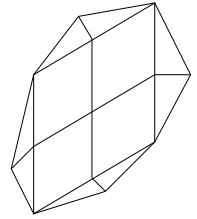


Abb. 8.3.7. Kombination von Dreiecken und Parallelogrammen

Bemerkung 8.3.18 Die Dreiecks- und Parallelogrammzerlegungen lassen sich auch kombiniert anwenden. Ein Polygonegebiet Q kann in Dreiecke und Parallelogramme zerlegt werden (vgl. Abbildung 8.3.7). V_N ist in diesem Fall durch $\{u \in C^0(\Omega) : u \text{ linear auf den Dreiecken, bilinear auf den Parallelogrammen}\}$ (mit eventueller Nullrandbedingung) zu definieren.

8.3.4 Quadratische Elemente für $\Omega \subset \mathbb{R}^2$

\mathcal{T} sei eine zulässige Triangulation eines Polygonegebietes Ω . Wir wollen die Dimension des Finite-Element-Unterraumes dadurch vergrößern, dass wir anstelle linearer nun quadratische Funktionen

$$u(x, y) = a_{T,1} + a_{T,2}x + a_{T,3}y + a_{T,4}x^2 + a_{T,5}xy + a_{T,6}y^2 \quad \text{auf } T \in \mathcal{T} \quad (8.3.13)$$

zulassen:

$$V_N = \{u \in C^0(\bar{\Omega}) : u = 0 \text{ auf } \Gamma = \partial\Omega, \text{ auf jedem } T_i \in \mathcal{T} \text{ stimmt } u \text{ mit einer quadratischen Funktion überein}\}. \quad (8.3.14)$$

Lemma 8.3.19 a) $\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3$ seien die Ecken eines Dreiecks $T \in \mathcal{T}$, während $\mathbf{x}^4, \mathbf{x}^5, \mathbf{x}^6$ die Seitenmittelpunkte⁶ seien (vgl. Abbildung 8.3.8a). Jede auf T quadratische Funktion ist eindeutig durch die Werte $\{u(\mathbf{x}^j) : j = 1, \dots, 6\}$ beschrieben.

b) Die Beschränkung der Funktion (8.3.13) auf eine Seite von $T \in \mathcal{T}$ ergibt eine eindimensionale quadratische Funktion, die eindeutig durch die drei auf dieser Seite liegenden Knotenwerte (z.B. $u(\mathbf{x}^1)$, $u(\mathbf{x}^4)$, $u(\mathbf{x}^2)$ in Abbildung 8.3.8a) bestimmt wird.

c) Ist u auf jedem $T \in \mathcal{T}$ quadratisch und stetig in allen Knotenpunkten (d.h. Dreiecksecken und Seitenmittelpunkte), so ist u stetig auf Ω .

Beweis. a) Die quadratische Funktion ergibt sich als eindeutig definiertes Interpolationspolynom der Form (8.3.13). Auch Teil b) ist elementar.

c) T und \hat{T} seien die benachbarten Dreiecke aus Abbildung 8.3.8a. Da $u|_T$ und $u|_{\hat{T}}$ in $\mathbf{x}^1, \mathbf{x}^4, \mathbf{x}^2$ übereinstimmen, stellen sie nach Teil b) die gleiche quadratische Funktion auf der gemeinsamen Seite von T und \hat{T} dar. ■

⁶Statt der Seitenmittelpunkte könnte jeder andere Punkt der Seite (ohne die Eckpunkte) gewählt werden. Die Wahl muss aber für beide an dieser Seite anliegenden Dreiecke übereinstimmen.

Als (innere) Knoten bezeichnen wir jetzt alle (inneren) Dreiecksecken und (inneren) Seitenmittelpunkte. Nach Lemma 8.3.19a können wir zu jedem inneren Knoten \mathbf{x}^i eine Basisfunktion b_i finden, die auf jedem $T \in \mathcal{T}$ quadratisch ist und $b_i(\mathbf{x}^i) = 1$, $b_i(\mathbf{x}^j) = 0$ für $j \neq i$ erfüllt. Gemäß Lemma 8.3.19c gehört b_i zu V_N . Dies beweist:

Bemerkung 8.3.20 Die Anzahl der inneren Knoten ist die Dimension von V_N aus (8.3.14).

Jedes $u \in V_N$ hat die Darstellung $u = \sum_{i=1}^N u(\mathbf{x}^i) b_i$, wobei die zum Knoten \mathbf{x}^i gehörende Basisfunktion durch (8.3.9a) charakterisiert ist.

Im Weiteren wollen wir annehmen, dass wie in Abbildung 8.3.7 sowohl Dreiecke als auch Parallelogramme bei der Zerlegung verwendet werden. Auf den Dreiecken seien die Funktionen $u \in V_N$ quadratisch. Die Ansatzfunktion auf den Parallelogrammen P muss folgenden Bedingungen genügen:

- (α) $u(x, y)|_P$ muss durch die Werte in den 4 Ecken und 4 Seitenmitten eindeutig bestimmt sein (vgl. Abbildung 8.3.8b).
- (β) Die Beschränkung auf eine Seite von P stellt eine (eindimensionale) quadratische Funktion (der Bogenlänge) dar.

Wegen Bedingung (α) muss der Ansatz genau 8 Koeffizienten besitzen. Die quadratische Funktion (8.3.13) besitzt nur 6 Koeffizienten, während ein biquadratischer Ansatz $\sum_{0 \leq i, j \leq 2} a_{ij} \xi^i \eta^j$ einen Parameter zuviel enthält. Lässt man im biquadratischen Ansatz den Term $\xi^2 \eta^2$ weg, ergibt sich für das Einheitsquadrat die Funktion

$$v(\xi, \eta) := a_1 + a_2 \xi + a_3 \eta + a_4 \xi^2 + a_5 \xi \eta + a_6 \eta^2 + a_7 \xi^2 \eta + a_8 \xi \eta^2, \quad (8.3.15)$$

der sogenannte quadratische Ansatz der *Serendipity-Klasse*. Die Beschränkungen auf die vier Seiten $\xi = 0$, $\xi = 1$ bzw. $\eta = 0$, $\eta = 1$ ergeben eine quadratische Funktion in η bzw. in ξ . Die Abbildung Φ aus (8.3.11b) (vgl. Abbildung 8.3.6a,b) ergibt die auf dem Parallelogramm P definierte Funktion $u(x, y) = v(\Phi^{-1}(x, y))$, die weiterhin die Forderungen (α), (β) erfüllt.

Würde man den vollen, biquadratischen Ansatz verwenden, braucht man einen weiteren Knoten, den man als Parallelogrammschwerpunkt wählen kann. Ebenso lassen sich kubische Ansätze in Dreiecken und Parallelogrammen durchführen (vgl. Schwarz [75]).

8.3.5 Elemente für $\Omega \subset \mathbb{R}^3$

Im dreidimensionalen Fall sei anstelle von (8.3.6) vorausgesetzt:

$$\Omega \subset \mathbb{R}^3 \quad \text{sei Polyeder.} \quad (8.3.16)$$

Der Triangulation aus §8.3.2 entspricht jetzt eine Zerlegung von Ω in Tetraeder. Die Zerlegung heißt *zulässig*, falls (8.3.7a-d) sinngemäß gelten: (8.3.7a) wird zu “ T_i ($1 \leq i \leq t$) sind offene Tetraeder”, in (8.3.7d) muss es nun heißen “Für $i \neq j$ ist $\bar{T}_i \cap \bar{T}_j$ entweder leer oder eine gemeinsame Ecke, Kante oder Seite von T_i und T_j ”. Jede lineare Funktion $a_1 + a_2 x + a_3 y + a_4 z$ ist eindeutig durch die Werte in den 4 Ecken des Tetraeders bestimmt. Als Basis des Raumes

$$V_N = \{u \in C^0(\bar{\Omega}) : u = 0 \text{ auf } \partial\Omega, u \text{ linear auf jedem Tetraeder } T_i (1 \leq i \leq t)\}$$

wählt man b_i mit der Eigenschaft (8.3.9a): $b_i(\mathbf{x}^i) = 1$, $b_i(\mathbf{x}^j) = 0$ ($j \neq i$). Der Träger von b_i besteht aus allen Tetraedern, die \mathbf{x}^i als Eckpunkt teilen. Die Dimension $N = \dim V_N$ ist wieder die Anzahl der inneren Knotenpunkte (d.h. Tetraederecken). Wie im zweidimensionalen Fall lässt sich der lineare Ansatz durch einen quadratischen ersetzen. Anstelle des Tetraeders kann auch ein Parallelepipiped oder ein dreieckiges Prisma mit entsprechenden Funktionsansätzen verwendet werden (vgl. Schwarz [75]).

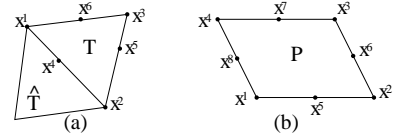


Abb. 8.3.8. Knoten für einen quadratischen Ansatz (a) auf einem Dreieck, (b) der Serendipity-Klasse auf einem Parallelogramm

8.3.6 Behandlung von Nebenbedingungen

Mitunter ist der zugrundeliegende Raum V eine Teilmenge eines einfacher diskretisierbaren Raumes $W \supset V$. Dabei gehöre eine Funktion $w \in W$ zu V , falls gewisse Nebenbedingungen erfüllt sind. Bevor diese Situation in allgemeiner Form beschrieben wird, seien zwei Beispiele als Illustration vorgeführt.

Beispiel 8.3.21 Will man die Neumann-Randwertaufgabe $-\Delta u = g$ in Ω und $\partial u / \partial n = \varphi$ auf Γ durch die Nebenbedingung $\int_{\Omega} u(x) dx = 0$ eindeutig lösbar machen, kann man den Raum V als

$$V = \left\{ u \in H^1(\Omega) : \int_{\Omega} u(x) dx = 0 \right\} \quad (8.3.17a)$$

wählen. Bei beschränktem Ω ist die Form $a(u, v) := \int_{\Omega} \langle \nabla u, \nabla v \rangle dx$ V -elliptisch. Der schwachen Formulierung (8.1.1) mit $f(v) := \int_{\Omega} g v dx + \int_{\Gamma} \varphi v d\Gamma$ entspricht die Gleichung $-\Delta u = g$, $\partial u / \partial n = \varphi$, falls g und φ der Integrabilitätsbedingung $f(1) = 0$ genügen (vgl. (3.4.2)). Aber auch wenn $f(1) \neq 0$, existiert eine schwache Lösung, die zur korrigierten Gleichung $-\Delta u = \tilde{g}$, $\tilde{g}(x) := g(x) - [\int_{\Omega} g dx + \int_{\Gamma} \varphi d\Gamma] / \int_{\Omega} dx$ gehört.

Beispiel 8.3.22 Das Adler-Problem aus Übungsaufgabe 7.4.15 benutzt

$$V = \{ u \in H^1(\Omega) : u \text{ konstant auf } \Gamma \} \quad (8.3.17b)$$

In beiden Fällen ist $W = H^1(\Omega)$ eine echte Obermenge von V . Sei \mathcal{T} eine Triangulation von Ω mit I_{in} inneren Knoten und I_R Randknoten. $W_h \subset W$ sei der Raum der linearen Dreieckselemente⁷ (vgl. Bemerkung 8.3.15). Seine Dimension ist

$$N_h := \dim W_h = I_{in} + I_R. \quad (8.3.18a)$$

Die Basisfunktionen $\{b_i : 1 \leq i \leq N_h\} \subset W_h$ seien wie üblich durch $b_i(\mathbf{x}^j) = \delta_{ij}$ beschrieben (vgl. (8.3.9a)). Als Finite-Element-Raum zu V definieren wir $V_h := W_h \cap V$, der eine kleinere Dimension besitzt. Die Differenz M_h entspricht die Anzahl der Nebenbedingungen:

$$V_h := W_h \cap V, \quad M_h := N_h - \dim V_h. \quad (8.3.18b)$$

In den Beispielen 8.3.21 und 8.3.22 gilt $M_h = 1$ bzw. $M_h = I_R - 1$. Die Schwierigkeit bei der numerischen Lösung des diskreten Problems

$$\text{bestimme } u^h \in V_h \quad \text{mit} \quad a(u^h, v) = f(v) \quad \text{für alle } v \in V_h \quad (8.3.19)$$

beginnt bei der Wahl einer Basis von V_h . Es ist nicht möglich, eine Teilmenge der Funktionen b_1, \dots, b_{N_h} zu verwenden. So gehört etwa in Beispiel 8.3.21 kein b_i zu V und damit auch nicht zu V_h , denn $\int_{\Omega} b_i dx > 0$.

Im Prinzip ist es möglich, im Falle von (8.3.17a) Linearkombinationen $b_i := \alpha b_{i_1} + \beta b_{i_2} \in V_h$ als neue Basisfunktionen zu konstruieren, die wieder einen lokalen (aber etwas verbreiterten) Träger besitzen. Im Falle von (8.3.17b) ist es noch relativ einfach, eine praktikable Basis von V_h zu finden: Die Basisfunktionen $b_i \in W_h$, die zu inneren Knoten gehören, liegen auch in V_h , da $b_i \in H_0^1(\Omega) \subset V$. Als weiteres Element verwende man $b_0 := \sum' b_j$ wobei \sum' die Summe über alle Randknoten bezeichnet. Trotzdem wäre es auch in diesem Fall prinzipiell einfacher, könnte man mit der Standardbasis von W_h arbeiten.

Um die Aufgabe (8.3.19) mit Hilfe der $b_i \in W_h$ zu behandeln, führen wir wieder die Schreibweise $w^h = P\mathbf{w}$ ein ($\mathbf{w} \in \mathbb{R}^{N_h}$ Koeffizientenvektor, $w^h \in W_h$ und P aus (8.1.6)). Jedes $v \in V_h \subset W_h$ hat die Darstellung $P\mathbf{v}$ mit \mathbf{v} aus $\mathbf{V} := \{\mathbf{v} \in \mathbb{R}^{N_h} : P\mathbf{v} \in V_h\} = P^{-1}V_h$. Aufgabe (8.3.19) ist daher äquivalent zu

$$\text{bestimme } \mathbf{u} \in \mathbf{V} \quad \text{mit} \quad a(P\mathbf{u}, P\mathbf{v}) = f(P\mathbf{v}) \quad \text{für alle } \mathbf{v} \in \mathbf{V}. \quad (8.3.19')$$

Nach (8.3.18a,b) gilt $\dim \mathbf{V} = \dim V_h = N_h - M_h = \dim W_h - M_h$. V_h bzw. \mathbf{V} seien durch M_h lineare Nebenbedingungen

$$\sum_{j=1}^{N_h} c_{ij} w_j = 0 \quad (1 \leq i \leq M_h) \quad (8.3.20)$$

⁷Wir verwenden hier die (maximale) Schrittweite h als Index von V_h, W_h und M_h , da die Dimension in den Fällen mit und ohne Nebenbedingung verschieden ist.

beschrieben:

$$\mathbf{V} = \text{Kern}(\mathbf{C}) = \{\mathbf{w} \in \mathbb{R}^{N_h} : \mathbf{C}\mathbf{w} = \mathbf{0}\}, \quad (8.3.21)$$

wobei $\mathbf{C} = (c_{ij})$ die $M_h \times N_h$ -Matrix der Koeffizienten aus (8.3.20) darstellt. Im Falle des Beispiels 8.3.21 ist

$$M_h = 1, \quad c_{1j} = \int_{\Omega} b_j(x) dx \quad (1 \leq j \leq N_h).$$

Im Falle des Beispiels 8.3.22 seien $\mathbf{x}^0, \dots, \mathbf{x}^{M_h}$ mit $M_h = I_R - 1$ die Randknoten. Dann kann \mathbf{C} wie folgt definiert werden:

$$M_h = I_R - 1, \quad c_{ii} = 1 \quad (1 \leq i \leq M_h), \quad c_{i,i+1} = -1 \quad (1 \leq i < M_h), \quad c_{M_h,0} = -1, \quad c_{ij} = 0 \text{ sonst.}$$

Die Variation über $\mathbf{v} \in \mathbf{V}$ in (8.3.19') kann man durch die Variation über $\mathbf{w} \in \mathbb{R}^{N_h}$ ersetzen, wenn man die Nebenbedingungen (8.3.20) mit Hilfe von *Lagrange-Faktoren*⁸ λ_i ($1 \leq i \leq M_h$) ankoppelt, die zum Vektor $\lambda = (\lambda_1, \dots, \lambda_{M_h})$ zusammengefasst seien. Die resultierende Aufgabe lautet:

$$\text{bestimme } \mathbf{u} \in \mathbb{R}^{N_h} \quad \text{und} \quad \lambda \in \mathbb{R}^{M_h} \quad \text{mit} \quad \mathbf{C}\mathbf{u} = \mathbf{0} \quad \text{und} \quad (8.3.22a)$$

$$a(P\mathbf{u}, P\mathbf{w}) + \langle \lambda, \mathbf{C}\mathbf{w} \rangle = f(P\mathbf{w}) \quad \text{für alle } \mathbf{w} \in \mathbb{R}^{M_h}, \quad (8.3.22b)$$

wobei $\langle \lambda, \mu \rangle = \sum \lambda_i \mu_i$ das Skalarprodukt des \mathbb{R}^{M_h} ist.

Satz 8.3.23 Die Aufgaben (8.3.19) und (8.3.22a,b) sind in folgendem Sinne äquivalent. Ist \mathbf{u}, λ ein Lösungspaar von (8.3.22a,b), so löst \mathbf{u} auch (8.3.19') und $P\mathbf{u}$ ist Lösung von (8.3.19). Ist umgekehrt $u^h = P\mathbf{u}$ eine Lösung von (8.3.19), so existiert genau ein $\lambda \in \mathbb{R}^{M_h}$, sodass \mathbf{u} und λ die Aufgabe (8.3.22a,b) lösen.

Bevor wir den Satz beweisen, geben wir eine zu (8.3.22a,b) äquivalente Matrixformulierung an.

Bemerkung 8.3.24 \mathbf{L} und \mathbf{f} seien wie in (8.1.8a,b) definiert. Ferner sei $\mathbf{B} := \mathbf{C}^\top$ mit \mathbf{C} aus (8.3.21). Dann ist (8.3.22a,b) äquivalent zum Gleichungssystem

$$\begin{bmatrix} \mathbf{L} & \mathbf{B} \\ \mathbf{B}^\top & \mathbf{O} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{0} \end{bmatrix}. \quad (8.3.22')$$

Beweis. $\mathbf{L}\mathbf{u} + \mathbf{B}\lambda = \mathbf{f}$ ist äquivalent zu $a(P\mathbf{u}, P\mathbf{w}) + \langle \lambda, \mathbf{C}\mathbf{w} \rangle = \langle \mathbf{L}\mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{B}\lambda, \mathbf{w} \rangle = \langle \mathbf{f}, \mathbf{w} \rangle = f(P\mathbf{w})$ für alle $\mathbf{w} \in \mathbb{R}^{N_h}$. $\mathbf{B}^\top \mathbf{u} = \mathbf{0}$ stimmt mit der Nebenbedingung $\mathbf{C}\mathbf{u} = \mathbf{0}$ überein. ■

Beweis des Satzes 8.3.23. a) \mathbf{u}, λ löse (8.3.22a,b). $\mathbf{C}\mathbf{u} = \mathbf{0}$ impliziert $\mathbf{u} \in \mathbf{V}$ und $u^h := P\mathbf{u} \in V_h$. (8.3.22b) gilt insbesondere für alle $\mathbf{w} \in \mathbf{V}$, sodass wegen $\mathbf{C}\mathbf{w} = \mathbf{0}$ (8.3.19') folgt.

b) (8.3.19') impliziert $\mathbf{f} - \mathbf{L}\mathbf{u} \in \mathbf{V}^\perp$. Da $\mathbf{V}^\perp = (\text{Kern}(\mathbf{C}))^\perp = (\text{Kern}(\mathbf{B}^\top))^\perp = \text{Bild}(\mathbf{B})$, existiert ein $\lambda \in \mathbb{R}^{M_h}$ mit $\mathbf{f} - \mathbf{L}\mathbf{u} = \mathbf{B}\lambda$, sodass (8.3.22') und damit auch (8.3.22a,b) erfüllt sind. Für die $N_h \times M_h$ -Matrix \mathbf{B} mit $\text{Rang}(\mathbf{B}) = M_h$ gilt aber auch $\text{Kern}(\mathbf{B}) = \{0\}$, sodass die Lösung λ eindeutig ist. ■

Man beachte, dass im Falle des Beispiels 8.3.21 das Gleichungssystem (8.3.22') im Wesentlichen mit (4.7.10a,b) identisch ist.

8.4 Fehlerabschätzungen bei Finite-Elemente-Verfahren

In diesem Unterkapitel beschränken wir uns auf die linearen Elemente des Abschnitts 8.3.2 und setzen deshalb voraus:

$$\mathcal{T} \quad \text{sei eine zulässige Triangulation von } \Omega \subset \mathbb{R}^2, \quad (8.4.1a)$$

$$V_N \quad \text{sei durch (8.3.8) definiert, falls } V = H_0^1(\Omega), \quad (8.4.1b)$$

$$V_N \quad \text{sei wie in Bemerkung 8.3.15b, falls } V = H^1(\Omega). \quad (8.4.1c)$$

⁸ Joseph-Louis Lagrange, geb. 25. Jan. 1736 in Turin, gest. 10. April 1813 in Paris

8.4.1 Vorbereitungen

Nach Satz 8.2.1 hat man die charakteristische Größe $d(u, V_N) = \inf\{|u - w|_1 : w \in V_N\}$ zu bestimmen. Dazu untersucht man zunächst das Referenzdreieck aus Abbildung 8.3.4. Die rechte Seite in (8.4.2) besteht aus den Werten von u in den Eckpunkten von T und aus der L^2 -Norm der zweiten partiellen Ableitungen.

Lemma 8.4.1 Sei $T = \{(\xi, \eta) : \xi, \eta \geq 0, \xi + \eta \leq 1\}$ das Einheitsdreieck. Für alle $u \in H^2(T)$ gilt

$$\|u\|_{H^2(T)}^2 \leq C \left[|u(0,0)|^2 + |u(1,0)|^2 + |u(0,1)|^2 + \sum_{|\alpha|=2} \|D^\alpha u\|_{L^2(T)}^2 \right]. \quad (8.4.2)$$

Beweis. Zunächst soll gezeigt werden, dass die Bilinearform

$$a(u, v) := u(0,0)v(0,0) + u(1,0)v(1,0) + u(0,1)v(0,1) + \sum_{|\alpha|=2} (D^\alpha u, D^\alpha v)_{L^2(T)}$$

auf $H^2(T) \times H^2(T)$ stetig und $H^2(T)$ -koerziv ist. Die Stetigkeit ergibt sich aus der stetigen Einbettung $H^2(T) \subset C^0(T)$, die zu $|u(x)| \leq \hat{C}|u|_2$ für alle $x \in T$ führt (vgl. Satz 6.2.31). Somit folgt die Schranke $|a(u, v)| \leq (1 + 3\hat{C})|u|_2|v|_2$. Da $T \in C^{0,1}$, ist Satz 6.4.8b anwendbar: $H^2(T)$ ist kompakt eingebettet in $H^1(T)$. Gemäß Lemma 6.4.13 existiert eine Konstante $C_{1/2}$, sodass

$$|u|_1^2 \leq \left(\frac{1}{2}|u|_2 + C_{1/2}|u|_0 \right)^2 \leq \frac{1}{2}|u|_2^2 + 2C_{1/2}^2|u|_0^2 \quad \text{für alle } u \in H^2(T). \quad (*)$$

Da $\sum_{|\alpha|=2} |D^\alpha u|_0^2 = |u|_2^2 - |u|_1^2$, zeigt die Abschätzung $a(u, u) \geq |u|_2^2 - |u|_1^2 \geq \frac{1}{2}|u|_2^2 - 2C_{1/2}^2|u|_0^2$, dass $a(\cdot, \cdot)$ $H^2(T)$ -koerziv ist.

2) Da auch die Einbettung $H^2(T) \subset L^2(T)$ kompakt ist, lässt sich Satz 6.5.15 anwenden: Der zu $a(\cdot, \cdot)$ gehörende Operator $A \in L(H^2(T), H^2(T)')$ hat entweder eine Inverse $A^{-1} \in L(H^2(T)', H^2(T))$ oder $\lambda = 0$ ist ein Eigenwert, zu dem eine Eigenfunktion $0 \neq e \in H^2(T)$ existiert. Im letztgenannten Fall muss e insbesondere die Gleichung $a(e, e) = 0$ erfüllen. Hieraus folgt aber $D^\alpha e = 0$ für alle $|\alpha| = 2$, sodass e linear sein muss: $e(x, y) = \alpha + \beta x + \gamma y$. Ferner schließt man aus $a(e, e) = 0$ auch auf $e(0,0) = e(1,0) = e(0,1) = 0$, sodass $e = 0$ im Widerspruch zur Annahme. Über Lemma 6.5.3 und Übungsaufgabe 6.5.6c beweist man die $H^2(T)$ -Elliptizität: $a(u, u) \geq C_E|u|_2^2$ mit $C_E := 1/[\|A^{-1}\|_{H^2(T) \leftarrow H^2(T)'}^2(1 + 3\hat{C})] > 0$ und der Konstanten \hat{C} aus dem ersten Beweisteil. Somit ergibt sich Behauptung (8.4.2) mit $C := 1/C_E$. ■

Das Lemma 8.4.1 ist auf den Fall der linearen Elemente zugeschnitten, da nach einer geeigneten linearen Interpolation der Fehler in $(0,0)$, $(1,0)$, $(0,1)$ verschwindet. Die Verallgemeinerung auf höhere Elemente lautet wie folgt: Seien $\mathbf{x}^i \in T$ ($1 \leq i \leq q$) Stützstellen, die ein Interpolationspolynom vom Grad $\leq t-1$ eindeutig bestimmen. Dann sind die Normen $\|u\|_{H^t(T)}$ und $\sqrt{\sum_{i=1}^q |u(\mathbf{x}^i)|^2 + \sum_{|\alpha|=t} \|D^\alpha u\|_{L^2(T)}^2}$ äquivalent.

Abschätzung (8.4.2) ist auch für andere Dreiecke T als das Einheitsdreieck gültig, aber die Konstante C hängt von T ab. In Lemma 8.4.2 wird zunächst nur ein skaliertes Einheitsdreieck untersucht. Man beachte, dass die verschiedenen Ableitungsterme unterschiedlich skalieren. Darum wird nicht $\|u\|_{H^2(T_h)}^2$, sondern ein homogener Ausdruck $\|D^\beta u\|_{L^2(T)}^2$ abgeschätzt.

Lemma 8.4.2 Seien $h > 0$ und $T_h := hT := \{(x, y) : x, y \geq 0, x + y \leq h\}$. Für jedes $u \in H^2(T_h)$ gilt

$$\|D^\beta u\|_{L^2(T)}^2 \leq C \left\{ h^{2-2|\beta|} \left[|u(0,0)|^2 + |u(1,0)|^2 + |u(0,1)|^2 \right] + h^{4-2|\beta|} \sum_{|\alpha|=2} \|D^\alpha u\|_{L^2(T)}^2 \right\} \quad (8.4.3)$$

für alle $\beta \in \mathbb{N}_0^2$ mit $|\beta| \leq 2$. C ist die Konstante aus (8.4.2) und damit unabhängig von u, β und h .

Beweis. Sei $v(\xi, \eta) := u(\xi h, \eta h) \in H^2(T)$. Die Ableitungen nach (x, y) und $(\xi, \eta) = (x, y)/h$ sind durch $D_{x,y}^\beta = h^{-|\beta|} D_{\xi,\eta}^\beta$ verknüpft. Für jedes $|\beta| \leq 2$ ergibt die Substitutionsregel

$$\begin{aligned} \|D^\beta u\|_{L^2(T_h)}^2 &= \iint_{T_h} |D_{x,y}^\beta u|^2 dx dy = \iint_T |h^{-|\beta|} D_{\xi,\eta}^\beta v|^2 h^2 d\xi d\eta = h^{2-2|\beta|} \|D^\beta v\|_{L^2(T)}^2 \quad (**) \\ &\leq h^{2-2|\beta|} \|v\|_{H^2(T)}^2 \stackrel{\text{Lemma 8.4.1}}{\leq} C h^{2-2|\beta|} \left[|v(0,0)|^2 + |v(1,0)|^2 + |v(0,1)|^2 + \sum_{|\alpha|=2} \|D^\alpha v\|_{L^2(T)}^2 \right]. \end{aligned}$$

Da $v(0,0) = u(0,0)$, $v(1,0) = u(h,0)$, $v(0,1) = u(0,h)$ und $\|D_{\xi,\eta}^\alpha v\|_{L^2(T)}^2 = h^2 \|D_{x,y}^\alpha u\|_{L^2(T_h)}^2$ wegen (**) mit $\beta := \alpha$ und $|\alpha| = 2$, erhält man die Behauptung (8.4.3). ■

Lemma 8.4.3 \tilde{T} sei ein beliebiges Dreieck mit

$$\text{Seitenlänge} \leq h_{\max}, \quad \text{Innenwinkel} \geq \alpha_0 > 0. \quad (8.4.4)$$

Für jedes $u \in H^2(\tilde{T})$ gilt

$$\|D^\beta u\|_{L^2(\tilde{T})}^2 \leq C(\alpha_0) \left\{ h_{\max}^{2-2|\beta|} \sum_{\mathbf{x} \text{ Ecke von } \tilde{T}} |u(\mathbf{x})|^2 + h_{\max}^{4-2|\beta|} \sum_{|\alpha|=2} \|D^\alpha u\|_{L^2(\tilde{T})}^2 \right\} \quad \text{für alle } |\beta| \leq 2, \quad (8.4.5)$$

wobei $C(\alpha_0)$ nur von α_0 , nicht aber von u, β oder h abhängt.

Beweis. Seien $h \leq h_{\max}$ eine der Seitenlängen von \tilde{T} und T_h das Dreieck aus Lemma 8.4.2. Ähnlich wie in Übungsaufgabe 8.3.14a sei $\Phi : T_h \rightarrow \tilde{T}$ die lineare Abbildung, die T_h auf \tilde{T} abbildet. $v(\xi, \eta) := u(\Phi(\xi, \eta))$ gehört zu $H^2(T_h)$. Unter der Bedingung (8.4.4) ist die Determinante $|\det \Phi| \in [1/K(\alpha_0), K(\alpha_0)]$ nach oben und unten beschränkt. Aus $\|D_{x,y}^\beta u\|_{L^2(\tilde{T})}^2 \leq C_1(\alpha_0) \sum_{|\beta'|=|\beta|} \|D_{\xi,\eta}^{\beta'} v\|_{L^2(T_h)}^2$, (8.4.3) und $\sum_{|\alpha|=2} \|D_{\xi,\eta}^\alpha v\|_{L^2(T_h)}^2 \leq C_2(\alpha_0) \sum_{|\alpha|=2} \|D_{x,y}^\alpha u\|_{L^2(\tilde{T})}^2$ folgt (8.4.5). ■

Aus Lemma 8.4.3 ergibt sich Satz 8.4.4 als Hauptresultat. In (8.4.6) tritt $H^2(\Omega) \cap V$ auf. Für $V = H^1(\Omega)$ ist dies $H^2(\Omega)$, nur bei $V = H_0^1(\Omega)$ sind in $H^2(\Omega) \cap H_0^1(\Omega)$ die zusätzlichen Nullrandbedingungen zu beachten.

Satz 8.4.4 Voraussetzung (8.4.1a-c) gelte für \mathcal{T}, V_N und V . $\alpha_0 > 0$ sei der kleinste Innenwinkel aller $T \in \mathcal{T}$, während h die maximale Seitenlänge aller $T \in \mathcal{T}$ sei. Dann ist

$$\inf_{v \in V_N} \|u - v\|_{H^k(\Omega)} \leq C'(\alpha_0) h^{2-k} \|u\|_{H^2(\Omega)} \quad \text{für alle } k \in \{0, 1, 2\} \text{ und } u \in H^2(\Omega) \cap V. \quad (8.4.6)$$

Beweis. Zu $u \in H^2(\Omega)$ wählt man $v := \sum_i u(\mathbf{x}^i) b_i \in V_N$, d.h. $v \in V_N$ mit $v(\mathbf{x}^i) = u(\mathbf{x}^i)$ an den Knoten \mathbf{x}^i . Da $w := u - v$ an den Ecken jedes $T_i \in \mathcal{T}$ verschwindet (wegen Interpolationseigenschaft bzw. eventueller Nullrandbedingung) und $D^\alpha w = D^\alpha u$ für $|\alpha| = 2$ (wegen $D^\alpha v = 0$ für alle linearen(!) Funktionen $v \in V_N$), liefert (8.4.5) die Abschätzung

$$\begin{aligned} \|D^\beta w\|_{L^2(\Omega)}^2 &= \sum_{T_i \in \mathcal{T}} \|D^\beta w\|_{L^2(T_i)}^2 \stackrel{(8.4.5)}{\leq} \sum_{T_i \in \mathcal{T}} C(\alpha_0) h^{4-2|\beta|} \sum_{|\alpha|=2} \|D^\alpha u\|_{L^2(T_i)}^2 \\ &= C(\alpha_0) h^{4-2|\beta|} \sum_{|\alpha|=2} \|D^\alpha u\|_{L^2(\Omega)}^2 \leq C(\alpha_0) h^{4-2|\beta|} \|u\|_{H^2(\Omega)}^2. \end{aligned}$$

Summation über $|\beta| \leq k$ beweist die Ungleichung (8.4.6). ■

Die Interpolation $v(\mathbf{x}^i) = u(\mathbf{x}^i)$ aus dem vorhergehenden Beweis ist nur für stetige Funktionen sinnvoll. Da die Sobolev-Einbettung $H^2(\Omega) \subset C^0(\bar{\Omega})$ für die (physikalisch realistischen) Dimensionen $d \leq 3$ richtig ist (vgl. Satz 6.2.31), lässt sich die Aussage von Satz 8.4.4 auch noch auf stückweise lineare Funktionsansätze in Tetraedern erweitern. Will man dagegen die Ungleichung (8.4.6) für Raumdimensionen $d \geq 4$ erreichen oder will man auf der rechten Seite von (8.4.6) die $H^2(\Omega)$ -Norm durch eine schwächere ersetzen, muss man die Interpolationsabbildung durch den Clément-Operator ersetzen (vgl. Clément [23] und Braess [12, §II.6,9]).

8.4.2 Eigenschaften von Folgen von Finite-Element-Räumen

8.4.2.1 Uniforme, formreguläre Triangulationsfolgen und K -Gitter

Für die Diskussion der Konvergenz von Finite-Element-Lösungen benötigt man eine Folge von Finite-Element-Räumen. Im Fall der Differenzenverfahren ist die Schrittweite h ein eindeutiger Parameter, der gegen null strebt. Für Finite-Element-Räume ist die Dimension N dagegen nur eine partielle Information.

Sei $\{\mathcal{T}_\nu : \nu \in \mathbb{N}\}$ eine Folge von Triangulationen ("Triangulation" ist hier wörtlich gemeint: \mathcal{T}_ν enthält Dreiecke $\subset \Omega \subset \mathbb{R}^2$). Die zugehörigen Finite-Element-Räume seien V_{N_ν} mit der Dimension $N_\nu \rightarrow \infty$. Zu jedem $T \in \mathcal{T}_\nu$ sei

$$\rho_{T,\text{ausen}} := \inf\{\rho : T \subset K_\rho(\mathbf{x}), \rho > 0, \mathbf{x} \in \mathbb{R}^2\} \quad (8.4.7a)$$

der Radius des Umkreises zu T . Entsprechend sei

$$\rho_{T,\text{innen}} := \sup\{\rho : T \supset K_\rho(\mathbf{x}), \rho > 0, \mathbf{x} \in \mathbb{R}^2\} \quad (8.4.7b)$$

der Radius des Innenkreises. Ferner seien

$$h_T := \text{diam}(T) \quad \text{und} \quad \begin{cases} h_{\max,\nu} := h_{\max}(\mathcal{T}_\nu) := \max\{h_T : T \in \mathcal{T}_\nu\}, \\ h_{\min,\nu} := h_{\min}(\mathcal{T}_\nu) := \min\{h_T : T \in \mathcal{T}_\nu\}. \end{cases} \quad (8.4.7c)$$

Offenbar gilt $\rho_{T,\text{innen}} \leq h_T \leq \rho_{T,\text{ausen}}$ für alle $T \in \mathcal{T}_\nu$.

Für spätere Schlussfolgerungen benötigt man verschiedene Eigenschaften:

- Die Folge $\{\mathcal{T}_\nu\}_{\nu \in \mathbb{N}}$ heißt *uniform*, wenn $\sup\{h_{\max,\nu}/h_{\min,\nu} : \nu \in \mathbb{N}\} < \infty$.
- Die Folge $\{\mathcal{T}_\nu\}_{\nu \in \mathbb{N}}$ heißt *formregulär*, wenn $\sup\{\rho_{T,\text{ausen}}/\rho_{T,\text{innen}} : T \in \mathcal{T}_\nu, \nu \in \mathbb{N}\} < \infty$.
- Eine Triangulation \mathcal{T} hat die *K -Gittereigenschaft*, wenn die Ungleichung

$$h_{\max,T_1} \leq K h_{\max,T_2} \quad \text{für alle Paare von benachbarten Dreiecken } T_1, T_2 \in \mathcal{T}, T_1 \cap T_2 \neq \emptyset$$

gilt. Eine Folge $\{\mathcal{T}_\nu\}_{\nu \in \mathbb{N}}$ erfüllt die K -Gittereigenschaft, wenn alle \mathcal{T}_ν K -Gitter sind.

Im Falle der uniformen Folge $\{\mathcal{T}_\nu\}_{\nu \in \mathbb{N}}$ haben alle Dreiecke $T \in \mathcal{T}_\nu$ einen Durchmesser, der von h_ν nur um den festen Faktor $C := \sup\{h_{\max,\nu}/h_{\min,\nu} : \nu \in \mathbb{N}\}$ abweichen kann. Lokale Verfeinerungen wie in der rechten Abbildung 8.3.2 sind damit nur eingeschränkt möglich.

Dagegen braucht eine lokale Verfeinerung die Formregularität nicht zu verschlechtern. In der rechten Abbildung 8.3.2 gilt $\rho_{T,\text{ausen}}/\rho_{T,\text{innen}} \leq 2$ auch bei weiteren Verfeinerungen. In §8.6.3.3 wird der Zusammenhang zwischen Formregularität und K -Gittereigenschaft diskutiert werden.

Die Formregularität kann statt mit den Um- und Innenkreisen auch mit dem kleinsten Innenwinkel α_T eines Dreiecks T beschrieben werden: $\inf\{\alpha_T : T \in \mathcal{T}_\nu, \nu \in \mathbb{N}\} < \infty$. Die letzte Bedingung ist in (8.4.4) verwendet worden.

8.4.2.2 Maximalwinkelbedingung

Die Bedingung " $\alpha_T \geq \alpha_0 > 0$ " aus (8.4.4) ist stärker als notwendig. Da die Summe aller Dreieckswinkel π beträgt, folgt aus (8.4.4), dass

$$\alpha_{T,\max} := \text{maximaler Innenwinkel von } T \leq \bar{\alpha} < \pi. \quad (8.4.8)$$

Der Umkehrschluss gilt nicht: Die Winkel $\alpha_1 = \pi/2$, $\alpha_2 = \pi/2 - \varepsilon$, $\alpha_3 = \varepsilon$ eines flachen, rechtwinkligen Dreiecks erfüllen (8.4.8) mit $\bar{\alpha} = \pi/2 < \pi$, trotzdem kann α_3 beliebig klein werden. Eine genauere Abschätzung der Transformationen in Lemma 8.4.3 zeigt, dass schon die schwächere Maximalwinkelbedingung (8.4.8) ausreicht, eine Abschätzung (8.4.6) zu beweisen, in der $C'(\alpha_0)$ durch eine Funktion $C(\bar{\alpha})$ ersetzt werden kann, die nur von $\bar{\alpha}$ abhängt (vgl. [52, §3.4.2]).

8.4.3 H^1 -Abschätzungen für lineare Elemente

Sei h_T die längste Seite von $T \in \mathcal{T}$ gemäß (8.4.7c). $h := \max\{h_T : T \in \mathcal{T}\}$ ist die größte in der Triangulation \mathcal{T} auftretende Kantenlänge. h wird als Parameter zur Indizierung verwendet: $\mathcal{T} = \mathcal{T}_h$.

Bei der Konstruktion der Triangulationen ist darauf zu achten, dass mit zunehmender Verfeinerung ($h_\nu \rightarrow 0$) die Formregularität erhalten bleibt. Strategien zur systematischen Konstruktion von formregulären Folgen V_{h_ν} findet man in §8.6.3.3.

Der Fall $k = 1$ aus Satz 8.4.4 lässt sich nun wie folgt schreiben:

Satz 8.4.5 *Voraussetzung (8.4.1a-c) gelte für eine Folge formregulärer Triangulationen \mathcal{T}_ν . Dann existiert eine Konstante C , sodass für alle $h = h_\nu$ und $V_h = V_{h_\nu}$ die folgende Abschätzung gilt:*

$$\inf_{v \in V_h} |u - v|_1 \leq Ch |u|_2 \quad \text{für alle } u \in H^2(\Omega) \cap V. \quad (8.4.6')$$

Die Kombination dieses Satzes mit Satz 8.2.1 ergibt

Satz 8.4.6 *Voraussetzung (8.4.1a-c) gelte für eine Folge formregulärer Triangulationen \mathcal{T}_ν . Die Bilinearform erfülle (8.1.2) und (8.1.17). Die Konstanten $\varepsilon_N =: \varepsilon_{h_\nu} > 0$ aus (8.1.17) seien nach unten durch $\varepsilon_{h_\nu} \geq \tilde{\varepsilon} > 0$ beschränkt (vgl. Zusatz 8.2.3). Aufgabe (8.1.1) habe eine Lösung $u \in H^2(\Omega) \cap V$. $u^h \in V_h$ sei die Finite-Element-Lösung. Dann gibt es eine von u , $h = h_\nu$ und v unabhängige Konstante C , sodass*

$$|u^h - u|_1 \leq Ch |u|_2. \quad (8.4.9)$$

Die Kombination der Sätze 8.4.5 und 8.2.2 lautet:

Satz 8.4.7 *Voraussetzung (8.4.1a-c) gelte für eine Folge formregulärer Triangulationen \mathcal{T}_ν mit $h_\nu \rightarrow 0$. Die Bilinearform genüge (8.1.2) und (8.1.17) mit $\varepsilon_{h_\nu} \geq \tilde{\varepsilon} > 0$. Dann hat die Aufgabe (8.1.1) eine eindeutige Lösung $u \in V$, und die Finite-Element-Lösung $u^{h_\nu} \in V_{h_\nu}$ konvergiert gegen u :*

$$|u^{h_\nu} - u|_1 \rightarrow 0 \quad \text{für } \nu \rightarrow \infty. \quad (8.4.10)$$

Beweis. $u \in V$ und $\varepsilon > 0$ seien beliebig. Da $H^2(\Omega) \cap V$ dicht in V , existiert $u_\varepsilon \in H^2(\Omega) \cap V$ mit $|u - u_\varepsilon|_1 \leq \varepsilon/2$. Wegen (8.4.6) und $h_\nu \rightarrow 0$ gibt es einen Index ν und eine Funktion $v_\varepsilon \in V_{h_\nu}$ mit $|u_\varepsilon - v_\varepsilon|_1 \leq \varepsilon/2$, also $|u - v_\varepsilon|_1 \leq \varepsilon$. Dies beweist (8.2.4a). ■

Satz 8.4.7 beweist Konvergenz ohne einschränkende Voraussetzungen an u . Die Konvergenzordnung $\mathcal{O}(h)$ in Abschätzung (8.4.9) benötigt dagegen die Annahme, dass die Lösung $u \in V$ in $H^2(\Omega)$ liegt. Wie sich herausstellen wird, ist diese Annahme keineswegs in jedem Fall erfüllt. Eine schwächere Voraussetzung ist $u \in V \cap H^s(\Omega)$ mit $s \in (1, 2)$. Das entsprechende Resultat lautet wie folgt.

Satz 8.4.8 *In den Voraussetzungen des Satzes 8.4.6 sei $u \in V \cap H^2(\Omega)$ durch $u \in V \cap H^s(\Omega)$ mit einem $s \in [1, 2]$ ersetzt. Ω sei hinreichend glatt. Dann gilt*

$$|u^h - u|_1 \leq Ch^{s-1} |u|_s. \quad (8.4.11)$$

Der Beweis benutzt eine Verallgemeinerung des Satzes 8.4.4 mit $k = 1$:

Lemma 8.4.9 *Unter den Voraussetzungen des Satzes 8.4.4 und geeigneten Voraussetzungen an Ω gilt*

$$\inf_{v \in V_h} |u - v|_1 \leq C''(\alpha_0) h^{s-1} \|u\|_{H^s(\Omega)} \quad \text{für alle } s \in [1, 2] \text{ und } u \in H^s(\Omega) \cap V. \quad (8.4.12)$$

Der Beweis beruht auf einem hier nicht näher erläuterten Interpolationsargument. (8.4.12) gilt für $s = 2$ (vgl. (8.4.9)) und $s = 1$, da $\inf |u - v|_1 \leq |u - 0|_1 = |u|_1$. Hieraus folgt (8.4.12) für $s \in (1, 2)$ mit der Norm $|\cdot|_s$ des Interpolationsraumes $[H^1(\Omega) \cap V, H^2(\Omega) \cap V]_{s-1}$ (vgl. Lions-Magenes [58]), der unter geeigneten Voraussetzungen an Ω mit $H^s(\Omega) \cap V$ übereinstimmt.

Für $s = 1$ wird die rechte Seite in (8.4.12) zu $\text{const} \cdot |u|_1$. In der Tat ist die Abschätzung durch $\inf\{|u - v|_1 : v \in V_h\} \leq |u|_1$ die bestmögliche. Zum Beweis wähle man $u \perp V_h$ (Orthogonalität bezüglich $|\cdot|_1$). Auf der anderen Seite ist $s = 2$ der maximale Wert, für den die Abschätzungen (8.4.9) und (8.4.12) gültig sind. Auch $u \in C^\infty(\Omega)$ erlaubt keine bessere Approximationsordnung als $\mathcal{O}(h)$!

Die in (8.2.5) eingeführte Ritz-Projektion $S_N : V \rightarrow V_N$ sei jetzt als $S_h : V \rightarrow V_h$ geschrieben. Ungleichung (8.4.11) wird zu $|u - S_h u|_1 / |u|_s \leq Ch^{s-1}$ und beweist die

Folgerung 8.4.10 *Voraussetzung (8.4.1a-c) gelte für \mathcal{T} , V_h und V . Die Bilinearform erfülle (8.1.2) und (8.1.17). Dann genügt die Ritz-Projektion S_h der Abschätzung*

$$\|I - S_h\|_{H^1(\Omega) \leftarrow H^2(\Omega) \cap V} \leq Ch^1. \quad (8.4.13)$$

Unter der Voraussetzung von Lemma 8.4.9 gilt

$$\|I - S_h\|_{H^1(\Omega) \leftarrow H^s(\Omega) \cap V} \leq Ch^{s-1} \quad \text{für alle } s \in [1, 2]. \quad (8.4.13')$$

8.4.4 L^2 -Abschätzungen für lineare Elemente

Nach Satz 8.4.6 ist $\mathcal{O}(h)$ die optimale Konvergenzordnung. Dies Ergebnis scheint im Widerspruch zur $\mathcal{O}(h^2)$ -Konvergenz der Fünfpunktformel zu stehen (vgl. Abschnitt 4.5), denn das Finite-Element-Verfahren mit spezieller Triangulation ist fast identisch mit der Fünfpunktformel (vgl. Übungsaufgabe 8.3.13). Der Grund hierfür ist, dass der Finite-Element-Fehler $u^h - u$ in der $|\cdot|_1$ -Norm gemessen wird. Die Abschätzung $\inf\{|u - v|_0 : v \in V_h\} \leq Ch^2 |u|_2$ aus Satz 8.4.4 legt die Vermutung nahe, dass die $|\cdot|_0$ -Norm des Fehlers von der Ordnung $\mathcal{O}(h^2)$ ist: $|u - u_h|_0 \leq Ch^2 |u|_2$. Diese Aussage ist jedoch ohne weitere Voraussetzungen falsch.

Bisher wurde nur die Existenz einer schwachen Lösung $u \in V$ (zum Beispiel $V = H_0^1(\Omega)$ oder $V = H^1(\Omega)$) garantiert. In Satz 8.4.6 benötigen wir jedoch die stärkere Annahme $u \in H^2(\Omega) \cap V$. Eine analoge Regularitätsbedingung stellen wir auch an das zu (8.1.1) *adjungierte Problem*

$$\text{suche } u \in V \quad \text{mit} \quad a^*(u, v) = f(v) \quad \text{für alle } v \in V, \quad (8.4.14)$$

das die adjungierte Bilinearform $a^*(u, v) := a(v, u)$ benutzt. Für $f \in L^2(\Omega) \subset V'$ wird $f(v)$ zu $(f, v)_{L^2(\Omega)}$. Die benötigte *Regularitätsannahme* ist die H^2 -Regularität für das adjungierte Problem:

$$\text{Für alle } f \in L^2(\Omega) \text{ habe das Problem (8.4.14) eine Lösung } u \in H^2(\Omega) \cap V \text{ mit } |u|_2 \leq C_R |f|_0. \quad (8.4.15)$$

In Kapitel 9 werden wir sehen, dass diese Aussage für hinreichend glatte Gebiete Ω zutrifft.

Satz 8.4.11 (Aubin-Nitsche) *Es gelte (8.4.15), (8.1.2), (8.1.17) mit $\varepsilon_N = \varepsilon_h \geq \tilde{\varepsilon} > 0$ und*

$$\inf_{v \in V_h} |u - v|_1 \leq C_0 h |u|_2 \quad \text{für alle } u \in H^2(\Omega) \cap V. \quad (8.4.16)$$

Die Aufgabe (8.1.1) habe eine Lösung $u \in V$. $u^h \in V_h \subset V$ sei die Finite-Element-Lösung. Dann gilt

$$|u^h - u|_0 \leq C_1 h |u|_1 \quad (8.4.17a)$$

mit einer von u und h unabhängigen Konstante C_1 . Falls die Lösung u auch zu $H^2(\Omega) \cap V$ gehört, gibt es eine von u und h unabhängige Konstante C_2 , sodass

$$|u^h - u|_0 \leq C_2 h^2 |u|_2. \quad (8.4.17b)$$

[Hinreichend für Bedingung (8.4.16) ist nach Satz 8.4.4, dass V_h der Finite-Element-Raum zu einer zulässigen und formregulären Triangulation ist.]

Beweis (vgl. Nitsche [66]). Zu $e := u^h - u \in L^2(\Omega)$ definiere man $w \in H^2(\Omega) \cap V$ als Lösung von (8.4.14) für $f := e$:

$$a(v, w) = (e, v)_{L^2(\Omega)} \quad \text{für alle } v \in V. \quad (8.4.18a)$$

Zu w gibt es auf Grund von (8.4.16) ein $w^h \in V_h$ mit

$$|w^h - w|_1 \leq C_0 h |w|_2 \stackrel{(8.4.15)}{\leq} C_0 C_R h |e|_0. \quad (8.4.18b)$$

Gleichung (8.2.3) lautet $a(e, v) = 0$ für alle $v \in V_h$, also gilt insbesondere

$$a(e, w^h) = 0. \quad (8.4.18c)$$

Die bisherigen Resultate ergeben

$$|e|_0^2 = (e, e)_{L^2(\Omega)} \stackrel{(8.4.18a)}{=} a(e, w) \stackrel{(8.4.18c)}{=} a(e, w - w^h) \stackrel{(8.1.2)}{\leq} C_S |e|_1 |w^h - w|_1 \stackrel{(8.4.18b)}{\leq} C_S |e|_1 C_0 C_R h |e|_0,$$

also

$$|e|_0 \leq C_S C_0 C_R h |e|_1. \quad (8.4.18d)$$

Aus (8.2.1) leitet man

$$|e|_1 = \|u^h - u\|_V \leq \left(1 + \frac{C_S}{\varepsilon_N}\right) \inf_{v \in V_N} \|u - v\|_V \leq \left(1 + \frac{C_S}{\tilde{\varepsilon}}\right) \|u\|_V$$

ab, sodass (8.4.17a) mit $C_1 := C_S C_0 C_R (1 + C_S/\tilde{\varepsilon})$ folgt. Falls $u \in H^2(\Omega) \cap V$, kann man die Ungleichung (8.4.9), das ist $|u^h - u|_1 \leq Ch |u|_2$, benutzen und erhält (8.4.17b) mit $C_2 := C_S C_0 C_R C$. ■

Zusatz 8.4.12 Die Ungleichungen (8.4.17a) und (8.4.17b) sind gleichbedeutend mit den Eigenschaften (8.4.19) der Ritz-Projektion S_h :

$$\|I - S_h\|_{L^2(\Omega) \leftarrow V} \leq C_1 h, \quad \|I - S_h\|_{L^2(\Omega) \leftarrow H^2(\Omega) \cap V} \leq C_2 h^2. \quad (8.4.19)$$

Die Abschätzungen (8.4.19) lassen sich auch direkt beweisen. Die Definition von \hat{S}_h und der Zusammenhang zwischen S_h und \hat{S}_h finden sich in

Übungsaufgabe 8.4.13 Zu $a(\cdot, \cdot)$ gehöre der Operator $L : V \rightarrow V'$, die Ritz-Projektion S_h und die Systemmatrix \mathbf{L} . Man zeige: a) Zur adjungierten Bilinearform $a^*(\cdot, \cdot)$ gehören die Systemmatrix \mathbf{L}^\top und die Ritz-Projektion

$$\hat{S}_h = P(\mathbf{L}^\top)^{-1} P^* L^*. \quad (8.4.20a)$$

b) Es gilt der Zusammenhang

$$S_h = L^{-1} \hat{S}_h L \quad (8.4.20b)$$

Zweiter Beweis zu Satz 8.4.11. Sei

$$H_\star^2(\Omega) := \{u \in H^2(\Omega) : u \text{ ist Lösung von (8.4.14) zu einem } f \in L^2(\Omega)\} \subset V$$

das Bild von $L^{\star-1} \in L(L^2(\Omega), H^2(\Omega))$. $H_\star^2(\Omega)$ wird mit der Norm $|\cdot|_2$ versehen. $L^{\star-1} \in L(L^2(\Omega), H_\star^2(\Omega))$ ist wegen $L^2(\Omega)' = L^2(\Omega)$ mit $L^{-1} \in L(H_\star^2(\Omega)', L^2(\Omega))$ äquivalent (vgl. Lemma 6.3.2), sodass es ein C_α mit

$$\|L^{-1}\|_{L^2(\Omega) \leftarrow H_\star^2(\Omega)'} \leq C_\alpha \quad (8.4.21a)$$

gibt. Die Voraussetzungen (8.1.2) und (8.1.17) aus Satz 8.4.11 übertragen sich ohne Änderung der Konstanten auf das adjungierte Problem, sodass sich die Aussage des Satzes 8.2.1 in der Form $\|I - \hat{S}_h\|_{V \leftarrow H_\star^2(\Omega)} \leq C_\beta h$ schreiben lässt. Für den adjungierten Operator ergibt sich

$$\|(I - \hat{S}_h)^*\|_{H_\star^2(\Omega)' \leftarrow V'} \leq C_\beta h. \quad (8.4.21b)$$

Die Grundvoraussetzung (8.1.2) schreibt sich als

$$\|L\|_{V' \leftarrow V} \leq C_S. \quad (8.4.21c)$$

Der Gleichung (8.4.20b) entnimmt man die Darstellung

$$I - S_h = I - L^{-1} \hat{S}_h L = L^{-1} (I - \hat{S}_h^*) L = L^{-1} (I - \hat{S}_h)^* L.$$

Auf Grund von (8.4.21a-c) erhält man hieraus

$$\|I - S_h\|_{L^2(\Omega) \leftarrow V} \leq \|L^{-1}\|_{L^2(\Omega) \leftarrow H_\star^2(\Omega)'} \|(I - \hat{S}_h)^*\|_{H_\star^2(\Omega)' \leftarrow V'} \|L\|_{V' \leftarrow V} \leq C_\alpha C_\beta h C_S. \quad (8.4.21d)$$

Damit ist die erste Ungleichung in (8.4.19) mit $C_1 = C_\alpha C_\beta C_S$ bewiesen. Mit S_h ist auch $I - S_h$ Projektion, sodass

$$\|I - S_h\|_{L^2(\Omega) \leftarrow H^2(\Omega) \cap V} = \|(I - S_h)^2\|_{L^2(\Omega) \leftarrow H^2(\Omega) \cap V} \leq \|I - S_h\|_{L^2(\Omega) \leftarrow V} \|I - S_h\|_{V \leftarrow H^2(\Omega) \cap V}.$$

Die erste Ungleichung in (8.4.19) und (8.4.13) ergeben die zweite Ungleichung in (8.4.19). ■

Die Regularitätsbedingung (8.4.15) wird im folgenden Satz abgeschwächt bzw. verallgemeinert.

Satz 8.4.14 Es gelte (8.1.2), (8.1.17) mit $\varepsilon_N = \varepsilon_h \geq \tilde{\varepsilon} > 0$ sowie Ungleichung (8.4.12) für alle $1 \leq s \leq 2$. Anstelle von (8.4.15) setzen wir für ein $t \in [0, 1]$ die H^{2-t} -Regularität voraus: Falls $V = H^1(\Omega)$, gelte $L^{\star-1} \in L(H^t(\Omega)', H^{2-t}(\Omega))$; falls $V = H_0^1(\Omega)$, gelte $L^{\star-1} \in L(H^{-t}(\Omega), H^{2-t}(\Omega) \cap V)$. Dann genügt die Finite-Element-Lösung der Ungleichung

$$|u^h - u|_t \leq C_{t,s} h^{s-t} |u|_s \quad (0 \leq t \leq 1 \leq s \leq 2). \quad (8.4.22)$$

Die Ritz-Projektion erfüllt

$$\|I - S_h\|_{H^t(\Omega) \leftarrow H^s(\Omega) \cap V} \leq C_{t,s} h^{s-t} \quad (0 \leq t \leq 1 \leq s \leq 2). \quad (8.4.23)$$

Beweis. Zur Vereinfachung wird nur der Fall $V = H^1(\Omega)$ behandelt. In (8.4.21a-c) ersetze man $L^2(\Omega)$ durch $H^t(\Omega)$, $H_\star^2(\Omega)'$ durch $H_\star^{2-t}(\Omega)'$ mit $H_\star^{2-t}(\Omega) := \{u \in H^{2-t}(\Omega) : u \text{ Lösung zu } f \in H^t(\Omega)'\}$. (8.4.21b) wird wegen (8.4.13') [mit s ersetzt durch $2-t$] zu

$$\|(I - \hat{S}_h)^*\|_{H_\star^{2-t}(\Omega)' \leftarrow V'} = \|I - S_h\|_{V \leftarrow H_\star^{2-t}(\Omega)} \leq C_\beta h^{1-t}.$$

Wie in (8.4.21d) zeigt man $\|I - S_h\|_{H^t(\Omega) \leftarrow V} \leq Ch^{1-t}$. Zusammen mit $\|I - S_h\|_{V \leftarrow H^s(\Omega) \cap V} \leq Ch^{s-1}$ (vgl. (8.4.13')) folgt (8.4.23) und hieraus (8.4.22). ■

Im Zusammenhang mit nichtlinearen Randwertproblemen werden Fehlerabschätzungen bezüglich weiterer Normen (z.B. $L^\infty(\Omega)$, $L^p(\Omega)$) interessant. Hierzu sei auf Ciarlet [22, §3.3] und Schatz [73] verwiesen.

8.5 Verallgemeinerungen

8.5.1 Fehlerabschätzungen für andere Elemente

Die in Abschnitt 8.4 für lineare Funktionen auf Dreieckselementen bewiesenen Fehlerabschätzungen gelten ebenso für die bilinearen Funktionen auf Parallelogrammen und die Kombination beider Elemente (vgl. Abbildung 8.3.7). Die Beweise verlaufen analog. Auch für die Tetraederelemente eines dreidimensionalen Gebietes $\Omega \subset \mathbb{R}^3$ lassen sich diese Resultate übertragen. Beim Beweis beachte man, dass $u \in H^2(\Omega)$ auch für $\Omega \subset \mathbb{R}^3$ noch wohldefinierte Knotenwerte $u(\mathbf{x}^i)$ besitzt, da $2 > n/2$ die Sobolev-Einbettung sichert ($n = 3$: Dimension von Ω ; vgl. Satz 6.2.31).

Bei quadratischen Elementen (vgl. Abschnitt 8.3.4) erwartet man eine entsprechend gesteigerte Konvergenzordnung. Allgemein lässt sich nachweisen: Ist die Ansatzfunktion in jedem $T \in \mathcal{T}$ ein Polynom vom Grad $k \geq 1$ (d.h. $u(\mathbf{x}) = \sum_{|\nu| \leq k} \alpha_\nu \mathbf{x}^\nu$ auf T), so gilt

$$d(u, V_h) \equiv \inf \{|u - v|_1 : v \in V_h\} \leq Ch^k |u|_{k+1} \quad \text{für alle } u \in H^{k+1}(\Omega) \cap V \quad (8.5.1)$$

(vgl. Ciarlet [22, Theorem 3.2.1]). Legt man Parallelogramme zugrunde und verwendet Ansatzfunktionen, die mindestens Polynome vom Grad k enthalten, gilt (8.5.1) ebenfalls. Zum Beispiel erfüllen biquadratische Elemente oder quadratische Ansätze der Serendipity-Klasse diese Bedingung für $k = 2$. Das folgende dem Satz 8.4.6 entsprechende Resultat folgt auf Grund von Satz 8.2.1:

Satz 8.5.1 V_h erfülle (8.5.1). Die Bilinearform erfülle (8.1.2) und (8.1.17) mit $\varepsilon_N =: \varepsilon_h \geq \tilde{\varepsilon} > 0$. Aufgabe (8.1.1) habe eine Lösung $u \in V \cap H^{k+1}(\Omega)$. Dann genügt die Finite-Element-Lösung $u^h \in V_h$ der Abschätzung

$$|u^h - u|_1 \leq C' h^k |u|_{k+1}. \quad (8.5.2)$$

Die Ritz-Projektion erfüllt $\|I - S_h\|_{V \leftarrow H^{k+1}(\Omega) \cap V} \leq Ch^k$.

Die Abschätzung (8.4.11) gilt jetzt für $s \in [1, k]$, wenn $u \in H^s(\Omega) \cap V$. Unter geeigneten Regularitätsbedingungen ergeben sich wie in Satz 8.4.11 die Fehlerschranken

$$|u^h - u|_0 \leq Ch^{k+1} |u|_{k+1}, \quad (8.5.3a)$$

$$\|u^h - u\|_{H^{k+1}(\Omega)'} \leq Ch^{2k} |u|_{k+1}. \quad (8.5.3b)$$

Für (8.5.3b) benötigt man beispielsweise die H^{k+1} -Regularität: Für jedes $f \in H^{k-1}(\Omega)$ habe das adjungierte Problem (8.4.14) eine Lösung $u \in H^{k+1}(\Omega)$.

Bemerkung 8.5.2 Unter geeigneten Voraussetzungen gilt Ungleichung (8.4.22) für $1 - k \leq t \leq 1 \leq s \leq k + 1$. Für negative t ist darin $|\cdot|_t$ als Dualnorm von $H^{-t}(\Omega)$ zu interpretieren.

8.5.2 Finite Elemente für Gleichungen höherer Ordnung

8.5.2.1 Einführung: Eindimensionale biharmonische Gleichung

Alle bisher konstruierten Finite-Element-Räume V_h sind für Gleichungen der Ordnung $2m > 2$ unbrauchbar, da dann $V_h \not\subset H^m(\Omega)$. Gemäß Beispiel 6.2.5 ist es für $V_h \subset H^2(\Omega)$ notwendig, dass sich nicht nur die

Funktion u , sondern auch deren Ableitungen u_{x_i} ($1 \leq i \leq n$) zwischen den Elementen stetig ändern. Die Ansatzfunktionen müssen daher stückweise glatt und insgesamt aus $C^1(\bar{\Omega})$ sein.

Als Modellproblem stellen wir die eindimensionale biharmonische Gleichung

$$u''''(x) = g(x) \quad \text{für } 0 < x < 1, \quad u(0) = u'(0) = u(1) = u'(1) = 0$$

vor, deren schwache Formulierung wie folgt lautet:

$$a(u, v) = f(v) \quad \text{für alle } v \in H_0^2(0, 1), \quad \text{wobei} \quad a(u, v) := \int_0^1 u'' v'' dx, \quad f(v) := \int_0^1 g v dx. \quad (8.5.4)$$

Das Intervall $\Omega = (0, 1)$ sei äquidistant in Teilintervalle der Länge h zerlegt. Die bisher verwendeten stückweise linearen Funktionen (siehe Abbildung 8.3.1) können als lineare Splinefunktion angesehen werden, so dass es nun nahe liegt, V_h als Raum der *kubischen Splines* (mit $u = u' = 0$ bei $x = 0$ und $x = 1$) anzusetzen (vgl. Stoer [77, §2.4]). Als Basisfunktionen können die *B-Splines* dienen, deren Träger im Allgemeinen aus vier Teilintervallen bestehen (vgl. Abbildung 8.5.1a). Da kubische Splinefunktionen zu $C^2(0, 1)$ gehören, liegen sie nicht nur in $H_0^2(0, 1)$, sondern sogar in $H^3(0, 1) \cap H_0^2(0, 1)$. Einfacher als mit Splinefunktionen arbeitet man mit der *kubischen Hermite-Interpolation*:

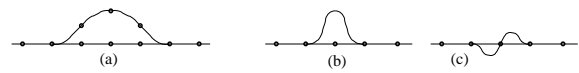


Abb. 8.5.1. a) B-Spline, b,c) Hermite-Basisfunktionen

Zu jedem inneren Knoten $x_j = jh$ gibt es zwei Basisfunktionen b_{1i} und b_{2i} mit $b_{1i}(x_i) = 1$, $b'_{1i}(x_i) = 0$, $b_{2i}(x_i) = 0$, $b'_{2i}(x_i) = 1$, $b_{ki}(x_j) = b'_{ki}(x_j) = 0$ für $k = 1, 2$, $j \neq i$ (vgl. Abbildung 8.5.1b,c). Der Träger besteht aus nur zwei Teilintervallen. Ihre Darstellung lautet

$$\begin{aligned} b_{1i}(x) &= (h - |x - x_i|)^2 (h + 2|x - x_i|) / h^3 & \text{für } x_{i-1} \leq x \leq x_{i+1}, \\ b_{2i}(x) &= (h - |x - x_i|)^2 (x - x_i) / h^2 & \text{für } x_{i-1} \leq x \leq x_{i+1}, \\ b_{1i}(x) &= b_{2i}(x) = 0 & \text{für } x \notin [x_{i-1}, x_{i+1}]. \end{aligned}$$

Übungsaufgabe 8.5.3 u_{1i} und u_{2i} seien die Koeffizienten der Darstellung $u^h = \sum_{i=1}^{1/h-1} [u_{1i} b_{1i} + u_{2i} b_{2i}] \in V_h$. Man zeige: Die Koeffizienten der Finite-Element-Lösung von Aufgabe (8.5.4) sind durch die Gleichungen

$$\begin{aligned} h^{-3} [-12u_{1,i-1} + 24u_{1,i} - 12u_{1,i+1}] &+ h^{-2} [-6u_{2,i-1} + 6u_{2,i+1}] &= f_{1,i}, \\ h^{-2} [6u_{1,i-1} - 6u_{1,i+1}] &+ h^{-1} [2u_{2,i-1} + 8u_{2,i} + 2u_{2,i+1}] &= f_{2,i} \end{aligned} \quad (8.5.6)$$

gegeben, wobei $f_{1i} := \int_{x_{i-1}}^{x_{i+1}} b_{1i} g dx$, $f_{2i} := \int_{x_{i-1}}^{x_{i+1}} b_{2i} g dx$, $x_i = ih$.

Das System (8.5.6) unterscheidet sich völlig von den Differenzengleichungen (vgl. Abschnitt 5.3.3), denn in Gleichung (8.5.6) erscheinen Knotenwerte u_{1i} der Funktion gemeinsam mit Knotenwerten u_{2i} der Ableitung.

8.5.2.2 Der zweidimensionale Fall

Der Ansatz (8.5.5) lässt sich auf $\Omega \subset \mathbb{R}^2$ übertragen, wenn man eine Zerlegung in Rechteckselemente zugrunde legt (wie z.B. links in Abbildung 8.3.5 auf Seite 128). Die Ansatzfunktion ist *bikubisch*:

$$V_h := \left\{ u \in C^1(\bar{\Omega}) : u = \frac{\partial u}{\partial n} = 0 \text{ auf } \Gamma, u = \sum_{\nu, \mu=0}^3 \alpha_{\nu\mu} x^\nu y^\mu \text{ auf jedem Teilrechteck der Zerlegung} \right\}.$$

Zu jedem inneren Knotenpunkt (Ecken $\mathbf{x} \in \Omega$ der Rechteckselemente) kann man die vier Werte u , u_x , u_y , u_{xy} vorschreiben. Entsprechend gehören zu jedem Knoten vier Unbekannte und vier Basisfunktionen $b_{1i}(x, y), \dots, b_{4i}(x, y)$. Letztere ergeben sich als die Produkte $b_{ji}(x) b_{ki}(y)$ ($j, k = 1, 2$) der in §8.5.2.1 beschriebenen, eindimensionalen Basisfunktionen (vgl. Meis-Marcowitz [62, p. 248], Schwarz [75, p. 126]).

⁹Charles Hermite, geb. am 24. Dez. 1822 in Dieuze, Lorraine, gest. am 14. Januar 1901 in Paris

Geht man von einer Triangulation \mathcal{T} aus, so führt ein *Ansatz fünften Grades* zum Ziel: $u(\mathbf{x}) = \sum_{|\nu| \leq 5} \alpha_\nu \mathbf{x}^\nu$ auf allen $T \in \mathcal{T}$. Die Anzahl der Freiheitsgrade ist 21 (Anzahl der $\nu \in \mathbb{N}_0^2$ mit $|\nu| \leq 5$). Als Knoten wählt man die Punkte $\mathbf{x}^1, \dots, \mathbf{x}^6$ aus Abbildung 8.3.8a (auf Seite 130). In den Ecken $\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3$ schreibt man pro Punkt die 6 Werte $\{D^\mu u(\mathbf{x}^j) : |\mu| \leq 2\}$ vor. Die 3 verbleibenden Freiheitsgrade ergeben sich aus der Vorgabe der Normalenableitung $\partial u(\mathbf{x}^j)/\partial n$ in den Seitenmitten $\mathbf{x}^4, \mathbf{x}^5, \mathbf{x}^6$. Man überlegt sich: Wenn zwei benachbarte Dreiecke (siehe T und \hat{T} in Abbildung 8.3.8a) gemeinsame Eckwerte $\{D^\mu u(\mathbf{x}^j) : |\mu| \leq 2, j = 1, 2\}$ und übereinstimmende Normalenableitungen in \mathbf{x}^4 besitzen, so sind u und ∇u an der gemeinsamen Seite stetig, d.h. $V_h \subset H^2(\Omega)$. Je nachdem ob $V = H_0^2(\Omega)$, $V = H^2(\Omega) \cap H_0^1(\Omega)$ oder $V = H^2(\Omega)$, hat man u bzw. deren erste Ableitungen an den Randknoten \mathbf{x}^i auf null zu setzen.

Bemerkung 8.5.4 Die hier beschriebenen Finite-Element-Räume $V_h \subset H^2(\Omega)$ lassen sich selbstverständlich auch für Differentialgleichungen der Ordnung $2m = 2$ verwenden.

8.5.2.3 Fehlerabschätzung

Anstelle von (8.5.1) erhält man

$$\inf \{|u - v|_m : v \in V_h\} \leq Ch^{k+1-m} |u|_{k+1} \quad \text{für alle } u \in H^{k+1}(\Omega) \cap V, \quad (8.5.7)$$

wobei $k \geq m$ vom Grad des Polynomansatzes abhängt (z.B. $k = 3$ für kubische Splines, kubische bzw. bikubische Hermite-Interpolation). Dabei ist $m = 2$ im Falle der biharmonischen Gleichung. Wie in Satz 8.5.1 folgt aus (8.5.7) die Fehlerabschätzung

$$|u^h - u|_m \leq Ch^{k+1-m} |u|_{k+1} \quad (8.5.8a)$$

für die Finite-Element-Lösung $u^h \in V_h$. Unter geeigneten Regularitätsvoraussetzungen erhält man

$$|u^h - u|_t \leq C_{t,s} h^{s-t} |u|_s \quad (2m - k - 1 \leq t \leq m \leq s \leq k + 1) \quad (8.5.8b)$$

(vgl. Bemerkung 8.5.2). Die maximale Konvergenzordnung $2(k-m)+2$ ergibt sich für $s = k+1, t = 2m-k-1$ und erfordert $u \in H^{k+1}(\Omega) \cap V$. Außerdem muss jede Lösung des adjungierten Problems (8.4.14) mit $f \in H^{k+1-2m}(\Omega)$ zu $H^{k+1}(\Omega)$ gehören.

8.5.3 Finite Elemente für Nichtpolygon-Gebiete

Da die Vereinigung von Dreiecken oder Parallelogrammen nur Polygonegebiete erzeugt, war in (8.3.6) die Polygongestalt vorausgesetzt worden. Die Finite-Element-Methode ist aber keineswegs auf diese Gebiete beschränkt. Im Gegenteil, finite Elemente lassen sich krummlinigen Rändern leicht anpassen. Hierzu diskutieren wir zwei Zugänge. Der erste versucht, Ω mit üblichen finiten Elementen auszuschöpfen. Um die gewohnten Fehlerabschätzungen zu erhalten, muss man allerdings im Allgemeinen die isoparametrischen Elemente des zweiten Zuganges verwenden.

Seien Ω beliebig und $V = H^1(\Omega)$. Die Triangulation \mathcal{T} kann so gewählt werden, dass (8.3.7a,b,d) gelten und die "äußeren" Dreiecke wie in Abbildung 8.5.2 zwei Ecken auf $\Gamma = \partial\Omega$ haben. Im konvexen Fall von Abbildung 8.5.2a erweitert man die auf T definierte lineare Funktion auf $\tilde{T} := \bar{T} \cup B$. Für den Fall, dass der Rand konkav ist (vgl. Abbildung 8.5.2b), ist T durch $\tilde{T} := T \setminus B$ zu ersetzen. Entsprechend geht man in der Situation von Abbildung 8.5.2c vor. Die Knotenpunkte und die Darstellung der Basisfunktionen werden durch diese Modifikation nicht berührt. $\bar{\Omega}$ ist die Vereinigung der Abschlüsse aller inneren Dreiecke $T_i \in \mathcal{T}$ und aller modifizierten, äußeren Dreiecke \tilde{T} . Alle Eigenschaften und Resultate aus den Abschnitten 8.3-8.4 übertragen sich auf die neue Situation. Die einzige Schwierigkeit ist von praktischer Art: Zur Berechnung von \mathbf{L} und \mathbf{f} hat man Integrale über die Bogendreiecke \tilde{T} auszuwerten.

Sei nun $V = H_0^1(\Omega)$. Die vorherige Konstruktion führt nicht zum Ziel, da die Fortsetzung der linearen Funktion auf \tilde{T} nicht auf dem Randstück $\Gamma \cap \partial\tilde{T}$ von \tilde{T} verschwindet. Somit wäre $V_h \subset H_0^1(\Omega)$ nicht erfüllt. Solange das Gebiet Ω konvex ist, tritt nur die Situation von Abbildung 8.5.2a auf, und u^h wird auf $B \subset \tilde{T}$

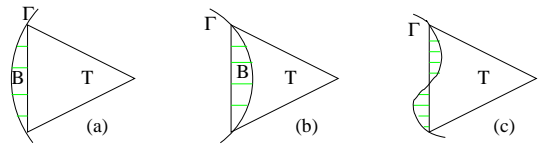
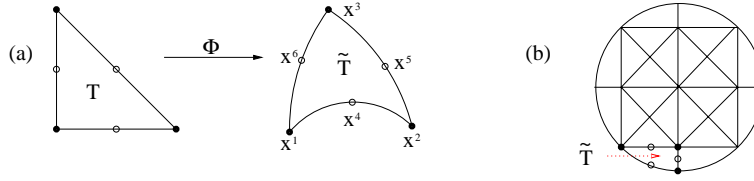


Abb. 8.5.2. Krummlinige Ränder

Abb. 8.5.3. a) Abbildung des Referenzdreiecks T auf das Bogendreieck \tilde{T} ; b) isoparametrische Dreieckszerlegung

durch $u^h = 0$ fortgesetzt. Im Falle von Abbildung 8.5.2b muss man auch den inneren Knoten von T auf den Wert null setzen, sodass $u^h = 0$ auf $\tilde{T} = T \setminus B$ und insbesondere $u^h = 0$ auf $\Gamma \cap \partial(T \setminus B)$. Insgesamt ergibt sich, dass der Träger aller $u^h \in V_h$ ein Ω einbeschriebenes Polygonegebiet darstellt. Eine Interpretation wäre die folgende: In der Randwertaufgabe wird Ω durch ein approximierendes Polygonegebiet $\Omega_h \subset \Omega$ ersetzt (vgl. Satz 2.4.6). Die beschriebene Finite-Element-Lösung stimmt mit der überein, die sich für das kleinere Polygonegebiet Ω_h ergäbe. Die Fehlerabschätzung aus Satz 8.4.4 gilt aber nur für das Teilgebiet Ω_h : $\inf_{v \in V_h} \|u - v\|_{H^1(\Omega_h)} \leq Ch \|u\|_{H^2(\Omega)}$. Wegen $v = 0$ auf $\Omega \setminus \Omega_h$ für $v \in V_h$, ist zusätzlich $\|u\|_{H^1(\Omega \setminus \Omega_h)}$ abzuschätzen. $\Omega \setminus \Omega_h$ ist in dem Streifen

$$S_\delta = \{x \in \Omega : \text{dist}(x, \Gamma) \leq \delta\} \quad \text{der Breite} \quad \delta := \max\{\text{dist}(x, \Gamma) : x \in \Omega \setminus \Omega_h\}$$

enthalten. Man kann wie folgt abschätzen:

$$\inf_{v \in V_h} \|u - v\|_{H^1(\Omega \setminus \Omega_h)} = \|u\|_{H^1(\Omega \setminus \Omega_h)} \leq \|u\|_{H^1(S_\delta)} \leq C\sqrt{\delta} \|u\|_{H^2(\Omega)} \quad \text{für } u \in H^2(\Omega) \cap H_0^1(\Omega).$$

Besteht $\Omega \setminus \Omega_h$ nur aus den Bogenstücken B der Abbildung 8.5.2a (wie zum Beispiel im Falle eines konvexen Gebietes) und ist $\Omega \in C^{1,1}$, so ist $\delta = \mathcal{O}(h^2)$. Hieraus folgt die gleichlautende Abschätzung $\inf_{v \in V_h} \|u - v\|_{H^1(\Omega)} \leq Ch^1 \|u\|_{H^2(\Omega)}$ wie für ein Polygonegebiet. Falls aber wie in Abbildung 8.5.2b das gesamte Dreieck T zu $\Omega \setminus \Omega_h$ gehört, wird δ zu $\mathcal{O}(h)$ und die Approximation verschlechtert sich zu $\inf_{v \in V_h} \|u - v\|_{H^1(\Omega)} \leq Ch^{1/2} \|u\|_{H^2(\Omega)}$ (vgl. Strang-Fix [78, p. 192]).

Um die Dreiecks- oder Parallelogramm-Elemente dem krummlinigen Rand besser anzupassen, kann man die Technik der *isoparametrischen finiten Elemente* verwenden. Aus Abbildung 8.3.4 (Seite 128) geht hervor, dass eine affine Transformation $\Phi : (\xi, \eta) \mapsto \mathbf{x}^1 + \xi(\mathbf{x}^2 - \mathbf{x}^1) + \eta(\mathbf{x}^3 - \mathbf{x}^1)$ das Referenzdreieck T in ein beliebiges Dreieck $\tilde{T} \in \mathcal{T}$ abbildet. Die lineare [bzw. im Falle von §8.3.4 quadratische] Funktion $u(x)$ über \tilde{T} kann durch $v(\xi, \eta) = u(\Phi(\xi, \eta))$ als Bild einer linearen [bzw. quadratischen] Funktion v über T dargestellt werden. Wir ersetzen die affine Dreieckstransformation Φ nun durch eine allgemeinere, quadratische Abbildung

$$\Phi(\xi, \eta) := \begin{bmatrix} a_1 + a_2\xi + a_3\eta + a_4\xi^2 + a_5\xi\eta + a_6\eta^2 \\ a_7 + a_8\xi + a_9\eta + a_{10}\xi^2 + a_{11}\xi\eta + a_{12}\eta^2 \end{bmatrix} : T \longrightarrow \tilde{T} \subset \mathbb{R}^2.$$

Das Bild \tilde{T} ist ein Bogendreieck. Die Koeffizienten a_1, \dots, a_{12} sind eindeutig durch die Knoten $\mathbf{x}^1, \dots, \mathbf{x}^6$ von \tilde{T} festgelegt, die sich als Bilder der Eckpunkte und Seitenmitten des Referenzdreiecks T ergeben sollen (vgl. Abbildung 8.5.3a). Die bisherige Triangulation \mathcal{T} kann man durch eine "Triangulation" $\tilde{\mathcal{T}}$ mit Bogendreiecken ersetzen, wenn benachbarte Elemente den gleichen Bogen als gemeinsamen Rand besitzen und auch die Mittelknoten zusammenfallen. Die Ansatzfunktionen auf $\tilde{T} \in \tilde{\mathcal{T}}$ haben die Gestalt $u(x) = v(\Phi^{-1}(x))$, wobei $v(\xi, \eta)$ linear [bzw. quadratisch] in ξ, η ist. Der entstehende Finite-Element-Raum ist der Raum der isoparametrischen, linearen [bzw. quadratischen] Elemente (vgl. Jung-Langer [47, §4.5.6], Knabner-Angermann [52, §3.8], Strang-Fix [78, p. 192], Ciarlet [22], Schwarz [75], Zienkiewicz [94]). Im Allgemeinen besteht kein Anlass, im Inneren von Ω gekrümmte Bogendreiecke zu verwenden. Wie in Abbildung 8.5.3b wählt man übliche Dreiecke im Inneren (d.h. die quadratische Transformation Φ wird wieder affin gewählt: $a_4 = a_5 = a_6 = a_{10} = a_{11} = a_{12} = 0$). Ein Randdreieck, wie zum Beispiel \tilde{T} aus Abbildung 8.5.3b, wird dagegen wie folgt definiert: \mathbf{x}^1 und \mathbf{x}^2 seien die Ecken von \tilde{T} , die auf Γ liegen. Man wählt einen weiteren Randpunkt \mathbf{x}^4 zwischen \mathbf{x}^1 und \mathbf{x}^2 und verlangt, dass der Dreiecksrand $\partial\tilde{T}$ den Gebietsrand Γ in $\mathbf{x}^1, \mathbf{x}^4, \mathbf{x}^2$ interpoliert.

Bemerkung 8.5.5 Abschnitt 5.2.2 zeigte, dass im Falle von anderen als Dirichlet-Randbedingungen die Konstruktion von Differenzenschemata zunehmend kompliziert wird. Stattdessen kann man die üblichen Differenzenverfahren auch auf die inneren Gitterpunkte beschränken und in Randnähe das Randwertproblem durch (z.B. isoparametrische) finite Elemente diskretisieren.

8.6 A-posteriori-Fehlerabschätzungen, Adaptivität

Die folgenden Überlegungen betreffen zwei unterschiedlichen Aufgabenstellungen. Der erste Wunsch ist eine Aussage über die Qualität der erzielten Finite-Element-Lösung. Diese Frage wird in §8.6.1 diskutiert werden. Die zweite Frage zielt auf die Effizienz der Finite-Element-Diskretisierung und wird in §8.6.2 behandelt. Wie sich herausstellen wird, sind beide Aufgaben miteinander verbunden; zumindest versucht man die Effizienz mit Hilfe der A-posteriori-Fehlerabschätzungen zu erzielen.

8.6.1 A-posteriori-Fehlerabschätzungen

8.6.1.1 Kritik der bisherigen Fehlerbetrachtungen

Die bisher beschriebenen Fehlerabschätzungen für die Finite-Element-Lösung (aber auch für die Lösungen der Differenzenverfahren) werden sämtlich *a priori* bestimmt, d.h. ohne dass Information über die berechnete Lösung einfließt. Da die Abschätzungen von der Form $\|\text{Fehler}\| \leq \alpha(h) \cdot \|u\|$ sind, erfährt man hieraus etwas über das asymptotische Verhalten, wenn h gegen null strebt: $\|\text{Fehler}\| \leq \mathcal{O}(\alpha(h))$. Ferner deutet die Ungleichung (auch für festes h) an, wie der Fehler von der Glattheit von u abhängt, wenn diese durch die Norm $\|\cdot\|$ ausgedrückt wird. Obwohl dies interessante Aussagen sind, bleiben andere Fragen offen.

- Die Fehlerschranke ist im Allgemeinen nur qualitativ gegeben, d.h. Konstanten bleiben unbekannt. Damit ist die folgende konkrete Frage nicht beantwortbar:

$$\text{gilt } \|\text{Fehler}\| \leq 0.001 ? \quad (8.6.1)$$

Selbst wenn der Faktor $\alpha(h)$ quantitativ gegeben wäre, bliebe das folgende Problem:

- Der Faktor $\|u\|$ ist die Norm einer (zumindest vor der Berechnung, also *a priori*) unbekannten Funktion und somit selbst unbekannt.

Generell gilt für jede Art der Abschätzung, auch für (8.6.1), dass sie viel zu pessimistisch sein kann (es wäre nicht befriedigend, wenn (8.6.1) nachgewiesen wird, aber in Wirklichkeit $\|\text{Fehler}\| = 10^{-6}$ gilt). Schließlich kann man die Frage nach der

- Wahl der Normen

stellen. Die bisher verwandten Normen sind gewählt worden, weil sich die Abschätzungen dann am einfachsten ergeben. Man kann Gründe angeben, warum die verwendeten Sobolev-Normen nicht optimal sind: Regularitätsanalysen in §9.1.6 werden zeigen, dass Lösungen von $Lu = f$ (f glatt) im Inneren des Gebietes (entfernt vom Rand) glatter sind als in Randnähe. Es wäre daher angemessen, Normen für u zu verwenden, die auf dieses Verhalten eingehen und die Glattheit nicht wie die Sobolev-Raumes örtlich gleichmäßig messen.

Man kann noch weiter gehen und fragen, warum überhaupt Normen verwendet werden. In vielen Fällen ist man nicht an der Lösung u als Funktion interessiert, sondern möchte nur Funktionswerte an speziellen Stellen oder Mittelwerte etc. Die allgemeine Formulierung wäre ein (lineares) Funktional $\phi(u)$. Statt des allgemeinen Fehlers $u - u^h$ möchte man dann etwas über $\phi(u - u^h)$ erfahren.

8.6.1.2 Konzept der A-posteriori-Fehlerabschätzungen

Der Name besagt, dass man den Fehler $e := u - u^h$ untersuchen möchte, *nachdem* man u^h bestimmt hat.

Eine mögliche Lösung dieser Frage könnte wie folgt aussehen. Seien u die Lösung von $Lu = f$ (z.B. mit Nullrandbedingung) und u^h die Finite-Element-Lösung (ebenfalls mit Nullrandbedingung). Dann gilt

$$Le = L(u - u^h) = Lu - Lu^h = f - Lu^h =: r. \quad (8.6.2)$$

Die rechte Seite r (Defekt oder Residuum genannt) gehört zum Dualraum $H_0^{-1}(\Omega)$. Falls es gelänge, $\|r\|_{-1} \leq \varepsilon$ mit einem konkreten Wert ε abzuschätzen, erhielte man $\|e\|_1 \leq \|L^{-1}\|_{1 \leftarrow -1} \varepsilon$. Falls weiterhin $\|L^{-1}\|_{1 \leftarrow -1} \leq C$ mit einer berechenbaren Größe C nachweisbar wäre, hätte man das Ziel erreicht: $\|e\|_1 \leq C\varepsilon$ stellt eine Fehlerabschätzung mit bekannter Schranke auf der rechten Seite dar.

Die Abschätzung von r durch $\|r\|_{-1} \leq \varepsilon$ ist im strikten Sinne unmöglich: $\|r\|_{-1} = \sup_{v \in H_0^1(\Omega)} (r, v) / \|v\|_1$ ist nicht mit endlichem Rechenaufwand erhältlich. Der naheliegende Versuch, die Testfunktionen $v \in H_0^1(\Omega)$ auf den Finite-Element-Raum V_h zu beschränken, scheitert völlig, da r senkrecht auf V_h steht (vgl. (8.2.3)).

Schon die Berechnung der L^2 -Norm einer rechten Seite r ist mit endlich vielen Daten unmöglich. Jede Abschätzung von $\|r\|_{-1}$ erfordert daher A-priori-Annahmen über f und über die Koeffizienten von L . Dagegen ist $\|L^{-1}\|_{1 \leftarrow -1} \leq C$ mit gewissen Einschränkungen berechenbar.¹⁰

Im folgenden Abschnitt wird die Abschätzung von $\|e\|_1$ mittels r direkt ausgeführt.

8.6.1.3 Beispiel eines Residuum-Fehlerschätzers

Es sei der Modellfall

$$-\Delta u = f \quad \text{in } \Omega, \quad u = 0 \quad \text{auf } \Gamma, \quad (8.6.3)$$

angenommen, sodass $a(u, v) = f(v)$ aus Beispiel 7.2.10 für alle $v \in V := H_0^1(\Omega)$ die Variationsformulierung darstellt. Sei \mathcal{T} eine zulässige Triangulation (vgl. Definition 8.3.7). Wir führen die folgenden Notationen ein: Die drei Dreiecksseiten von $T \in \mathcal{T}$ bilden die Kantenmenge $\mathcal{E}(T)$, während die drei Ecken des Dreiecks $T \in \mathcal{T}$ die Knotenmenge $\mathcal{N}(T)$ ausmachen. Ihre Vereinigung liefert

$$\mathcal{E} := \bigcup_{T \in \mathcal{T}} \mathcal{E}(T), \quad \mathcal{N} := \bigcup_{T \in \mathcal{T}} \mathcal{N}(T).$$

Die Anteile des Dirichlet-Randes sind $\mathcal{E}_D := \{E \in \mathcal{E} : E \subset \Gamma\}$, das Komplement ist $\mathcal{E}_\Omega := \{E \in \mathcal{E} : E \subset \Omega\}$. Zu einer Kante $E \in \mathcal{E}$ gehören zwei Knoten als Endpunkte: sie bilden die Menge $\mathcal{N}(E)$.

Die Dreiecksmenge, die die Nachbarschaft eines Dreiecks $T \in \mathcal{T}$, einer Kante $E \in \mathcal{E}$ und eines Knotens $x \in \mathcal{N}$ ausmacht, wird wie folgt bezeichnet:

$$\omega_T := \bigcup_{T' : \mathcal{E}(T) \cap \mathcal{E}(T') \neq \emptyset} T', \quad \omega_E := \bigcup_{T' : E \in \mathcal{E}(T')} T'$$

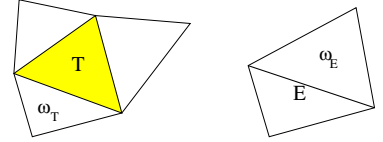


Abb. 8.6.1. ω_T und ω_E

(vgl. Abbildung 8.6.1). Jeder Kante $E \in \mathcal{E}$ wird eine Normalenrichtung \vec{n}_E zugeordnet (das Vorzeichen darf beliebig fixiert werden; wenn T und T' die gemeinsame Kante $E \in \mathcal{E}(T) \cap \mathcal{E}(T')$ besitzen, ist \vec{n}_E somit nur bezüglich eines Dreiecks die übliche äußere Normalenrichtung). Da die Ableitungen der Finite-Element-Funktionen an den Kanten $E \in \mathcal{E}_\Omega$ unstetig sein können, treten zwei Grenzwerte $\varphi(x \pm t\vec{n}_E)$ für $t \searrow 0$ auf. Ihre Differenz wird mit dem Symbol $[\cdot]$ bezeichnet:

$$[\varphi]_E(x) := \lim_{t \searrow 0} \varphi(x + t\vec{n}_E) - \lim_{t \searrow 0} \varphi(x - t\vec{n}_E) \quad \text{für } x \in E \in \mathcal{E}_\Omega.$$

Auf beiden Seiten der kontinuierlichen Gleichung $a(u, v) = f(v)$ subtrahiere man $a(u^h, v)$:

$$\int_\Omega \langle \nabla(u - u^h), \nabla v \rangle dx = \int_\Omega f v dx - \int_\Omega \langle \nabla u^h, \nabla v \rangle dx \quad \text{für alle } v \in V = H_0^1(\Omega). \quad (8.6.4)$$

Die rechte Seite repräsentiert das Residuum r aus (8.6.2) mittels $r(v) := \int_\Omega f v dx - \int_\Omega \langle \nabla u^h, \nabla v \rangle dx$.

Lemma 8.6.1 *Sei c_Ω die Poincaré-Friedrichs-Konstante aus $\|v\|_{L^2(\Omega)} \leq c_\Omega |v|_{1,0}$ für $v \in H_0^1(\Omega)$ (vgl. Lemma 6.2.12). Dann ist der Diskretisierungsfehler $\|u - u^h\|_{H^1(\Omega)}$ beidseitig mittels r abschätzbar:*

$$\|r\|_{-1} \leq \|u - u^h\|_{H^1(\Omega)} \leq (1 + c_\Omega^2) \|r\|_{-1}, \quad \|r\|_{-1} = \sup_{\substack{v \in V = H_0^1(\Omega) \\ \|v\|_{H^1(\Omega)} = 1}} \left| \int_\Omega f v dx - \int_\Omega \langle \nabla u^h, \nabla v \rangle dx \right|. \quad (8.6.5)$$

Beweis. Sei $w \in H_0^1(\Omega)$. $\sup_{v \in H_0^1(\Omega), \|v\|_{H^1(\Omega)} = 1} \left| \int_\Omega \langle \nabla w, \nabla v \rangle dx \right| = \|\nabla w\|_{L^2(\Omega)} = |w|_{1,0}$ erfüllt nach Voraussetzung $\|w\|_{L^2(\Omega)} \leq c_\Omega |w|_{1,0}$, sodass

$$|w|_{1,0}^2 \leq \|w\|_{H^1(\Omega)}^2 \leq (1 + c_\Omega^2) |w|_{1,0}^2. \quad (8.6.6)$$

¹⁰Die Norm $\|L^{-1}\|_{1 \leftarrow -1}$ ist als Eigenwertproblem formulierbar. Seine approximative Lösung ist zwar nicht notwendigerweise eine obere Schranke, aber eine gute Näherung der rechten Seite in $\|L^{-1}\|_{1 \leftarrow -1} \leq C$ ist genauso brauchbar.

Da $|u - u^h|_{1,0} = \|\nabla(u - u^h)\|_{L^2(\Omega)} = \sup_{\|v\|_{H^1(\Omega)}=1} \left| \int_{\Omega} \langle \nabla(u - u^h), \nabla v \rangle dx \right|$, ist einerseits

$$\begin{aligned} \|r\|_{-1} &= \sup_{0 \neq v \in H_0^1(\Omega)} \frac{|r(v)|}{\|v\|_{H^1(\Omega)}} \stackrel{(8.6.4)}{=} \sup_{0 \neq v \in H_0^1(\Omega)} \frac{\left| \int_{\Omega} \langle \nabla(u - u^h), \nabla v \rangle dx \right|}{\|v\|_{H^1(\Omega)}} \\ &\stackrel{(8.6.6)}{\leq} \sup_{0 \neq v \in H_0^1(\Omega)} \frac{\left| \int_{\Omega} \langle \nabla(u - u^h), \nabla v \rangle dx \right|}{\|\nabla v\|_{L^2(\Omega)}} = \|\nabla(u - u^h)\|_{L^2(\Omega)} \leq \|u - u^h\|_{H^1(\Omega)} \end{aligned}$$

und andererseits

$$\begin{aligned} \|u - u^h\|_{H^1(\Omega)} &\stackrel{(8.6.6)}{\leq} \sqrt{1 + c_{\Omega}^2} |u - u^h|_{1,0} = \sqrt{1 + c_{\Omega}^2} \sup_{0 \neq v \in H_0^1(\Omega)} \frac{\left| \int_{\Omega} \langle \nabla(u - u^h), \nabla v \rangle dx \right|}{\|\nabla v\|_{L^2(\Omega)}} \\ &\stackrel{(8.6.4)}{=} \sqrt{1 + c_{\Omega}^2} \sup_{0 \neq v \in H_0^1(\Omega)} \frac{|r(v)|}{|v|_{1,0}} \stackrel{(8.6.6)}{\leq} (1 + c_{\Omega}^2) \sup_{0 \neq v \in H_0^1(\Omega)} \frac{|r(v)|}{\|v\|_{H^1(\Omega)}} = (1 + c_{\Omega}^2) \|r\|_{-1}. \end{aligned}$$

Zusammen ergibt sich die Behauptung. ■

In (8.6.5) wird $\int_{\Omega} f v dx - \int_{\Omega} \langle \nabla u^h, \nabla v \rangle dx$ partiell integriert. Für jedes T gilt

$$\int_T \langle \nabla u^h, \nabla v \rangle dx = - \int_T (\Delta u^h) v dx - \sum_{E \in \mathcal{E}(T)} \int_E \frac{\partial u^h}{\partial n} v dx \quad \text{für alle } v \in H^1(T).$$

Nach der Aufsummation über alle $T \in \mathcal{T}$ ergibt sich

$$\int_{\Omega} f v dx - \int_{\Omega} \langle \nabla u^h, \nabla v \rangle dx = \sum_{T \in \mathcal{T}} \left[\int_T (f - \Delta u^h) v dx - \sum_{E \in \mathcal{E}_{\Omega}(T)} \int_E [\langle \nabla u^h, \vec{n}_E \rangle]_E v dx \right] \quad \text{für alle } v \in H_0^1(T) \quad (8.6.7)$$

(man beachte, dass das Skalarprodukt $[\langle \nabla u^h, \vec{n}_E \rangle]_E$ zwar von der Vorzeichenwahl von \vec{n}_E abhängt, aber auch die Differenz $[\cdot]_E$, sodass der Ausdruck $[\langle \nabla u^h, \vec{n}_E \rangle]_E$ unabhängig von der Wahl von \vec{n}_E ist). Die Kanten $E \subset \Gamma$ werden ausgenommen, da dort $v = 0$ (außerdem wäre dort die Differenz $[\cdot]_E$ undefiniert).

Für stückweise lineare finite Elemente gilt in unserem Beispiel $\Delta u^h = 0$ in jedem T . In den Abschätzungen wird f pro Dreieck durch eine Konstante f_T ersetzt, die im optimalen Fall als Mittelwert

$$f_{T,\text{exakt}} := \frac{1}{\text{Fläche}(T)} \int_T f dx \quad (\text{konstante Funktion auf } T)$$

zu wählen ist. Die lokale Größe

$$\eta_T := \sqrt{h_T^2 \|f_T\|_{L^2(T)}^2 + \frac{1}{2} \sum_{E \in \mathcal{E}_{\Omega}(T)} h_E \|[\langle \nabla u^h, \vec{n}_E \rangle]_E\|_{L^2(E)}^2} \quad \text{mit } \begin{cases} h_T = \text{diam}(T), \\ h_E = \text{Länge}(E) \end{cases} \quad (8.6.8)$$

wird im Folgenden eine zentrale Rolle spielen.

Satz 8.6.2 *Seien u die Lösung von (8.6.3) und u^h die Lösung für stückweise lineare Elemente zur Triangulation \mathcal{T} . Dann gibt es Konstanten \bar{c} und \underline{c} , die nur von Ω und der Formregularität der Triangulation \mathcal{T} abhängen, sodass*

$$\|u - u^h\|_{H^1(\Omega)} \leq \bar{c} \sqrt{\sum_{T \in \mathcal{T}} \eta_T^2 + \sum_{T \in \mathcal{T}} h_T^2 \|f - f_T\|_{L^2(T)}^2} \quad (8.6.9a)$$

$$\text{und} \quad \eta_T \leq \underline{c} \sqrt{\|u - u^h\|_{H^1(\omega_T)}^2 + \sum_{T' \subset \omega_T} \|f - f_{T'}\|_{L^2(T')}^2}. \quad (8.6.9b)$$

Beweis. Wir verwenden die Darstellung (8.6.7) von $r(v)$ und verwenden die Galerkin-Gleichung $r(v^h) = 0$ für alle $v^h \in V_h$.

Zusammen mit $\Delta u^h = 0$ liefert die Identität

$$r(v) = r(v - v^h) = \sum_{T \in \mathcal{T}} \left[\int_T f(v - v^h) dx - \sum_{E \in \mathcal{E}_\Omega(T)} \int_E [\langle \nabla u^h, \vec{n}_E \rangle]_E (v - v^h) dx \right] \quad \text{für alle } v^h \in V_h.$$

Für v^h verwendet man die Clément-Interpolierende von v (vgl. [23], Ende von §8.4.1). Hierfür gilt

$$\|v - v^h\|_{L^2(T)} \leq C_1 h_T \|u\|_{H^1(\widetilde{\omega}_T)}, \quad \|v - v^h\|_{L^2(E)} \leq C_2 \sqrt{h_E} \|u\|_{H^1(\widetilde{\omega}_E)}$$

(siehe [23]), wobei $\widetilde{\omega}_T := \bigcup_{T': \mathcal{N}(T) \cap \mathcal{N}(T') \neq \emptyset} T' \supset \omega_T$ und $\widetilde{\omega}_E := \bigcup_{T': \mathcal{N}(E) \cap \mathcal{N}(T') \neq \emptyset} T' \supset \omega_E$. Die Konstanten C_1, C_2 hängen nur von der Formregularität (d.h. dem kleinsten Innenwinkel) ab. Wir erhalten die Abschätzung

$$\begin{aligned} |r(v)| &\leq C_1 \sum_{T \in \mathcal{T}} h_T \|f\|_{L^2(T)} \|v\|_{H^1(\widetilde{\omega}_T)} + C_2 \sum_{E \in \mathcal{E}_\Omega(T)} \sqrt{h_E} \|[\langle \nabla u^h, \vec{n}_E \rangle]_E\|_{L^2(E)} \|v\|_{H^1(\widetilde{\omega}_E)} \\ &\stackrel{\text{Cauchy-Schwarz-Ungleichung}}{\leq} C_3 \|v\|_{H^1(\Omega)} \left[\sum_{T \in \mathcal{T}} h_T^2 \|f\|_{L^2(T)}^2 + \sum_{E \in \mathcal{E}_\Omega(T)} h_E \|[\langle \nabla u^h, \vec{n}_E \rangle]_E\|_{L^2(E)}^2 \right] \end{aligned}$$

für alle $v \in H_0^1(\Omega)$. Der Term $\|f\|_{L^2(T)}^2$ ist durch $2\|f_T\|_{L^2(T)}^2 + 2\|f - f_T\|_{L^2(T)}^2$ abschätzbar. Zusammen mit (8.6.5) und (8.6.8) erhalten wird die Aussage (8.6.9a).

Die zweite Ungleichung (8.6.9b) verwendet eine spezielle Wahl von v . Details können bei Verfürth [85, Seiten 15-17] nachgelesen werden. ■

Der Satz bietet mit (8.6.9a) eine Fehlerabschätzung an, die im Wesentlichen die lokalen Größen η_T verwendet. Wichtig ist, dass η_T *a posteriori*, d.h. nach der Berechnung von u^h vollständig berechenbar ist. Es verbleibt $f - f_T$, deren Schranke $\|f - f_T\|_{L^2(T)}$ *a priori* bekannt sein muss. Annahmen über f sind nach der Diskussion in §8.6.1.2 unvermeidbar.

Die zweite Ungleichung (8.6.9b) zeigt, dass η_T bis auf den Term $\mathcal{O}(h^2 \|f - f_T\|_{L^2(\omega_T)})$ auch nach oben durch den Fehler $\|u - u^h\|_{H^1(\omega_T)}^2$ mit einer etwas größeren Umgebung $\omega_T \supset T$ abschätzbar ist. Summation aller η_T^2 liefert

$$\sum_{T \in \mathcal{T}} \eta_T^2 \leq C \left[\|u - u^h\|_{H^1(\Omega)}^2 + \sum_{T \in \mathcal{T}} \|f - f_T\|_{L^2(T)}^2 \right].$$

Damit sind modulo $\sum_{T \in \mathcal{T}} \|f - f_T\|_{L^2(T)}^2$ der Fehlerschätzer $\eta := \sqrt{\sum_{T \in \mathcal{T}} \eta_T^2}$ und der Fehler $\|u - u^h\|_{H^1(\Omega)}$ äquivalent.

Einen Fehlerschätzer η nennt man *zuverlässig*, wenn $\|u - u^h\|_{H^1(\Omega)} \leq \text{const} \cdot \eta + (\dots)$ mit einem A-priori-Term (\dots) wie z.B. $\sum_{T \in \mathcal{T}} \|f - f_T\|_{L^2(T)}^2$. Um eine Überschätzung infolge von $\|u - u^h\|_{H^1(\Omega)} \ll \eta$ zu vermeiden, benötigt man die umgekehrte Ungleichung. Ein Schätzer η mit $\eta \leq \text{const} \cdot \|u - u^h\|_{H^1(\Omega)} + (\dots)$ heißt *effizient*. Damit eine Aussage wie (8.6.1) beantwortbar wird, muss die Größe von const in (8.6.9a) bekannt sein. Der optimale Fall liegt vor, wenn $\text{const} = 1$ in beiden Ungleichungen gilt oder zumindest asymptotisch $\|u - u^h\|_{H^1(\Omega)}/\eta \rightarrow 1$ für $h \rightarrow 0$ gilt. Dann heißt η *asymptotisch exakt*.

Mehr zu Fehlerschätzern erfährt man in der Einführung von Verfürth [85]. Man vergleiche auch [29], [36, §4.8] und Knabner-Angermann [52, §4.2]. Die Frage der Konvergenz wird von Dörfler [28] behandelt.

Fehlerschätzer für Funktionale $\phi(u - u^h)$ anstelle von Normen werden zum Beispiel von Becker-Rannacher [9] diskutiert.

8.6.2 Effizienz der Finite-Element-Methode

8.6.2.1 Maß für die Güte einer Diskretisierung

Jede (konsistente und konvergente) Diskretisierung bestimmt auf die eine oder andere Art eine Näherung u^h von u . Die Frage liegt nahe, wie man zwei Diskretisierungen vergleicht. Bei den zu vergleichenden Diskretisierungen kann es sich um verschiedene Diskretisierungsmethoden handeln, aber auch um gleiche Finite-Element-Ansätze auf verschiedenen Triangulationen.

Zwei unterschiedliche Ziele bieten sich als Kriterium an: (a) möglichst kleiner Rechenaufwand (z.B. gemessen als Laufzeit des Computerprogramms) oder (b) möglichst kleiner Fehler (gemessen in irgendeiner Norm). Da beide Ziele gegenläufig sind, müssen Aufwand und Genauigkeit verknüpft werden. Dies führt auf zwei Fragestellungen, die die Effizienz der Diskretisierung vergleichen:

- (i) Gegeben ein $\varepsilon > 0$. Welches Verfahren liefert eine Näherung \tilde{u} mit $\|u - \tilde{u}\| \leq \varepsilon$ mit dem kleinsten Rechenaufwand?
- (ii) Seien u^I und u^{II} die Resultate zweier Verfahren bei gleichem Rechenaufwand. Welche Näherung ist genauer?

Der Rechenaufwand als Maß bringt eine neue Schwierigkeit mit sich, da er sich notwendigerweise auf einen Algorithmus zur Berechnung der Lösung bezieht und dieser nicht durch die Diskretisierungsmethode festgeschrieben ist. Letztere beschreibt das zu lösende Gleichungssystem. Hierfür stehen viele Verfahren zur Verfügung, die sich in der Rechenzeit wesentlich unterscheiden können (vgl. [42]). Nachdem aber gute Gleichungslöser eine Rechenzeit aufweisen, die proportional zur Dimension n des Gleichungssystems ist, bietet sich der Ausweg an, die Dimension n als Maß des Rechenaufwandes zu verwenden.¹¹

8.6.2.2 Konsistenzordnung

Wenn wir die obige Formulierung (ii) aufnehmen, haben wir $\varepsilon = \varepsilon(n)$ als Funktion der Dimension n zu diskutieren. Stückweise lineare finite Elemente führten auf einen Fehler $\varepsilon = \mathcal{O}(h^2)$ in der L^2 -Norm. Für eine uniforme Triangulation und die Raumdimension d ($\Omega \subset \mathbb{R}^d$) gilt $h = \mathcal{O}(n^{-1/d})$, sodass $\varepsilon = \mathcal{O}(n^{-2/d})$. Verwendet man allgemeiner stückweise Polynome¹² vom Grad p , so ergibt sich für den Fehler

$$\varepsilon = \mathcal{O}(n^{-(p+1)/d}).$$

Hierbei ist jeweils $u \in H^{p+1}(\Omega) \cap V$ (z.B. mit $V = H_0^1(\Omega)$) unterstellt. Die Frage (ii) lässt sich damit asymptotisch (für große n) eindeutig beantworten: Verfahren hoher Konsistenzordnung sind besser als solche mit niedrigerer Ordnung.

8.6.2.3 Wahl der Triangulation

Legt man den lokalen Polynomgrad p fest, bleibt die Frage, wie die Triangulation \mathcal{T} zu wählen ist. Da sich das Fehlerquadrat $\|e\|_{L^2(\Omega)}^2$ ($e := u - u^h$) als Summe der $\|e\|_{L^2(T)}^2$ ($T \in \mathcal{T}$) zusammensetzt, liegt das heuristische *Prinzip der Fehleräquilibrierung* nahe: Alle Fehlerbeiträge $\|e\|_{L^2(T)}$ sollten von vergleichbarer Größe sein. Dieses Prinzip wird als konstruktives Kriterium verwendet, um in §8.6.3 die Triangulation adaptiv zu konstruieren.

8.6.3 Adaptive Finite-Element-Methode

8.6.3.1 Aufgabenstellung

Wir gehen von der Zielstellung (i) aus: Gegeben eine gewünschte Fehlertoleranz $\varepsilon > 0$ in einer gewissen Norm $\|\cdot\|$ ist eine Finite-Element-Diskretisierung mit möglichst kleiner Dimension n gesucht, die diese Fehlertoleranz einhält.

¹¹Bei dieser Vereinfachung darf nicht vergessen werden, dass die Rechenzeit für die Gleichungsauflösung nicht nur von der Dimension, sondern auch von der Struktur abhängt (wie schwachbesetzt ist die Matrix?). Ein anderes Maß für den Aufwand wäre der Speicherbedarf. Man mache sich aber klar, dass ein Differenzenverfahren bei gleicher Dimension wesentlich weniger Speicher verbraucht als eine Finite-Element-Methode.

¹²Aus numerischen Gründen sind passende Polynombasen zu verwenden. Mögliche Probleme demonstrieren die Polynome $b_\nu(x) = x^\nu/\nu$ ($\nu = 1, \dots, N$) für das eindimensionale Randwertproblem $u'' = f$ in $\Omega = (0, 1)$, $u(0) = 0$, $u'(1) = 0$ (natürliche Randbedingung). Die Systemmatrix ist die *Hilbert-Matrix* $A_{ij} = 1/(i+j-1)$, die ein Beispiel für extrem schlechte Kondition ist (vgl. Maëß [60, S. 108]).

8.6.3.2 Adaptive Gitterverfeinerung

Beispielsweise mit Hilfe der Fehlerschätzer aus §8.6.1.3 seien Größen $\eta(T)$ definiert, die jedem Element von \mathcal{T} eine Fehlerschätzgröße zuordnen. Die heuristische Annahme ist, dass eine Gitterverfeinerung in Elementen T mit großen Werten von $\eta(T)$ den Gesamtfehler stärker verkleinert, als wenn man Elemente mit kleineren $\eta(T)$ verfeinert. Da man im Sinne der Effizienz den Fehler verkleinern, aber n möglichst wenig vergrößern möchte, wählt man die Teilmenge der zu verfeinernden Elemente wie folgt aus:

- Schritt I: Sei $\vartheta \in (0, 1)$ ein fester Wert. Bestimme $\eta_{\max} := \max \{\eta(T) : T \in \mathcal{T}\}$ und setze $\mathcal{T}_{\text{verfeinern}} := \{T \in \mathcal{T} : \eta(T) \geq \vartheta \eta_{\max}\}$.
- Schritt II: Man verfeinere alle Elemente aus $\mathcal{T}_{\text{verfeinern}}$ gemäß des Verfahrens aus §8.6.3.3.
- Schritt III: Man bestimme die Lösung u^h und die $\eta(T)$ zur neuen Triangulation \mathcal{T}_{neu} .
- Schritt IV: Wenn die A-posteriori-Fehlerschätzung eine hinreichende Genauigkeit anzeigt, akzeptiere man u^h als Resultat, sonst wiederhole man den Prozess bei Schritt I.

Falls der Fehlerschätzer $\eta(T)$ schon äquilibriert ist, d.h. alle $\eta(T)$ die ähnliche Größe haben, könnte für $\vartheta = 1/2$ die Teilmenge $\mathcal{T}_{\text{verfeinern}} = \mathcal{T}$ resultieren. Die Verfeinerung wäre dann uniform. Sind die Werte $\eta(T)$ dagegen ungleich verteilt, werden eventuell nur wenige Elemente verfeinert (d.h. $\#\mathcal{T}_{\text{verfeinern}} \ll \#\mathcal{T}$).

Das obige Verfahren unterscheidet nur zwischen den Alternativen Verfeinern und Nicht-Verfeinern. Alternativ kann eine lokal zu erzielende Schrittweite pro T angegeben werden, sodass Elemente eventuell mehrfach unterteilt werden.

8.6.3.3 Gitterverfeinerungstechniken

Die Grundaufgabe lautet: Ausgehend von einer Teilmenge \mathcal{T}_v von \mathcal{T} möchte man eine neue Triangulation \mathcal{T}_{neu} berechnen, die anstelle aller $T \in \mathcal{T}_v$ kleinere Elemente aufweist und ansonsten mit \mathcal{T} möglichst übereinstimmt. Dabei ist Folgendes zu bedenken:

- Je mehr Elemente in $\mathcal{T} \cap \mathcal{T}_{\text{neu}}$ liegen, umso weniger Neuberechnungen von Matrixkoeffizienten sind nötig.
- Wenn keine hängenden Knoten in Kauf genommen werden sollen, erzwingt die Verfeinerung eines Dreiecks $T \in \mathcal{T}_v$ unter Umständen die seines Nachbarn.
- Die Formregularität erfordert Vorkehrungen, die verhindern, dass zum Beispiel die Innenwinkel der Dreiecke bei jedem Verfeinerungsschritt kleiner werden.
- Wenn die K -Gittereigenschaft noch nicht aus der Formregularität folgt, ist ihre Einhaltung zusätzlich zu beachten.

Zunächst eine Anmerkung zu d). Im eindimensionalen Fall, wenn die Elemente Intervalle sind, gibt es keine Formregularität. Abbildung 8.6.2 zeigt eine Verfeinerungsfolge in $\Omega = (0, 1)$, bei der jeweils alle Teilintervalle in $(0, 1/4)$ verfeinert werden sollen. Die erste Verfeinerung führt auf ein K -Gitter mit $K = 2$. Die zweite, rein lokale Halbierung der Teilintervalle in $(0, 1/4)$ liefert $K = 4$, da die Intervalle $[3/16, 4/16]$ und $[1/4, 1/2]$ mit dem Längenverhältnis 4 benachbart sind. Will man $K = 2$ erhalten, muss auch $[1/4, 1/2]$ halbiert werden.



Abb. 8.6.2. K -Gitter mit $K = 2$, c) wird verworfen

Zur Einhaltung der Formregularität gibt es ein einfaches hinreichendes Kriterium. Wenn während des Verfeinerungsalgorithmus nur Dreiecke erzeugt werden, die zu Dreiecken der Ausgangstriangulation T kongruent sind, verändern sich die Winkel nicht – insbesondere bleibt der minimale Winkel konstant. Daher ist die Standardverfeinerung (‘‘Typ A’’) eine Zerlegung von $T \in \mathcal{T}_v$ in vier kongruente Teildreiecke halber Seitenlänge (vgl. Abbildung 8.6.3a).

Ein Problem ergibt sich, wenn $T \in \mathcal{T}_v$ ein Nachbardreieck $T' \in \mathcal{T} \setminus \mathcal{T}_v$ besitzt (vgl. Abbildung 8.6.3b), da ein hängender Knoten auftritt (vgl. §8.8.2). Würde man auch T' gemäß Typ A zerlegen, breitet sich die Verfeinerung in der gesamten Zusammenhangskomponente aus und man erhielte eine uniforme Verfeinerung. Deshalb wird ein $T' \in \mathcal{T} \setminus \mathcal{T}_v$ nur dann gemäß Typ A zerlegt, wenn es zwei verfeinerte Nachbarn besitzt (vgl. gestricheltes Dreieck in Abbildung 8.6.4b). Andernfalls sieht die Verfeinerungsstrategie nach Bank [7] vor, dass T' gemäß Abbildung 8.6.4a halbiert wird (Verfeinerung vom ‘‘Typ B’’). Dieser Schritt verkleinert

eventuell den kleinsten Innenwinkel. Damit diese Winkelverkleinerung nicht wiederholt wird, werden in einem nächsten Verfeinerungsschritt zunächst die Verfeinerungen vom Typ B rückgängig gemacht.

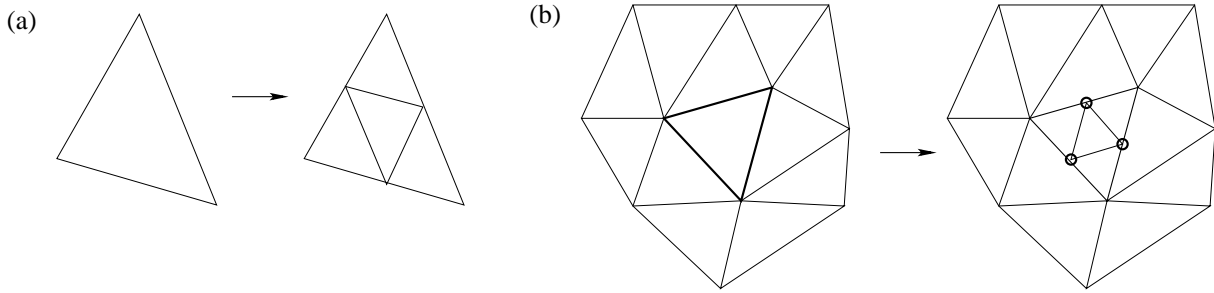


Abb. 8.6.3. (a) Typ A: Regelmäßige Zerlegung eines Dreiecks, (b) Hängende Knoten nach lokaler Verfeinerung.

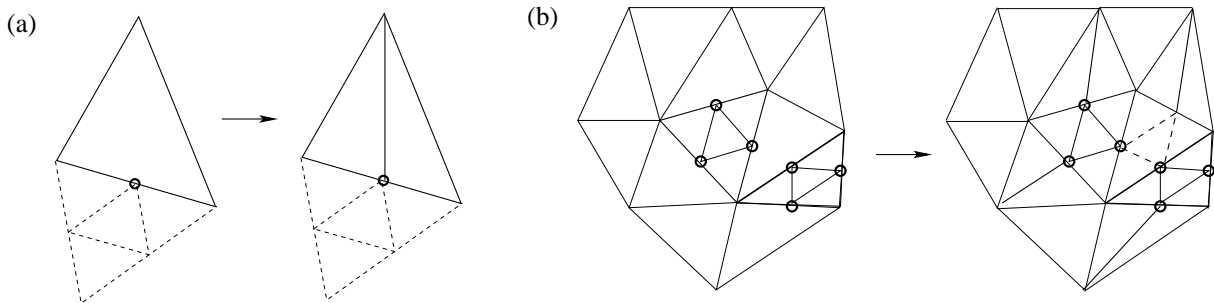


Abb. 8.6.4. (a) Typ B: Zerlegung eines Dreiecks durch eine Seitenhalbierende, (b) Resultat der Verfeinerung. Man vergleiche auch Knabner-Angermann [52, §4.1].

8.6.3.4 Adaptive Gittervergrößerung

Bisher sind wir davon ausgegangen, dass die Ausgangstriangulation grob ist und verfeinert werden muss. Bei einigen Aufgabenstellungen hat man dagegen eine Ausgangstriangulation, in der Elemente kleiner sind als erforderlich. In Analogie zum obigen Verfeinerungsverfahren kann man eine Teilmenge $\mathcal{T}_{\text{vergrößern}}$ ermitteln (z.B. mittels $\eta(T) \leq \eta_{\min}$) und dort lokal das Gitter vergrößern. Dazu ist es notwendig, die vorher ausgeführten Verfeinerungsschritte in einer Baumstruktur zu speichern. Dabei sind die Teildreiecke jeweils die Söhne des Vaterdreiecks. Die Verfeinerung bedeutet dann, dass ein Blattknoten (d.h. ein bisher unverfeinertes Dreieck) neue Söhne bekommt, während bei der Vergrößerung Blätter, die Söhne eines Vaterknotens sind, entfernt werden. Wie bei der Verfeinerung muss darauf geachtet werden, dass keine hängenden Knoten übrig bleiben.

8.6.3.5 hp -Methoden

Bisher war der Polynomgrad p der stückweisen Finite-Element-Funktionen fixiert, und die Elementgröße h war als die Größe angesehen worden, die theoretisch gegen null streben soll. Bei einer konkreten Finite-Element-Berechnung sind naturgemäß h und p fixiert. Im Zug der Verfeinerung kann man aber die Frage stellen, was günstiger ist: die (lokale) Verkleinerung von h oder die (lokale) Anhebung des Polynomgrades p . Gemäß §8.6.2.2 sollte die Vergrößerung von p vorteilhaft sein, wenn die Lösung hinreichend glatt ist. Das Vorgehen aus §8.6.3.2 ist in Schritt II dadurch zu modifizieren, dass zusätzlich vorgeschlagen werden muss, ob h oder p oder beide verändert werden sollen. (Vgl. Großmann-Roos [36, §4.9.2]).

8.7 Eigenschaften der Systemmatrix

8.7.1 Zusammenhang von \mathbf{L} und L_h

Bei Differenzenverfahren wurden die Eigenschaften der Matrix eingehend untersucht, da hiervon die Lösbarkeit der Differenzengleichungen und die Konvergenzanalyse abhängen. Im Fall der Finite-Element-

Diskretisierung erhalten wir die entsprechenden Aussagen auf andere Weise. Maßgebend für Lösbarkeit und Konvergenz ist der Unterraum V_h bzw. der zur Bilinearform $a(\cdot, \cdot) : V_h \times V_h \rightarrow \mathbb{R}$ gehörende Operator $L_h : V_h \rightarrow V_h$, der in (8.1.11a) als L_N eingeführt wurde.

Die Systemmatrix \mathbf{L} hängt bei festem V_h von der gewählten Basis $\{b_1, \dots, b_{N_h}\}$ ab. Dabei sei

$$N_h := \dim(V_h).$$

Der in (8.1.6) definierte Isomorphismus $P : \mathbb{R}^{N_h} \rightarrow V_h$ bildet den Koeffizientenvektor \mathbf{v} in $P\mathbf{v} = \sum_{i=1}^{N_h} v_i b_i \in V_h$ ab. Bei finiten Elementen stimmen die Koeffizienten im Allgemeinen mit den Knotenwerten überein: $v_i = (P\mathbf{v})(\mathbf{x}^i)$, $1 \leq i \leq N_h$. Der Zusammenhang zwischen der Matrix \mathbf{L} und dem Operator $L_h : V_h \rightarrow V_h$ ist gemäß Lemma 8.1.9 durch

$$\mathbf{L} = P^* L_h P, \quad L_h = P^{*-1} \mathbf{L} P^{-1} \quad (8.7.1)$$

gegeben. Die Definition von P^* durch (8.1.11b): $\langle P^* u, v \rangle = (u, Pv)_{L^2(\Omega)}$ hängt noch von der Wahl des Skalarproduktes $\langle \cdot, \cdot \rangle$ im \mathbb{R}^{N_h} ab. Hier sei wie üblich gewählt:

$$\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^{N_h} u_i v_i. \quad (8.7.2a)$$

In Übungsaufgabe 8.1.11 wurden bereits die folgenden Eigenschaften von \mathbf{L} festgestellt: Wenn $a(\cdot, \cdot)$ symmetrisch [und V-elliptisch] ist, so ist \mathbf{L} symmetrisch [und positiv definit]. Man beachte, dass Differenzverfahren diese Eigenschaft nicht immer haben. Zum Poisson-Problem gehört eine symmetrische Form, trotzdem ist die Matrix L_h aus Abschnitt 4.8.1 nicht symmetrisch (vgl. Satz 4.8.4).

8.7.2 Normäquivalenzen und Massematrix

Für die Gleichungslösung von $\mathbf{L}\mathbf{u} = \mathbf{f}$ spielt die Kondition $\text{cond}(\mathbf{L})$ eine wichtige Rolle. Wir wollen zeigen, dass unter Standardannahmen $\text{cond}(\mathbf{L}) = \mathcal{O}(h^{-2m})$ gilt (vgl. Bemerkung 5.3.10). Zuvor ist aber die zugrundeliegende Norm des \mathbb{R}^{N_h} festzulegen. Bis auf den Skalierungsfaktor $h^{n/2}$ ist

$$\|\mathbf{u}\|_h := \sqrt{h^n \sum_{i=1}^{N_h} |u_i|^2} \quad (n: \text{Dimension von } \Omega \subset \mathbb{R}^n) \quad (8.7.2b)$$

die zum Skalarprodukt (8.7.2a) gehörende Euklidische Norm. Die zugehörige Matrixnorm

$$\|\mathbf{L}\| := \sup\{\|\mathbf{L}\mathbf{v}\|_h / \|\mathbf{v}\|_h : \mathbf{0} \neq \mathbf{v} \in \mathbb{R}^{N_h}\} \quad (8.7.2c)$$

ist die übliche Spektralnorm von \mathbf{L} (sie ist unabhängig vom Skalierungsfaktor in $\|\cdot\|_h$). Als Alternative zu $\|\cdot\|_h$ aus (8.7.2b) sei die Norm

$$\|\mathbf{u}\|_P := \|P\mathbf{u}\|_{L^2(\Omega)} \quad \text{für } \mathbf{u} \in \mathbb{R}^{N_h} \quad (8.7.3)$$

eingeführt.

Für wichtige Fälle sind $\|\cdot\|_P$ und $\|\cdot\|_h$ äquivalent (gleichmäßig bezüglich h). Als Beispiel wählen wir die linearen Elemente aus.

Satz 8.7.1 $\{\mathcal{T}_h\}$ sei eine formreguläre, uniforme Folge von Triangulationen. V_h sei der in (8.3.8) definierte Raum der linearen Elemente mit der üblichen Knotenbasis (vgl. (8.3.9a)). Dann gibt es eine von h unabhängige Konstante C_P , sodass

$$\frac{1}{C_P} \|\mathbf{u}\|_P \leq \|\mathbf{u}\|_h \leq C_P \|\mathbf{u}\|_P. \quad (8.7.4)$$

Konkrete Schranken in der Ungleichung (8.7.4) werden von Wathen [87] für verschiedene Elemente angegeben. Der Beweis des Satzes verwendet als Grundlage das folgende Lemma.

Lemma 8.7.2 $T = \{(\xi, \eta) : \xi, \eta \geq 0, \xi + \eta \leq 1\}$ sei das Einheitsdreieck (vgl. Abbildung 8.3.4). Ist u linear auf T , so gilt

$$\frac{1}{24} [u(0,0)^2 + u(1,0)^2 + u(0,1)^2] \leq \iint_T u(\xi, \eta)^2 d\xi d\eta \leq \frac{1}{6} [u(0,0)^2 + u(1,0)^2 + u(0,1)^2]. \quad (8.7.5)$$

Beweis. $u(\xi, \eta) = u_1 + (u_2 - u_1)\xi + (u_3 - u_1)\eta$ mit den Knotenwerten $u_1 := u(0, 0)$, $u_2 := u(1, 0)$, $u_3 := u(0, 1)$ führt auf

$$\iint_T u(\xi, \eta)^2 d\xi d\eta = \frac{1}{24} [u_1 \ u_2 \ u_3] \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}.$$

Die Eigenwerte der symmetrischen 3×3 -Matrix sind $\lambda_1 = \lambda_2 = 1$ und $\lambda_3 = 4$. Daher ist die rechte Seite $\geq \sum_{i=1}^3 u_i^2/24$ und $\leq 4 \sum u_i^2/24$. ■

Beweis des Satzes 8.7.1. Man stellt $\|\mathbf{u}\|_P^2$ als $\int_\Omega (P\mathbf{u})^2 dx = \sum_{T_i \in \mathcal{T}_h} \int_{T_i} (P\mathbf{u})^2 dx$ dar. Sei $\Phi_i : T \rightarrow T_i$ die lineare Transformation (aus Übungsaufgabe 8.3.14a) von T auf T_i . u_1, u_2, u_3 seien die Werte von $u = P\mathbf{u}$ in den Ecken von T_i . Die Ungleichung $0 < C_1 h^2 \leq |\det \Phi'_i| \leq C_2 h^2$ und Lemma 8.7.2 zeigen

$$\frac{C_1 h^2}{24} (u_1^2 + u_2^2 + u_3^2) \leq \int_{T_i} (P\mathbf{u})^2 dx = |\det \Phi'_i| \iint_T [(P\mathbf{u})(\Phi_i(\xi, \eta))]^2 d\xi d\eta \leq \frac{C_2 h^2}{6} (u_1^2 + u_2^2 + u_3^2). \quad (8.7.6)$$

Nach Definition einer formregulären Triangulation sind alle Dreieckswinkel $\geq \alpha_0 > 0$. Sei $\mathbb{N} \ni M \geq 2\pi/\alpha_0$. Jeder Knoten gehört deshalb zu mindestens einem und höchstens M Dreiecken. Da u_1, u_2, u_3 Koeffizienten des Vektors \mathbf{u} sind, liefert die Summation von Gleichung (8.7.6) über alle $T_i \in \mathcal{T}_h$

$$\frac{C_1}{24} \|\mathbf{u}\|_h^2 \leq \|\mathbf{u}\|_P^2 = \sum_{T_i \in \mathcal{T}_h} \int_{T_i} (P\mathbf{u})^2 dx \leq \frac{MC_2}{6} \|\mathbf{u}\|_h^2,$$

sodass Ungleichung (8.7.4) mit $C_P := \sqrt{\max(MC_2/6, 24/C_1)}$ erfüllt ist. ■

Übungsaufgabe 8.7.3 $V_h \subset H_0^1(\Omega)$ basiere auf der Quadratgittertriangulation (vgl. Übungsaufgabe 8.3.13). Man zeige: $\sqrt{4/3} \|\mathbf{u}\|_P \leq \|\mathbf{u}\|_h \leq \sqrt{8} \|\mathbf{u}\|_P$.

Die Matrix

$$\mathbf{M} := P^* P, \quad \text{d.h.} \quad M_{ij} = \int_\Omega b_i(\mathbf{x}) b_j(\mathbf{x}) d\mathbf{x}, \quad (8.7.7)$$

heißt in ingenieurtechnischen Anwendungen *Massematrix*.

Bemerkung 8.7.4 a) Ungleichung (8.7.4) ist äquivalent zu

$$\|\mathbf{M}\| \leq C_P^2 h^n, \quad \|\mathbf{M}^{-1}\| \leq C_P^2 h^{-n}.$$

b) Es gelten die Ungleichungen $\|\mathbf{u}\|_P \leq \|h^{-n} \mathbf{M}\|^{1/2} \|\mathbf{u}\|_h$ und $\|\mathbf{u}\|_h \leq \|h^n \mathbf{M}^{-1}\|^{1/2} \|\mathbf{u}\|_P$.

Beweis. 1) $\|\mathbf{u}\|_P^2 = \int_\Omega (P\mathbf{u})^2 dx = (P\mathbf{u}, P\mathbf{u})_0 = \langle P^* P \mathbf{u}, \mathbf{u} \rangle = \langle \mathbf{M} \mathbf{u}, \mathbf{u} \rangle \leq \|\mathbf{M}\| \langle \mathbf{u}, \mathbf{u} \rangle = \|\mathbf{M}\| h^{-n} \|\mathbf{u}\|_h^2$ beweist die erste Ungleichung in b). Die zweite folgt aus

$$\begin{aligned} \|\mathbf{u}\|_h^2 &= h^n \langle \mathbf{u}, \mathbf{u} \rangle = h^n \langle \mathbf{M}^{1/2} \mathbf{u}, \mathbf{M}^{-1} \mathbf{M}^{1/2} \mathbf{u} \rangle \\ &\leq \|h^n \mathbf{M}^{-1}\| \langle \mathbf{M}^{1/2} \mathbf{u}, \mathbf{M}^{1/2} \mathbf{u} \rangle = \|h^n \mathbf{M}^{-1}\| \langle \mathbf{M} \mathbf{u}, \mathbf{u} \rangle = \|h^n \mathbf{M}^{-1}\| \|\mathbf{u}\|_P^2, \end{aligned}$$

da \mathbf{M} positiv definit ist.

2) $h^{-n} \|\mathbf{M}\| \leq C_P^2$ und $h^n \|\mathbf{M}^{-1}\| \leq C_P^2$ implizieren wegen b) die Ungleichung (8.7.4). Umgekehrt zeigt $\langle \mathbf{M} \mathbf{u}, \mathbf{u} \rangle^{1/2} = \|\mathbf{u}\|_P \leq C_P \|\mathbf{u}\|_h = C_P h^{n/2} \langle \mathbf{u}, \mathbf{u} \rangle^{1/2}$, dass $h^{-n} \|\mathbf{M}\| \leq C^2$, da \mathbf{M} positiv definit ist (vgl. Lemma 4.3.25). Ebenso folgt $h^n \|\mathbf{M}^{-1}\| \leq C_P^2$ aus $\|\mathbf{u}\|_h \leq C_P \|\mathbf{u}\|_P$. ■

8.7.3 Inverse Abschätzung und Kondition von \mathbf{L}

Die H^1 -Norm kann nicht durch die L^2 -Norm abgeschätzt werden, d.h. $\sup\{|u|_1 / |u|_0 : 0 \neq u \in V\} = \infty$. Für einen endlich-dimensionalen Raum V_h ist dagegen $C_h := \sup\{|u|_1 / |u|_0 : 0 \neq u \in V_h\}$ endlich. Man sagt, V_h erfülle die *inverse Abschätzung*, wenn $C_h = \mathcal{O}(h^{-1})$, d.h.

$$|u|_1 \leq C_I h^{-1} |u|_0 \quad \text{für alle } u \in V_h, \quad (8.7.8)$$

wobei C_I unabhängig von h ist, wenn $h \rightarrow 0$.

Satz 8.7.5 $\{\mathcal{T}_h\}$ sei eine Folge formregulärer, uniformer Triangulationen. Der in (8.3.8) eingeführte Raum V_h der linearen finiten Elemente erfüllt die inverse Abschätzung.

Beweis. Wie in Satz 8.7.1 erfolgt der Nachweis durch Transformation der Integrale $\int_{T_i} |D^\alpha(Pu)|^2 dx$ für $T_i \in \mathcal{T}_h$ und $|\alpha| \leq 1$ auf Integrale über dem Einheitsdreieck T . Hierbei ist zusätzlich $D = D_x$ auf $D_{\xi,\eta}$ zu transformieren (vgl. Beweis zu Lemma 8.4.2). ■

Satz 8.7.6 Sei $\Omega \subset \mathbb{R}^n$. Es gelte (8.1.2), (8.7.8) und $\|\mathbf{u}\|_P \leq C_P \|\mathbf{u}\|_h$. Dann gilt

$$\|\mathbf{L}\| \leq h^{n-2} C_S C_P^2 C_I^2. \quad (8.7.9)$$

Beweis. Die Ungleichung

$$\langle \mathbf{u}, \mathbf{L}\mathbf{v} \rangle = a(P\mathbf{u}, P\mathbf{v}) \leq C_S |P\mathbf{u}|_1 |P\mathbf{v}|_1 \leq C_S C_I^2 h^{-2} |P\mathbf{u}|_0 |P\mathbf{v}|_0 \leq C_S C_I^2 h^{-2} C_P^2 \|\mathbf{u}\|_h \|\mathbf{v}\|_h$$

multipliziert man mit h^n und setze $\mathbf{u} := \mathbf{L}\mathbf{v}$:

$$\|\mathbf{L}\mathbf{v}\|_h^2 = h^n \langle \mathbf{L}\mathbf{v}, \mathbf{L}\mathbf{v} \rangle \leq C_S C_I^2 h^{n-2} C_P^2 \|\mathbf{L}\mathbf{v}\|_h \|\mathbf{v}\|_h,$$

also $\|\mathbf{L}\mathbf{v}\|_h \leq C_S C_I^2 h^{-2} C_P^2 \|\mathbf{v}\|_h$ für alle \mathbf{v} , d.h. (8.7.9). ■

Satz 8.7.7 Sei $\Omega \subset \mathbb{R}^n$. Es gelte (8.1.17) mit $\varepsilon_h \geq \varepsilon > 0$ und $\|\cdot\|_h \leq C_P \|\cdot\|_P$. Dann gilt

$$\|\mathbf{L}^{-1}\| \leq C_P^2 h^{-n} / \varepsilon. \quad (8.7.10)$$

Beweis. Es ist $\mathbf{L}^{-1} = P^{-1} L_h^{-1} P^{*-1}$ und $\|L_h^{-1}\|_{V_h \leftarrow V'_h} \leq 1/\varepsilon_h \leq 1/\varepsilon$ (vgl. Abschnitt 8.1), sodass

$$\|\mathbf{L}^{-1}\mathbf{f}\|_h \leq C_P \|\mathbf{L}^{-1}\mathbf{f}\|_P = C_P |P\mathbf{L}^{-1}\mathbf{f}|_0 = C_P |L_h^{-1} P^{*-1}\mathbf{f}|_0 \leq C_P \|L_h^{-1} P^{*-1}\mathbf{f}\|_{V_h} \leq (C_P/\varepsilon) \|P^{*-1}\mathbf{f}\|_{V'_h}.$$

Da

$$\begin{aligned} \|P^{*-1}\mathbf{f}\|_{V'_h} &= \sup_{0 \neq v \in V_h} |(v, P^{*-1}\mathbf{f})_0| / |v|_1 = \sup_{0 \neq v \in V_h} |\langle P^{-1}v, \mathbf{f} \rangle| / |v|_1 = \sup_{\mathbf{v} \neq 0} |\langle \mathbf{v}, \mathbf{f} \rangle| / |P\mathbf{v}|_1 \\ &\leq \sup_{\mathbf{v} \neq 0} h^{-n} \|\mathbf{v}\|_h \|\mathbf{f}\|_h / |P\mathbf{v}|_0 \leq C_P \sup_{\mathbf{v} \neq 0} h^{-n} \|\mathbf{v}\|_h \|\mathbf{f}\|_h / \|\mathbf{v}\|_h = C_P h^{-n} \|\mathbf{f}\|_h, \end{aligned}$$

ergibt sich $\|\mathbf{L}^{-1}\mathbf{f}\|_h \leq C_P^2 h^{-n} \|\mathbf{f}\|_h / \varepsilon$ für alle \mathbf{f} , also (8.7.10). ■

Die Kombination der Sätze 8.7.6 und 8.7.7 führt auf

Satz 8.7.8 Es gelte (8.1.2), (8.1.17) mit $\varepsilon_h \geq \varepsilon > 0$, (8.7.4), $m = 1$ und (8.7.8). Dann gilt

$$\text{cond}(\mathbf{L}) \leq h^{-2} C_S C_P^4 C_I^2 / \varepsilon. \quad (8.7.11)$$

Die Beweisideen lassen sich im Prinzip auf den Fall $2m > 2$, d.h. für Randwertaufgaben höherer Ordnung übertragen. Die inverse Abschätzung wird zu

$$|u|_m \leq C_I h^{-m} |u|_0 \quad \text{für alle } u \in V_h. \quad (8.7.12)$$

Damit die Normäquivalenz (8.7.4) gilt, muss man bei der Definition der Norm $\|\cdot\|_h$ vorsichtig vorgehen. Wenn alle Koeffizienten u_i von \mathbf{u} Knotenwerte $(P\mathbf{u})(\mathbf{x}^i)$ darstellen (wie beispielsweise beim Splineansatz für $V_h \subset H^2(\Omega)$), lässt sich $\|\cdot\|_h$ wie in (8.7.2b) erklären. Sobald aber Koeffizienten u_i von \mathbf{u} Ableitungen $(D^\alpha P\mathbf{u})(\mathbf{x}^i)$ darstellen, ist u_i^2 in (8.7.2b) durch $(h^{|\alpha|} u_i)^2$ zu ersetzen. Beispielsweise hat \mathbf{u} für die Hermite-Funktionen aus (8.5.5) die Koeffizienten u_{1i} und u_{2i} (vgl. Übungsaufgabe 8.5.3), wobei $u_{1i} = (P\mathbf{u})(x^i)$ und $u_{2i} = \frac{d}{dx}(P\mathbf{u})(x^i)$. Die geeignete Definition von $\|\cdot\|_h$ lautet $\|u\|_h^2 := h^n \sum_i (u_{1i}^2 + h^2 u_{2i}^2)$ mit $n = 1$, da $(0,1) \subset \mathbb{R}^1$.

Übungsaufgabe 8.7.9 $V \subset U \subset V'$ sei ein Gelfand-Dreier. $V_h \subset V$ sei ein endlich-dimensionaler Unterraum. Man zeige: Die inverse Abschätzung

$$\|u\|_V \leq C_I h^{-m} \|u\|_U \quad \text{für alle } u \in V_h \quad (8.7.13a)$$

impliziert

$$\|u\|_U \leq C_I h^{-m} \|u\|_{V'}, \quad \text{und} \quad \|u\|_V \leq C_I^2 h^{-2m} \|u\|_{V'}, \quad \text{für alle } u \in V_h. \quad (8.7.13b)$$

8.7.4 Elementmatrizen

Die noch zu definierenden Elementmatrizen dienen zur Darstellung der Systemmatrix \mathbf{L} . Abgesehen davon, dass hierdurch \mathbf{L} berechnet werden kann, hat die Darstellung von \mathbf{L} über die Elementmatrizen verschiedene Vorteile im Zusammenhang mit Parallelisierungsalgorithmen oder der Gebietszerlegungsmethode.

Sei $T \in \mathcal{T}_h$ ein Element (im einfachsten Fall ein Dreieck) mit den Knotenpunkten $\mathbf{x}^{T,1}, \mathbf{x}^{T,2}, \dots, \mathbf{x}^{T,m} \in T$ (die Zahl m der Knotenpunkte darf von T abhängen). Die Integranden in $L_{ij} = a(b_j, b_i) = \int_{\Omega} \varphi_{\Omega}(b_j, b_i) d\mathbf{x} + \int_{\Gamma} \varphi_{\Gamma}(b_j, b_i) d\Gamma$ (vgl. (8.1.8a)) seien durch (7.4.10) definiert. Die Integrale lassen sich als Summen von Elementbeiträgen schreiben:

$$L_{ij} = \sum_{T \in \mathcal{T}_h} \left\{ \int_T \varphi_{\Omega}(b_j, b_i) d\mathbf{x} + \int_{\Gamma \cap \overline{T}} \varphi_{\Gamma}(b_j, b_i) d\Gamma \right\} \quad (1 \leq i, j \leq n). \quad (8.7.14)$$

Dabei tritt das Randintegral nur für die $T \in \mathcal{T}_h$ auf, die den Rand berühren.

Die (einfachen) Knotenwerte von b_i sind die Funktionswerte $b_i(\mathbf{x}^{T,\nu})$ für $1 \leq \nu \leq m$. Für kompliziertere Finite-Element-Ansätze können auch Ableitungen $D^{\alpha} b_i(\mathbf{x}^{T,\nu})$ oder Richtungsableitungen auftreten. Zur Vereinfachung der Darstellung lassen wir im Folgenden nur Funktionswerte $b_i(\mathbf{x}^{T,\nu})$ als Funktionale zu. Nach Definition der Finite-Element-Funktionen b_i gilt für den Träger $Tr(b_i) \cap T \neq \emptyset$ genau dann, wenn einer der Knotenwerte $b_i(\mathbf{x}^{T,\nu})$ von null verschieden ist. Damit bietet sich das folgende Vorgehen an.

Die Beschränkung der Finite-Element-Funktionen aus V_h auf T erzeuge den Raum V_T (z.B. bestehend aus linearen Funktionen in (8.3.8), bilinearen in (8.3.12a,b), quadratischen in (8.3.13) und dem speziellen Unterraum in (8.3.15)). Die Interpolation in den Knotenpunkten $\mathbf{x}^{T,\nu}$ sei eindeutig (notwendig: $\dim V_T = m$). Zu $T \in \mathcal{T}_h$ und $1 \leq \nu \leq m$ sei ϕ_{ν}^T die Lagrange-Basisfunktion, d.h.

$$\phi_{\nu}^T \in V_T, \quad \phi_{\nu}^T(\mathbf{x}^{T,\mu}) = \delta_{\nu,\mu} \quad (1 \leq \nu, \mu \leq m).$$

Damit gilt insbesondere für die Basisfunktion $b_i \in V_h$, dass

$$b_i|_T = \sum_{\nu=1}^m b_i(\mathbf{x}^{T,\nu}) \phi_{\nu}^T. \quad (8.7.15)$$

Zu jedem $T \in \mathcal{T}_h$ definiert man die $m \times m$ -Matrix \mathbf{E}_T mit den Koeffizienten

$$\mathbf{E}_{T,\nu\mu} = \int_T \varphi_{\Omega}(\phi_{\mu}^T, \phi_{\nu}^T) d\mathbf{x} + \int_{\Gamma \cap \overline{T}} \varphi_{\Gamma}(\phi_{\mu}^T, \phi_{\nu}^T) d\Gamma \quad (1 \leq \nu, \mu \leq m). \quad (8.7.16)$$

Zu ihrer praktischen Berechnung verwendet man die diskutierte Abbildung auf das Einheitsselement (vgl. Übungsaufgabe 8.3.14). Man beachte, dass der Teil $\int_{\Gamma \cap \overline{T}} \dots$ für Dirichlet-Randwertaufgaben entfällt und auch sonst nur für Elemente T auftritt, die Γ berühren.

Die Kombination von (8.7.15) und (8.7.16) zusammen mit der Definition (8.7.14) von \mathbf{L} liefert die Darstellung

$$\mathbf{L}_{ij} = \sum_{T \in \mathcal{T}_h} \sum_{\nu=1}^m \sum_{\mu=1}^m b_i(\mathbf{x}^{T,\nu}) b_j(\mathbf{x}^{T,\mu}) \mathbf{E}_{T,\nu\mu}. \quad (8.7.17)$$

Die dreifache Summe sieht abschreckend aus, aber die Summe über $T \in \mathcal{T}_h$ kann auf $T \subset Tr(b_i) \cap Tr(b_j)$ beschränkt werden. Die ν - und μ -Summen reduzieren sich im Allgemeinen auf einen einzigen Term, da $b_i(\mathbf{x}^{T,\nu})$ üblicherweise nur für ein ν den Wert 1 und sonst den Wert 0 annimmt.¹³

Bemerkung 8.7.10 a) Wenn die Matrix \mathbf{L} zum Zweck der Matrixvektormultiplikation $\mathbf{v} \in \mathbb{R}^n \mapsto \mathbf{L}\mathbf{v}$ benötigt wird, kann man stattdessen direkt von den Elementmatrizen Gebrauch machen.

b) Die Elementmatrizen $(\mathbf{E}_T)_{T \in \mathcal{T}_h}$ brauchen mehr Speicherplatz als \mathbf{L} . Im Fall der Quadratgittertriangulation von $\Omega = (0,1)^2$ (vgl. Abbildung 8.3.2) ist die Zahl der Dreiecke etwa $2n$. Bei stückweise linearen Elementen ist $m = 3$ die Zahl der Knoten, sodass die Elementmatrizen $18n$ Speicherplätze brauchen, während \mathbf{L} wegen der 5 Nichtnullelemente pro Zeile mit $5n$ auskommt.

c) Während man aus $(\mathbf{E}_T)_{T \in \mathcal{T}_h}$ die Matrix \mathbf{L} berechnen kann, ist es nicht möglich, aus \mathbf{L} die Elementbeiträge $(\mathbf{E}_T)_{T \in \mathcal{T}_h}$ zurückzugewinnen.

¹³Bei Konstruktionen mit hängenden Knoten (vgl. §8.8.2) gibt es Abweichungen von diesem Normalfall.

Aus verschiedenen Gründen (z.B. bei Gebietszerlegungsverfahren, siehe [42, §11.2]) kann es vorkommen, dass man das Gebiet Ω als disjunkte Vereinigung von zwei (oder mehr) Teilgebieten Ω_1 und Ω_2 (vgl. Abbildung 10.1.1 auf Seite 189) beschreiben und auf beiden Teilgebieten die Randwertaufgaben und die zugehörigen Matrizen \mathbf{L}_1 und \mathbf{L}_2 verwenden möchte, die sich formal aus der Ersetzung von \int_{Ω} in der Definition der Bilinearform durch \int_{Ω_1} und \int_{Ω_2} ergeben. Unter der stillschweigenden Voraussetzung, dass die Elemente $T \in \mathcal{T}_h$ nur jeweils einem Teilgebiet angehören, lässt sich \mathbf{L}_1 aus (8.7.17) gewinnen, indem man $\sum_{T \in \mathcal{T}_h}$ durch $\sum_{T \in \mathcal{T}_h, T \subset \Omega_1}$ ersetzt. Wegen Bemerkung 8.7.10 ist es aber nicht möglich, \mathbf{L}_1 und \mathbf{L}_2 aus \mathbf{L} zu gewinnen.

8.7.5 Positivität, Maximumprinzip

Es sei vorausgesetzt, dass das Maximumprinzip für die (kontinuierliche) Differentialgleichung gelte. Bei der Konstruktion der Differenzenverfahren wird häufig versucht, die Vorzeichenbedingungen (4.3.1a) einzuhalten, um über die M-Matrixeigenschaft zum einen Stabilität und zum anderen das (diskrete) Maximumprinzip zu sichern. Bisher wurden die Vorzeichen des Finite-Element-Matrixelements \mathbf{L}_{ij} noch nicht diskutiert. Hierzu gibt es zunächst eine negative Aussage: Das kontinuierliche Maximumprinzip impliziert nicht automatisch das diskrete Maximumprinzip. Insbesondere können Außerdiagonalkoeffizienten \mathbf{L}_{ij} das “falsche” Vorzeichen besitzen.

Für die Finite-Element-Diskretisierung von $L = -\Delta$ mit stückweise linearen Elementen lässt sich die folgende Darstellung finden: Zu den Indizes i, j mögen zwei verbundene Dreieckspunkte \mathbf{x}^i und \mathbf{x}^j gehören. Damit besteht der Schnitt der Träger der Basisfunktionen b_i und b_j aus den beiden Dreiecken T' und T'' aus Abbildung 8.7.1. Seien α' und α'' die Winkel in T' und T'' , die der Seite $\overline{\mathbf{x}^i \mathbf{x}^j}$ jeweils gegenüberliegen. Dann gilt

$$\mathbf{L}_{ij} = \int_{T' \cup T''} \langle \nabla b_i, \nabla b_j \rangle d\mathbf{x} = -\frac{1}{2} \frac{\sin(\alpha' + \alpha'')}{\sin(\alpha') \sin(\alpha'')},$$

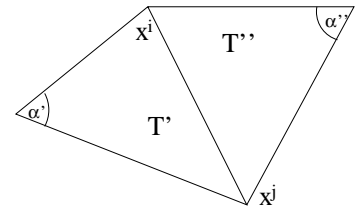


Abb. 8.7.1. Winkel α und α'

wie in Knabner-Angermann [52, §3.9] bewiesen wird. Damit liegt das gewünschte Vorzeichen $\mathbf{L}_{ij} \leq 0$ genau dann vor, wenn $\alpha' + \alpha'' \leq \pi$. Für Dreiecke, die den Rand schneiden, werden in [52, Seite 161] noch zusätzliche Bedingungen genannt.

8.8 Weitere Hinweise

Auf viele Details und Modifikationen der Finite-Element-Methode kann hier nicht eingegangen werden. Im Folgenden seien einzelne Punkte stichwortartig erwähnt.

8.8.1 Nichtkonforme Elemente

Die Bedingung (8.1.3), $V_h \subset V$, charakterisiert die konformen Finite-Element-Methoden. Diskretisierungen, die auf $V_h \not\subset V$ beruhen, nennt man *nichtkonform*. Ein Beispiel für ein nichtkonformes Element ist das “Wilsonsche Rechteck”. Dabei wird eine Zerlegung in Rechtecke $R_i = (x_{i1}, x_{i2}) \times (y_{1i}, y_{2i})$ zu Grunde gelegt und die Funktion auf R_i quadratisch angesetzt:

$$V_h := \{u \text{ quadratisch auf jedem } R_i, u \text{ stetig in allen Ecken der Rechtecke } R_i\}.$$

Man beachte, dass $u \in V_h$ nur in den Knotenpunkten (Ecken der R_i), aber nicht in Ω stetig zu sein braucht. Eine Funktion, die entlang einer Seite von R_i unstetig ist, kann nicht zu $H^1(\Omega)$ gehören: $V_h \not\subset V = H^1(\Omega)$. Eine mögliche Basis für V_h ist die folgende. Zu jedem der 4 Knoten \mathbf{x}^j sei b_j die zu den bilinearen Elementen gehörende Basisfunktion (z.B. $b_i^{(1)}(x, y) := \frac{(x_{2i}-x)(y_{2i}-y)}{(x_{2i}-x_{1i})(y_{2i}-y_{1i})}$, vgl. §8.3.3). Der hier gewählte quadratische Ansatz enthält zwei zusätzliche Freiheitsgrade. Deshalb wählt man zu jedem Rechteck $R_i = (x_{1i}, x_{2i}) \times (y_{1i}, y_{2i})$ die zwei weiteren Basisfunktionen

$$b_i^{(5)}(x, y) := (x - x_{1i})(x_{2i} - x), \quad b_i^{(6)}(x, y) := (y - y_{1i})(y_{2i} - y),$$

die in allen vier Eckpunkten von R_i verschwinden, sodass dort die Stetigkeit gesichert ist, wenn $b_i^{(5)}$ und $b_i^{(6)}$ außerhalb von R_i durch null fortgesetzt werden.

Das erste Problem, das sich für nichtkonforme Elemente stellt, ist die Definition der Systemmatrix \mathbf{L} . Eine Erklärung durch $L_{ij} = a(b_j, b_i)$ wie in (8.1.8a) ist wegen $b_j, b_i \notin V$ zunächst nicht definiert! Man muss daher neben der Diskretisierung von V durch V_h auch $a(\cdot, \cdot)$ durch eine V_h -abhängige Bilinearform $a_h(\cdot, \cdot) : V_h \times V_h \rightarrow \mathbb{R}$ ersetzen und $L_{ij} := a_h(b_j, b_i)$ definieren. Im Falle der Zerlegung von Ω in Rechtecke R_i wird $a_h(\cdot, \cdot)$ durch

$$a_h(u, v) := \sum_{|\alpha| \leq m} \sum_{|\beta| \leq m} \sum_i \int_{R_i} a_{\alpha\beta}(x) [D^\alpha u(x)] [D^\beta v(x)] dx$$

erklärt, wenn $a(\cdot, \cdot)$ durch (7.2.6) gegeben ist, d.h. die Integration \int_Ω wird durch $\sum_i \int_{R_i}$ ersetzt. Man beachte $a(u, v) = a_h(u, v)$ für $u, v \in V$.

Konforme Finite-Element-Verfahren ergeben stets konsistente Diskretisierungen. Bei nichtkonformen Verfahren ist Konsistenz nicht in jedem Fall gewährleistet. Zu diesem Problem, insbesondere zu dem sogenannten "Patchtest" vergleiche man Strang-Fix [78, p. 174], Ciarlet [22, §4.2], Stummel [80], Gladwell-Wait [33, pp. 83-92], Thomasset [81]. Dass sinnvoll erscheinende Ansätze für V_h völlig falsche Diskretisierungen erzeugen können, zeigt

Übungsaufgabe 8.8.1 \mathcal{T} sei eine zulässige Triangulation. Es sei $V_h := \{u \text{ konstant auf jedem } T \in \mathcal{T}\}$. Man zeige: a) $V_h \not\subset V := H^1(\Omega)$.

b) $\inf\{|u - v|_0 : v \in V_h\} \leq Ch|u|_1$ für alle $u \in V$ (eine Art von Konsistenz, die hier aber nicht hilft).

c) Seien $a(u, v) := \int_\Omega [\langle \nabla u, \nabla v \rangle + uv] dx$ und $a_h(u, v) := \sum_{T \in \mathcal{T}} \int_T [\langle \nabla u, \nabla v \rangle + uv] dx$. Was ergibt sich für $L_{ij} = a_h(b_j, b_i)$?

Nichtkonforme Elemente sind insbesondere für Gleichungen höherer Ordnung von Interesse. Konforme finite Elemente aus $V = H^2(\Omega)$ müssen stetig differenzierbar aneinanderstossen, was sie recht kompliziert macht. Vereinfachungen führen notwendigerweise zu $V_h \not\subset V$.

Genaugenommen gehören auch die isoparametrischen finiten Elemente aus Abschnitt 8.5.3 im Falle von $V = H_0^1(\Omega)$ zu den nichtkonformen Methoden, da die Elemente aus V_h auf Γ zwar null approximieren, aber im Allgemeinen nicht verschwinden, sodass $V_h \not\subset V = H_0^1(\Omega)$. Allerdings gilt $V_h \subset H^1(\Omega)$, sodass die Bilinearform $a(\cdot, \cdot)$ nicht in $a_h(\cdot, \cdot)$ verändert werden muss.

Für die folgende Fehlerabschätzung nichtkonformer Galerkin-Verfahren muss a_h auf $Z_h \times Z_h$ stetig definiert sein, wobei $Z_h := V + V_h = \{v + v_h : v \in V, v_h \in V_h\}$.

Satz 8.8.2 (zweites Lemma von Strang) Die Bilinearform a_h sei stetig, symmetrisch und positiv definit auf $Z_h \times Z_h$, sodass $\|z\|_h := \sqrt{a_h(z, z)}$ eine Norm auf Z_h definiert. Seien $u^h \in V_h$ und $u \in V$ die Lösungen von $a_h(u^h, v) = f_h(v)$ ($v \in V_h$) bzw. (8.1.1). Dann gilt

$$\|u - u^h\|_h \leq \text{const} \left\{ \|f_h - a_h(u, \cdot)\|_{V'_N} + \inf_{z^h \in V_h} \|u - z^h\|_h \right\}.$$

Beweis. Für alle $z^h, v^h \in V_h$ gilt

$$a_h(u^h - z^h, v^h) = a_h(u - z^h, v^h) + a_h(u^h, v^h) - a_h(u, v^h) = a_h(u - z^h, v^h) + f_h(v^h) - a_h(u, v^h).$$

Mit $v^h := u^h - z^h$ erhält man

$$\begin{aligned} \|u^h - z^h\|_h^2 &= a_h(u^h - z^h, u^h - z^h) = a_h(u - z^h, u^h - z^h) + f_h(u^h - z^h) - a_h(u, u^h - z^h) \\ &\leq C_h \|u - z^h\|_h \|u^h - z^h\|_h + \|f_h - a_h(u, \cdot)\|_{V'_N} \|u^h - z^h\|_h, \end{aligned}$$

wobei C_h die Schranke von a_h ist. Nach Kürzen folgt $\|u^h - z^h\|_h \leq C_h \|u - z^h\|_h + \|f_h - a_h(u, \cdot)\|_{V'_N}$. In

$$\|u - u^h\|_h \leq \|u - z^h\|_h + \|z^h - u^h\|_h \leq (C_h + 1) \|u - z^h\|_h + \|f_h - a_h(u, \cdot)\|_{V'_N}$$

werden die Dreiecksungleichung und die vorherige Abschätzung verwendet, was die Behauptung beweist. ■

8.8.2 Nichtzulässige Triangulationen

Triangulationen mit unzulässigen Dreiecken seien hier an Hand des Beispiels von Abbildung 8.8.1 diskutiert. Die gezeigte Triangulation entsteht, indem man in der regelmäßigen und zulässigen Triangulation die mittleren Dreiecke in jeweils vier kleinere zerlegt. Die Zulässigkeitsbedingungen (8.3.7a-d) werden zum Beispiel verletzt, da $\overline{T_0} \cap \overline{T_1}$ zwar eine Seite von T_1 , aber nicht von T_0 ist. Eine andere Verletzung der Zulässigkeit betrifft den Punkt $P = \overline{T_0} \cap \overline{T_2}$, der Eckpunkt von T_2 , aber nicht von T_0 ist. Die Punkte P, Q, R, S , die die Zulässigkeitsbedingungen verletzen, nennt man "hängende Knoten". Die anderen Knoten \mathbf{x}^i sind in der Abbildung durch $i = 1, \dots, 17$ bezeichnet.

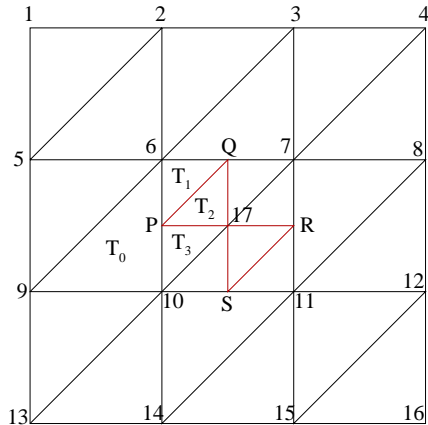


Abb. 8.8.1. Unzulässige Triangulation

Wie bisher lässt sich der Finite-Element-Raum durch (8.3.8) definieren (hier ohne Nullrandbedingung):

$$V_h = \{u \in C^0(\overline{\Omega}) : u|_T \text{ linear für alle } T \in \mathcal{T}_h\}.$$

Der Unterschied zu einer zulässigen Triangulation wird bei der Frage nach der Dimension $\dim V_h$ deutlich. Da insbesondere $u|_{T_0}$ linear ist, ergibt sich der Knotenwert $u(P)$ eindeutig aus $u(\mathbf{x}^6)$ und $u(\mathbf{x}^{10})$:

$$u(P) = \frac{1}{2} (u(\mathbf{x}^6) + u(\mathbf{x}^{10})). \quad (8.8.1)$$

Anders als in Bemerkung 8.3.10 kann man somit nicht beliebige Werte von u in allen Knoten der Triangulation vorgeben, sondern nur in den regulären, d.h. nicht-hängenden Knoten $\mathbf{x}^1, \dots, \mathbf{x}^{17}$. Dementsprechend ist die Basis von V_h durch $\{b_1, \dots, b_{17}\}$ gegeben, wobei $b_i(\mathbf{x}^j) = \delta_{ij}$ ($1 \leq i, j \leq 17$). Bei der Berechnung der Systemmatrix aus den Elementmatrizen tritt hier der Fall auf, dass die Gewichte

$b_i(\mathbf{x}^{T,\nu})$ in (8.7.17) einen Wert verschieden von 0 und 1 annehmen.

Bemerkung 8.8.3 a) Auch wenn die Triangulation \mathcal{T}_h unzulässig ist, lässt sich der (konforme) Finite-Element-Raum V_h konstruieren. Bei der Definition der Basisfunktionen hat man zwischen regulären und hängenden Knoten zu unterscheiden.

b) Für Finite-Element-Räume V_h , bei denen keine Stetigkeit zwischen den Elementen verlangt wird¹⁴ (zum Beispiel bei stückweise konstanten Funktionen), braucht man nicht zwischen regulären und hängenden Knoten zu unterscheiden, d.h. in diesem Fall kann die Zulässigkeitsbedingung entfallen.

Bei zulässigen Triangulationen lassen sich einfache Aussagen über die Träger der Basisfunktionen machen. Bemerkung 8.3.11 besagt, dass alle am Träger von b_i beteiligten Dreiecke am Knotenpunkt \mathbf{x}^i hängen. Dies gilt nicht in der Gegenwart hängender Knoten. Abbildung 8.8.2 zeigt einen Ausschnitt der Triangulation, wobei P, Q, R hängende Knoten sind. Die Basisfunktion b_1 im regulären Knoten \mathbf{x}^1 hat in P auf Grund von (8.8.1) den Wert $b_1(P) = 1/2$. Dies induziert in Q den Wert $b_1(Q) = 1/4$ und weiter in R den Wert $b_1(R) = 1/8$. Damit gehören die Dreiecke T_1, T_2, T_3 zum Träger von b_1 , obwohl diese Dreiecke nur indirekt mit \mathbf{x}^1 verbunden sind. Den hiermit verbundenen Komplikationen bei der Implementierung des Finite-Element-Verfahrens entgeht man mit der Zulässigkeitsvoraussetzung. Gelegentlich wird ein Kompromiss gemacht: Man erlaubt nur hängende Knoten ersten Grades (d.h. das Dreieck T_1 darf keinen weiteren hängenden Knoten Q enthalten, der ein hängender Knoten zweiten Grades wäre).

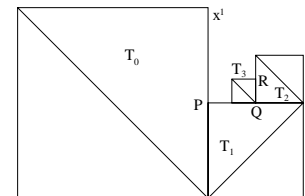


Abb. 8.8.2.

8.8.3 Trefftz-Verfahren

In Satz 6.5.12 wurde festgestellt, dass für symmetrische und V -elliptische Bilinearformen die schwache Formulierung (8.1.1) des Randwertproblems zum Minimierungsproblem

$$\text{suche } u \in V \quad \text{mit} \quad J(u) := \min\{J(v) : v \in V\} \quad (8.8.2)$$

¹⁴Da stückweise konstante Elemente nicht in $H^1(\Omega)$ liegen, treten sie nicht für die hier behandelten Differentialgleichungen auf. Verwendung finden sie dagegen bei Finite-Element-Approximationen von Integralgleichungen.

äquivalent ist (vgl. auch Satz 7.2.9, Übungsaufgabe 7.3.8). Das Maximierungsproblem

$$\text{suche } w \in W \quad \text{mit} \quad K(w) := \max\{K(\hat{w}) : \hat{w} \in W\} \quad (8.8.3)$$

nennt man ein zu (8.8.2) *duales* (oder *komplementäres*) *Variationsproblem*, falls beide die gleiche Lösung $u = w \in V \cap W$ besitzen. Beispielsweise führt das Poisson-Problem $-\Delta u = f \in L^2(\Omega)$ in Ω , $u = 0$ auf Γ auf $J(v) = \int_{\Omega} |\nabla v|^2 dx$ für $v \in V := H_0^1(\Omega)$. Ein spezielles duales Variationsproblem stammt von Trefftz (1926):

$$K(w) := - \int_{\Omega} |\nabla w|^2 dx \quad \text{für } w \in W := \{v \in H^1(\Omega) : -\Delta v = f\}$$

(vgl. Velte [84, p. 91]). Man beachte, dass die Funktionen $v \in V$ die Randbedingung $v = 0$ auf Γ erfüllen, während $w \in W$ keine Randbedingung, aber dafür die Differentialgleichung $Lw = f$ erfüllen muss.

8.8.4 Finite-Element-Verfahren für singuläre Lösungen

Die bisherigen Fehlerabschätzung wie zum Beispiel (8.4.9) beruhen auf der Annahme $u \in H^2(\Omega) \cap V$, die in Abschnitt 9.1 näher diskutiert werden wird. Dass diese Voraussetzung nicht stets zu gelten braucht, zeigt

Beispiel 8.8.4 a) Die Laplace-Gleichung $\Delta u = 0$ im L-Gebiet Ω aus Beispiel 2.1.4 hat die Lösung¹⁵ $u = r^{2/3} \sin((2\varphi - \pi)/3)$. Diese Funktion gehört nur zu $H^s(\Omega)$ mit $s < 1 + 2/3$.

b) Die Lösung $u = r^{1/2} \sin(\varphi/2)$ aus Beispiel 5.2.3 gehört nur zu $H^s(\Omega)$ mit $s < 3/2$.

Ersetzt man zum Beispiel im L-Gebiet Ω die Laplace-Gleichung durch die allgemeinere Poisson-Gleichung $\Delta u = f$ in Ω , $u = 0$ auf Γ , so lässt sich noch zeigen, dass sich für $f \in L^2(\Omega)$ die Lösung u zerlegen lässt in

$$u = u_0 + \alpha \chi(r) r^{2/3} \sin((2\varphi - \pi)/3) \quad \text{mit } u_0 \in H^2(\Omega) \cap H_0^1(\Omega) \text{ und } \alpha \in \mathbb{R}$$

(vgl. Strang-Fix [78, p. 257ff]). Dabei ist $\chi(r) \in C^\infty(\Omega)$ eine vorgebbare “Abschneidefunktion” mit $\chi(r) = 1$ in $0 \leq r \leq 1/4$, $\chi(r) = 0$ für $r \geq 1/2$, sodass $\chi(r) r^{2/3} \sin((2\varphi - \pi)/3) \in H_0^1(\Omega)$ gewährleistet ist.

Eine übliche Diskretisierung durch lineare Dreieckselemente erreicht nur $|u - u^h|_1 = \mathcal{O}(h^s)$, $s < 2/3$. Nimmt man jedoch die Funktion $\chi(r) r^{2/3} \sin(\dots)$ oder ähnliche Funktionen noch zu V_h hinzu, lässt sich die Approximation von u in V_h verbessern: $d(u, V_h) = \mathcal{O}(h)$, sodass $|u - u^h|_1 = \mathcal{O}(h)$. Zu diesem Themenkomplex sei auf Babuška-Rosenzweig [6] und Blum-Dobrowolski [10] verwiesen. Man vergleiche auch Gladwell-Wait [33, p. 119], Hackbusch [40].

Ein anderer Zugang, der traditionelle finite Elemente verwendet und ohne explizite Kenntnisse der Singularität auskommt, ist der folgende. \mathcal{T} sei eine Triangulation, deren Dreiecke sich dort verfeinern, wo die Singularität der Lösung erwartet wird, z.B. an der einspringenden Ecke (vgl. Abbildungen 2.1.1 und 5.2.1). Zur Fehleranalyse vergleiche man Schatz-Wahlbin [74].

8.8.5 Hierarchische Basen

Die zugrundeliegende Idee sei am eindimensionalen Fall vorgestellt. $V_h \subset H_0^1(0,1)$ sei wie in (8.3.2) der Raum der stückweise linearen Funktionen bei äquidistanter Unterteilung des Intervalles $\Omega = (0,1)$ in Teile der Länge h . Für $h = 1/2$ ist $\dim(V_{1/2}) = 1$ und b_1 aus Abbildung 8.8.3a ist einzige Basisfunktion. Offenbar gilt $V_{1/4} \subset V_{1/2}$. Als Basisfunktion in $V_{1/4}$ übernimmt man $b_1 \in V_{1/2}$ und fügt die Funktionen b_2, b_3 aus Abbildung 8.8.3b hinzu. $\{b_1, b_2, b_3\}$ ist Basis des $V_{1/4}$. Ebenso gilt $V_{1/8} \subset V_{1/4}$. b_1, b_2, b_3 ergänzt durch b_4, \dots, b_7 (vgl. Abbildung 8.8.3c) ergeben eine Basis des $V_{1/8}$, die sogenannte *hierarchische Basis*. Im Gegensatz zu den bisher benutzten Basisfunktionen sind die Träger der b_i nicht notwendig von der Länge $\mathcal{O}(h)$, zum Beispiel ist $\text{Tr}(b_1) = [0,1]$. Hieraus folgt, dass die Systemmatrix \mathbf{L} stärker aufgefüllt ist als sonst üblich. Trotzdem hat diese Basiswahl wichtige Vorteile:

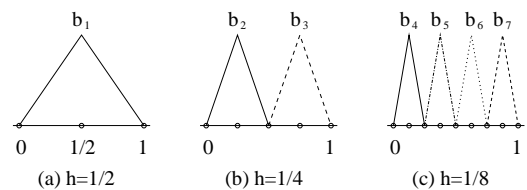


Abb. 8.8.3. Hierarchische Basis

¹⁵Ein Gebiet mit Innenwinkel $\omega \in (0, 2\pi)$ führt zu einer Lösung mit dem Faktor $r^{\pi/\omega}$. Sie gehört zu $H^s(\Omega)$ für alle $s < 1 + \pi/\omega$.

(i) Eine zuvor berechnete Matrix $\mathbf{L} = \mathbf{L}^{(2h)}$ zu V_{2h} bildet eine Hauptuntermatrix der Systemmatrix $\mathbf{L} = \mathbf{L}^{(h)}$ zu V_h , sodass eine (globale oder lokale) Verfeinerung der Zerlegung nur minimale zusätzliche Rechnungen erfordert.

(ii) Die Kondition $\text{cond}(\mathbf{L}) = \|\mathbf{L}\| \|\mathbf{L}^{-1}\|$ der Systemmatrix ist wesentlich besser als im Standardfall (z.B. $\text{cond}(\mathbf{L}) = \mathcal{O}(\log h)$ statt $\mathcal{O}(h^{-2})$ im zweidimensionalen Fall $\Omega \subset \mathbb{R}^2$; vgl. Abschnitt 8.7).

(iii) Zur Auflösung des Gleichungssystems $\mathbf{L}\mathbf{u} = \mathbf{f}$ stehen geeignete Verfahren bereit. Näheres hierzu findet man bei Yserentant [92], Bank-Dupont-Yserentant [8] und Hackbusch [42].

8.8.6 Superkonvergenz

Bei Differenzenverfahren wurde der Fehler durch die Gitterfunktion $u_h - R_h u$ (u_h : Differenzenlösung, u : exakte Lösung) beschrieben, während bei der Finite-Element-Methode die Lösung u^h auf dem gesamten Gebiet Ω definiert ist und deshalb $u^h - u$ als Fehler verwendet wird.

Eine Gitterfunktion u_h könnte man im Prinzip zu einer auf Ω definierten Funktion u^h fortsetzen, indem man die Werte $u_h(\mathbf{x}^i)$ (\mathbf{x}^i : Gitterpunkte von Ω_h) als Knotenwerte zum Beispiel der stückweise linearen oder bilinearen finiten Elemente wählt und somit u_h durch $u^h := Pu_h$ interpoliert (zu P vergleiche man (8.1.6)). Zu den Fehlern $u_h - R_h u$ in den Gitterpunkten kämen dann die Interpolationsfehler $u - PR_h u$ hinzu:

$$u^h - u = Pu_h - u = P(u_h - R_h u) - (u - PR_h u).$$

Sobald die Interpolationsfehler mindestens die gleiche Größenordnung haben wie $u_h - R_h u$, erhält man für $Pu_h - u$ die gleiche Konvergenzaussage, andernfalls zeigt $Pu_h - u$ schlechtere Konvergenz als $u_h - R_h u$.

Umgekehrt kann man bei Finite-Element-Verfahren danach fragen, welche Fehlerabschätzungen man für die Knotenwerte $u_h = \{u^h(\mathbf{x}^i) : \mathbf{x}^i \text{ Knoten}\}$ oder geeignet gemittelte Werte bekommt. Erhält man bessere Konvergenz als für $u^h - u$, spricht man von *Superkonvergenz*. Dieser Begriff kann sich auch auf Differenzenquotienten von Knotenwerten beziehen, wenn diese genauer sind als $|D^\alpha u^h - D^\alpha u|_0 \leq |u^h - u|_1$ ($|\alpha| = 1$). Man vergleiche hierzu Wahlbin [86] und z.B. Lesaint-Zlámal [57], Bramble-Schatz [17], Thomée [83], Louis [59], Großmann-Roos [36, §4.9.1]. Superkonvergenzresultate für Gradienten bilden u.a. die Basis für Fehler-schätzer (vgl. Verfürth [85, Seiten 36ff]).

Gelegentlich wird auch die Fehleraussage (8.5.8b) für negatives t , die um den Faktor h^{-t} besser als die Aussage für $t = 0$ ist, als Superkonvergenz bezeichnet.

8.8.7 Die Mörtelmethode (“mortar finite elements”)

8.8.7.1 Einleitung

Bisher wurde stets vorausgesetzt, dass das gesamte Gebiet Ω einheitlich von einem zusammenhängenden Finite-Elemente-Gitter überdeckt wird. Der Zusammenhang ist zum Beispiel durch die Bedingung der Zulässigkeit der Triangulation gegeben. Selbst wenn hängende Knoten auftreten, gibt es zum Beispiel noch gemeinsame Eckpunkte. Abbildung 8.8.4 zeigt ein Gebiet Ω , das in zwei Teilgebiete Ω_1 und Ω_2 zerlegt ist. Der gemeinsame interne Rand (“interface”) sei mit $\Gamma_{12} := \overline{\Omega_1} \cap \overline{\Omega_2}$ bezeichnet. In beiden Teilgebieten werden Triangulationen \mathcal{T}_1 und \mathcal{T}_2 unabhängig voneinander konstruiert. Da \mathcal{T}_i zulässige Triangulation von Ω_i sei soll, folgt aus (8.3.7c), dass $\bigcup_{T \in \mathcal{T}_i} \overline{T} = \overline{\Omega_i}$ (hierzu muss angenommen werden, dass Ω_i ein Polygon ist). Damit muss Γ_{12} in der Vereinigung der Dreiecksseiten von $T \in \mathcal{T}_i$ enthalten sein. Da dies für $i = 1$ und $i = 2$ gilt, besteht zwischen \mathcal{T}_1 und \mathcal{T}_2 der Zusammenhang, dass beide Γ_{12} mit ihren Dreiecksseiten überdecken. Dagegen gibt es keinen Zusammenhang zwischen den auf Γ_{12} liegenden Eckpunkten beider Triangulationen.

Gründe, auf eine durchgehend zulässige Triangulation zu verzichten, könnten sein:

- Ω_1 und Ω_2 werden unabhängig von einander modelliert und diskretisiert und erst in einem späteren Konstruktionsschritt zusammengesetzt.
- Ω_1 und Ω_2 sind Teilgebiete, deren Daten auf einem Parallelrechner unterschiedlichen Prozessoren zugewiesen sind. Mögliche Gitterverfeinerungen würden Kommunikation erfordern, wenn ein global zulässiges Gitter resultieren soll. Der Verzicht auf diese Bedingung erspart die Kommunikation (an dieser Stelle).

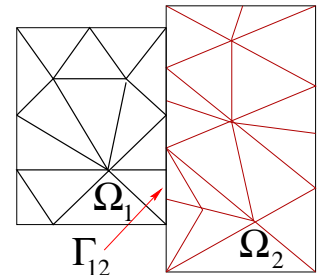


Abb. 8.8.4. Unabhängige Triangulationen in zwei Teilgebieten

• Bei instationären Problemen kann sich das Gebiet Ω zeitlich verändern. Beispielsweise könnte sich das Teilgebiet Ω_2 vertikal bewegen, während Ω_1 fixiert ist. Ein global zulässiges Gitter müsste dann in jedem Zeitschritt neu konstruiert und die Daten transferiert werden. Diese Schwierigkeit entfällt, wenn die Zulässigkeitsbedingung nur lokal für Ω_1 und Ω_2 gefordert wird.

Um die Notation nicht zunötig aufzublähen, sei die Poisson-Gleichung $-\Delta u = f$ in $\Omega \subset \mathbb{R}^2$ mit der Dirichlet-Randbedingung $u = 0$ auf $\partial\Omega$ angenommen. Auf kontinuierlicher Stufe genügen die Übergangsbedingungen $u^{(1)}|_{\Gamma_{12}} = u^{(2)}|_{\Gamma_{12}}$ und $\partial u^{(1)}/\partial n|_{\Gamma_{12}} = \partial u^{(2)}/\partial n|_{\Gamma_{12}}$, um zwei Lösungen $u^{(i)}$ von $-\Delta u^{(i)} = f|_{\Omega_i}$, $u^{(i)}|_{\partial\Omega_i \setminus \Gamma_{12}} = 0$ so zu verbinden, dass u mit $u|_{\Omega_i} = u^{(i)}$ ($i = 1, 2$) die Lösung des globalen Problems ist.

Auf der Finite-Element-Ebene werden keine direkten Übergangsbedingungen an die Dirichlet- und Neumann-Randwerte formuliert. Statt dessen wird die Verbindung durch eine weitere Variationsformulierung hergestellt, die der Methode den Namen gegeben hat. Die Variationsformulierung wird zunächst für das kontinuierliche Problem definiert.

8.8.7.2 Variationsformulierung im kontinuierlichen Fall

Seien $X := \{v \in L^2(\Omega) : v|_{\Omega_i} \in H^1(\Omega_i) \text{ für } i = 1, 2 \text{ mit } v|_{\partial\Omega} = 0\}$ und $M = (H^{1/2}(\Gamma_{12}))'$. Man beachte, dass Funktionen $v \in X$ zwei Arten von Randwerten auf Γ_{12} besitzen: Da die Beschränkung $v^{(1)} := v|_{\Omega_1}$ zu $H^1(\Omega_1)$ gehört, definiert sie Randwerte $v^{(1)}|_{\Gamma_{12}} \in H^{1/2}(\Gamma_{12})$. Die Randwerte $v^{(2)}|_{\Gamma_{12}} \in H^{1/2}(\Gamma_{12})$ von $v^{(2)} := v|_{\Omega_2}$ werden im Allgemeinen von $v^{(1)}|_{\Gamma_{12}}$ verschieden sein. Der Sprung

$$[v] := v^{(1)}|_{\Gamma_{12}} - v^{(2)}|_{\Gamma_{12}} \in H^{1/2}(\Gamma_{12})$$

wird in der zweiten der folgenden Bilinearformen verwendet:

$$\begin{aligned} a(u, v) &:= \sum_{k=1}^2 \int_{\Omega_k} \langle \nabla u(x), \nabla v(x) \rangle \, dx, & \text{für } u, v \in X, \\ b(v, \mu) &:= (\mu, [v])_{L^2(\Gamma_{12})} = \langle \mu, [v] \rangle_{(H^{1/2}(\Gamma_{12}))' \times H^{1/2}(\Gamma_{12})}. & \text{für } v \in X \text{ und } \mu \in M. \end{aligned}$$

Wir suchen nun das Paar $(u, \lambda) \in X \times M$, sodass

$$\begin{aligned} a(u, v) + b(v, \lambda) &= (f, v)_{L^2(\Omega)} & \text{für alle } v \in X, \\ b(u, \mu) &= 0 & \text{für alle } \mu \in M, \end{aligned} \tag{8.8.4}$$

Nehmen wir $u \in H^2(\Omega) \cap H_0^1(\Omega)$ an, so liefert der erste Teil von (8.8.4) nach partieller Integration

$$-\int_{\Omega} v \Delta u \, dx + \int_{\Gamma_{12}} \left(v^{(1)} \frac{\partial u^{(1)}}{\partial n} - v^{(2)} \frac{\partial u^{(2)}}{\partial n} + [v] \lambda \right) d\Gamma = \int_{\Omega} f v \, dx$$

($u^{(i)}$ sind wie oben als $u|_{\Omega_i}$ definiert; die Normalenrichtung \vec{n} ist bezüglich Ω_1 festgelegt). Hieraus liest man $-\Delta u = f$ in Ω ab. Das Γ_{12} -Integral liefert

$$\begin{aligned} 0 &= \int_{\Gamma_{12}} \left(v^{(1)} \frac{\partial u^{(1)}}{\partial n} - v^{(2)} \frac{\partial u^{(2)}}{\partial n} + (v^{(1)} - v^{(2)}) \lambda \right) d\Gamma \\ &= \int_{\Gamma_{12}} \left\{ v^{(1)} \left(\frac{\partial u^{(1)}}{\partial n} - \frac{\partial u^{(2)}}{\partial n} \right) + (v^{(1)} - v^{(2)}) \left(\frac{\partial u^{(2)}}{\partial n} + \lambda \right) \right\} d\Gamma. \end{aligned}$$

Da $v^{(1)}, v^{(2)}$ unabhängig voneinander gewählt werden können, folgen die Gleichungen

$$\frac{\partial u^{(1)}}{\partial n} = \frac{\partial u^{(2)}}{\partial n} \quad \text{und} \quad \lambda = -\frac{\partial u^{(2)}}{\partial n} \quad \text{auf } \Gamma_{12}.$$

Der zweite Teil in (8.8.4) liefert $[u] = 0$, d.h. $u^{(1)} = u^{(2)}$ auf Γ_{12} .

8.8.7.3 Diskrete Variationsformulierung

X_h sei der Teilraum von X , der beschränkt auf Ω_i der Finite-Element-Raum der stückweise linearen Polynome auf der Triangulation \mathcal{T}_i ist.

Sei \mathcal{T}_{12} die Beschränkung der Triangulation \mathcal{T}_1 auf Γ_{12} . Das eindimensionale Gitter \mathcal{T}_{12} besteht aus allen Strecken $E \cap \Gamma_{12}$ (E : Dreieckskante von $T \in \mathcal{T}_1$) und allen Knotenpunkten von \mathcal{T}_1 , die auf Γ_{12} liegen. Der diskrete Raum M_h besteht aus allen auf \mathcal{T}_{12} stückweise linearen Funktionen, die in den Endpunkten von Γ_{12} verschwinden. (Man beachte, dass die Konstruktion nicht symmetrisch ist: \mathcal{T}_{21} als Beschränkung von \mathcal{T}_2 auf Γ_{12} liefert ein anderes Gitter auf Γ_{12}).

Die Variationsformulierung (8.8.4) mit X_h , M_h statt X , M liefert das zu lösende diskrete Problem.

8.8.7.4 Verallgemeinerung

Statt der einfachen Poisson-Gleichung können allgemeinere Differentialgleichungen mit allgemeineren Randbedingungen behandelt werden. Insbesondere können statt zwei der Teilgebiete eine Zerlegung von Ω in k Teilgebiete erlaubt werden. Der Leser sei auf Wohlmuth [90] (zweidimensionaler Fall) und Kim-Lazarov-Pasciak-Vassilevski [51] verwiesen. Zur Stabilität der Methode vergleiche man u.a. Braess-Dahmen [13] und Dahmen-Faermann-Graham-Hackbusch-Sauter [27] bewiesen.

8.8.8 Verwandte Diskretisierungen

Neben der Standard-Galerkin-Methode gibt es weitere Varianten, die hier stichwortartig erwähnt seien.

8.8.8.1 Galerkin-Verfahren für Kleinste-Quadrate-Formulierung

Die Standarddiskretisierung minimiert die Energienorm, wenn das Problem symmetrisch und V -koerziv ist. Wie bei der Kleinste-Quadrate-Minimierung kann man von dem Ansatz $Lu - f = 0$ ausgehen, den man auf den Ansatzraum $u \in V_N$ beschränkt (z.B. mit $V_N \subset H_0^1(\Omega)$). Damit stellt sich die Minimierungsaufgabe: Minimiere $\|Lu - f\|_{L^2(\Omega)}$ über alle $u \in V_N$ (vgl. [71, S. 249]).

8.8.8.2 Unstetiges Galerkin-Verfahren

Finite-Element-Funktionen aus $H^1(\Omega)$ sind stetig. Auch im Falle der nichtkonformen Elementen gibt es noch stetige Übergänge in speziellen Knoten. Ansätze mit vollständig unstetigen Ansätzen pro Dreieck erlauben eine größere Flexibilität und führen auf die mit DGFEM ("discontinuous Galerkin finite element method") bezeichnete Methode (vgl. den Konferenzband [24] und z.B. [71, S. 255]).

8.8.8.3 Petrov-Galerkin- und Finite-Volumen-Verfahren

Ein Petrov-Galerkin-Verfahren liegt vor, wenn verschiedene Unterräume für die Test- und Ansatzfunktionen verwendet werden:

$$\text{Finde } u^N \in V_N, \quad \text{sodass } a(u^N, w) = f(w) \quad \text{für alle } w \in W_N,$$

wobei $\dim V_N = \dim W_N$ zunächst der einzige Zusammenhang zwischen dem Ansatzraum V_N und dem Testraum W_N ist. Damit diese Diskretisierung sinnvoll ist, muss die Stabilität nachgewiesen werden.

Ist $\omega \subset \Omega$ ein Teilstück von Ω und $Lu = \operatorname{div} a(x) \operatorname{grad} u$ der Differentialoperator, liefert die partielle Integration $\int_{\omega} w Lu \, dx = \int_{\partial\omega} a(x) w(x) \frac{\partial u}{\partial n} \, d\Gamma - \int_{\omega} a(x) \langle \operatorname{grad} w, \operatorname{grad} u \rangle \, dx$. Man wählt nun $w(x)$ als die charakteristische Funktion von ω : $w = 1$ auf ω und $w = 0$ außerhalb:

$$\int_{\partial\omega} a(x) \frac{\partial u}{\partial n} \, d\Gamma = \int_{\omega} f \, dx \quad \text{für alle Volumina } \omega \in \mathcal{T}_{fV},$$

wobei die Menge \mathcal{T}_{fV} der *finiten Volumen* noch zu definieren ist (traditionell spricht man Volumen, auch wenn Flächen vorliegen).

Ist V_N der übliche Raum der stückweise linearen Elemente auf Dreiecken der Triangulation \mathcal{T} . Die Dimension $N = \dim V_N$ ist durch die Anzahl der Knoten von \mathcal{T} gegeben. Wegen $\dim V_N = \dim W_N$ wählt man ein duales Gitter \mathcal{T}_{fV} , das zu jedem Knoten $\mathbf{x} \in \mathcal{T}$ ein "finites Volumen" $\omega_{\mathbf{x}}$ enthält (dies ist im Allgemeinen ein Polygon; vgl. Knabner-Angermann [52, §8] und Großmann-Roos [36, §2.4.2]).

Für $\Omega \subset \mathbb{R}^2$ ist der Name "Box-Methode" ein Synonym für Finite-Volumen-Methode. Für Differentialoperatoren $L = \operatorname{div} a(x) \operatorname{grad}$, die nur aus dem Hauptteil bestehen, und unter geeigneten Voraussetzungen an den Koeffizienten a stellt sich heraus, dass die Box-Methode und das Standard-Galerkin-Verfahren für stückweise lineare Ansatzfunktionen identische Systemmatrizen liefern (vgl. Hackbusch [41]).

9

Regularität

9.1 Lösungen der Randwertaufgabe in $H^s(\Omega)$, $s > m$

9.1.1 Das Regularitätsproblem

Die schwache Formulierung des Randwertproblems

$$Lu = g \quad \text{in } \Omega, \quad Bu = \varphi \quad \text{auf } \Gamma \quad (2m \text{ Ordnung des Differentialoperators } L) \quad (9.1.1)$$

durch

$$u \in V, \quad a(u, v) = f(v) \quad \text{für alle } v \in V \quad (9.1.2)$$

war in Kapitel 7 die Grundlage, auf der die Frage nach der Existenz und Eindeutigkeit der Lösung beantwortet werden konnte. Mit der Existenz einer Lösung ist dabei die Existenz einer schwachen Lösung $u \in V$ gemeint.

Die Fehlerabschätzungen in Abschnitt 8.4 haben deutlich gemacht, dass die Aussage $u \in V$ für numerische Zwecke nicht ausreichend ist. Unter dieser Voraussetzung lässt sich nur $\|u^h - u\|_V \rightarrow 0$ zeigen. Die interessantere quantitative Abschätzung $\|u^h - u\|_V = \mathcal{O}(h)$ wie z.B. in Satz 8.4.6 für $V = H_0^1(\Omega)$ oder $V = H^1(\Omega)$ erfordert die Annahme $u \in H^2(\Omega) \cap V$. Die Bedingung $u \in H^2(\Omega)$ oder allgemeiner $u \in H^s(\Omega)$ ist eine *Regularitätsaussage*, d.h. eine Aussage über die Glattheit der Lösung, die in diesem Abschnitt näher untersucht werden soll.

Die Regularitätsbeweise der folgenden Abschnitte sind sehr technisch. Um die Beweisstrategie durchsichtiger zu machen, sei zunächst der Beweis der nachfolgenden Ungleichung (9.1.4), $|u|_s \leq C'_s |f|_{s-m}$, für die Helmholtz-Gleichung $-\Delta u + u = f$ in $\Omega \subset \mathbb{R}^2$, $u = 0$ auf Γ , skizziert. Man beachte die Unterscheidung nach $\Omega = \mathbb{R}^n$, $\Omega = \mathbb{R}_+^n$, Ω beliebig (hier: $n = 2$).

1. *Schritt:* $\Omega = \mathbb{R}^2$. Da die zugehörige Bilinearform $a(u, v) = \int_{\Omega} \langle \nabla u, \nabla v \rangle d\mathbf{x} + \int_{\Omega} uv d\mathbf{x}$ $H^1(\mathbb{R}^2)$ -elliptisch ist, gilt (9.1.4) für $s = m = 1$. Durch Induktion soll (9.1.4) für $s = 1, 2, 3, \dots$ bewiesen werden. Dazu leiten wir die Differentialgleichung nach x ab, wobei $\mathbf{x} = (x, y)$: $-\Delta u_x + u_x = f_x$. Falls $f \in H^{s-2}(\Omega)$, ist $f_x \in H^{s-3}(\Omega)$ und die Gleichung $-\Delta v + v = f_x$ hat nach Induktionsannahme eine eindeutige Lösung $v \in H^{s-1}(\Omega)$ mit $|v|_{s-1} \leq C'_{s-1} |f_x|_{s-3} \leq C'_{s-1} |f|_{s-2}$. Stellt man diese Ungleichung für $v = u_x$ und analog für $v = u_y$ auf, ergibt sich $|u|_s \leq |u|_{s-1} + |u_x|_{s-1} + |u_y|_{s-1} \leq 3C'_{s-1} |f|_{s-2}$. Also ist (9.1.4) auch für s gezeigt.

2. *Schritt:* $\Omega = \mathbb{R}_+^2 := \mathbb{R} \times (0, \infty)$. Wie oben kann man zur Abschätzung $|u_x|_{s-1} \leq C'_{s-1} |f|_{s-2}$ gelangen, denn auch u_x erfüllt $-\Delta u_x + u_x = f_x$ und die Randbedingung $u_x = 0$ auf Γ . Für u_y trifft dies nicht zu. Aber $u_x \in H^{s-1}(\mathbb{R}_+^2)$ impliziert $u_{xx} \in H^{s-2}(\mathbb{R}_+^2)$ und $u_{xy} \in H^{s-2}(\mathbb{R}_+^2)$. $u \in H^s(\mathbb{R}_+^2)$ würde gelten, wenn auch $u_{yy} \in H^{s-2}(\mathbb{R}_+^2)$ gezeigt werden könnte. Diese Eigenschaft ergibt sich jedoch aus der Differentialgleichung: $u_{yy} = \Delta u - u_{xx} = u - f - u_{xx} \in H^{s-2}(\mathbb{R}_+^2)$.

3. *Schritt:* Ω sei beliebig, aber hinreichend glatt. Wie in Abbildung 6.2.1 wird Ω in (überlappende) Teilgebiete Ω_i zerlegt, die jeweils in \mathbb{R}^2 bzw. \mathbb{R}_+^2 abgebildet werden können. Entsprechend wird die Lösung u in $\sum \chi_i u$ aufgespalten (χ_i : Zerlegung der Eins). Die Überlegungen aus den Schritten 1 und 2 beweisen dann Abschätzungen für $|\chi_i u|_s$, die zusammen (9.1.4) ergeben.

Die obigen Beweisschritte sind noch unvollständig. Kann z.B. die Gleichung $-\Delta u_x + u_x = f_x$ in \mathbb{R}^2 nicht eine Lösung $u_x \in L^2(\mathbb{R}^2)$ besitzen, die nicht zu $H^1(\mathbb{R}^2)$ gehört und damit nicht mit der Lösung $v \in H^1(\mathbb{R}^2)$ von $-\Delta v + v = f_x$ zusammenfällt?

Sei im Folgenden stets $s \geq m$. Das Randwertproblem (9.1.2) mit $V = H_0^1(\Omega)$ heißt H^s -regulär, wenn jede Lösung $u \in H_0^1(\Omega)$ der Aufgabe (9.1.2) mit $f \in H^{s-2m}(\Omega)$ zu $H^s(\Omega) \cap H_0^1(\Omega)$ gehört und der Abschätzung

$$|u|_s \leq C_s [|f|_{s-2m} + |u|_m] \quad (9.1.3)$$

genügt. Ist L der zu $a(\cdot, \cdot)$ gehörende Operator, so heißt auch L H^s -regulär.

Bemerkung 9.1.1 m sei wie in (9.1.1) definiert. a) H^m -Regularität gilt immer.

b) Das Variationsproblem (9.1.2) habe für alle $f \in H^{-m}(\Omega)$ eine eindeutige Lösung $u \in H_0^m(\Omega)$ mit $|u|_m \leq C_0 |f|_{-m}$. Ist das Randwertproblem H^s -regulär, so erfüllt die schwache Lösung von (9.1.2) mit $f \in H^{s-2m}(\Omega)$ die Ungleichung

$$|u|_s \leq C'_s |f|_{s-m} . \quad (9.1.4)$$

c) L sei der zu $a(\cdot, \cdot)$ gehörende Operator. (9.1.4) ist zu $L^{-1} \in L(H^{s-2m}(\Omega), H^s(\Omega))$ und (9.1.4') äquivalent:

$$\|L^{-1}\|_{H^s(\Omega) \leftarrow H^{s-2m}(\Omega)} \leq C'_s . \quad (9.1.4')$$

Beweis. a) Ungleichung (9.1.3) gilt stets mit $C_m = 1$. b) In (9.1.3) schätze man $|u|_m$ durch $C_0 |f|_{-m}$ ab und benutze, dass wegen $s \geq m$ die Einbettung $H^{s-2m}(\Omega) \subset H^{-m}(\Omega)$ stetig ist, sodass $|f|_{-m} \leq C' |f|_{s-2m}$. Damit folgert man (9.1.4) mit der Konstanten $C'_s = C_s(1 + C_0 C')$. ■

Die folgende Bemerkung zeigt, dass eine Störung von $a(\cdot, \cdot)$ durch einen glatten Term der Ordnung $< 2m$ die H^s -Regularität nicht ändert.

Bemerkung 9.1.2 Seien $m \leq s$ und $m, s \in \mathbb{N}$. L sei für alle $t \in \{m, m+1, \dots, s\}$ H^t -regulär. δL sei ein Operator der Ordnung $\leq 2m-1$, d.h. $\delta L \in L(H^r(\Omega) \cap H_0^m(\Omega), H^{r+1-2m}(\Omega))$ für alle $r \in \{m, m+1, \dots, s-1\}$. Dann ist auch $L + \delta L$ H^t -regulär für $t = m, m+1, \dots, s$. Die Annahme über δL trifft insbesondere dann zu, wenn δL zur Bilinearform $a''(\cdot, \cdot)$ aus Lemma 7.2.12 gehört und deren Koeffizienten $a_{\alpha\beta}$ hinreichend glatt sind: $a_{\alpha\beta} \in C^{\max\{0, s-2m+|\beta|\}}$.

Beweis. Wir verwenden Induktion. Für $s = m$ folgt die Aussage aus Bemerkung 9.1.1a. Die Behauptung sei für $s-1$ angenommen. u sei schwache Lösung von $(L + \delta L)u = f \in H^{s-2m}(\Omega)$, also auch Lösung von $Lu = \tilde{f} := f - \delta Lu$. Nach Induktionsannahme gilt bereits $|u|_{s-1} \leq C'_{s-1} [|f|_{s-2m-1} + |u|_m]$, sodass

$$|\tilde{f}|_{s-2m} \leq |f|_{s-2m} + \|\delta L\|_{s-2m \leftarrow s-1} |u|_{s-1} \leq C''_{s-1} [|f|_{s-2m} + |u|_m].$$

Auf Grund der H^s -Regularität von L gehört u zu $H^s(\Omega)$ und erfüllt $|u|_s \leq C'_s [|\tilde{f}|_{s-2m} + |u|_m]$. Zusammen ergeben diese Ungleichungen die Behauptung (9.1.4). ■

9.1.2 Regularitätssätze für $\Omega = \mathbb{R}^n$

Das Gebiet $\Omega = \mathbb{R}^n$ zeichnet sich dadurch aus, dass es keinen Rand besitzt, sodass auch keine Randbedingungen auftreten.

Satz 9.1.3 (Regularitätssatz) Seien $m \in \mathbb{N}$ und $\Omega = \mathbb{R}^n$. Die Bilinearform

$$a(u, v) := \sum_{|\alpha| \leq m} \sum_{|\beta| \leq m} \int_{\Omega} a_{\alpha\beta}(\mathbf{x}) [D^\alpha u(\mathbf{x})] [D^\beta v(\mathbf{x})] d\mathbf{x} \quad (9.1.5)$$

sei $H^m(\mathbb{R}^n)$ -koerziv. Für ein $k \in \mathbb{N}$ gelte

$$D^\gamma a_{\alpha\beta} \in L^\infty(\Omega) \quad \text{für alle } \alpha, \beta, \gamma \text{ mit } |\gamma| \leq \max\{0, k + |\beta| - m\}. \quad (9.1.6)$$

Dann gehört jede schwache Lösung $u \in H^m(\mathbb{R}^n)$ der Aufgabe

$$a(u, v) = (f, v)_0 \quad \text{für alle } v \in H^m(\mathbb{R}^n) \quad (9.1.7)$$

mit $f \in H^{-m+k}(\mathbb{R}^n)$ zu $H^{m+k}(\mathbb{R}^n)$ und erfüllt die Abschätzung

$$|u|_{m+k} \leq C_k [|f|_{-m+k} + |u|_m]. \quad (9.1.8)$$

Zusatz 9.1.4 Anstelle der Ungleichung (9.1.8) könnte auch (9.1.8') verwendet werden:

$$|u|_{m+k} \leq C_{k,k-1} [|f|_{-m+k} + |u|_{m+k-1}]. \quad (9.1.8')$$

Beweis. a) (9.1.8) \Rightarrow (9.1.8') ergibt sich aus $|u|_m \leq |u|_{m+k-1}$ wegen $k \geq 1$.

b) Gilt Satz 9.1.3 für k , so auch für $1, \dots, k-1$ anstelle von k . Die Kombination der Aussagen (9.1.8') für $k, k-1, \dots, 1$ liefert

$$\begin{aligned} |u|_{m+k} &\leq C_{k,k-1} \{|f|_{-m+k} + |u|_{m+k-1}\} \leq C_{k,k-1} \{|f|_{-m+k} + C_{k-1,k-2} [|f|_{-m+k-1} + |u|_{m+k-1}]\} \\ &\leq \dots \leq C' \sum_{\ell=0}^k |f|_{-m+k-\ell} + C'' |u|_m \end{aligned}$$

und damit (9.1.8) wegen $|f|_{-m+k-\ell} \leq |f|_{-m+k}$ für $\ell \geq 0$. ■

Lemma 9.1.5 Sei $a(\cdot, \cdot)$ die Bilinearform (9.1.5) mit Koeffizienten $a_{\alpha\beta}$, die (9.1.6) erfüllen. Seien $\ell \in \{0, \dots, k\}$ und $u \in H^{m+\ell}(\mathbb{R}^n)$ fest gewählt.

a) Dann ist $F := a(u, \cdot)$ ein Funktional aus $H^{-m+\ell}(\mathbb{R}^n)$ mit der Abschätzung

$$a(u, v) \leq C |u|_{m+\ell} |v|_{m-\ell}.$$

b) Falls $a_0(u, v) := \sum_{|\alpha|+|\beta| < 2m} \int_{\Omega} a_{\alpha\beta} D^\alpha u D^\beta v d\mathbf{x}$ (d.h. $a(\cdot, \cdot)$ ohne den Hauptteil), gilt die stärkere Abschätzung

$$a_0(u, v) \leq C_0 |u|_{m+\ell} |v|_{m-\ell-1}.$$

c) Sei D^γ mit $|\gamma| = 1$ eine erste Ableitung. Dann definiert $a(D^\gamma u, v) = -a(u, D^\gamma v) + b_\gamma(u, v)$ eine Bilinearform $b_\gamma(\cdot, \cdot)$ bzw. ein Funktional $b_\gamma(u, \cdot)$ mit

$$b_\gamma(u, v) \leq C |u|_{m+\ell} |v|_{m-\ell}.$$

Beweis. a) Sei $b(u, v) := \int_{\mathbb{R}^n} a_{\alpha\beta} D^\alpha u D^\beta v d\mathbf{x}$ einer der Terme von $a(u, v)$ und $u \in H^{m+\ell}(\mathbb{R}^n)$. Wir unterscheiden drei Fälle a α) bis a γ).

a α) Sei $|\beta| \leq m - \ell$. Es gilt bereits $|b(u, v)| \leq \|a_{\alpha\beta}\|_\infty |u|_{|\alpha|} |v|_{|\beta|} \leq |u|_{m+\ell} |v|_{m-\ell}$.

a β) Sei $m - |\beta| + 1 \leq \ell \leq m$. Sei γ ein Multiindex mit $|\gamma| = |\beta| - m + \ell \in [1, |\beta|]$ und $\gamma \leq \beta$. Partielle Integration liefert

$$b(u, v) = (-1)^{|\gamma|} \int_{\mathbb{R}^n} D^\gamma (a_{\alpha\beta} D^\alpha u) D^{\beta-\gamma} v d\mathbf{x}.$$

Nach Voraussetzung (9.1.6) bleiben die Ableitungen von $a_{\alpha\beta}$ beschränkt, sodass

$$|b(u, v)| \leq C |u|_{|\alpha|+|\gamma|} |v|_{|\beta-\gamma|} \leq C |u|_{m+\ell} |v|_{m-\ell}.$$

a γ) Sei $\ell > m$. Mit $\gamma := \beta$ liefert die obige Darstellung

$$b(u, v) = (-1)^{|\beta|} \int_{\mathbb{R}^n} D^\beta (a_{\alpha\beta} D^\alpha u) v d\mathbf{x} = (g, v)_0 \quad \text{mit } g := (-1)^{|\beta|} D^\beta (a_{\alpha\beta} D^\alpha u) \in H^{m+\ell-|\alpha|-|\beta|}(\mathbb{R}^n).$$

Da $|\alpha| \leq m$ und $\ell > m \geq |\beta|$, folgt $m + \ell - |\alpha| - |\beta| \geq \ell - m$, sodass $|b(u, v)| \leq |g|_{-m+\ell} |v|_{m-\ell}$. Mit

$$|g|_{-m+\ell} \leq |D^\beta (a_{\alpha\beta} D^\alpha u)|_{-m+\ell} \leq C |u|_{|\alpha|+|\beta|-m+\ell} \stackrel{|\alpha|+|\beta| \leq 2m}{\leq} C |u|_{m+\ell}$$

hat man auch in diesem Fall $|b(u, v)| \leq C|u|_{m+\ell}|v|_{m-\ell}$ nachgewiesen.

Summation über alle α, β liefert das gewünschte Resultat $|a(u, v)| \leq C|u|_{m+\ell}|v|_{m-\ell}$.

b) Gelte beispielsweise $|\alpha| \leq m$ und $|\beta| \leq m-1$ für $b(u, v) := \int_{\mathbb{R}^n} a_{\alpha\beta} D^\alpha u D^\beta v d\mathbf{x}$. Falls $|\beta| \leq m-\ell-1$, folgt direkt $|b(u, v)| \leq C_0|u|_{m+\ell}|v|_{m-\ell-1}$. Falls $m-|\beta| \leq \ell \leq m-1$, werden $|\beta|-m+\ell+1 \in [1, |\beta|]$ partielle Integrationen durchgeführt und liefern wegen $|\alpha|+|\beta|+1-m+\ell \leq 2m-m+\ell=m+\ell$ die Abschätzung $|b(u, v)| \leq C|u|_{m+\ell}|v|_{m-\ell}$. Der Fall $|\alpha| \leq m-1$ und $|\beta| \leq m$ ist dem Leser überlassen.

c) Sei $b_{\alpha\beta}(u, v) := \int_{\mathbb{R}^n} a_{\alpha\beta} D^\alpha u D^\beta v d\mathbf{x}$ ein Summand der Bilinearform $a(u, v)$. Partielle Integration zeigt

$$b_{\alpha\beta,\gamma}(u, v) = - \int_{\mathbb{R}^n} (D^\gamma a_{\alpha\beta}) (D^\alpha u) (D^\beta v) d\mathbf{x} \quad \text{für } b_{\alpha\beta,\gamma}(u, v) := b_{\alpha\beta}(D^\gamma u, v) + b_{\alpha\beta}(u, D^\gamma v).$$

Summation über α, β und die zu Teil a) analogen Überlegungen zeigen $b_\gamma(u, v) \leq C|u|_{m+\ell}|v|_{m-\ell}$.

Hinsichtlich der Glattheit der Koeffizienten $a_{\alpha\beta}$ gilt: Die Bilinearform $b_\gamma(\cdot, \cdot)$ enthält neue Koeffizienten $a_{\alpha\beta,\gamma} := D^\gamma a_{\alpha\beta}$, die (9.1.6) für $k-1$ statt k erfüllen. ■

In Satz 9.1.3 ist $a(\cdot, \cdot)$ nur als $H^m(\mathbb{R}^n)$ -koerziv angenommen, sodass eine existierende Lösung nicht eindeutig zu sein braucht. Um Eindeutigkeit zu sichern, benötigt man z.B. $H^m(\mathbb{R}^n)$ -Elliptizität, die im Teil b) des nächsten Lemmas diskutiert wird.

Lemma 9.1.6 *Es gelte (9.1.5) und (9.1.6). Sei $a_H(u, v) := \sum_{|\alpha|, |\beta|=m} \int_{\mathbb{R}^n} a_{\alpha\beta} D^\alpha u D^\beta v d\mathbf{x}$.*

a) *Gilt Satz 9.1.3 für $a_H(\cdot, \cdot)$ und $k=1$ so auch für die allgemeine Bilinearform (9.1.5) und $k=1$.*

b) *O.B.d.A. darf in Satz 9.1.3 $a(\cdot, \cdot)$ durch $a_\Lambda(u, v) = a(u, v) + \Lambda \int_{\mathbb{R}^n} u v d\mathbf{x}$ und $H^m(\mathbb{R}^n)$ -Koerzivität durch $H^m(\mathbb{R}^n)$ -Elliptizität ersetzt werden. Die gleiche Ergänzung ist in Teil a) möglich, d.h. $a_H(u, v)$ kann durch $a_H(u, v) + \Lambda(u, v)_0$ ersetzt werden.*

Beweis. b) Für $\Lambda \geq C_K$ (C_K aus der Koerzitivitätsungleichung (6.5.10)) wird die $H^m(\mathbb{R}^n)$ -koerzive Bilinearform $a(\cdot, \cdot)$ zu einer $H^m(\mathbb{R}^n)$ -elliptischen Form $a_\Lambda(\cdot, \cdot)$. Sei Satz 9.1.3 für $a_\Lambda(\cdot, \cdot)$ angenommen. Ferner sei u eine schwache Lösung von (9.1.7): $a(u, v) = (f, v)_0$. Offenbar ist u die (wegen der $H^m(\mathbb{R}^n)$ -Elliptizität) eindeutige Lösung von

$$a_\Lambda(u, v) = F_\Lambda(v) := (f + \Lambda u, v)_0.$$

Wegen $|F_\Lambda(v)| \leq |f|_{-m+k}|v|_{m-k} + |\Lambda| |u|_{-m+k}|v|_{m-k} \leq [|f|_{-m+k} + |\Lambda| |u|_{-m+k}] |v|_{m-k}$ gehört auch F_Λ zu $H^{-m+k}(\mathbb{R}^n)$. Nach Annahme (Satz 9.1.3 für $a_\Lambda(\cdot, \cdot)$) folgt die Regularitätsabschätzung (9.1.8') in der Form $|u|_{m+k} \leq C_{k,k-1} [|F_\Lambda|_{-m+k} + |u|_{m+k-1}]$. Mit $|F_\Lambda|_{-m+k} \leq |f|_{-m+k} + |\Lambda| |u|_{-m+k}$ schließt man auf

$$|u|_{m+k} \leq C_{k,k-1} [|f|_{-m+k} + |\Lambda| |u|_{-m+k} + |u|_{m+k-1}] \leq C'_{k,k-1} [|f|_{-m+k} + |u|_{m+k-1}]$$

wegen $|u|_{-m+k} \leq |u|_{m+k-1}$. Damit ist die Aussage des Satzes 9.1.3 für $a(\cdot, \cdot)$ erreicht.

Ersetzung von $a(\cdot, \cdot)$ durch $a_H(\cdot, \cdot)$ zeigt, dass auch in Teil a) o.B.d.A. $\Lambda(u, v)_0$ addiert werden darf.

a) Sei die Gültigkeit von Satz 9.1.3 für $a_H(\cdot, \cdot)$ und $k=1$ angenommen, wobei gemäß Teil b) die $H^m(\mathbb{R}^n)$ -Elliptizität von $a_H(\cdot, \cdot)$ gesichert sei. Ferner sei $a_0(\cdot, \cdot) := a(\cdot, \cdot) - a_H(\cdot, \cdot)$. Eine Lösung $u \in H^m(\mathbb{R}^n)$ von $a_H(u, v) = (f, v)_0$ ist die eindeutige Lösung von

$$a_H(u, v) = F_H(v) := (f, v)_0 + a_0(u, v) \quad \text{für alle } v \in H^m(\mathbb{R}^n).$$

Lemma 9.1.5b mit $\ell=0$ und $|(f, v)_0| \leq |f|_{-m+1}|v|_{m-1}$ ergeben $|F_H(v)| \leq [|f|_{-m+1} + C_0|u|_m] |v|_{m-1}$ und damit $F_0 \in H^{-m+1}(\mathbb{R}^n)$. Dank

$$|u|_{m+1} \leq C_1 [|F_H|_{-m+1} + |u|_m] \leq C_1 [|f|_{-m+1} + C_0|u|_m + |u|_m] \leq C'_1 [|f|_{-m+1} + |u|_m]$$

ist (9.1.8) mit $k=1$ für $a(\cdot, \cdot)$ anstelle von $a_H(\cdot, \cdot)$ gezeigt. ■

Lemma 9.1.7 *Es gelte (9.1.5) und (9.1.6). Gilt (9.1.8) für ein $\ell \in \{1, \dots, k-1\}$ anstelle von k , so auch für $\ell+1$.*

Beweis. Sei u Lösung von (9.1.7). Nach Voraussetzung gilt $u \in H^{m+\ell}(\mathbb{R}^n)$ und $|u|_{m+\ell} \leq C_\ell [|f|_{-m+\ell} + |u|_m]$. Sei D^γ mit $|\gamma|=1$ eine erste Ableitung und b_γ die Bilinearform aus Lemma 9.1.5. Offenbar ist $w := D^\gamma u$ Lösung von

$$a(w, v) = a(D^\gamma u, v) = -a(u, D^\gamma v) + b_\gamma(u, v) = (f, D^\gamma v)_0 + b_\gamma(u, v) =: F_\gamma(v).$$

Da $|(f, D^\gamma v)_0| \leq |f|_{-m+k} |D^\gamma v|_{m-k} \leq |f|_{-m+k} |v|_{m-k+1} \leq |f|_{-m+k} |v|_{m-\ell}$ und $|b_\gamma(u, v)| \leq C |u|_{m+\ell} |v|_{m-\ell}$, folgt $F_\gamma \in H^{-m+\ell}(\mathbb{R}^n)$ mit $|F_\gamma|_{-m+\ell} \leq |f|_{-m+k} + C |u|_{m+\ell}$.

Satz 9.1.3 mit ℓ und F_γ anstelle von k und f zeigt $w \in H^{m+\ell}(\mathbb{R}^n)$ und

$$|w|_{m+\ell} \leq C_\ell \left[|F_\gamma|_{-m+\ell} + |w|_m \right] \leq C_\ell [|f|_{-m+k} + C |u|_{m+\ell} + |D^\gamma u|_m] \leq C_\ell [|f|_{-m+k} + C' |u|_{m+\ell}]$$

Da der Multiindex γ mit $|\gamma| = 1$ in $w := D^\gamma u$ beliebig ist, folgt $u \in H^{m+\ell+1}(\mathbb{R}^n)$ und die Ungleichung

$$|u|_{m+\ell+1} \leq (n+1) C_\ell [|f|_{-m+k} + C' |u|_{m+\ell}].$$

Zusammen mit der vorausgesetzten Regularitätsabschätzung $|u|_{m+\ell} \leq C_\ell [|f|_{-m+\ell} + |u|_m]$ erhalten wir $|u|_{m+\ell+1} \leq C_{\ell+1} [|f|_{-m+k} + |u|_m]$, d.h. die Regularitätsaussage für $\ell+1$. ■

Beweis des Satzes 9.1.3. 1) Induktionsanfang $k=1$. Gemäß Lemma 9.1.6a darf $a(\cdot, \cdot)$ durch den Hauptteil $a_H(u, v) := \sum_{|\alpha|, |\beta|=m} \int_{\mathbb{R}^n} a_{\alpha\beta} D^\alpha u D^\beta v d\mathbf{x}$ ersetzt werden. Sei $\partial_{h,i}$ der Differenzenoperator¹

$$\partial_{h,i} u(\mathbf{x}) := \frac{1}{h} \left[u \left(\mathbf{x} + \frac{h}{2} \mathbf{e}_i \right) - u \left(\mathbf{x} - \frac{h}{2} \mathbf{e}_i \right) \right], \quad \mathbf{e}_i: i\text{-ter Einheitsvektor}, 1 \leq i \leq n.$$

u sei die Lösung von (9.1.7). Für $v \in H^m(\mathbb{R}^n)$ setze man

$$d_i(u, v) := d(u, v) := a(u, \partial_{h,i} v) + a(\partial_{h,i} u, v) \quad \text{für alle } v \in H^m(\mathbb{R}^n). \quad (9.1.9a)$$

Man beachte, dass $d_i(u, v)$ wegen $\partial_{h,i} v, \partial_{h,i} u \in H^m(\mathbb{R}^n)$ wohldefiniert ist. Es sei $w^\pm(\mathbf{x}) := w(\mathbf{x} \pm h\mathbf{e}_i/2)$.

$$\begin{aligned} \int_{\mathbb{R}^n} a u \partial_{h,i} v d\mathbf{x} &= \frac{1}{h} \int_{\mathbb{R}^n} a u [v^+ - v^-] d\mathbf{x} = \frac{1}{h} \int_{\mathbb{R}^n} [a^- u^- - a^+ u^+] v d\mathbf{x} \\ &= - \int_{\mathbb{R}^n} a v \partial_{h,i} u d\mathbf{x} + \frac{1}{h} \int_{\mathbb{R}^n} [(a^- - a) u^- + (a - a^+) u^+] v d\mathbf{x} \end{aligned}$$

(mit den Ersetzungen $a \rightsquigarrow a_{\alpha\beta}$, $u \rightsquigarrow D^\alpha u$, $v \rightsquigarrow D^\beta v$) zeigt

$$d(u, v) = - \sum_{|\alpha|=m} \sum_{|\beta|=m} \int_{\mathbb{R}^n} \frac{1}{h} \left\{ (a_{\alpha\beta} - a_{\alpha\beta}^-) D^\alpha u^- + (a_{\alpha\beta}^+ - a_{\alpha\beta}) D^\alpha u^+ \right\} D^\beta v d\mathbf{x}. \quad (9.1.9b)$$

Da $|a_{\alpha\beta} - a_{\alpha\beta}^\pm|/h \leq \|\partial a_{\alpha\beta}/\partial x_i\|_{L^\infty} \leq \text{const}$ auf Grund von $\max\{0, k + |\beta| - m\} = k \geq 1$ (vgl. (9.1.6)), gilt

$$|d(u, v)| \leq C |u|_m |v|_m \quad \text{für alle } v \in H^m(\mathbb{R}^n). \quad (9.1.9c)$$

Wegen der $H^m(\mathbb{R}^n)$ -Koerzivität gibt es Konstanten $C_E > 0$ und C_K (vgl. (6.5.10)) mit

$$C_E |\partial_{h,i} u|_m^2 \leq a(\partial_{h,i} u, \partial_{h,i} u) + C_K |\partial_{h,i} u|_0^2 = -a(u, \partial_{h,i}^2 u) + d(u, \partial_{h,i} u) + C_K |\partial_{h,i} u|_0^2 \quad (9.1.10a)$$

(hier wird $v := \partial_{h,i} u \in H^m(\mathbb{R}^n)$ ausgenutzt!). Der erste Summand kann gemäß (9.1.7) umgeformt werden: $a(u, \partial_{h,i}^2 u) = (f, \partial_{h,i}^2 u)_0$ und durch $|f|_{-m+1} |\partial_{h,i}^2 u|_{m-1}$ abgeschätzt werden. Die Ungleichung $|\partial_{h,i} v|_s \leq |v|_{s+1}$ für $s = m-1$ und $v = \partial_{h,i} u$ liefert

$$|a(u, \partial_{h,i}^2 u)| \leq |f|_{-m+1} |\partial_{h,i} u|_m. \quad (9.1.10b)$$

Wegen (9.1.9c) lautet die Abschätzung des zweiten Summanden in (9.1.10a):

$$|d(u, \partial_{h,i} u)| \leq C_d |u|_m |\partial_{h,i} u|_m. \quad (9.1.10c)$$

Schließlich ist

$$|\partial_{h,i} u|_0^2 \leq |\partial_{h,i} u|_{m-1} |\partial_{h,i} u|_m \leq |u|_m |\partial_{h,i} u|_m. \quad (9.1.10d)$$

¹Eigentlich würde man lieber die Ableitung $\partial u/\partial x_i$ verwenden. Da wir aber nicht wissen, ob diese zu $V = H^m(\mathbb{R}^n)$ gehört, wird die Differenz $\partial_{h,i} u$ als Ersatz genommen.

Die Schranken (9.1.10a-d) liefern

$$|\partial_{h,i} u|_m \leq [|f|_{-m+1} + (C_d + C_K) |u|_m] / C_E \quad \text{für alle } h > 0, 1 \leq i \leq n.$$

Lemma 6.2.25 beweist $u \in H^{m+1}(\mathbb{R}^n)$ und Ungleichung (9.1.8) für $k = 1$.

2) Der Induktionsschritt ist durch Lemma 9.1.7 gegeben. ■

Zusatz 9.1.8 Die $H^m(\mathbb{R}^n)$ -Koerzivität in Satz 9.1.3 lässt sich ersetzen: (a) durch die hinreichenden Bedingungen aus Satz 7.2.11 oder (b) durch die Annahme, dass für ein λ die Bilinearform $a_H(u, v) + \lambda(u, v)_0$ die Voraussetzung (6.5.4a,b) erfüllt.

Zusatz 9.1.9 Ist zusätzlich $a(\cdot, \cdot)$ $H^m(\mathbb{R}^n)$ -elliptisch oder erfüllt $a(\cdot, \cdot)$ die Bedingung (6.5.4a,b), so ist in Satz 9.1.3 die Abschätzung (9.1.8) durch (9.1.11) ersetzbar:

$$|u|_{m+k} \leq C_k |f|_{-m+k}. \quad (9.1.11)$$

Beweis. Folgt nach Bemerkung 9.1.2. ■

Zusatz 9.1.10 $a(\cdot, \cdot)$ sei $H^m(\mathbb{R}^n)$ -koerziv. Die Bedingungen (9.1.6) und $f \in H^{k-m}(\mathbb{R}^n)$ seien für ein $k \in \mathbb{N}$ mit $k > s + n/2 > n/2$ erfüllt. Dann gehört die schwache Lösung u von (9.1.7) zu $C^s(\mathbb{R}^n)$. Für $s > 2m$ ist damit die schwache Lösung auch eine klassische Lösung.

Beweis. Die Behauptung ergibt sich aus dem Sobolevschen Einbettungssatz 6.2.31. ■

Zusatz 9.1.11 $a(\cdot, \cdot)$ sei $H^m(\mathbb{R}^n)$ -koerziv. Die Bedingungen (9.1.6) und $f \in H^{k-m}(\mathbb{R}^n)$ seien für alle $k \in \mathbb{N}$ erfüllt. Dann gehört die schwache Lösung der Aufgabe (9.1.7) zu $C^\infty(\mathbb{R}^n)$. Die Voraussetzungen sind insbesondere dann erfüllt, wenn f zu $C_0^\infty(\mathbb{R}^n)$ gehört und die Koeffizienten $a_{\alpha\beta}$ konstant sind.

Die Verallgemeinerung von $u \in H^{m+k}(\mathbb{R}^n)$ mit $k \in \mathbb{N}$ auf $u \in H^{m+s}(\mathbb{R}^n)$ mit reellem $s > 0$ lautet wie folgt:

Satz 9.1.12 $a(\cdot, \cdot)$ aus (9.1.5) sei $H^m(\mathbb{R}^n)$ -koerziv. Seien

$$s = k + \Theta, \quad k \in \mathbb{N}_0, \quad 0 < \Theta < \vartheta < 1, \quad t := k + \vartheta.$$

Für die Koeffizienten gelte (mit $\Omega = \mathbb{R}^n$):

$$a_{\alpha\beta} \in \begin{cases} C^{t+|\beta|-m}(\Omega) & \text{falls } k + |\beta| \geq m, \\ L^\infty(\Omega) & \text{sonst.} \end{cases} \quad (9.1.12)$$

Dann gehört jede schwache Lösung von (9.1.7) mit $f \in H^{-m+s}(\mathbb{R}^n)$ zu $H^{m+s}(\mathbb{R}^n)$ und erfüllt die Abschätzung

$$|u|_{m+s} \leq C_s [|f|_{-m+s} + |u|_m]. \quad (9.1.13)$$

Beweis. Der Induktionsanfang ist durch $k = 0$, d.h. $0 < s < t < 1$ gegeben. Der Differenzenquotient $\partial = \partial_{h,i}$ ist durch eine Approximation seiner Potenz ∂^s zu ersetzen:

$$Ru(\mathbf{x}) := R_{h,j}u(\mathbf{x}) := h^{-s} \sum_{\mu=0}^{\infty} e^{-\mu h} (-1)^\mu \binom{s}{\mu} u(\mathbf{x} + \mu h \mathbf{e}_j) \quad (\mathbf{e}_j: j\text{-ter Einheitsvektor}).$$

Dabei ist $\binom{s}{0} = 1$, $\binom{s}{\mu} = (-1)^\mu (-s)(-s-1)\dots(-s-\mu+1)/\mu!$ der Binomialkoeffizient.

Übungsaufgabe 9.1.13 Sei $0 < s < 1$. Man zeige: a) Der zu $R_{h,j}$ adjungierte Operator ist

$$R^*u(\mathbf{x}) = R_{h,j}^*u(\mathbf{x}) = h^{-s} \sum_{\mu=0}^{\infty} e^{-\mu h} (-1)^\mu \binom{s}{\mu} u(\mathbf{x} - \mu h \mathbf{e}_j).$$

b) Für alle $z \in \mathbb{C}$ mit $|z| < 1$ ist $(1-z)^s = \sum_{\mu=0}^{\infty} \binom{s}{\mu} (-z)^\mu$.

c) Für die Fourier-Transformierten gilt

$$(\widehat{R_{h,j}u})(\xi) = [(1 - e^{-h+i\xi_j h})/h]^s \widehat{u}(\xi), \quad (\widehat{R_{h,j}^*u})(\xi) = [(1 - e^{-h-i\xi_j h})/h]^s \widehat{u}(\xi).$$

Hinweis: $u(\cdot + \delta \mathbf{e}_j)(\xi) = e^{i\xi_j \delta} \widehat{u}(\xi)$.

d) $|R_{h,j}u|_\tau \leq C_\tau |u|_{\tau+s}$ für alle $\tau \in \mathbb{R}$, $h > 0$, $1 \leq j \leq n$, $u \in H^{\tau+s}(\mathbb{R}^n)$; ebenso $|R_{h,j}^*u|_\tau \leq C_\tau |u|_{\tau+s}$.

Hinweis: Man verwende die Normen $|\cdot|_\tau$, $|\cdot|_{\tau+s}$ (vgl. (6.2.15)) und zeige $|(\widehat{Ru})(\xi)| \leq (1 + |\xi|^2)^{s/2} |\widehat{u}(\xi)|$.

In Analogie zu (9.1.9a) erhält man (o.B.d.A. ist die Bilinearform auf den Hauptteil beschränkt)

$$\begin{aligned} d(u, v) &:= a(u, R_{h,j}v) + a(R_{h,j}^*u, v) \\ &= - \sum_{|\alpha|=|\beta|=m} h^{-s} \sum_{\mu=1}^{\infty} e^{-\mu h} (-1)^\mu \binom{s}{\mu} \int_{\mathbb{R}^n} [a_{\alpha\beta}(\mathbf{x} - \mu h \mathbf{e}_j) - a_{\alpha\beta}(\mathbf{x})] D^\alpha u(\mathbf{x} - \mu h \mathbf{e}_j) D^\beta v(\mathbf{x}) d\mathbf{x}, \end{aligned}$$

da der erste Summand in

$$(a_{\alpha\beta} D^\alpha u)(\mathbf{x} - \mu h \mathbf{e}_j) = a_{\alpha\beta}(\mathbf{x}) D^\alpha u(\mathbf{x} - \mu h \mathbf{e}_j) + [a_{\alpha\beta}(\mathbf{x} - \mu h \mathbf{e}_j) - a_{\alpha\beta}(\mathbf{x})] D^\alpha u(\mathbf{x} - \mu h \mathbf{e}_j)$$

zu $a_H(R^*u, v)$ gehört. Da $|a_{\alpha\beta}(\mathbf{x} - \mu h \mathbf{e}_j) - a_{\alpha\beta}(\mathbf{x})| \leq C(\mu h)^t$, folgt

$$|d(u, v)| \leq C |u|_m |v|_m \left[1 + h^{t-s} \sum_{\mu=1}^{\infty} e^{-\mu h} (-1)^\mu \binom{s}{\mu} \mu^t \right] \leq C' |u|_m |v|_m,$$

denn es gilt $\binom{s}{\mu} = \mathcal{O}(\mu^{-s-1})$ und $\sum_{\mu=1}^{\infty} e^{-\mu h} \mu^{t-s-1} = \mathcal{O}(h^{s-t})$.

Die gleichen Überlegungen wie im Beweis des Satzes 9.1.3 ergeben

$$|R_{h,j}^*u|_m \leq C [|R_{h,j}f|_{-m} + |u|_m] \leq C' [|f|_{-m+s} + |u|_m] \quad \text{für alle } h > 0, 1 \leq j \leq n$$

(vgl. Übungsaufgabe 9.1.13d). Um dieselbe Abschätzung für $|u|_{m+s}$ zu erreichen, schreiben wir $(|u|_{m+s}^\wedge)^2$ als

$$\int_{|\xi| \leq 1/h} (1 + |\xi|^2)^{m+2} |\widehat{u}(\xi)|^2 d\xi + \int_{|\xi| \geq 1/h} (1 + |\xi|^2)^{m+2} |\widehat{u}(\xi)|^2 d\xi.$$

Das zweite Integral strebt für $h \rightarrow 0$ gegen null. Das erste lässt sich durch $|u|_m^2$ plus

$$\begin{aligned} C \sum_{j=1}^n |R_{h,j}^*u|_m^2 &\geq C' \int_{|\xi| \leq 1/h} (1 + |\xi|^2)^m \sum_{j=1}^n |\widehat{R_{h,j}^*u}(\xi)|^2 d\xi \\ &\stackrel{\text{Übungsaufgabe 9.1.13c}}{=} C' \int_{|\xi| \leq 1/h} (1 + |\xi|^2)^m \sum_{j=1}^n |(1 - e^{-h-i\xi_j h})/h|^{2s} |\widehat{u}(\xi)|^2 d\xi \\ &= C' \int_{|\xi| \leq 1/h} (1 + |\xi|^2)^m \sum_{j=1}^n \left| \frac{\sin(\xi_j h)}{h} \right|^{2s} |\widehat{u}(\xi)|^2 d\xi \\ &\geq C'' \int_{|\xi| \leq 1/h} (1 + |\xi|^2)^m |\xi|^{2s} |\widehat{u}(\xi)|^2 d\xi \end{aligned}$$

abschätzen. Wie in Lemma 6.2.25 folgen $u \in H^{m+s}(\mathbb{R}^n)$ und (9.1.13).

2) Die Induktion über k verläuft analog zum Beweis des Satzes 9.1.3. ■

Übungsaufgabe 9.1.14 Man verallgemeinere Zusatz 9.1.10 mit Hilfe des Satzes 9.1.12.

9.1.3 Regularitätssätze für $\Omega = \mathbb{R}_+^n$

Der Halbraum \mathbb{R}_+^n aus (6.2.18) ist durch $x_n > 0$ charakterisiert. Wie in Kapitel 7 beschränken wir uns auf die folgenden zwei Fälle: Entweder ist ein Dirichlet-Problem für beliebiges $m \geq 1$ gegeben oder die natürliche Randbedingung ist für $m = 1$ gestellt.

Satz 9.1.15 (homogenes Dirichlet-Problem) *Der Satz 9.1.3 gilt sinngemäß für das Dirichlet-Problem*

$$u \in H_0^m(\mathbb{R}_+^n), \quad a(u, v) = (f, v)_0 \quad \text{für alle } v \in H_0^m(\mathbb{R}_+^n). \quad (9.1.14)$$

Auch Satz 9.1.12 lässt sich übertragen, wenn man die Werte $s = 1/2, 3/2, \dots, m - 1/2$ ausnimmt.

Zum Beweis benötigen wir die folgende, sehr technische Hilfsbehauptung:

Lemma 9.1.16 *Sei $s > 0$, $s \notin \{1/2, 3/2, \dots, m - 1/2\}$. Die Norm $|\cdot|_s$ von $H^s(\mathbb{R}_+^n)$ ist äquivalent zu*

$$\|u\|_s := \sqrt{|u|_0^2 + \sum_{|\alpha|=m} |D^\alpha u|_{s-m}^2}. \quad (9.1.15)$$

Beweis. Der relativ elementare Fall $s \geq m$ ist dem Leser überlassen. Auch für $0 < s < m$ wäre der Beweis wesentlich einfacher, wenn in (9.1.15) die Dualnorm $|\cdot|_{s-m}$ von $(H_0^{m-s}(\mathbb{R}_+^n))'$ durch diejenige von $(H^{m-s}(\mathbb{R}_+^n))'$ ersetzt würde.

1) Zunächst soll die Behauptung für $\Omega = \mathbb{R}^n$ statt $\Omega = \mathbb{R}_+^n$ bewiesen werden. Gemäß Satz 6.2.26a ist für $\Omega = \mathbb{R}^n$ die Norm $\|\cdot\|_s$ zu

$$\|u\|_s := \sqrt{|u|_0^2 + \sum_{|\alpha|=m} (|D^\alpha u|_{s-m})^2}$$

äquivalent. Da $(\|u\|_s)^2 = \int_{\mathbb{R}^n} [1 + (\sum_{|\alpha|=m} |\xi^\alpha|^2)(1 + |\xi|^2)^{s-m}] |\hat{u}(\xi)|^2 d\xi$ und

$$0 < C_0 (1 + |\xi|^2)^s \leq [1 + \sum_{|\alpha|=m} |\xi^\alpha|^2 (1 + |\xi|^2)^{s-m}] \leq C_1 (1 + |\xi|^2)^s,$$

sind auch $\|u\|_s$ und $|\cdot|_s$ äquivalent.

2) Zum Übergang auf $\Omega = \mathbb{R}_+^n$ ist die folgende Fortsetzung $\phi : H^s(\mathbb{R}_+^n) \rightarrow H^s(\mathbb{R}^n)$ zu untersuchen, wobei $\mathbf{x} = (\mathbf{x}', x_n) \in \mathbb{R}^n$:

$$\begin{aligned} (\phi u)(\mathbf{x}) &:= u(\mathbf{x}) \quad \text{für } \mathbf{x} \in \mathbb{R}_+^n \text{ (d.h. } x_n > 0) \\ (\phi u)(\mathbf{x}', x_n) &:= \sum_{\nu=1}^L a_\nu [u(\mathbf{x}', -\nu x_n) + u(\mathbf{x}', -x_n/\nu)] \quad \text{für } x_n < 0, \end{aligned}$$

wobei die Koeffizienten a_ν in der folgenden Übungsaufgabe definiert sind.

Übungsaufgabe 9.1.17 *Die Koeffizienten a_ν von ϕ seien als Lösung des linearen Gleichungssystems*

$$\sum_{\nu=1}^L a_\nu (\nu^k + \nu^{-k}) = (-1)^k \quad (0 \leq k \leq L-1)$$

gewählt. Man zeige:

- a) $u \in C^{L-1}(\mathbb{R}_+^n)$ ergibt $\phi u \in C^{L-1}(\mathbb{R}^n)$.
- b) $\phi \in L(H^k(\mathbb{R}_+^n), H^k(\mathbb{R}^n))$ für $k = 0, 1, \dots, L$.
- c) Der zu ϕ adjungierte Operator lautet

$$(\phi^* u)(\mathbf{x}', x_n) = u(\mathbf{x}', x_n) + \sum_{\nu=1}^L a_\nu \left[\frac{1}{\nu} u(\mathbf{x}', -x_n/\nu) + \nu u(\mathbf{x}', -\nu x_n) \right] \quad \text{für } x_n > 0.$$

- d) $(\partial/\partial x_n)^k (\phi^* u)(\mathbf{x}', 0) = 0$ und $u \in C^k(\mathbb{R}^n)$ für $k = 0, 1, \dots, L-2$.
- e) $\phi^* \in L(H^k(\mathbb{R}^n), H_0^k(\mathbb{R}_+^n))$ für $k = 0, 1, \dots, L-1$. Hinweis: vgl. Korollar 6.2.44.
- f) $\phi \in L(H^k(\mathbb{R}_+^n), H^k(\mathbb{R}^n))$ für $s = 1-L, 2-L, \dots, 0, 1, \dots, L$.

3) $\|\cdot\|_s \leq C \|\cdot\|_s$ ergibt sich aus $D^\alpha \in L(H_0^s(\mathbb{R}_+^n), H^{s-m}(\mathbb{R}_+^n))$ (durch Fortsetzungsargumente aus Bemerkung 6.3.14b beweisbar), sodass $\|\cdot\|_s \leq \|\cdot\|_s$ zu zeigen bleibt.

4) $|u|_s \leq |\phi u|_s \leq \|\phi u\|_s$ gilt nach Beweisteil 1). Die Ungleichung $\|\phi u\|_s \leq C \|u\|_s$, die den Beweis abschließen würde, reduziert sich auf

$$|D^\alpha \phi u|_{s-m} \leq C |D^\alpha u|_{s-m} \quad \text{für } |\alpha| = m, u \in H^k(\mathbb{R}_+^n). \quad (9.1.16)$$

Für

$$(\phi_\alpha u)(\mathbf{x}', x_n) := \begin{cases} u(\mathbf{x}', x_n) & \text{für } x_n > 0, \\ \sum_{\nu=1}^L a_\nu [(-\nu)^{\alpha_n} u(\mathbf{x}', -\nu x_n) + \left(\frac{-1}{\nu}\right)^{\alpha_n} u(\mathbf{x}', -x_n/\nu)] & \text{sonst} \end{cases} \quad \text{mit } |\alpha| = m$$

verifiziert man $D^\alpha \phi = \phi_\alpha D^\alpha$. Wie in Übungsaufgabe 9.1.17f zeigt man $\phi_\alpha \in L(H^s(\mathbb{R}_+^n), H^s(\mathbb{R}^n))$ für $s = 1 + m - L, 2 + m - L, \dots, L - m$. Dieses Resultat lässt sich auf reelle $s \in [1 + m - L, L - m]$ bis auf die Ausnahmen $\frac{1}{2} - s \in \mathbb{N}$ (d.h. $s = -1/2, -3/2, \dots$) übertragen (vgl. Lions-Magenes [58, p. 54ff]). Sei $L \geq 2m + 1$ und $v \in H^{m-s}(\mathbb{R}^n)$. Da $\phi_\alpha^* v \in H_0^{m-s}(\mathbb{R}_+^n)$, folgert man aus $(D^\alpha \phi u, v)_{L^2(\mathbb{R}^n)} = (D^\alpha u, \phi_\alpha^* v)_{L^2(\mathbb{R}_+^n)}$ die Abschätzung

$$(D^\alpha \phi u, v)_0 = |D^\alpha u|_{s-m} \|\phi_\alpha^*\|_{H_0^{m-s}(\mathbb{R}_+^n) \leftarrow H^{m-s}(\mathbb{R}^n)} |v|_{m-s} \quad \text{für alle } v \in H^{m-s}(\mathbb{R}^n),$$

also (9.1.16) mit $C := \|\phi_\alpha^*\|_{H_0^{m-s}(\mathbb{R}_+^n) \leftarrow H^{m-s}(\mathbb{R}^n)} = \|\phi_\alpha\|_{H^{s-m}(\mathbb{R}^n) \leftarrow H^{s-m}(\mathbb{R}_+^n)}$. ■

Beweis des Satzes 9.1.15. 1) Sei zunächst $k = s = 1$. Der Beweis des Satzes 9.1.3 ist für die Differenzen $\partial_{h,j} u$ ($j = 1, \dots, n-1$) wiederholbar und impliziert die Existenz der Ableitungen $\partial u / \partial x_j \in H_0^m(\mathbb{R}_+^n)$, $j \neq n$. Damit hat man $D^\alpha u \in H^1(\mathbb{R}_+^n)$ für alle $|\alpha| = m$ bis auf $\alpha = (0, \dots, 0, m)$.

2) Wir setzen

$$\hat{\alpha} := (0, \dots, 0, m) \in \mathbb{Z}^n, \quad w := a_{\hat{\alpha}\hat{\alpha}} D^{\hat{\alpha}} u, \quad F_\alpha(v) := \int_{\mathbb{R}_+^n} w(\mathbf{x}) D^\alpha v(\mathbf{x}) d\mathbf{x} \quad \text{für } |\alpha| = m,$$

wobei $a_{\hat{\alpha}\hat{\alpha}}$ der Koeffizient aus der Bilinearform (9.1.5) ist. Der weitere Beweisgang verläuft wie folgt. Wir werden in Teil 3) zeigen, dass

$$|F_\alpha(v)| \leq C |v|_{m-1} \quad \text{für } |\alpha| = m, v \in H^{m-s}(\mathbb{R}_+^n). \quad (9.1.17)$$

Da $F_\alpha(v) = (w, D^\alpha v)_0 = (-1)^m (D^\alpha w, v)_0$, bedeutet (9.1.17), dass $D^\alpha w \in H^{1-m}(\mathbb{R}_+^n)$ und $|D^\alpha w|_{1-m} \leq C$ für $|\alpha| = m$. Gemäß Lemma 9.1.16 folgt $w \in H^1(\mathbb{R}_+^n)$. Die Koerzitivität von $a(\cdot, \cdot)$ impliziert die gleichmäßige Elliptizität von $L = \sum_{|\alpha|=|\beta|=m} (-1)^m D^\beta a_{\alpha\beta} D^\alpha$, d.h. $\sum a_{\alpha\beta} \xi^{\alpha+\beta} \geq \varepsilon |\xi|^{2m}$ (vgl. Satz 7.2.13). Für $\xi = (0, \dots, 0, 1)$ erhält man $a_{\hat{\alpha}\hat{\alpha}}(\mathbf{x}) \geq \varepsilon$. Damit folgt $D^{\hat{\alpha}} u \in H^1(\mathbb{R}_+^n)$ aus $w \in H^1(\mathbb{R}_+^n)$ und $D^\gamma a_{\hat{\alpha}\hat{\alpha}} \in L^\infty(\mathbb{R}_+^n)$, $|\gamma| = 1$. Nach Teil 1) gehören alle übrigen Ableitungen $D^\alpha u$ ($|\alpha| \leq m$, $\alpha \neq \hat{\alpha}$) ohnehin zu $H^1(\mathbb{R}_+^n)$, sodass $u \in H^{m-1}(\mathbb{R}_+^n)$ bewiesen ist.

3) *Nachweis von (9.1.17).* Für jedes $\alpha \neq \hat{\alpha}$ gibt es ein γ mit $|\gamma| = 1$, $\gamma_n = 0$, $0 \leq \gamma \leq \alpha$ (komponentenweise Ungleichungen). Partielle Integration liefert

$$F_\alpha(v) = - \int_{\mathbb{R}_+^n} [(D^\gamma a_{\hat{\alpha}\hat{\alpha}})(D^{\hat{\alpha}} u)(D^{\alpha-\gamma} v) + a_{\hat{\alpha}\hat{\alpha}}(D^{\hat{\alpha}+\gamma} u)(D^{\alpha-\gamma} v)] d\mathbf{x} \quad \text{für alle } v \in C_0^\infty(\mathbb{R}_+^n),$$

also $|F_\alpha(v)| \leq C_\alpha |v|_{m-1}$ mit $C_\alpha := C [|u|_m + |D^\gamma u|_m]$ für alle $v \in C_0^\infty(\mathbb{R}_+^n)$. Hierbei wurde ausgenutzt, dass $D^\gamma u \in H^m(\mathbb{R}_+^n)$ nach Teil 1). Da $C_0^\infty(\mathbb{R}_+^n)$ dicht in $H_0^{m-1}(\mathbb{R}_+^n)$ liegt, folgt (9.1.17) für $\alpha \neq \hat{\alpha}$.

Es bleibt $\alpha = \hat{\alpha}$ zu untersuchen. Wir schreiben

$$F_{\hat{\alpha}}(v) = a(u, v) - \hat{a}(u, v) \quad \text{mit} \quad \hat{a}(u, v) := \sum' \int_{\mathbb{R}_+^n} a_{\alpha\beta} (D^\alpha u) (D^\beta v) d\mathbf{x},$$

wobei \sum' die Summation über alle Paare $(\alpha, \beta) \neq (\hat{\alpha}, \hat{\alpha})$ darstellt. Zu jedem $(\alpha, \beta) \neq (\hat{\alpha}, \hat{\alpha})$ gibt es ein γ mit

$$|\gamma| = 1, \quad 0 \leq \gamma \leq \beta, \quad \alpha + \gamma \neq (0, \dots, 0, m+1).$$

Wie oben integriert man jeden Summanden mit $|\beta| = m$ partiell:

$$\int_{\mathbb{R}_+^n} a_{\alpha\beta} (D^\alpha u) (D^\beta v) \, d\mathbf{x} = - \int_{\mathbb{R}_+^n} (D^\gamma a_{\alpha\beta}) (D^\alpha u) (D^{\beta-\gamma} v) \, d\mathbf{x} - \int_{\mathbb{R}_+^n} a_{\alpha\beta} (D^{\alpha+\gamma} u) (D^{\beta-\gamma} v) \, d\mathbf{x}$$

für $v \in C_0^\infty(\mathbb{R}_+^n)$, und schätzt durch $C|v|_{m-1}$ mit $C = C(u)$ ab. Insgesamt erhält man $|\hat{a}(u, v)| \leq C|v|_{m-1}$. Zusammen mit $|a(u, v)| = |(f, v)_0| \leq |f|_{-m+1} |v|_{m-1}$ folgt (9.1.17) auch für $\alpha = \hat{\alpha}$.

4) Durch Induktion ($k = 2, \dots$) beweist man ebenso $|F_\alpha(v)| \leq C|v|_{m-k}$ und hieraus $u \in H^{k+m}(\mathbb{R}_+^n)$. Für reelle $s > 0, s \neq 1/2, \dots, m-1/2$, beweist man entsprechend $|F_\alpha(v)| \leq C|v|_{m-s}$ und $u \in H^{s+m}(\mathbb{R}_+^n)$. ■

Die Verallgemeinerung des Satzes 9.1.15 auf inhomogene Randwerte lautet wie folgt.

Satz 9.1.18 Die Bilinearform $a(\cdot, \cdot)$ aus (9.1.5) sei $H_0^m(\mathbb{R}_+^n)$ -koerziv. Für ein $s > 0, s \notin \{1/2, \dots, m-1/2\}$, gelte entweder (9.1.6), falls $s = k \in \mathbb{N}$, oder (9.1.12), falls $s \notin \mathbb{N}$. $u \in H^m(\mathbb{R}_+^n)$ sei schwache Lösung des inhomogenen Dirichlet-Problems

$$a(u, v) = (f, v)_0 \quad \text{für alle } v \in H_0^m(\mathbb{R}_+^n), \quad (9.1.18a)$$

$$\partial^\ell u / \partial n^\ell = \varphi_\ell \quad \text{auf } \Gamma = \partial\mathbb{R}_+^n \text{ für } \ell = 0, 1, \dots, m-1, \quad (9.1.18b)$$

wobei

$$f \in H^{-m+s}(\mathbb{R}_+^n), \quad \varphi_\ell \in H^{m+s-\ell-1/2}(\Gamma) \quad (0 \leq \ell \leq m-1). \quad (9.1.19)$$

Dann gehört u zu $H^{m+s}(\mathbb{R}_+^n)$ und genügt der Ungleichung

$$|u|_{m+s} \leq C_s \left[|f|_{-m+s} + \sum_{\ell=0}^{m-1} |\varphi_\ell|_{m+s-\ell-1/2} + |u|_m \right]. \quad (9.1.20)$$

Beweis. Für $m = 1$ garantiert Satz 6.2.33 die Existenz eines $u_0 \in H^{m+s}(\mathbb{R}_+^n)$, das die Randbedingungen (9.1.18b) erfüllt (für $m > 1$ vgl. Wloka [89, p. 133]). $w := u - u_0$ ist die Lösung des homogenen Problems $a(w, v) = F(v) := (f, v)_0 - a(u_0, v)$ (vgl. Bemerkung 7.3.2). Satz 9.1.15 lässt sich auch für die hier vorliegende rechte Seite $F(v)$ (anstelle von $(f, v)_0$) übertragen und liefert $w \in H^{m+s}(\mathbb{R}_+^n)$. ■

Mit ähnlichen Mitteln beweist man

Satz 9.1.19 (natürliche Randbedingungen) Die Bilinearform $a(\cdot, \cdot)$ aus (9.1.5) sei $H^1(\mathbb{R}_+^n)$ -koerziv. Für ein $s > 0$ gelte entweder (9.1.6), falls $s = k \in \mathbb{N}$, oder (9.1.12), falls $s \notin \mathbb{N}$. $u \in H^1(\mathbb{R}_+^n)$ sei schwache Lösung der Aufgabe

$$a(u, v) = f(v) := \int_{\mathbb{R}_+^n} g(\mathbf{x}) v(\mathbf{x}) \, d\mathbf{x} + \int_{\mathbb{R}_+^n} \varphi(\mathbf{x}) v(\mathbf{x}) \, d\Gamma \quad \text{für alle } v \in H^1(\mathbb{R}_+^n),$$

wobei

$$g \in H^{s-1} := \left\{ \begin{array}{ll} H^{s-1}(\mathbb{R}_+^n) & \text{für } s \geq 1 \\ (H^{1-s}(\mathbb{R}_+^n))' & \text{für } s < 1 \end{array} \right\}, \quad \varphi \in H^{s-1/2}(\Gamma).$$

Dann gehört u zu $H^{1+s}(\mathbb{R}_+^n)$ und genügt der Abschätzung

$$|u|_{1+s} \leq C_s \left[\|g\|_{H^{s-1}} + |\varphi|_{s-1/2} + |u|_1 \right]. \quad (9.1.21)$$

Ist ein Randwertproblem in der Form (9.1.1): $Lu = g, Bu = \varphi$ mit $B = b^T \Delta + b_0, |b_n(\mathbf{x})| \geq \varepsilon > 0$ auf $\Gamma = \partial\mathbb{R}_+^n$ gegeben, so kann man gemäß Satz 7.4.11 eine zugehörige Variationsformulierung finden und Satz 9.1.19 anwenden.

9.1.4 Regularitätssätze für allgemeines $\Omega \subset \mathbb{R}^n$

Die folgenden Sätze zeigen, dass die bisherigen Regularitätsaussagen auch für $\Omega \subset \mathbb{R}^n$ gültig sind, wenn Ω hinreichend glatt berandet ist.

Satz 9.1.20 Sei $\Omega \in C^{t+m}$ für ein $t \geq 0$. Die Bilinearform (9.1.5) sei $H_0^m(\Omega)$ -koerziv. $s \geq 0$ erfülle

$$s + 1/2 \notin \{1, 2, \dots, m\}, \quad 0 \leq s \leq t, \text{ falls } t \in \mathbb{N}, \quad 0 \leq s < t, \text{ falls } t \notin \mathbb{N}.$$

Für die Koeffizienten gelte

$$D^\gamma a_{\alpha\beta} \in L^\infty(\Omega) \quad \text{für alle } \alpha, \beta, \gamma \text{ mit } |\gamma| \leq \max\{0, t + |\beta| - m\}, \quad \text{falls } t \notin \mathbb{N},$$

$$a_{\alpha\beta} \in \begin{cases} C^{t+|\beta|-m}(\overline{\Omega}) & \text{für } |\beta| > m - t \\ L^\infty(\Omega) & \text{sonst} \end{cases}, \quad \text{falls } t \in \mathbb{N}. \quad (9.1.22)$$

Dann gehört jede schwache Lösung $u \in H_0^m(\Omega)$ der Aufgabe

$$a(u, v) = \int_{\Omega} f(\mathbf{x})v(\mathbf{x})d\mathbf{x} \quad \text{für alle } v \in H_0^m(\Omega)$$

mit $f \in H^{-m+s}(\Omega)$ zu $H^{m+s}(\Omega) \cap H_0^m(\Omega)$ und erfüllt die Abschätzung

$$|u|_{m+s} \leq C_s [|f|_{-m+s} + |u|_m]. \quad (9.1.23)$$

Bei inhomogenen Randbedingungen

$$\partial^\ell u / \partial n^\ell = \varphi_\ell \quad \text{mit } \varphi_\ell \in H^{m+s-\ell-1/2}(\Gamma) \quad \text{für } \ell = 0, 1, \dots, m-1$$

anstelle von $u \in H_0^m(\Omega)$, impliziert $u \in H^m(\Omega)$ die Aussage $u \in H^{m+s}(\Omega)$ und die Abschätzung (9.1.20).

Beweis. 1) $\{U^i : i = 0, 1, \dots, N\}$ mit $U^i \subset \Omega$ sei eine Überdeckung von Ω wie in Lemma 6.2.37. $\{\chi_i\}$ mit $\chi_i = \sigma_i^2 \in C^\infty(\Omega)$, $Tr(\chi_i) \subset U^i$, sei die zugehörige Partition der Eins aus Lemma 6.2.38. Es gibt Abbildungen $\alpha^i \in C^t(U^i)$, die U^i ($i \geq 1$) in \mathbb{R}_+^n abbilden, sodass $\alpha^i(\partial U^i \cap \Gamma) \subset \partial \mathbb{R}_+^n$. Dagegen liegt U^0 im Inneren von Ω , sodass $\Gamma \cap \partial U^0 = \emptyset$.

Die Lösung u lässt sich als $\sum_i \chi_i u$ darstellen. In Teil 2) werden wir $\chi_0 u$, in Teil 3) $\chi_i u$ für $i \geq 1$ untersuchen.

2) Wir setzen

$$d_i(u, v) := a(\chi_i u, v) - a(u, \chi_i v) \quad (i = 0, 1, \dots, N)$$

und wollen die Abschätzung

$$|d_0(u, v)| \leq C_d |u|_m |v|_{m-s} \quad (u \in H_0^m(\Omega), v \in H_0^{m-s}(\Omega), s \geq 1) \quad (9.1.24)$$

für $s = 1$ zeigen. Jeder Summand von $d_0(u, v)$ hat die Form

$$\int_{U^0} a_{\alpha\beta}(\mathbf{x}) [(D^\alpha(\chi_0 u)) (D^\beta v) - (D^\alpha u) (D^\beta(\chi_0 v))] d\mathbf{x}.$$

Da $D^\alpha(\chi_0 u) = \chi_0 D^\alpha u$ + niedrigere Ableitungen von u , gilt

$$[(D^\alpha(\chi_0 u)) (D^\beta v) - (D^\alpha u) (D^\beta(\chi_0 v))] = \sum_{\gamma, \delta} c_{\gamma\delta} D^\gamma u D^\delta v \quad \text{mit } |\gamma|, |\delta| \leq m, |\gamma| + |\delta| \leq 2m - 1.$$

Den Term $\int_{U^0} a_{\alpha\beta} c_{\gamma\delta} D^\gamma u D^\delta v dx$ mit $|\delta| = m$ integriert man partiell und erhält eine Abschätzung durch $C |u|_m |v|_{m-1}$, woraus (9.1.24) folgt.

Die Koeffizienten $a_{\alpha\beta}$ lassen sich so auf \mathbb{R}^n fortsetzen, dass sie die entsprechenden Bedingungen (9.1.22) auf \mathbb{R}^n erfüllen. Die entstehende Bilinearform sei mit $\bar{a}_0(u, v)$ bezeichnet. Da $\chi_0 \in C^\infty(\Omega)$ einen Träger $Tr(\chi_0) \subset U^0$ besitzt, bereitet die Fortsetzung von $\chi_0 u$ durch $(\chi_0 u)(\mathbf{x}) = 0$ für $\mathbf{x} \in \mathbb{R}^n \setminus U^0$ keine Schwierigkeiten. $d_0(u, v)$ kann formal für $v \in H^{m-s}(\mathbb{R}^n)$ erklärt werden, da nur die Beschränkung von v auf U^0 eine Rolle spielt. Wegen (9.1.24) lässt sich $d_0(u, v)$ für festes $u \in H_0^m(\Omega)$ in der Form

$$d_0(u, v) = (d_0, v)_{L^2(\mathbb{R}^n)} \quad \text{mit } d_0 \in H^{s-m}(\mathbb{R}^n), |d_0|_{s-m} \leq C_d |u|_m$$

schreiben. $\chi_0 u$ ist schwache Lösung von

$$\bar{a}_0(\chi_0 u, v) = a(\chi_0 u, v) = a(u, \chi_0 v) + d_0(u, v) = (f, \chi_0 v)_0 + (d_0, v)_0 = (\chi_0 f + d_0, v)_{L^2(\mathbb{R}^n)} \quad (v \in H^m(\mathbb{R}^n)).$$

Sätze 9.1.3 bzw. 9.1.12 beweisen $\chi_0 u \in H^{m+s}(\mathbb{R}^n)$ (damit auch $\chi_0 u \in H^{m+s}(\Omega)$) und

$$\begin{aligned} |\chi_0 u|_{m+s} &\leq C [|\chi_0 f|_{s-m} + |d_0|_{s-m} + |\chi_0 u|_m] \\ &\leq C' [|f|_{s-m} + C_d |u|_m + C_0 |u|_m] \leq C'' [|f|_{s-m} + |u|_m], \end{aligned} \quad (9.1.25a)$$

wobei s zunächst noch durch $s \leq 1$ beschränkt ist.

3) Die gleichen Überlegungen für $\chi_i u$ ($i = 1, \dots, N$) zeigen

$$a(\chi_i u, v) = (\chi_i f + d_i, v)_0 \quad \text{für alle } v \in H^m(\mathbb{R}^n) \text{ mit } |d_i|_{s-m} \leq C_d |u|_m.$$

Nach Voraussetzung gehören die Abbildung $\alpha^i : U^i \rightarrow \mathbb{R}_+^n$ und ihre Inverse $(\alpha^i)^{-1}$ zu $C^{t+m}(\bar{U}^i)$ bzw. $C^{t+m}(\alpha^i(\bar{U}^i))$. Setze $\tilde{u}(\tilde{\mathbf{x}}) := u(\mathbf{x})$ für $\tilde{\mathbf{x}} = \alpha^i(\mathbf{x})$, d.h. $\tilde{u} = u \circ (\alpha^i)^{-1}$. Ebenso seien $\tilde{a}_{\alpha\beta}, \tilde{\chi}_i, \tilde{f}, \tilde{d}_i$ definiert. In

$$a(\chi_i u, v) = \sum_{\alpha, \beta} \int_{\alpha^i(U^i)} \tilde{a}_{\alpha\beta} D_x^\alpha (\tilde{\chi}_i \tilde{u}) D_x^\beta \tilde{v} \left| \det (\alpha^i)' \right|^{-1} d\tilde{\mathbf{x}}$$

kann man wegen $t \geq 0$ die Ableitungen D_x^α, D_x^β durch Ableitungen nach den neuen Koordinaten \tilde{x} ausdrücken und erhält damit eine Form

$$a_i(\tilde{\chi}_i \tilde{u}, \tilde{v}) = \sum_{\alpha, \beta} \int_{\alpha^i(U^i)} \hat{a}_{\alpha\beta} D^\alpha (\tilde{\chi}_i \tilde{u}) D^\beta \tilde{v} d\tilde{\mathbf{x}},$$

die wiederum $H_0^m(\alpha^i(U^i))$ -koerziv ist und deren neue Koeffizienten $\hat{a}_{\alpha\beta}$ die (9.1.22) entsprechenden Bedingungen erfüllen. Wie in 2) lassen sich $\hat{a}_{\alpha\beta}$ so auf $\mathbb{R}_+^n \supset \alpha^i(U^i)$ fortsetzen, dass die entstehende Bilinearform $\bar{a}_i(\cdot, \cdot)$ $H_0^m(\mathbb{R}_+^n)$ -koerziv ist. Auf

$$\bar{a}_i(\tilde{\chi}_i \tilde{u}, \tilde{v}) = a_i(\tilde{\chi}_i \tilde{u}, \tilde{v}) = a(\chi_i u, v) = (\chi_i f + d_i, v)_0 = \left(\frac{\tilde{\chi}_i \tilde{f} + \tilde{d}_i}{|\det (\alpha^i)'|}, \tilde{v} \right)_{L^2(H_0^m(\mathbb{R}_+^n))} \quad \text{für } \tilde{v} \in H_0^m(\mathbb{R}_+^n)$$

ist Satz 9.1.15 anwendbar und liefert $\tilde{\chi}_i \tilde{u} \in H^{m+s}(\mathbb{R}_+^n)$ und

$$|\tilde{\chi}_i \tilde{u}|_{m+s} \leq \tilde{C}_s \left[|\tilde{f}|_{s-m} + |\tilde{d}_i|_{s-m} + |\tilde{\chi}_i \tilde{u}|_m \right].$$

Rücktransformation (vgl. Sätze 6.2.18, 6.2.26g) ergibt

$$|\chi_i u|_{m+s} \leq C_s [|f|_{s-m} + |d_i|_{s-m} + |\chi_i u|_m] \leq C_s [|f|_{s-m} + C_d |u|_m + C |u|_m],$$

also

$$|\chi_i u|_{m+s} \leq C_s [|f|_{s-m} + |u|_m] \quad \text{für alle } 1 \leq i \leq N, \quad 0 \leq s \leq 1. \quad (9.1.25b)$$

4) (9.1.25a,b) gelten für $s \leq 1$. Da $|u|_{m+s} = |\sum_i \chi_i u|_{m+s} \leq \sum_i |\chi_i u|_{m+s}$, ist Abschätzung (9.1.23) für $s \leq 1$ bewiesen. Falls die Voraussetzungen des Satzes ein $s \in (1, 2]$ zulassen, beweist man (9.1.23) wie folgt. Da schon $u \in H^{m+1}(\Omega)$ nachgewiesen ist, kann man die Formen $d_i(u, v)$ nach einer weiteren partiellen Integration durch $|d_i(u, v)| \leq C_d |u|_{m+1} |v|_{m-s}$ abschätzen. Dementsprechend gilt $d_i(u, v) = (d_i, v)_0$ mit $d_i \in H^{-m+s}(\Omega)$ und $|d_i|_{-m+s} \leq C_d |u|_{m+1}$. Setzt man die bisherige Abschätzung (9.1.23) für $s = 1$ ein, erhält man (9.1.25a,b) und damit auch (9.1.23) für $1 < s \leq 2$. Weitere Induktion liefert (9.1.23) für die zulässigen $s \in (k, k+1]$.

5) Der Fall inhomogener Randwerte wird wie in Satz 9.1.18 behandelt. ■

Analog beweist man

Satz 9.1.21 (natürliche Randbedingungen) Sei $\Omega \in C^{t+1}$ mit $t \geq 0$. Solange $0 \leq s \leq t \in \mathbb{N}$ oder $0 \leq s < t \notin \mathbb{N}$, kann in Satz 9.1.19 \mathbb{R}_+^n durch Ω ersetzt werden.

Übungsaufgabe 9.1.22 Man übertrage die Zusätze 9.1.9, 9.1.10, 9.1.11 auf die Situation der Sätze 9.1.20 und 9.1.21. Welche Bedingungen an Ω kommen hinzu?

Zusatz 9.1.23 Ω und die Koeffizienten von $a(\cdot, \cdot)$ mögen die Voraussetzungen aus Satz 9.1.20 bzw. 9.1.21 erfüllen. Eine Eigenfunktion, d.h. eine Lösung $u \in V$ ($V = H_0^m(\Omega)$ im Falle von Satz 9.1.20, $V = H^1(\Omega)$ im Falle von Satz 9.1.21) von

$$a(u, v) = 0 \quad \text{für alle } v \in V, \quad u \neq 0 \quad (9.1.26)$$

gehört zu $H^{m+s}(\Omega)$ für alle $0 \leq s \leq t \in \mathbb{N}$ oder $0 \leq s < t \notin \mathbb{N}$.

Beweis. Man wende Satz 9.1.20 (9.1.21) für $f = 0$ (und $\varphi = 0$) an. ■

Gemäß Satz 6.2.31 (Lemma von Sobolev) erhält man über $C^{k+\lambda}(\overline{\Omega}) \supset H^{k+\lambda+n/2}(\Omega)$ hinreichende Bedingungen dafür, dass u eine klassische Lösung aus $C^{k+\lambda}(\overline{\Omega})$ ist.

Die minimalen Bedingungen für $u \in C^{k+\lambda}(\overline{\Omega})$ ergeben sich aus einem anderen theoretischen Zugang, der im Wesentlichen auf Schauder zurückgeht. Der folgende Satz findet sich z.B. bei Miranda [63, p. 166]. Er zeigt die $C^{k+\lambda}$ -Regularität des Operators L aus (5.1.1a).

Satz 9.1.24 Seien $k \geq 2$ und $0 < \lambda < 1$. $\Omega \in C^{k+\lambda}$ sei ein beschränktes Gebiet. Der Differentialoperator $L = \sum a_{ij} \partial^2 / \partial x_i \partial x_j + \sum a_i \partial / \partial x_i + a$ sei gleichmäßig elliptisch in Ω (d.h. (5.1.3a) gelte). Es gelte $a_{ij}, a_i, a \in C^{k-2+\lambda}(\overline{\Omega})$ und $f \in C^{k-2+\lambda}(\overline{\Omega})$, $\varphi \in C^{k+\lambda}(\Gamma)$. Dann hat die Randwertaufgabe $Lu = f$ in Ω , $u = \varphi$ auf Γ entweder eine eindeutige (klassische) Lösung $u \in C^{k+\lambda}(\overline{\Omega})$, oder es existiert ein endlich-dimensionaler Eigenraum $\{0\} \neq E \subset C^{k+\lambda}(\overline{\Omega})$, sodass für alle $e \in E$ gilt: $Le = 0$ in Ω , $e = 0$ auf Γ . Falls $a \leq 0$, trifft stets die erste Alternative zu.

Die Voraussetzung $\Omega \in C^{t+m}$ aus Satz 9.1.20 ist stärker als notwendig. Für das Dirichlet-Problem reicht schon die Lipschitz-Stetigkeit von Γ , um das folgende Resultat zu erhalten.

Satz 9.1.25 (Nečas [64]) $\Omega \in C^{0,1}$ sei ein beschränktes Gebiet. Die Bilinearform (9.1.5) sei $H_0^m(\Omega)$ -koerziv. Es gelte

$$1/2 \geq t > s > 0.$$

Die Koeffizienten $a_{\alpha\beta} \in L^\infty(\Omega)$ müssen zu $C^t(\overline{\Omega})$ gehören, falls $|\beta| = m$. Dann gehört die schwache Lösung $u \in H_0^m(\Omega)$ der Aufgabe

$$a(u, v) := \int_{\Omega} f(\mathbf{x}) v(\mathbf{x}) d\mathbf{x} \quad \text{für alle } v \in H_0^m(\Omega)$$

mit $f \in H^{-m+s}(\Omega)$ zu $H_0^{m+s}(\Omega)$ und erfüllt die Abschätzung (9.1.23).

Die Voraussetzung der $H_0^m(\Omega)$ -Koerzivität kann durch die gleichmäßige Elliptizität (7.2.3) ersetzt werden (vgl. Sätze 7.2.11, 7.2.13). Die Aussage des Satzes 9.1.25 lässt sich nicht auf $s \geq 1/2$ erweitern, da sonst $u \in H_0^{m+s}(\Omega)$ eine weitere Randbedingung beinhalten würde.

Der Beweis des Satzes 9.1.25 benutzt einen mit $R_{h,j}$ (vgl. Beweis von Satz 9.1.12) verwandten Isomorphismus $R = R^*$ zwischen $H_0^{m+s}(\Omega)$ und $H_0^m(\Omega)$ sowie zwischen $H_0^m(\Omega)$ und $H_0^{m-s}(\Omega)$, sodass die Form $b(u, v) := a(Ru, Rv)$ $H_0^{m+s}(\Omega)$ -koerziv ist. Man hat nun nachzuweisen, dass auch $b(u, v) := a(u, R^2v)$ $H_0^{m+s}(\Omega)$ -koerziv ist. $f \in H^{-m+s}(\Omega)$ impliziert $\tilde{f} := R^2f \in H^{-m-s}(\Omega)$. Jede Lösung von $a(u, v) = (f, v)_0$ ist auch Lösung von $a(u, R^2\tilde{v}) = \tilde{b}(u, \tilde{v}) = (\tilde{f}, \tilde{v})_0 = (f, R^2\tilde{v})_0$, sodass $u \in H_0^{m+s}(\Omega)$ folgt.

9.1.5 Regularität bei konvexem Gebiet und Gebieten mit Ecken

Ein Gebiet Ω ist konvex, falls mit $\mathbf{x}', \mathbf{x}'' \in \Omega$ auch $\mathbf{x}' + t(\mathbf{x}'' - \mathbf{x}')$ für alle $0 \leq t \leq 1$ zu Ω gehört. Konvexe Gebiete gehören insbesondere zu $C^{0,1}$, erlauben aber stärkere Regularitätsaussagen als Satz 9.1.25.

Satz 9.1.26 (Kadlec [48]) Ω sei beschränkt und konvex. Die Bilinearform (9.1.5) sei $H_0^1(\Omega)$ -koerziv. Die Koeffizienten des Hauptteiles seien Lipschitz-stetig:

$$a_{\alpha\beta} \in C^{0,1}(\overline{\Omega}) \quad \text{für alle } |\alpha| = |\beta| = 1,$$

für die übrigen gelte:

$$a_{\alpha\beta} \in L^\infty(\Omega) \quad \text{für alle } \alpha, \beta, \gamma \text{ mit } \gamma \leq |\beta|, |\alpha| + |\beta| \leq 1.$$

Dann gehört jede schwache Lösung $u \in H_0^1(\Omega)$ der Aufgabe

$$a(u, v) := \int_{\Omega} f(\mathbf{x})v(\mathbf{x})d\mathbf{x} \quad \text{für alle } v \in H_0^1(\Omega)$$

mit $f \in L^2(\Omega)$ zu $H^2(\Omega) \cap H_0^1(\Omega)$ und erfüllt die Abschätzung

$$|u|_2 \leq C_1 [|f|_0 + |u|_1]. \quad (9.1.27)$$

Die Konstante C_1 ist nur vom Durchmesser von Ω abhängig.

Zum Beweis verweisen wir auf die Originalarbeit oder Grivard [35]. Der Beweis für den Spezialfall aus Zusatz 9.1.27 wird explizit vorgeführt.

Die H^2 -Regularität wurde in Abschnitt 8.4.4 benötigt. Eine Verallgemeinerung der Aussage (9.1.27) in der Form der H^{m+1} -Regularität für die biharmonische Differentialgleichung mit $m = 2$ ist für konvexe Polygone bekannt (vgl. Blum-Rannacher [11]). Für den Fall der Poisson-Gleichung lässt sich die Ungleichung (9.1.27) hinsichtlich der Konstanten konkreter formulieren.

Zusatz 9.1.27 Für die Lösung der Poisson-Gleichung $-\Delta u = f \in L^2(\Omega)$ im konvexen Gebiet Ω mit $u = 0$ auf Γ gilt

$$\sqrt{\sum_{|\alpha|=2} |D^\alpha u|_0^2} \leq |f|_0. \quad (9.1.28)$$

Ähnlich wie in Lemma 8.4.1 zeigt man, dass die linke Seite von (9.1.28) eine zu $|\cdot|_2$ äquivalente Norm von $H^2(\Omega) \cap H_0^1(\Omega)$ ist. Somit folgt (9.1.27). Als Modell des Beweises von Satz 9.1.26 führen wir den Beweis zu Zusatz 9.1.27 für den Fall $\Omega \subset \mathbb{R}^2$ durch.

Beweis des Zusatzes 9.1.27. 1) Zunächst sei vorausgesetzt, dass das konvexe Gebiet glatt ist: $\Omega \in C^\infty$. Nach Satz 9.1.20 hat $-\Delta u = f \in L^2(\Omega)$ eine Lösung $u \in H^2(\Omega) \cap H_0^1(\Omega)$. Wir wollen

$$\|u\|_2^2 := \sum_{|\alpha|=2} |D^\alpha u|_0^2 \leq |\Delta u|_0^2 \quad \text{für alle } u \in H^2(\Omega) \cap H_0^1(\Omega) \quad (9.1.29)$$

zeigen. Dazu genügt es, (9.1.29) für alle u aus der dichten Teilmenge $\{u \in C^\infty(\Omega) : u = 0 \text{ auf } \Gamma\} = C^\infty(\Omega) \cap H_0^1(\Omega)$ nachzuweisen. Partielle Integration liefert

$$\begin{aligned} \iint_{\Omega} |\Delta u|^2 dx dy &= \iint_{\Omega} (u_{xx}^2 + u_{yy}^2 + 2u_{xx}u_{yy}) dx dy = \text{partielle Integration des dritten Terms bezüglich } y \\ &= \iint_{\Omega} (u_{xx}^2 + u_{yy}^2 - 2u_{xxy}u_y) dx dy + 2 \int_{\Gamma} u_{xx}u_y n_y d\Gamma = \text{partielle Integration bezüglich } x \\ &= \iint_{\Omega} (u_{xx}^2 + u_{yy}^2 + 2u_{xy}^2) dx dy + 2 \int_{\Gamma} u_{xx}u_y n_y d\Gamma - 2 \int_{\Gamma} u_{xy}u_y n_x d\Gamma \\ &= \|u\|_2^2 + 2 \int_{\Gamma} (u_{xx}n_y - u_{xy}n_x) u_y d\Gamma, \end{aligned}$$

wobei $\vec{n} = (n_x, n_y)$ der Normalenvektor ist. Die Tangentenrichtung ist durch $\vec{t} = (-n_y, n_x)$ gegeben. $u_{xx}n_y - u_{xy}n_x = -(u_x)_t$ ist die negative Tangentialableitung. u_x und u_y lassen sich durch u_t und u_n ausdrücken. Da mit u auch u_t auf Γ verschwindet, ist $u_x = n_x u_n$ und $u_y = n_y u_n$. Das Randintegral wird somit zu

$$2 \int_{\Gamma} (u_{xx}n_y - u_{xy}n_x) u_y d\Gamma = -2 \int_{\Gamma} (n_x u_n)_t n_y u_n d\Gamma = - \int_{\Gamma} [2u_n^2 n_y (n_x)_t + n_x n_y (u_n^2)_t] d\Gamma.$$

Partielle Integration des zweiten Summanden ergibt

$$- \int_{\Gamma} [2u_n^2 n_y (n_x)_t + n_x n_y (u_n^2)_t] d\Gamma = \int_{\Gamma} u_n^2 [(n_y)_t n_x - (n_x)_t n_y] d\Gamma.$$

Die eckige Klammer in der letzten Darstellung ist die Krümmung von Γ in $\mathbf{x} \in \Gamma$, die für ein konvexes Gebiet stets ≥ 0 ist. Damit ist das Integral nichtnegativ und (9.1.28) bewiesen.

2) Jedes konvexe Gebiet Ω kann durch konvexe $\Omega_\nu \in C^\infty$ monoton approximiert werden:

$$\Omega_1 \subset \Omega_2 \subset \dots \subset \Omega, \quad \bigcup_\nu \Omega_\nu = \Omega.$$

Wir fassen $V_\nu := \{u \in H_0^1(\Omega) : \text{Tr}(u) \subset \overline{\Omega}_\nu\}$ als Ritz-Galerkin-Raum $V_\nu \subset H_0^1(\Omega)$ auf. Jedes $u \in C_0^\infty(\Omega)$ liegt in V_μ für hinreichend großes μ . Mit $C_0^\infty(\Omega)$ ist deshalb auch $\bigcup_\nu V_\nu$ eine dichte Teilmenge von $H_0^1(\Omega)$. Für jedes ν liefert die Ritz-Galerkin-Aufgabe die Lösung $u_\nu \in H_0^1(\Omega_\nu)$ von $-\Delta u_\nu = f$ in Ω_ν , $u_\nu = 0$ auf $\partial\Omega_\nu$. In $\Omega \setminus \Omega_\nu$ wird u_ν durch $u_\nu = 0$ fortgesetzt. Satz 8.2.2, der auch im Falle $\dim V_\nu = \infty$ anwendbar ist, beweist $|u_\nu - u|_1 \rightarrow 0$, wobei $u \in H_0^1(\Omega)$ die Lösung von $-\Delta u = f$ in Ω ist. Satz 9.1.30 wird zeigen, dass für jedes μ die Beschränkung von u auf Ω_μ zu $H^2(\Omega_\mu)$ gehört. Für jedes $v \in V_\mu \subset V_\nu$, $\nu \geq \mu$, ist

$$\begin{aligned} \left| \int_{\Omega_\mu} u_{xx} v dx dy \right| &= \left| \int_{\Omega_\mu} u_x v_x dx dy \right| = \left| \lim_{\substack{\nu \rightarrow \infty \\ \nu \geq \mu}} \int_{\Omega_\mu} (u_\nu)_x v_x dx dy \right| = \left| \lim_{\nu \rightarrow \infty} \int_{\Omega_\mu} (u_\nu)_x v_x dx dy \right| \\ &= \left| \lim_{\nu \rightarrow \infty} \int_{\Omega_\mu} (u_\nu)_{xx} v dx dy \right| \leq \sup_{\nu \geq \mu} \|(u_\nu)_{xx}\|_{L^2(\Omega_\mu)} \|v\|_{L^2(\Omega_\mu)}. \end{aligned}$$

Hieraus schließt man $\sum_{|\alpha|=2} \|D^\alpha u\|_{L^2(\Omega_\mu)}^2 \leq \sup_{\nu \geq \mu} \sum_{|\alpha|=2} \|D^\alpha u_\nu\|_{L^2(\Omega_\mu)}^2 \leq \|f\|_{L^2(\Omega_\mu)}^2 \leq \|f\|_{L^2(\Omega)}^2$ (vgl. (9.1.29)) und erhält (9.1.28). ■

Welche Rolle die H^2 -Regularität für die $H_0^1(\Omega)$ -Projektion auf einen Unterraum $V_h \subset H_0^1(\Omega)$ spielt, zeigt

Übungsaufgabe 9.1.28 Der Unterraum $V_h \subset H_0^1(\Omega)$ erfülle

$$\inf\{|u - v|_1 : v \in V_h\} \leq C_0 h |u|_2 \quad \text{für alle } u \in H^2(\Omega) \cap H_0^1(\Omega)$$

(vgl. (8.4.16)). Das Poisson-Problem sei H^2 -regulär (hinreichend ist nach Satz 9.1.25, dass Ω konvex ist). $Q_V : H_0^1(\Omega) \rightarrow V_h \subset H_0^1(\Omega)$ sei die bezüglich $|\cdot|_1$ orthogonale Projektion auf V_h . Man zeige: Es gibt ein C_1 , sodass

$$|u - Q_V u|_0 \leq C_1 h |u|_1 \quad \text{für alle } u \in H_0^1(\Omega).$$

Hinweise: 1) Mit dem Poisson-Problem ist auch die Randwertaufgabe $-\Delta u + u = f$ in Ω und $u = 0$ auf Γ H^2 -regulär (vgl. Bemerkung 9.1.2). Die zugehörige Bilinearform $a(\cdot, \cdot)$ ist das Skalarprodukt in $H_0^1(\Omega)$.

2) Q_V stimmt mit der Ritz-Projektion S_h für $a(\cdot, \cdot)$ überein.

3) Man wende Zusatz 8.4.12 an.

Im Zusammenhang mit finiten Elementen betrachtet man häufig Polygone Ω . Da Polygone zu $C^{0,1}$ gehören, ist die Dirichlet-Aufgabe nach Satz 9.1.25 H^{1+s} -regulär mit $0 \leq s < 1/2$. Ist das Polygon konvex (d.h. sind die Innenwinkel $\leq \pi$), so trifft H^2 -Regularität zu. Resultate zwischen $H^{3/2}$ und H^2 erhält man, wenn der maximale Innenwinkel des Polygons zwischen π und 2π liegt (vgl. Schatz-Wahlbin [74]). Sobald Ω eine einspringende Ecke besitzt, kann das Randwertproblem (mit $m = 1$) nicht mehr H^2 -regulär sein (vgl. Beispiel 2.1.4; bei einem Innenwinkel α gehört die Lösung zu $H^{1+s}(\Omega)$ mit $s < \pi/\alpha$). Allerdings lassen sich höhere Regularitätseigenschaften erzielen, wenn in den Ecken spezielle Verträglichkeitsbedingungen erfüllt sind (vgl. Kondrat'ev [53]).

Beispiel 9.1.29 u sei Lösung der Poisson-Gleichung $-\Delta u = f$ im Quadrat $\Omega = (0, 1) \times (0, 1)$ mit $u = 0$ auf Γ . $f \in H^{s-2}(\Omega)$ führt nur für $s < 3$ zu $u \in H^s(\Omega)$. Unter der zusätzlichen Verträglichkeitsbedingung, dass f in allen Ecken verschwindet: $f(0, 0) = f(0, 1) = f(1, 0) = f(1, 1) = 0$, kann man dagegen auch für $f \in H^{s-2}(\Omega)$ mit $3 < s < 4$ auf $u \in H^s(\Omega)$ schließen.

9.1.6 Regularität im Inneren

Bisher bezogen sich die Regularitätsaussagen auf das Gesamtgebiet Ω . Stattdessen kann man auch nach der Regularität der Lösung u in $\Omega_0 \subset\subset \Omega$ fragen. Der folgende Satz zeigt, dass die Antwort weder von der Glattheit des Randes noch von der Art der Randbedingung abhängt.

Satz 9.1.30 Seien $\Omega_0 \subset\subset \Omega_1 \subset \Omega$ und $s \geq 0$. Die Bilinearform (9.1.5) sei $H_0^m(\Omega_1)$ -koerziv. Für die Koeffizienten gelte Bedingung (9.1.22) mit Ω_1 statt Ω und mit $t \geq s \in \mathbb{N}$ oder $t > s$. Sei $u \in V \subset H^m(\Omega)$

eine schwache Lösung der Aufgabe $a(u, v) = \int_{\Omega} f v d\mathbf{x}$ für alle $v \in V$, wobei die Beschränkung $f|_{\Omega_1}$ zu $H^{-m+s}(\Omega_1)$ gehöre. Dann gehört die Beschränkung von u auf Ω_0 zu $H^{m+s}(\Omega_0)$ und erfüllt

$$\|u\|_{H^{m+s}(\Omega_0)} \leq C(s, \Omega_0, \Omega_1, \Omega) \left[\|f\|_{H^{-m+s}(\Omega_1)} + \|u\|_{H^{m+s}(\Omega)} \right].$$

Beweis. Eine spezielle Überdeckung von Ω ist durch $U^0 = \Omega_1$, $U^1 = \Omega \setminus \Omega_0$ gegeben. Damit ergibt sich die Behauptung aus Teil 2) des Beweises zu Satz 9.1.20. ■

9.2 Regularitätseigenschaften der Differenzengleichungen

Die Konvergenzabschätzungen aus Abschnitt 4.4 für Differenzengleichungen lauteten zum Beispiel $\|u - u_h\|_{\infty} \leq Ch^2 \|u\|_{C^4(\overline{\Omega})}$ unter der Voraussetzung, dass $u \in C^4(\overline{\Omega})$ oder $u \in C^{3,1}(\overline{\Omega})$. Diese Regularitätsannahme ist oft nicht erfüllt (vgl. Beispiele 2.1.3-2.1.4). Ein Vergleich mit der für die Finite-Element-Methode bewiesene Fehlerabschätzung $|u - u^h|_k \leq Ch^{2-k} |u|_2$ ($k = 0, 1$) legt die Vermutung nahe, dass eine ähnliche Ungleichung auch für Differenzlösungen existiert. Um diese zu gewinnen, muss man die Stabilitätsabschätzung $\|L_h^{-1}\|_2 \leq C$ (oder $\|L_h^{-1}\|_{\infty} \leq C$), die $L^{-1} \in L(L^2(\Omega), L^2(\Omega))$ entspricht, durch stärkere Abschätzungen ersetzen, die $L^{-1} \in L(H^{-1}(\Omega), H_0^1(\Omega))$ oder $L^{-1} \in L(L^2(\Omega), H^2(\Omega))$ entsprechen.

9.2.1 Diskrete H^1 -Regularität

Für $\Omega \subset \mathbb{R}^n$ werden ein unendliches Gitter Q_h und ein in Ω enthaltenes Gitter Ω_h definiert:

$$Q_h := \{\mathbf{x} \in \mathbb{R}^n : x_i = \nu_i h, \nu_i \in \mathbb{Z}\}, \quad \Omega_h := \Omega \cap Q_h. \quad (9.2.1)$$

Eine auf Ω_h definierte Gitterfunktion v_h wird auf Q_h durch $v_h = 0$ erweitert:

$$v_h(\mathbf{x}) := 0 \quad \text{für alle } \mathbf{x} \in Q_h \setminus \Omega_h. \quad (9.2.2)$$

Der Vektorraum aller Gitterfunktionen v_h mit (9.2.2) wird mit $L_h^2 = L_h^2(\Omega_h)$ bezeichnet. Die *Euklidische Norm* kann jetzt als L_h^2 -Norm interpretiert werden:

$$|v_h|_0 := \|v_h\|_{L_h^2} := \sqrt{h^n \sum_{\mathbf{x} \in Q_h} |v_h(\mathbf{x})|^2}. \quad (9.2.3a)$$

Sie gehört zum L_h^2 -Skalarprodukt

$$(v_h, w_h)_0 := (v_h, w_h)_{L_h^2} := h^n \sum_{\mathbf{x} \in Q_h} v_h(\mathbf{x}) w_h(\mathbf{x}). \quad (9.2.3b)$$

Das diskrete Analogon von $H_0^1(\Omega)$ ist H_h^1 mit der Norm

$$|v_h|_1 := \|v_h\|_{H_h^1} := \sqrt{|v_h|_0^2 + \sum_{i=1}^n |\partial_i^+ v_h|_0^2} \quad (\text{man beachte } v_h = 0 \text{ auf } Q_h \setminus \Omega_h), \quad (9.2.3c)$$

wobei ∂_i^+ die Vorwärtsdifferenz in x_i -Richtung ist. Die Dualnorm lautet

$$|v_h|_{-1} := \|v_h\|_{H_h^{-1}} := \sup \{ |(v_h, w_h)_0| / |w_h|_1 : w_h \neq 0 \text{ erfülle (9.2.2)} \}. \quad (9.2.3d)$$

Die zugehörigen Matrixnormen $\|L_h\|_{H_h^1 \leftarrow H_h^{-1}} = |L_h|_{1 \leftarrow -1}$, $\|L_h\|_{H_h^1 \leftarrow L_h^2} = |L_h|_{1 \leftarrow 0}$, usw. sind durch

$$|L_h|_{i \leftarrow j} := \sup \left\{ |L_h v_h|_i / |v_h|_j : 0 \neq v_h \text{ erfülle (9.2.2)} \right\} \quad \text{für } i, j \in \{-1, 0, 1\} \quad (9.2.3e)$$

definiert.

Übungsaufgabe 9.2.1 Man zeige: a) $|L_h|_{0 \leftarrow 0}$ ist die Spektralnorm von L_h (vgl. §4.3).
b) Es gelten die inversen Abschätzungen

$$|v_h|_i \leq C_{ij} h^{j-i} |v_h|_j \quad \text{für } 1 \geq i \geq j \geq -1. \quad (9.2.3f)$$

Der Differenzenoperator L_h gibt Anlass zur Bilinearform

$$a_h(v_h, w_h) := (L_h v_h, w_h)_{L_h^2}. \quad (9.2.4a)$$

$a_h(\cdot, \cdot)$ heißt H_h^1 -elliptisch, falls ein $C_E > 0$ existiert, sodass

$$a_h(u_h, u_h) \geq C_E |u_h|_1^2 \quad \text{für alle } u_h \text{ und alle } h > 0. \quad (9.2.4b)$$

Entsprechend heißt $a_h(\cdot, \cdot)$ H_h^1 -koerziv, wenn $C_E > 0$ und $C_K \in \mathbb{R}$ mit

$$a_h(u_h, u_h) \geq C_E |u_h|_1^2 - C_K |u_h|_0^2 \quad \text{für alle } u_h \text{ und alle } h > 0 \quad (9.2.4c)$$

existieren. Wie in Abschnitt 4.5 definiert, heißt L_h (bzw. $a_h(\cdot, \cdot)$) L_h^2 -stabil, falls

$$|L_h^{-1}|_{0 \leftarrow 0} \leq C_0 \quad \text{für alle } h > 0. \quad (9.2.4d)$$

Wir nennen L_h darüber hinaus H_h^1 -regulär, falls

$$|L_h^{-1}|_{1 \leftarrow -1} \leq C_1 \quad \text{für alle } h > 0. \quad (9.2.4e)$$

Übungsaufgabe 9.2.2 a) H_h^1 -Regularität impliziert L_h^2 -Stabilität.

b) $|A_h|_{i \leftarrow j} = |A_h^\top|_{-j \leftarrow -i}$ für alle $i, j \in \{-1, 0, 1\}$.

c) Ist L_h H_h^1 -regulär, so auch L_h^\top .

d) Sind L_h und L_h^\top stabil bezüglich $|\cdot|_\infty$, d.h. $\|L_h\|_\infty \leq C_\infty$, $\|L_h^\top\|_\infty \leq C_1$, so folgt die L_h^2 -Stabilität (9.2.4d) mit $C_0 := \sqrt{C_1 C_\infty}$.

e) H_h^1 -Elliptizität impliziert H_h^1 -Regularität.

Die folgende Aussage ähnelt der Alternative aus Satz 6.5.15.

Satz 9.2.3 Ist $a_h(\cdot, \cdot)$ H_h^1 -koerziv, und ist L_h L_h^2 -stabil, so ist L_h auch H_h^1 -regulär.

Beweis. Im Folgenden ist u_h beliebig, aber fest. Sei $f_h := L_h^\top u_h$, sodass $a_h(u_h, u_h) = (u_h, L_h^\top u_h)_0 = (u_h, f_h)_0$. Die Koerzivitt liefert

$$|u_h|_1^2 \leq [a_h(u_h, u_h) + C_K |u_h|_0^2] / C_E = [(u_h, f_h)_0 + C_K |u_h|_0^2] / C_E \leq C' [|f_h|_0 + C_K |u_h|_0] |u_h|_0.$$

Auf Grund der Stabilittsabschtzung $|u_h|_0 \leq C_0 |f_h|_0$ ergibt sich $|u_h|_1^2 \leq C'' |f_h|_0^2$. Hieraus schliet man $|L_h^{-\top}|_{1 \leftarrow 0} \leq \sqrt{C''} =: C^*$ und somit $|L_h^{-1}|_{0 \leftarrow -1} \leq C^*$ (vgl. Übungsaufgabe 9.2.2b).

2) Sei nun $f_h := L_h u_h$. Nach Teil 1) gilt $|u_h|_0 \leq C^* |f_h|_{-1}$. Indem man

$$|a_h(u_h, u_h)| = |(f_h, u_h)_0| \leq |f_h|_{-1} |u_h|_1$$

gem (5.3.10) durch $\frac{1}{2} C_E |u_h|_1^2 + \frac{1}{2} C_E^{-1} |f_h|_{-1}^2$ abschtzt, ergibt sich die Koerzivitt

$$|u_h|_1^2 \leq \frac{a_h(u_h, u_h) + C_K |u_h|_0^2}{C_E} \leq \frac{\frac{1}{2} C_E |u_h|_1^2 + \frac{1}{2} C_E^{-1} |f_h|_{-1}^2 + C_K |u_h|_0^2}{C_E} = \frac{1}{2} |u_h|_1^2 + \frac{\frac{1}{2} C_E^{-1} + C_K (C^*)^2}{C_E} |f_h|_{-1}^2$$

und daher $|L_h^{-1}|_{1 \leftarrow -1} \leq C_1$ mit $C_1 := C_E^{-2} + 2C_K (C^*)^2 C_E^{-1}$. ■

Anstelle der L_h^2 -Stabilitt in Satz 9.2.3 kann man auch die Lsbarkeit des kontinuierlichen Problems und eine Konsistenzbedingung voraussetzen (vgl. Folgerung 11.3.5).

In Analogie zu Lemma 7.2.12 beweist man

Übungsaufgabe 9.2.4 Ist L_h H_h^1 -koerziv, und enthält $\delta L_h := a_0 + \sum_i b_i \partial_i^\pm + \sum_i \partial_i^\pm c_i$ mit $|a_0|, |b_i|, |c_i| \leq \text{const}$ höchstens erste Differenzen, so ist auch $L_h + \delta L_h$ H_h^1 -koerziv.

Gemäß Übungsaufgabe 9.2.4 reicht es, den Hauptteil eines Differenzenoperators auf Koerzivität zu untersuchen.

Beispiel 9.2.5 $\Omega \subset \mathbb{R}^2$ sei beschränkt. Auf Ω_h sei L_h durch das Differenzenverfahren (5.1.18) (mit $a = 0$) gegeben. Dazu seien die Koeffizienten a_{11}, a_{22} auf \mathbb{R}^2 fortgesetzt angenommen (wesentlich sind nur die Werte, die in (9.2.5) für $(x, y) \in \Omega_h$ auftreten):

$$(L_h u_h)(x, y) = \partial_x^- \left(a_{11} \left(x + \frac{h}{2}, y \right) \partial_x^+ u_h(x, y) \right) + \partial_y^- \left(a_{22} \left(x, y + \frac{h}{2} \right) \partial_y^+ u_h(x, y) \right), \quad (9.2.5)$$

wobei $u_h(x, y) = 0$ für $(x, y) \in Q_h \setminus \Omega_h$ gemäß (9.2.2). Dabei gelte $-a_{11}, -a_{22} \geq \varepsilon > 0$ in $\bar{\Omega}$. Dann ist $a_h(\cdot, \cdot)$ H_h^1 -elliptisch und L_h H_h^1 -regulär.

Beweis. 1) Für beliebige auf Q_h definierte Gitterfunktionen v_h, w_h gelten die partiellen Summationsregeln

$$(v_h, \partial_{x_i}^+ w_h)_0 = -(\partial_{x_i}^- v_h, w_h)_0, \quad (v_h, \partial_{x_i}^- w_h)_0 = -(\partial_{x_i}^+ v_h, w_h)_0. \quad (9.2.6)$$

2) Da im Skalarprodukt $(L_h u_h, w_h)_0$ die Werte $(L_h u_h)(x, y)$ für $(x, y) \in Q_h \setminus \Omega_h$ wegen (9.2.2) mit $w_h(x, y) = 0$ multipliziert werden, können die eventuell undefinierten Koeffizienten a_{11}, a_{22} nach Belieben erklärt werden. Es folgt

$$\begin{aligned} a_h(u_h, u_h) &= (L_h u_h, u_h)_0 = - \left(a_{11} \left(\cdot + \frac{h}{2}, \cdot \right) \partial_x^+ u_h, \partial_x^+ u_h \right)_0 - \left(a_{22} \left(\cdot, \cdot + \frac{h}{2} \right) \partial_y^+ u_h, \partial_y^+ u_h \right)_0 \\ &\geq \varepsilon \left[|\partial_x^+ u_h|_0^2 + |\partial_y^+ u_h|_0^2 \right]. \end{aligned}$$

Wie in Lemma 6.2.12 beweist man, dass die Normen $|\cdot|_1$ und $|u_h|_{1,0} := [|\partial_x^+ u_h|_0^2 + |\partial_y^+ u_h|_0^2]^{1/2}$ für beschränktes Ω äquivalent sind (gleichmäßig bezüglich h), sodass (9.2.4b) folgt. Die H_h^1 -Regularität ergibt sich aus Übungsaufgabe 9.2.2e. ■

Übungsaufgabe 9.2.6 $\Omega \subset \mathbb{R}^2$ sei beschränkt. Die Gleichung

$$(a_{11} u_x)_x + (a_{12} u_y)_x + (a_{12} u_x)_y + (a_{22} u_y)_y = f$$

sei auf Ω_h durch die Differenzensterne (5.1.18-5.1.18) gegeben, wobei $u_h = 0$ in $Q_h \setminus \Omega_h$. Die Differentialgleichung sei gleichmäßig elliptisch in $\bar{\Omega}$: $a_{ii} < 0$, $a_{11}a_{22} - a_{12}a_{21} \geq -\varepsilon(a_{11} + a_{22}) > 0$. Ferner gelte $a_{ij} \in C^0(\bar{\Omega})$. Man zeige: Für hinreichend kleines h ist die zugehörige Matrix L_h H_h^1 -regulär. Ferner ist L_h für alle $h > 0$ H_h^1 -koerziv. Hinweis: Für hinreichend kleines h gilt für alle $d_1, d_2 \in \mathbb{R}$, dass

$$-a_{11} \left(x + \frac{h}{2}, y \right) d_1^2 - a_{22} \left(x, y + \frac{h}{2} \right) d_2^2 - \left[a_{12} \left(x + \frac{h}{2}, y + h \right) + a_{12} \left(x + \frac{h}{2}, y \right) \right] d_1 d_2 \geq \frac{\varepsilon}{2} (d_1^2 + d_2^2).$$

Die bisher konstruierten H_h^1 -koerziven Differenzenverfahren bleiben H_h^1 -koerziv, wenn Differenzen niedriger Ordnung hinzugenommen werden (vgl. Übungsaufgabe 9.2.4) oder wenn der Hauptterm $(a_{11} u_x)_x + \dots$ durch $a_{11} u_{xx} + \dots$ mit $a_{11} \in C^1(\bar{\Omega})$ ersetzt wird. Die erwähnten Differenzenverfahren werden in randnahen wie randfernen Gitterpunkten durch ein und denselben Differenzenoperator beschrieben. Die homogene Dirichlet-Randbedingung wird durch (9.2.2): “ $u_h = 0$ auf $Q_h \setminus \Omega_h$ ” diskretisiert. Will man die Randbedingung genauer approximieren, muss man in den randnahen Punkten von Ω_h spezielle Diskretisierungen wählen (vgl. Abschnitte 4.8.1-2). Man erhält somit eine Unregelmäßigkeit, die einen Beweis der H_h^1 -Regularität wie in Beispiel 9.2.5 erschwert.

Wir beginnen mit dem eindimensionalen Fall.

Lemma 9.2.7 L_h sei die Matrix der eindimensionalen Shortley-Weller-Diskretisierungen von $-u'' = f$ auf Ω und der Randbedingung $u = 0$ auf $\partial\Omega$:

$$h^{-2} \left\{ \frac{2}{s_\ell s_r} u_h(x) - \frac{2}{s_\ell (s_\ell + s_r)} u_h(x - s_\ell h) - \frac{2}{s_r (s_\ell + s_r)} u_h(x + s_r h) \right\} = f(x) \quad \text{für } x \in \Omega_h, \quad (9.2.7)$$

wobei $0 < s_\ell, s_r \leq 1$ (vgl. (4.8.7)) im Allgemeinen den Wert 1 haben und nur für den ersten [letzten] Gitterpunkt $x \in \Omega_h$ die Werte so definiert sind, dass $x - s_\ell h \in \partial\Omega$ [$x + s_r h \in \partial\Omega$]. In den Randpunkten $\xi \in \partial\Omega$ wird $u_h(\xi) = 0$ gesetzt. Für beliebiges $\Omega \subset \mathbb{R}$ gilt $(v_h, L_h v_h)_0 \geq \frac{1}{2} |\partial^+ v_h|_0^2$, wenn $v_h = 0$ auf $Q_h \setminus \Omega_h$ und h hinreichend klein. Für beschränktes Ω ist L_h daher H_h^1 -regulär.

Beweis. 1) Zunächst sei Ω als zusammenhängend angenommen. Die Gitterpunkte von Ω_h seien

$$x_j := x_0 + jh \in \Omega \quad \text{für } j = 0, \dots, k > 0.$$

Die Randpunkte von Ω seien $x_0 - s_{\ell,0}h$ und $x_k + s_{r,k}h$ mit $s_{\ell,0}, s_{r,k} \in (0, 1]$. Die weiteren Faktoren der Gleichung (9.2.7) in $x = x_i$ sind $s_{\ell,i} = s_{r,i} = 1$. Beachtet man $v_h = 0$ auf $\mathbb{R} \setminus \Omega$, erhält man folgende Identität:

$$\begin{aligned} (v_h, L_h v_h)_0 &= h \sum_{j=0}^k v_h(x_j) (L_h v_h)(x_j) \\ &= |\partial^+ v_h|_0^2 + \frac{1}{h} \left\{ v_h(x_0) \left[\left(\frac{2}{s_\ell} - 2 \right) v_h(x_0) + \left(1 - \frac{2}{1+s_\ell} \right) v_h(x_1) \right] \right. \\ &\quad \left. + v_h(x_k) \left[\left(\frac{2}{s_r} - 2 \right) v_h(x_k) + \left(1 - \frac{2}{1+s_r} \right) v_h(x_{k-1}) \right] \right\}, \end{aligned} \quad (9.2.8a)$$

wobei $s_\ell := s_{\ell,0}$ und $s_r := s_{r,k}$. Da $v_h(x_0) = h\partial^+ v_h(x_{-1})$ und $v_h(x_1) = h[\partial^+ v_h(x_{-1}) + \partial^+ v_h(x_0)]$ wegen $v_h(x_{-1}) = 0$, lässt sich der erste Summand der geschweiften Klammer als

$$\begin{aligned} &\frac{1}{h} v_h(x_0) \left[\left(\frac{2}{s_\ell} - 2 \right) v_h(x_0) + \left(1 - \frac{2}{1+s_\ell} \right) v_h(x_1) \right] \\ &= h \frac{1-s_\ell}{1+s_\ell} v_h(x_0) \left[\frac{2+s_\ell}{s_\ell} (\partial^+ v_h(x_{-1}))^2 - \partial^+ v_h(x_{-1}) \partial^+ v_h(x_0) \right] \end{aligned} \quad (9.2.8b)$$

schreiben. Für $\alpha := \partial^+ v_h(x_{-1})$, $\beta := \partial^+ v_h(x_0)$ wendet man die Ungleichung $-\alpha\beta \geq -\frac{\lambda}{2}\alpha^2 - \frac{1}{2\lambda}\beta^2$ mit $\frac{\lambda}{2} = (2+s_\ell)/s_\ell$ an (vgl. (5.3.10)): $\frac{2+s_\ell}{s_\ell}\alpha^2 - \alpha\beta \geq -\frac{1/4}{\lambda/2}\beta^2 = -\frac{1}{4}\frac{s_\ell}{s_\ell+2}\beta^2$. Da

$$s(1-s)/[4(1+s)(2+s)] \leq 0.018$$

(Maximum bei $s = (\sqrt{3}-1)/2$), folgt, dass der Ausdruck (9.2.8b) durch $\geq 0.018h\beta^2$ nach unten beschränkt ist. Analog behandelt man den zweiten Summanden der geschweiften Klammer und erhält

$$(v_h, L_h v_h)_0 \geq |\partial^+ v_h|_0^2 - 0.018h \left[(\partial^+ v_h(x_0))^2 + (\partial^+ v_h(x_{k-1}))^2 \right]. \quad (9.2.8c)$$

2) In Teil 1) wurde $k > 0$ angenommen. Für $k = 0$ ist

$$(v_h, L_h v_h)_0 = \frac{1}{h} \frac{2}{s_\ell s_r} (v_h(x_0))^2 = \frac{h}{s_\ell s_r} \left[(\partial^+ v_h(x_0))^2 + (\partial^+ v_h(x_{-1}))^2 \right] \geq |\partial^+ v_h|_0^2,$$

sodass (9.2.8c) ebenfalls gilt.

3) Sei Ω beliebig. Die Zusammenhangskomponenten seien die Intervalle $I_i = (a_i, b_i)$ ($i \in \mathbb{Z}$) mit $b_i \leq a_{i+1}$. $L_h^{(i)}$ sei die zu I_i gehörende Diskretisierungsmatrix. Jedes v_h mit Träger in $\Omega_h = Q_h \cap \Omega$ kann als $\sum_i v_h^{(i)}$ geschrieben werden, wobei $v_h^{(i)} = 0$ außerhalb von I_i . Seien $x_0^{(i)}$ und $x_{k(i)}^{(i)}$ der erste und letzte Gitterpunkt von $\Omega_h^{(i)} := \Omega_h \cap I_i$. Sei h hinreichend klein, sodass $0.018h \leq 1/2$. Die Abschätzung

$$\begin{aligned} |\partial^+ v_h|_0^2 &\leq 2 \sum_i \left\{ |\partial^+ v_h|_0^2 - 0.018h \left[(\partial^+ v_h(x_0^{(i)}))^2 + (\partial^+ v_h(x_{k(i)-1}^{(i)}))^2 \right] \right\} \\ &\leq 2 \sum_i \left(v_h^{(i)}, L_h^{(i)} v_h^{(i)} \right)_0 = 2 (v_h, L_h v_h)_0 \end{aligned}$$

beweist die Behauptung. ■

Satz 9.2.8 $\Omega \subset \mathbb{R}^2$ sei beschränkt. L_h sei die zur Shortley-Weller-Diskretisierung der Poisson-Gleichung gehörende Matrix (vgl. Abschnitt 4.8.1). Dann ist L_h H_h^1 -regulär.

Beweis. Sei L_h^x (L_h^y) der Anteil der x -Differenzen (y -Differenzen), sodass $L_h = L_h^x + L_h^y$. Die Beschränkung von L_h auf eine "Gitterlinie" $\{(\nu h, y) \in \Omega : \nu \in \mathbb{Z}\}$ (y fest) entspricht der Matrix L_h aus Lemma 9.2.7. Somit erhält man $(v_h, L_h^x v_h)_0 \geq \frac{1}{2} |\partial_x^+ v_h|_0^2$. Mit der analogen Ungleichung $(v_h, L_h^y v_h)_0 \geq \frac{1}{2} |\partial_y^+ v_h|_0^2$ ergibt sich $2(v_h, L_h v_h)_0 \geq |\partial_x^+ v_h|_0^2 + |\partial_y^+ v_h|_0^2 = |v_h|_1^2 - |v_h|_0^2 \geq \varepsilon_\Omega |v_h|_1^2$, also $\|L_h^{-1}\|_{1 \leftarrow -1} \leq 1/(2\varepsilon_\Omega)$. ■

Satz 9.2.9 $\Omega \subset \mathbb{R}^2$ sei beschränkt. Die Poisson-Gleichung sei mit Hilfe der Fünfpunktformel (4.8.14c) in randfernen und durch die Interpolation (4.8.16) in randnahen Punkten diskretisiert. Die zugehörige Matrix ist H_h^1 -regulär.

Beweis. Die entsprechenden, eindimensionalen Formeln stimmen bis auf den Skalierungsfaktor $s_r + s_\ell \leq 2$ mit (9.2.7) überein, sodass sich der Beweis des Satzes 9.2.8 leicht übertragen lässt. ■

Übungsaufgabe 9.2.10 Man zeige: Für ein konvexes Gebiet $\Omega \subset \mathbb{R}^2$ erfüllt die Matrix L_h der Shortley-Weller-Diskretisierung die Ungleichung

$$(v_h, L_h v_h)_0 \geq 0.982 \left[|\partial_x^+ v_h|_0^2 + |\partial_y^+ v_h|_0^2 \right].$$

Die Sätze 9.2.8 und 9.2.9 lassen sich in folgender Weise verstärken:

Zusatz 9.2.11 L_h sei wie in Satz 9.2.8 oder 9.2.9. $D_h = \text{diag}\{d(\mathbf{x}) : \mathbf{x} \in \Omega_h\}$ sei die Diagonalmatrix mit

$$d(\mathbf{x}) = \min \{2s_\ell s_r, 2s_o s_u, 1\} \quad (\mathbf{x} \in \Omega_h),$$

wobei s_ℓ, s_r, s_o, s_u aus (4.8.7) bzw. (4.8.16) stammen. Für randferne $x \in \Omega_h$ gilt offenbar $d(\mathbf{x}) = 1$. Die Matrix

$$L'_h := D_h L_h,$$

die zum umskalierten System $L'_h u_h = f'_h := D_h f_h$ gehört, ist ebenfalls H_h^1 -regulär: $|L_h^{-1}|_{1 \leftarrow -1} \leq C$.

Beweis. Im Falle des Shortley-Weller-Verfahrens (4.8.7) hat man die Korrekturterme (9.2.8b) für die x - und y -Richtungen zusammenzufassen. Es ergibt sich, dass sie $\geq -\frac{1}{2} [|\partial_x^+ v_h|_0^2 + |\partial_y^+ v_h|_0^2]$ bleiben, sodass die Ungleichung $(v_h, L'_h v_h)_0 \geq \frac{1}{2} [|\partial_x^+ v_h|_0^2 + |\partial_y^+ v_h|_0^2] \geq \varepsilon |v_h|_1^2$ mit $\varepsilon > 0$ folgt. Eine analoge Abschätzung erhält man für das Differenzenverfahren aus Abschnitt 4.8.2. ■

9.2.2 Konsistenz

Im Folgenden wollen wir die für Finite-Element-Fehlerabschätzung $|u^h - u|_1 \leq Ch |u|_2 \leq C'h |f|_0$ auf Differenzenverfahren übertragen. Dazu ist die *Konsistenzbedingung*

$$\left| L_h R_h - \tilde{R}_h L \right|_{-1 \leftarrow 2} := \left\| L_h R_h - \tilde{R}_h L \right\|_{H_h^{-1} \leftarrow H^2(\Omega)} \leq C_K h \quad (9.2.9)$$

für geeignete *Restriktionen*

$$R_h : H^2(\Omega) \cap H_0^1(\Omega) \rightarrow H_h^2, \quad \tilde{R}_h : L^2(\Omega) \rightarrow L_h^2$$

zu beweisen. Zur Konstruktion der Restriktionen wird zunächst $u \in H^2(\Omega)$ in $\bar{u} := E_2 u \in H^2(\mathbb{R}^2)$ fortgesetzt. Gemäß Satz 6.2.41c wird

$$\bar{u} := E_2 u = u \quad \text{auf } \Omega, \quad \|E_2 u\|_{H^2(\mathbb{R}^2)} \leq C \|u\|_{H^2(\Omega)} \quad \text{für alle } u \in H^2(\Omega) \quad (9.2.10a)$$

angenommen. Eine analoge Fortsetzung $E_0 : L^2(\Omega) \rightarrow L^2(\mathbb{R}^2)$ mit

$$\bar{f} := E_0 f = f \quad \text{auf } \Omega, \quad \|E_0 f\|_{L^2(\mathbb{R}^2)} \leq C \|f\|_{L^2(\Omega)} \quad \text{für alle } f \in L^2(\Omega) \quad (9.2.10b)$$

ist z.B. durch $\bar{f} = 0$ auf $\mathbb{R}^2 \setminus \Omega$ gegeben. Die Operatoren $\sigma_h^x, \sigma_h^y : C_0^\infty(\mathbb{R}^2) \rightarrow C_0(\mathbb{R}^2)$ seien durch

$$(\sigma_h^x u)(x, y) := \frac{1}{h} \int_{-h/2}^{h/2} u(x + \xi, y) d\xi, \quad (\sigma_h^y u)(x, y) := \frac{1}{h} \int_{-h/2}^{h/2} u(x, y + \eta) d\eta \quad (9.2.11)$$

definiert. Die Restriktionen R_h, \tilde{R}_h werden wie folgt gewählt:

$$R_h := \sigma_h^x \sigma_h^y E_2, \quad \text{d.h.} \quad (R_h u) = h^{-2} \int_{-h/2}^{h/2} \int_{-h/2}^{h/2} \bar{u}(x + \xi, y + \eta) d\xi d\eta \quad \text{mit } \bar{u} := E_2 u, \quad (9.2.12a)$$

$$\tilde{R}_h := (\sigma_h^x \sigma_h^y)^2 E_0, \quad (9.2.12b)$$

$$\text{d.h.} \quad (\tilde{R}_h u) = h^{-4} \int_{-h/2}^{h/2} \int_{-h/2}^{h/2} \int_{-h/2}^{h/2} \int_{-h/2}^{h/2} \bar{f}(x + \xi + \xi', y + \eta + \eta') d\xi d\xi' d\eta d\eta' \quad \text{mit } \bar{f} := E_0 f.$$

Die charakteristischen Eigenschaften der Faltungen σ_h^x, σ_h^y sind Gegenstand der

Übungsaufgabe 9.2.12 $\hat{\partial}_x$ sei die symmetrische Differenz $(\hat{\partial}_x u)(x, y) := [u(x + h/2, y) - u(x - h/2, y)]/h$. $\hat{\partial}_y$ ist analog definiert. Man zeige:

- a) $\sigma_h^x \sigma_h^y = \sigma_h^y \sigma_h^x$, $\hat{\partial}_x = \frac{\partial}{\partial x} \sigma_h^x = \sigma_h^x \frac{\partial}{\partial x}$, $\hat{\partial}_y = \frac{\partial}{\partial y} \sigma_h^y = \sigma_h^y \frac{\partial}{\partial y}$.
- b) $\|\sigma_h^x\|_{H^k(\mathbb{R}^2) \rightarrow H^k(\mathbb{R}^2)} \leq C$, $\|\sigma_h^y\|_{H^k(\mathbb{R}^2) \rightarrow H^k(\mathbb{R}^2)} \leq C$ (insbesondere für $k = 0, \pm 1, 2$).
- c) $\|\sigma_h^x \sigma_h^y v\|_{L_h^2} \leq \sqrt{\sum_{\mathbf{x} \in \Omega_h} |(\sigma_h^x \sigma_h^y v)(\mathbf{x})|^2} \leq \|v\|_{L^2(\mathbb{R}^2)}$ für alle $v \in L^2(\mathbb{R}^2)$.
- d) Für $a \in C^{0,1}(\mathbb{R}^2)$ gilt $\|a \sigma_h^x \sigma_h^y - \sigma_h^x \sigma_h^y a\|_{L_h^2 \leftarrow L^2(\mathbb{R}^2)} \leq Ch \|a\|_{C^{0,1}(\mathbb{R}^2)}$. Dabei sind

$$(a \sigma_h^x u)(\mathbf{x}) = a(\mathbf{x}) ((\sigma_h^x u)(\mathbf{x})) \quad \text{und} \quad (\sigma_h^x a u)(\mathbf{x}) = (\sigma_h^x (a u))(\mathbf{x}).$$

- e) $\|a (\sigma_h^x)^\nu (\sigma_h^y)^\mu - (\sigma_h^x)^\nu (\sigma_h^y)^\mu a\|_{L_h^2 \leftarrow L^2(\mathbb{R}^2)} \leq C_{\nu\mu} h \|a\|_{C^{0,1}(\mathbb{R}^2)}$ für $\nu, \mu \in \mathbb{N}$.
- f) $\|u - (\sigma_h^x)^\nu (\sigma_h^y)^\mu u\|_{H^k(\mathbb{R}^2)} \leq C_{\nu\mu} h \|u\|_{H^{k+1}(\mathbb{R}^2)}$ für $u \in H^{k+1}(\mathbb{R}^2)$.

Der Differentialoperator sei

$$L = \sum_{i,j=1}^n a_{ij}(\mathbf{x}) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^n a_i(\mathbf{x}) \frac{\partial^2}{\partial x_i} + a(\mathbf{x}). \quad (9.2.13a)$$

Wir nehmen zunächst $\Omega = \mathbb{R}^n$ an und diskretisieren L durch den regelmäßigen Differenzenoperator

$$L_h = \sum_{i,j=1}^n a_{ij}(\mathbf{x}) \partial_{x_i}^\pm \partial_{x_j}^\pm + \sum_{i=1}^n a_i(\mathbf{x}) \partial_{x_i}^\pm + a(\mathbf{x}) \quad (9.2.13b)$$

bei beliebiger Kombination der \pm -Vorzeichen.

Lemma 9.2.13 Sei $\Omega = \mathbb{R}^n$. Es gelte $a_{ij}, a_j, a \in C^{0,1}(\mathbb{R}^n)$. L und L_h seien durch (9.2.13a,b) gegeben. Dann gilt die Konsistenzabschätzung (9.2.9).

Beweis. Zur Vereinfachung der Notation sei $n = 2$ angenommen. Es ist

$$\tilde{R}_h a_{11} u_{xx}(\mathbf{x}) = (\sigma_h^x \sigma_h^y)^2 a_{11} u_{xx}(\mathbf{x}) = a_{11}(\mathbf{x}) (\sigma_h^x \sigma_h^y)^2 u_{xx}(\mathbf{x}) - \delta_1$$

mit $\|\delta_1\|_{L_h^2} \leq Ch \|a_{11}\|_{C^{0,1}(\mathbb{R}^2)} |u|_2$ (vgl. Übungsaufgabe 9.2.12e,b). Der $a_{11} u_{xx}$ entsprechende Term in (9.2.13b) sei z.B. $a_{11}(\mathbf{x}) \partial_x^+ \partial_x^+$. Übungsaufgabe 9.2.12a zeigt

$$a_{11} (\sigma_h^x \sigma_h^y)^2 u_{xx}(x, y) = a_{11} \hat{\partial}_x \hat{\partial}_x \sigma_h^y \sigma_h^y u(x, y) = a_{11} \partial_x^+ \partial_x^+ \sigma_h^y \sigma_h^y u(x - h, y) = a_{11} \partial_x^+ \partial_x^+ \sigma_h^y \sigma_h^y u(x, y) - a_{11} \partial_x^+ \delta_2$$

mit $\delta_2(x, y) := -\partial_x^+ \sigma_h^y \sigma_h^y [u(x - h, y) - u(x, y)] = -\sigma_h^x \sigma_h^y \sigma_h^y [u_x(x - h, y) - u_x(x, y)]$ und

$$\|\delta_2\|_2 \leq \|\sigma_h^y [u_x(x - h, y) - u_x(x, y)]\|_{L^2(\mathbb{R}^2)} \leq h |u|_2$$

(vgl. Übungsaufgabe 9.2.12c,b). Da $\hat{\partial}_x \hat{\partial}_x = \partial_x^+ \partial_x^-$, entfele der Fehlerterm δ_2 , wenn man $a_{11} u_{xx}$ durch $a_{11}(x) \partial_x^+ \partial_x^-$ approximiert. Schließlich hat man

$$\begin{aligned} a_{11} \partial_x^+ \partial_x^+ \sigma_h^y \sigma_h^y u &= a_{11} \partial_x^+ \partial_x^+ R_h u + a_{11} \partial_x^+ \partial_x^+ [\sigma_h^y \sigma_h^y - \sigma_h^x \sigma_h^x] u \\ &= a_{11} \partial_x^+ \partial_x^+ R_h u + a_{11} \partial_x^+ \sigma_h^x \sigma_h^y [\sigma_h^y - \sigma_h^x] u_x = a_{11} \partial_x^+ \partial_x^+ R_h u - a_{11} \partial_x^+ \delta_3 \end{aligned}$$

mit $\|\delta_3\|_{L_h^2} = \|\sigma_h^x \sigma_h^y [\sigma_h^y - \sigma_h^x] u_x\|_{L_h^2} \leq \|[\sigma_h^y - \sigma_h^x] u_x\|_{L^2(\mathbb{R}^2)} \leq Ch |u|_2$ (vgl. Übungsaufgabe 9.2.12c,f). Zusammen ergibt sich

$$\left[a_{11} \partial_x^+ \partial_x^+ R_h - \tilde{R}_h a_{11} \frac{\partial^2}{\partial x^2} \right] u = \delta_1 + a_{11} \partial_x^+ (\delta_2 + \delta_3).$$

Für den ersten Fehlerterm gilt

$$\|\delta_1\|_{H_h^{-1}} \leq \|\delta_1\|_{L_h^2} \leq Ch \|a_{11}\|_{C^{0,1}(\mathbb{R}^2)} |u|_2. \quad (9.2.14a)$$

Für beliebiges $v_h \in H_h^1$ ist

$$(v_h, a_{11} \partial_x^+ (\delta_2 + \delta_3))_{L_h^2} = -(\partial_x^\mp [a_{11} v_h], \delta_2 + \delta_3)_{L_h^2}. \quad (9.2.14b)$$

Aus $a \in C^{0,1}(\mathbb{R}^2)$ folgt $\|\partial_x^\pm [av_h]\|_{L_h^2} \leq \|a\|_{C^{0,1}(\mathbb{R}^2)} \|v_h\|_{L_h^2} + \|a\|_{C^0(\mathbb{R}^2)} \|v_h\|_{H_h^1} \leq C \|v_h\|_{H_h^1}$, sodass

$$\|a_{11} \partial_x^+ (\delta_2 + \delta_3)\|_{H_h^{-1}} = \sup_{0 \neq v_h \in H_h^1} \left| (\partial_x^\mp [a_{11} v_h], \delta_2 + \delta_3)_{L_h^2} \right| / |v_h|_1 \leq C \|\delta_2 + \delta_3\|_{L_h^2} \leq C' h |u|_2. \quad (9.2.14c)$$

(9.2.14a,c) beweisen

$$\left\| a_{11} \partial_x^+ \partial_x^+ R_h - \tilde{R}_h a_{11} \frac{\partial^2}{\partial x^2} \right\|_{H_h^{-1} \leftarrow H^2(\mathbb{R}^2)} \leq Ch. \quad (9.2.14d)$$

Analog zeigt man

$$\left\| a_{ij} \partial_{x_i}^\pm \partial_{x_j}^\pm R_h - \tilde{R}_h a_{ij} \frac{\partial^2}{\partial x_i \partial x_j} \right\|_{H_h^{-1} \leftarrow H^2(\mathbb{R}^2)} \leq Ch. \quad (9.2.14e)$$

Für $a_i \partial_{x_i}^\pm R_h - \tilde{R}_h a_i \partial / \partial x_i$ ergeben die gleichen Überlegungen eine $\mathcal{O}(h)$ -Abschätzung für die Normen $\|\cdot\|_{H_h^{-1} \leftarrow H^1(\mathbb{R}^2)}$ und $\|\cdot\|_{L_h^2 \leftarrow H^2(\mathbb{R}^2)}$. Beide sind abschätzbar durch die größere Norm $\|\cdot\|_{H_h^{-1} \leftarrow H^2(\mathbb{R}^2)}$, sodass

$$\left\| a_i \partial_{x_i}^\pm R_h - \tilde{R}_h a_i \partial / \partial x_i \right\|_{H_h^{-1} \leftarrow H^2(\mathbb{R}^2)} \leq Ch. \quad (9.2.14f)$$

Ebenso ist

$$\left\| a R_h - \tilde{R}_h a \right\|_{H_h^{-1} \leftarrow H^2(\mathbb{R}^2)} \leq Ch. \quad (9.2.14g)$$

(9.2.14e,f,g) beweisen die Behauptung (9.2.9). ■

Bei der Verallgemeinerung der Konsistenzabschätzung auf allgemeinere Gebiete $\Omega \subset \mathbb{R}^2$ tritt die folgende Schwierigkeit auf. Die Koeffizienten der Matrix L_h nach (4.8.7) oder (4.8.16) sind nicht durch Ch^{-2} beschränkt. In randnahen Punkten geht vielmehr die Inverse des Abstandes zum Randpunkt ein, und dieser Abstand kann beliebig klein sein. Ein Ausweg besteht darin, die Diskretisierung so zu formulieren, dass die Entfernungen zwischen Rand- und randnahen Punkten z.B. $\geq h/2$ bleiben. Eine zweite Möglichkeit ist eine geeignete Definition von R_h , sodass das in (9.2.9) erscheinende Produkt $L_h R_h$ abschätzbar wird (vgl. Hackbusch [38]). Hier soll ein dritter Ausweg beschritten werden: L_h wird durch die umskalierte Matrix $L'_h = D_h L_h$ aus Zusatz 9.2.11 ersetzt.

Satz 9.2.14 *Sei $\Omega \in C^2$ (oder konvex) und beschränkt. Zur Diskretisierung von $Lu = f$ für $L = -\Delta$ auf Ω mit $u = 0$ auf Γ werde die Diskretisierung $L_h u_h = f_h$ gemäß (4.8.7) oder (4.8.16) verwendet. $L'_h = D_h L_h$ sei wie in Zusatz 9.2.11 definiert. Dann gilt die Konsistenzabschätzung*

$$\left| L'_h R_h - D_h \tilde{R}_h L \right|_{-1 \leftarrow 2} \leq Ch. \quad (9.2.15)$$

Die Matrix L_h aus (4.8.7) bzw. (4.8.16) ist hier nur als Beispiel genommen. Der Beweis wird zeigen, dass die Abschätzung (9.2.9) bzw. (9.2.15) auch für andere L_h gilt, wenn $(L_h u_h)(\mathbf{x})$, \mathbf{x} randnah, eine zweite Differenz darstellt. Zunächst werden zwei Lemmata benötigt.

$\gamma_h \subset \Omega_h$ sei die Menge der randnahen Punkte. Ist v_h eine auf Ω_h definierte Gitterfunktion, so bezeichnen wir mit $v_h|_{\gamma_h}$ die Beschränkung der Funktion auf γ_h :

$$(v_h|_{\gamma_h})(\mathbf{x}) = v_h(\mathbf{x}) \quad \text{für } \mathbf{x} \in \gamma_h, \quad (v_h|_{\gamma_h})(\mathbf{x}) = 0 \quad \text{für } \mathbf{x} \in \Omega_h \setminus \gamma_h.$$

Lemma 9.2.15 *Sei $\Omega \in C^{0,1}$ beschränkt. Dann gibt es ein von h unabhängiges $C = C(\Omega)$, sodass*

$$|v_h|_{\gamma_h}|_0 \leq C |v_h|_{\gamma_h}|_1. \quad (9.2.16)$$

Beweis. Aus $\Omega \in C^{0,1}$ folgt: Es gibt Zahlen $K \in \mathbb{N}$ und $h_0 > 0$, sodass für alle $\mathbf{x} \in \gamma_h$ mit $h \leq h_0$ nicht alle Gitterpunkte $\{\mathbf{x} + (\nu h, \mu h) : -K \leq \nu, \mu \leq K\}$ in Ω liegen. Sei zum Beispiel $\mathbf{x} + (\nu_0 h, \mu_0 h) \notin \Omega$. Dann definieren wir die Funktionen $w_h^{\nu\mu}(-K \leq \nu, \mu \leq K)$ in \mathbf{x} durch $w_h^{\nu_0\mu_0}(\mathbf{x}) := v_h(\mathbf{x})$, $w_h^{\nu\mu}(\mathbf{x}) = 0$ für $(\nu, \mu) \neq (\nu_0, \mu_0)$. Damit ist $\sum_{\nu, \mu=-K}^K w_h^{\nu\mu} = v_h|_{\gamma_h}$ eine Zerlegung mit $w_h^{\nu\mu}(\mathbf{x}) = v_h(\mathbf{x})$ oder $w_h^{\nu\mu}(\mathbf{x}) = 0$.

Ohne Beschränkung der Allgemeinheit seien $\nu > 0$ und $\mu > 0$ angenommen. Zu $\mathbf{x} \in \Omega_h$ definieren wir die Punktkette

$$\mathbf{x}^0 = \mathbf{x}, \mathbf{x}^1 = \mathbf{x} + (h, 0), \dots, \mathbf{x}^\nu = \mathbf{x} + (\nu h, 0), \mathbf{x}^{\nu+1} = \mathbf{x} + (\nu h, h), \dots, \mathbf{x}^{\nu+\mu} = \mathbf{x} + (\nu h, \mu h).$$

Definitionsgemäß gilt $w_h^{\nu\mu}(\mathbf{x}) = 0$ oder $\mathbf{x}^{\nu+\mu} \notin \Omega$, d.h. $w_h^{\nu\mu}(\mathbf{x}^{\nu+\mu}) = 0$ (vgl. (9.2.2)). In beiden Fällen hat man

$$\begin{aligned} |w_h^{\nu\mu}(\mathbf{x})| &\leq |w_h^{\nu\mu}(\mathbf{x}^0) - w_h^{\nu\mu}(\mathbf{x}^{\nu+\mu})| = h \left| \frac{1}{h} [w_h^{\nu\mu}(\mathbf{x}^0) - w_h^{\nu\mu}(\mathbf{x}^1)] + \frac{1}{h} [w_h^{\nu\mu}(\mathbf{x}^1) - w_h^{\nu\mu}(\mathbf{x}^2)] + \dots \right| \\ &\leq h \left[|\partial_x^- w_h^{\nu\mu}(\mathbf{x}^1)| + |\partial_x^- w_h^{\nu\mu}(\mathbf{x}^2)| + \dots + |\partial_x^- w_h^{\nu\mu}(\mathbf{x}^\nu)| + |\partial_y^- w_h^{\nu\mu}(\mathbf{x}^{\nu+1})| + \dots + |\partial_y^- w_h^{\nu\mu}(\mathbf{x}^{\nu+\mu})| \right] \end{aligned}$$

und daher $|w_h^{\nu\mu}|_0 \leq h[\sqrt{\nu}|\partial_x^+ w_h^{\nu\mu}|_0 + \sqrt{\mu}|\partial_y^+ w_h^{\nu\mu}|_0] \leq \sqrt{2K}h|w_h^{\nu\mu}|_1$. Summation über ν, μ liefert die Abschätzung (9.2.16) für $v_h|_{\gamma_h} = \sum w_h^{\nu\mu}$. ■

In den meisten Fällen ist schon ein direkter Nachbarpunkt von $x \in \gamma_h$ in $\mathbb{R}^2 \setminus \Omega$, sodass (9.2.16) mit $C = \sqrt{2}$ folgt. Dies gilt insbesondere für konvexe Gebiete.

Lemma 9.2.16 *R_h sei durch (9.2.12a) definiert. E_2 erfülle (9.2.10a). Dann gilt*

$$(R_h u)(\xi) \leq Ch \|E_2 u\|_{H^2(K_{h/2}(\xi))} \quad \text{für alle } \xi \in \Gamma, u \in H_0^1(\Omega) \cap H^2(\Omega), \quad (9.2.17)$$

wobei $K_{h/2}(\xi) := \{\xi + \mathbf{x} \in \mathbb{R}^2 : \mathbf{x} \in (-h/2, h/2) \times (-h/2, h/2)\}$.

Beweis. 1) Sei $Q = (-\frac{1}{2}, \frac{1}{2}) \times (-\frac{1}{2}, \frac{1}{2})$. Für $v \in H^2(Q)$ zeigt man

$$\left| v(\mathbf{0}) - \int_Q v(\mathbf{x}) d\mathbf{x} \right| \leq C \sqrt{\int_Q (v_{xx}^2 + 2v_{xy}^2 + v_{yy}^2) dx dy},$$

da die linke Seite für lineare Funktionen $u(x, y) = \alpha + \beta x + \gamma y$ verschwindet. Der Beweis ist ähnlich wie für (8.4.2). Transformation von Q auf $Q_h := (-\frac{h}{2}, \frac{h}{2}) \times (-\frac{h}{2}, \frac{h}{2})$ liefert

$$\left| h^2 v(\mathbf{0}) - \int_{Q_h} v(\mathbf{x}) d\mathbf{x} \right| \leq Ch^3 \sqrt{\int_{Q_h} (v_{xx}^2 + 2v_{xy}^2 + v_{yy}^2) dx dy}. \quad (9.2.18)$$

2) Sei $\xi \in \Gamma$. Die Behauptung (9.2.17) folgt aus (9.2.18) mit $v(\mathbf{x}) := \bar{u}(\xi + h\mathbf{x}) = (E_2 u)(\xi + h\mathbf{x})$, da $u(\xi) = 0$ wegen $\xi \in \Gamma$. ■

Beweis des Satzes 9.2.14. 1) Ungleichung (9.2.15) ist bewiesen, falls

$$\left| \left(v_h, \left[L_h' R_h - D_h \tilde{R}_h L \right] u \right)_0 \right| \leq Ch \quad (9.2.19a)$$

für alle $v_h \in H_h^1$ und $u \in H^2(\Omega) \cap H_0^1(\Omega)$ mit $|v_h|_1 = |u|_2 = 1$. Zu diesem Zweck wird v_h in

$$v_h = v'_h + v''_h \quad \text{mit} \quad v'_h := v_h|_{\gamma_h}$$

zerlegt. In Teil 2) wird

$$\left| \left(v''_h, \left[L'_h R_h - D_h \tilde{R}_h L \right] u \right)_0 \right| \leq C_1 h \quad (9.2.19b)$$

gezeigt werden. Die weiteren Beweisschritte 3) und 4) werden

$$\left| \left(v'_h, \left[L'_h R_h - D_h \tilde{R}_h L \right] u \right)_0 \right| \leq C_2 h \quad (9.2.19c)$$

ergeben, sodass (9.2.19a) mit $C = C_1 + C_2$ folgt.

2) Lemma 9.2.15 zeigt

$$|v'_h|_0 \leq C_3 h |v_h|_1 = C_3 h. \quad (9.2.19d)$$

Für v''_h erhält man $|v''_h|_1 \leq |v_h|_1 + |v'_h|_1 = 1 + |v'_h|_1$. Die inverse Abschätzung (9.2.3f) liefert $|v'_h|_1 \leq C h^{-1} |v'_h|_0 \leq C C_3$, also

$$|v''_h|_1 \leq C_4. \quad (9.2.19e)$$

\hat{L}_h sei der (regelmäßige) Differenzenoperator auf dem unendlichen Gitter $Q_h = \{(\nu h, \mu h) : \nu, \mu \in \mathbb{Z}\}$. Da der Träger von v''_h aus $Q_h \setminus \gamma_h$ besteht, gilt $(v''_h, L'_h w_h)_0 = (v''_h, \hat{L}_h w_h)_0$ für alle w_h . Außerdem ist $(v''_h, D_h w_h)_0 = (v''_h, w_h)_0$. Dies beweist die erste Gleichheit in

$$\left| \left(v''_h, \left[L'_h R_h - D_h \tilde{R}_h L \right] u \right)_0 \right| = \left| \left(v''_h, \left[\hat{L}_h R_h - D_h \tilde{R}_h L \right] \bar{u} \right)_0 \right| \leq C_5 h |v''_h|_1 |\bar{u}|_2 \leq C_5 h C_4 C_6 =: C_1 h$$

wobei $\bar{u} := E_2 u$ die Fortsetzung von u auf \mathbb{R}^2 ist. Die weiteren Ungleichungen ergeben sich aus Lemma 9.2.13, (9.2.19e) und

$$|\bar{u}|_2 = \|\bar{u}\|_{H^2(\mathbb{R}^2)} \leq C_6 \|u\|_{H^2(\Omega)} = C_6 \quad (9.2.19f)$$

(vgl. (9.2.10a)). Satz 6.2.41c garantiert die Existenz einer Fortsetzung \bar{u} mit (9.2.19f), falls $\Omega \in C^2$. Eine weitere hinreichende Bedingung für (9.2.19f) ist die Konvexität von Ω .

3) Die linke Seite von (9.2.19c) wird in $(v'_h, L'_h R_h u)_0$ und $(v'_h, D_h \tilde{R}_h L u)_0$ aufgespalten. Der erste Term wird in Teil 4) abgeschätzt. Übungsaufgabe 9.2.12c und (9.2.19d) ergeben für den zweiten Term

$$\left| (v'_h, D_h \tilde{R}_h L u)_0 \right| \leq |v'_h|_0 \left| D_h \tilde{R}_h L u \right|_0 \leq C_3 h C' |L u|_0 \leq C_7 h |u|_2 = C_7 h. \quad (9.2.19g)$$

4) Wir setzen $w_h := (L'_h R_h u)|_{\gamma_h}$. Da der Träger von v'_h in γ_h enthalten ist, gilt $(v'_h, L'_h R_h u)_0 = (v'_h, w_h)_0$. L'_h enthält Differenzen bezüglich der x - und y -Richtung. Dementsprechend schreiben wir $w_h = w_h^x + w_h^y$. Im Weiteren beschränken wir uns (a) auf die Shortley-Weller-Diskretisierung, (b) auf den Term w_h^x und (c) auf den Fall, dass

$$\mathbf{x} \in \gamma_h, \quad \mathbf{x}^r = \mathbf{x} + (s_r h, 0) \in \Omega_h \quad (\text{d.h. } s_r = 1), \quad \mathbf{x}^\ell = \mathbf{x} - (s_\ell h, 0) \in \Gamma.$$

Die anderen Fälle sind analog abzuhandeln. Wir setzen

$$\hat{u} := R_h u = \sigma_h^x \sigma_h^y E_2 u \in H^2(\mathbb{R}^2).$$

Die Shortley-Weller-Differenz in x -Richtung lautet

$$\hat{w}_h^x(\mathbf{x}) = d_{\mathbf{x}} \left[\frac{\hat{u}(\mathbf{x}) - \hat{u}(\mathbf{x}^r)}{h s_r} - \frac{\hat{u}(\mathbf{x}^\ell) - \hat{u}(\mathbf{x})}{h s_\ell} \right] / \frac{(s_\ell + s_r) h}{2},$$

wobei $d_{\mathbf{x}}$ das Diagonalelement von D_h ist. Da in der Gleichung $L'_h u_h = f_h$ die Variablen $u_h(\xi)$, $\xi \in \Gamma$ (z.B. $\xi = \mathbf{x}^\ell$), bereits eliminiert sind, hat $w_h^x(\mathbf{x})$ die Gestalt

$$w_h^x(\mathbf{x}) = \hat{w}_h^x(\mathbf{x}) + \frac{2 d_{\mathbf{x}}}{h^2 s_\ell (s_\ell + s_r)} \hat{u}(\mathbf{x}^\ell)$$

Der Faktor $2d_{\mathbf{x}}/[h^2 s_\ell(s_\ell + s_r)]$ bleibt nach Definition von D_h durch $4h^{-2}$ beschränkt. Aus Lemma 9.2.16 und $K_{h/2}(\mathbf{x}^\ell) \subset K_{3h/2}(\mathbf{x})$ schließt man

$$|w_h^x(\mathbf{x}) - \hat{w}_h^x(\mathbf{x})| \leq 4h^{-2} |\hat{u}(\mathbf{x}^\ell)| \leq Ch^{-1} \|\bar{u}\|_{H^2(K_{3h/2}(\mathbf{x}))}. \quad (9.2.19h)$$

Die zweite dividierte Differenz \hat{w}_h^x in $\mathbf{x} = (x, y)$ lässt sich (wegen $s_r = 1$) durch

$$\hat{w}_h^x(x, y) = d_{\mathbf{x}} \int_{-hs_\ell}^h g(t) \hat{u}_{xx}(x+t, y) dt \quad \text{mit} \quad g(t) = \begin{cases} 2(t-h)/(h^2 + s_\ell h^2) & \text{für } 0 \leq t \leq h, \\ -2(s_\ell h + t)/(s_\ell h^2(1 + s_\ell)) & \text{für } -s_\ell h \leq t \leq 0, \\ 0 & \text{sonst,} \end{cases}$$

darstellen. Hieraus schließt man

$$|\hat{w}_h^x(x, y)| \leq d_{\mathbf{x}} \sqrt{\int_{-hs_\ell}^h g^2(t) dt} \sqrt{\int_{-hs_\ell}^h \hat{u}_{xx}^2(x+t, y) dt} \leq \frac{2}{h} \sqrt{\int_{-h}^h \hat{u}_{xx}^2(x+t, y) dt} \leq \frac{C}{h} \|\bar{u}\|_{H^2(K_{\frac{3h}{2}}(\mathbf{x}))} \quad (9.2.19i)$$

(vgl. (6.2.5a)). Aus (9.2.19h,i) und der entsprechenden Abschätzung für w_h^y erhält man die Schranke $|w_h(\mathbf{x})| \leq C_8 h^{-1} \|\bar{u}\|_{H^2(K_{3h/2}(x))}$, sodass

$$|w_h|_0^2 = h^2 \sum_{\mathbf{x} \in \gamma_h} |w_h(\mathbf{x})|^2 \leq C_8^2 \sum_{\mathbf{x} \in \gamma_h} \|\bar{u}\|_{H^2(K_{3h/2}(x))}^2 \leq 9C_8^2 \|\bar{u}\|_{H^2(\mathbb{R}^2)}^2 \leq (3C_8 C_6)^2 =: C_9^2.$$

Damit ist

$$|(v'_h, L'_h R_h u)_0| \leq |(v'_h, w_h)_0| \leq |v'_h|_0 |w_h|_0 \leq C_3 h C_9 =: C_{10} h \quad (9.2.19j)$$

bewiesen. (9.2.19g) und (9.2.19j) ergeben die fehlende Ungleichung (9.2.19c). ■

Bemerkung 9.2.17 Die Beweisschritte zu Satz 9.2.14 lassen sich ebenso für allgemeinere Differenzengleichungen (z.B. mit variablen Koeffizienten wie in Lemma 9.2.13) durchführen.

9.2.3 Optimale Fehlerabschätzungen

Die diskrete Lösung $u_h = L_h^{-1} f_h$ wird im Folgenden mit der Restriktion $u_h^* := R_h u$ der exakten Lösung $u = L^{-1} f$ verglichen. Aus der Darstellung

$$\begin{aligned} u_h - u_h^* &= L_h^{-1} f_h - R_h u = L_h^{-1} (f_h - \tilde{R}_h f) + L_h^{-1} \tilde{R}_h f - R_h u \\ &= L_h^{-1} (f_h - \tilde{R}_h f) - L_h^{-1} (L_h R_h - \tilde{R}_h L) u \end{aligned} \quad (9.2.20)$$

erhält man sofort den

Satz 9.2.18 Es gelte $u \in H^2(\Omega)$ für die Lösung von $Lu = f$. Die rechte Seite f_h der diskreten Gleichung $L_h u_h = f_h$ sei so gewählt, dass

$$\|f_h - \tilde{R}_h f\|_{-1} \leq C_f h. \quad (9.2.21)$$

Ist außerdem L_h H_h^1 -regulär, und gilt die Konsistenzbedingung (9.2.9) mit der Konstanten C_K , so genügt u_h der Fehlerabschätzung

$$\|u_h - u_h^*\|_1 \leq C_1 (C_f + C_K \|u\|_2) h. \quad (9.2.22)$$

Beweis. $\|u_h - u_h^*\|_1 \leq \|L_h^{-1}\|_{1 \leftarrow -1} (\|f_h - \tilde{R}_h f\|_{-1} + \|L_h R_h - \tilde{R}_h L\|_{-1 \leftarrow 2} \|u\|_2).$ ■

Zusatz 9.2.19 a) Ungleichung (9.2.21) ist insbesondere dann erfüllt, wenn $f_h := \tilde{R}_h f$ gewählt wird.

b) Die Wahl $f_h(\mathbf{x}) := f(\mathbf{x})$ für $\mathbf{x} \in \Omega_h$ (vgl. (4.2.6b)) führt auf (9.2.21), falls $f \in C^{0,1}(\bar{\Omega})$ oder $f \in H^2(\Omega)$. In diesen Fällen gilt sogar

$$\|f_h - \tilde{R}_h f\|_0 \leq C \|f_h - \tilde{R}_h f\|_\infty \leq C' h \|f\|_{C^{0,1}(\bar{\Omega})} \quad \text{bzw.} \quad \|f_h - \tilde{R}_h f\|_0 \leq Ch^2 \|f\|_2. \quad (9.2.23)$$

c) In Satz 9.2.18 lassen sich die H_h^1 -Regularität von L_h durch jene von $L'_h = D_h L_h$ (vgl. Zusatz 9.2.11), (9.2.9) durch (9.2.15) und (9.2.21) durch

$$|D_h f_h - D_h \tilde{R}_h f|_{-1} \leq C_f h \quad (9.2.21')$$

ersetzen.

Beweis. Der Beweis von (9.2.23) basiert auf der Ungleichung (9.2.18). ■

Fehlerabschätzungen der Ordnung $\mathcal{O}(h^2)$ lassen sich auf die gleiche Weise ableiten, wenn man Konsistenzbedingungen zweiter Ordnung hat. Diese wären z.B. (9.2.24a) oder (9.2.24b):

$$|L_h R_h - \tilde{R}_h L|_{-2 \leftarrow 2} \leq Ch^2, \quad (9.2.24a)$$

$$|L_h R_h - \tilde{R}_h L|_{-1 \leftarrow 3} \leq Ch^2, \quad (9.2.24b)$$

Bemerkung 9.2.20 Falls $\Omega_h = Q_h$ (d.h. $\Omega = \mathbb{R}^2$) oder $\Omega = (x', x'') \times (y', y'')$, lassen sich die Ungleichungen (9.2.24a,b) ähnlich wie in Lemma 9.2.13 zeigen.

Beispiel 9.2.21 Das Differenzenverfahren aus Beispiel 4.5.8 zeigt quadratische Konvergenz. Dieser Fall lässt sich wie folgt analysieren. Gemäß Bemerkung 9.2.20 zeigt man (9.2.24a). Im folgenden Abschnitt beweisen wir die H_h^2 -Regularität $|L_h^{-1}|_{2 \leftarrow 0} \leq C$, die für die vorliegende symmetrische Matrix L_h mit $|L_h^{-1}|_{0 \leftarrow -2} \leq C$ äquivalent ist (vgl. (6.3.3)). Darstellung (9.2.20) führt auf

$$|u_h - u_h^*|_0 \leq Ch^2 |u|_2. \quad (9.2.25a)$$

Die entsprechende Abschätzung

$$|u_h - u_h^*|_1 \leq Ch^2 |u|_3,$$

die auf (9.2.24b) basiert, scheitert daran, dass die Lösung aus Beispiel 4.5.8 nicht zu $H^3(\Omega)$ gehört (vgl. Beispiel 9.1.29).

Die Verifikation von (9.2.24b) in Anwesenheit von unregelmäßigen Diskretisierungen am Rande wird komplizierter.

9.2.4 H_h^2 -Regularität

Unter geeigneten Voraussetzungen an Ω hat $Lu = f \in L^2(\Omega)$ eine Lösung $u \in H^2(\Omega) \cap H_0^1(\Omega)$: $|u|_2 \leq C |f|_0$ (vgl. z.B. Satz 9.1.22). Für die diskrete Lösung von $L_h u_h = f_h$ stellt sich die entsprechende Frage: Gilt $|u_h|_2 \leq C |f_h|_0$? Zunächst ist die Norm $|\cdot|_2$ von H_h^2 zu definieren.

Falls $\Omega = \mathbb{R}^2$ oder falls der Rand Γ mit Gitterlinien zusammenfällt, lässt sich

$$|u_h|_2 := \sqrt{|u_h|_1^2 + h^2 \sum_{\mathbf{x} \in \Omega_h} \left[|\partial_x^+ \partial_x^- u_h(\mathbf{x})|^2 + |\partial_y^+ \partial_y^- u_h(\mathbf{x})|^2 \right]}$$

definieren. Im Allgemeinen muss man dagegen am Rand unregelmäßige Differenzen verwenden. In randfernen Punkten $\mathbf{x} \in \Omega_h \setminus \gamma_h$ sei $D_{xx} u_h(x) = \partial_x^+ \partial_x^- u_h(\mathbf{x})$. In einem randnahen Gitterpunkt $\mathbf{x} = (x, y) \in \gamma_h$ zum Beispiel mit den Nachbarn $\mathbf{x}^\ell := (x - s_\ell h, y) \in \Gamma$, $\mathbf{x}^r := (x + h, y) \in \Omega_h$ läge es nahe, die dividierte Differenz (4.8.5) zu verwenden. Dann wäre jedoch die inverse Abschätzung $|u_h|_2 \leq Ch^{-1} |u_h|_1$ im Allgemeinen verletzt, da s_ℓ beliebig klein werden kann. Deshalb wird $D_{xx} u_h(\mathbf{x})$ mit Hilfe der Gitterpunkte $(x - s_\ell h, y)$, $(x + h, y)$, $(x + 2h, y)$ definiert:

$$D_{xx} u_h(\mathbf{x}) := \frac{2}{(2 + s_\ell) h} \left[\frac{u_h(x + 2h, y) - u_h(x + h, y)}{h} - \frac{u_h(x + h, y) - u_h(x - s_\ell h, y)}{(1 + s_\ell) h} \right],$$

wobei $u_h(x - s_\ell h, y) = 0$.

Falls Γ nicht nur zwischen $(x - s_\ell h, y)$ und (x, y) , sondern auch zwischen $(x + h, y)$ und $(x + 2h, y)$ verläuft, ist $x + 2h$ durch $x + h + s_r h$ zu ersetzen. Falls auch $(x + h, y)$ nicht mehr zu Ω gehört, setze man $D_{xx} u_h(\mathbf{x}) := h^{-2} u_h(\mathbf{x})$. Entsprechend wird die zweite Differenz D_{yy} definiert. Die gemischte Differenz lautet

im Inneren $D_{xy} := \partial_x^- \partial_y^-$. In Randnähe sind verschiedene Festlegungen von D_{xy} möglich. Die Definition von $|\cdot|_2$ lautet

$$|u_h|_2 := \sqrt{|u_h|_1^2 + h^2 \sum_{\mathbf{x} \in \Omega_h} \left[|D_{xx} u_h(\mathbf{x})|^2 + |D_{yy} u_h(\mathbf{x})|^2 \right]}.$$

Übungsaufgabe 9.2.22 Man beweise die inverse Abschätzung $|\cdot|_2 \leq Ch^{-1} |\cdot|_1$.

Wir nennen L_h H_h^2 -regulär, falls $|L_h^{-1}|_{2 \leftarrow 0} \leq C$. Eine äquivalente Formulierung ist $|u_h|_2 \leq C |f_h|_0$ für $u_h = L_h^{-1} f_h$, $f_h \in L_h^2$.

Übungsaufgabe 9.2.23 L_h sei H_h^2 -regulär, $L_h + \delta L_h$ H_h^1 -regulär und $|\delta L_h|_{0 \leftarrow 1} \leq C$ für alle h . Man zeige: Auch $L_h + \delta L_h$ ist H_h^2 -regulär.

Zum Nachweis der H_h^2 -Regularität kann man in Analogie zum Beweis des Zusatzes 9.1.27 das Skalarprodukt $(L_h u_h, L_h u_h)_0$ partiell summieren, um die Äquivalenz von $|f_h|_0^2 = |L_h u_h|_0^2$ und $|u_h|_2^2$ zu zeigen. Diese Technik ist allerdings nur für ein Rechteck Ω anwendbar. Hier wird ein einfacherer Beweis benutzt, der auch für allgemeine Gebiete durchführbar ist.

$P_h : L_h^2 \rightarrow L^2(\Omega)$ sei die folgende stückweise konstante Interpolation:

$$(P_h u_h)(\mathbf{x}) := \begin{cases} (\hat{P}_h u_h)(\mathbf{x}) & \text{falls } \mathbf{x} \in \Omega, \\ 0 & \text{falls } \mathbf{x} \notin \Omega, \end{cases} \quad (9.2.26a)$$

$$\hat{P}_h u_h(x, y) := u_h(x', y') \in \Omega_h \quad \text{und} \quad x' - h/2 < x \leq x' + h/2, \quad y' - h/2 < y \leq y' + h/2, \quad (9.2.26b)$$

wobei $u_h(\mathbf{x}) = 0$ für $\mathbf{x} \in Q_h \setminus \Omega_h$ (vgl. (9.2.2)).

Lemma 9.2.24 Es sei $\Omega \in C^{0,1}$. Dann gilt

$$|\tilde{R}_h P_h - I|_{-1 \leftarrow 0} \leq Ch, \quad (9.2.27a)$$

$$|P_h|_{0 \leftarrow 0} \leq C, \quad (9.2.27b)$$

$$|R_h|_{2 \leftarrow 2} \leq C. \quad (9.2.27c)$$

Beweis. 1) Die Abschätzung (9.2.27a) ist zu $|P_h^* \tilde{R}_h^* - I|_{0 \leftarrow 1} \leq Ch$ äquivalent (vgl. (6.3.3)). Daher ist

$$|w_h|_0 \leq Ch |u_h|_1 \quad \text{für} \quad w_h := (P_h^* \tilde{R}_h^* - I) u_h$$

zu zeigen. Aus der nachfolgenden Übungsaufgabe 9.2.25 ergibt sich die Darstellung

$$\hat{w}_h(\mathbf{x}) = \left(\sigma_h^x \sigma_h^y [\sigma_h^x \sigma_h^y - I] \hat{P}_h u_h \right)(\mathbf{x}) \quad \text{für} \quad \hat{w}_h := \left(\hat{P}_h^* R_h^* - I \right) u_h, \quad \mathbf{x} \in \Omega_h.$$

Übungsaufgabe 9.2.12c zeigt $|\hat{w}_h|_0 \leq |w|_{L^2(\mathbb{R}^2)}$ für $w := (\sigma_h^x \sigma_h^y - I) \hat{P}_h u_h$. Für jedes $\xi \in (0, 1) \times (0, 1)$ definiere man die Gitterfunktion

$$w_{h,\xi} \in L_h^2 \quad \text{mit} \quad w_{h,\xi}(\mathbf{x}) := w(\mathbf{x} + \xi) \quad \text{für } \mathbf{x} \in Q_h.$$

Man prüft nach, dass $w_{h,\xi}(x, y)$ eine gewichtete Summe von ersten Differenzen $u_h(x, y) - u_h(x', y')$ ist, wobei $(x', y') \in \{(x \pm h, y), (x, y \pm h), (x \pm h, y \pm h)\}$. Daher gilt $|w_{h,\xi}|_0 \leq h |u_h|_1$ für alle ξ , woraus man

$$|w|_0^2 = \int_{\mathbb{R}^2} |w(\mathbf{x})|^2 d\mathbf{x} = h^2 \int_{(0,1) \times (0,1)} \sum_{\mathbf{x} \in \Omega_h} |w(\mathbf{x} + \xi h)|^2 d\xi = \int_{(0,1) \times (0,1)} \sum_{\mathbf{x} \in \Omega_h} |w_{h,\xi}|_0^2 d\xi \leq h^2 |u_h|_1,$$

also $|\hat{w}_h|_0 \leq |w|_0 \leq h |u_h|_1$ schließt. Aus

$$w_h - \hat{w}_h = (P_h^* - \hat{P}_h^*) \tilde{R}_h^* u_h = (P_h^* - \hat{P}_h^*) \sigma_h^x \sigma_h^y \hat{P}_h u_h = (P_h^* - \hat{P}_h^*) \sigma_h^x \sigma_h^y \hat{P}_h (u_h|_{\hat{\gamma}_h})$$

mit $\hat{\gamma}_h = \{\mathbf{x} \in \Omega_h : K_{3h/2}(\mathbf{x}) \cap \Gamma \neq \emptyset\}$ folgert man

$$|w_h - \hat{w}_h|_0 \leq \left| (P_h^* - \hat{P}_h^*) \sigma_h^x \sigma_h^y \right|_{0 \leftarrow 0} \left| \hat{P}_h \right|_{0 \leftarrow 0} |u_h|_{\hat{\gamma}_h} \leq C |u_h|_{\hat{\gamma}_h} \leq C' h |u_h|_1$$

(vgl. zweiten Beweisteil und Lemma 9.2.15, in dem γ_h durch $\hat{\gamma}_h$ ersetzt werden darf). Zusammen mit $|\hat{w}_h|_0 \leq h|u_h|_1$ erhält man $|w_h|_0 \leq (C' + 1)h|u_h|_1$.

2) (9.2.27b) mit $C = 1$ folgt aus $|P_h u_h|_0 = |u_h|_0$.

3) Zweite Differenzen von $\hat{u} = R_h u$ sind bereits in (9.2.19i) abgeschätzt worden. (9.2.19i) liefert $|u_h|_2 \leq C|u|_2$, d.h. (9.2.27c). ■

Übungsaufgabe 9.2.25 Man zeige: Die adjungierten Operatoren zu \hat{P}_h und \tilde{R}_h sind $\hat{P}_h^* : L^2(\mathbb{R}^2) \rightarrow L_h^2$ und $\tilde{R}_h^* : L_h^2 \rightarrow L^2(\mathbb{R}^2)$ mit

$$\left(\hat{P}_h^* u\right)(\mathbf{x}) = \left(\sigma_h^x \sigma_h^y u\right)(\mathbf{x}) \quad \text{und} \quad \left(\tilde{R}_h^* u_h\right)(\mathbf{x}) = \left(\sigma_h^x \sigma_h^y \hat{P}_h u_h\right)(\mathbf{x}) \quad \text{für } \mathbf{x} \in Q_h.$$

Ferner gilt $(\sigma_h^x \sigma_h^y \hat{P}_h u_h)(\mathbf{x}) = u_h(\mathbf{x})$ für alle $\mathbf{x} \in Q_h$.

Die Identität

$$L_h^{-1} = R_h L^{-1} P_h - L_h^{-1} \left[\left(L_h R_h - \tilde{R}_h L \right) L^{-1} P_h + \left(\tilde{R}_h P_h - I \right) \right]$$

liefert die Abschätzung

$$\begin{aligned} |L_h^{-1}|_{2 \leftarrow 0} &\leq |R_h|_{2 \leftarrow 2} |L^{-1}|_{2 \leftarrow 0} |P_h|_{0 \leftarrow 0} \\ &+ |I|_{2 \leftarrow 1} |L_h^{-1}|_{1 \leftarrow -1} \left[\left| L_h R_h - \tilde{R}_h L \right|_{-1 \leftarrow 2} |L^{-1}|_{2 \leftarrow 0} |P_h|_{0 \leftarrow 0} + \left| \tilde{R}_h P_h - I \right|_{-1 \leftarrow 0} \right]. \end{aligned}$$

Die inverse Abschätzung aus Übungsaufgabe 9.2.22 ergibt $|I|_{2 \leftarrow 1} \leq Ch^{-1}$ für die Identität $I : H_h^1 \rightarrow H_h^2$. Zusammen mit den Ungleichungen (9.2.27a-c) folgt

Satz 9.2.26 L_h sei H_h^1 -regulär und erfülle die Konsistenzbedingung (9.2.9). Ω sei konvex oder aus C^2 , und L sei H^2 -regulär (d.h. $|L^{-1}|_{2 \leftarrow 0} \leq C$). Dann ist L_h auch H_h^2 -regulär.

Als Voraussetzung an Ω würde $\Omega \in C^{0,1}$ genügen, um (9.2.27a-c) zu garantieren. Andererseits entspricht die angegebene Bedingung den Voraussetzungen für Satz 9.2.14 und für die H^2 -Regularität von L .

Zusatz 9.2.27 In Satz 9.2.26 können (9.2.9) und die H_h^1 -Regularität von L_h durch (9.2.15) und die H_h^1 -Regularität von L'_h ersetzt werden. Resultat ist die H_h^2 -Regularität sowohl von L_h als auch L'_h .

Beweis. Zum Nachweis der H_h^2 -Regularität von L'_h definiere man $P_h u_h(x) := 0$ für $\mathbf{x} \in K_{h/2}(\hat{\mathbf{x}})$, $\hat{\mathbf{x}} \in \gamma_h$, und beachte $\left| D_h(\tilde{R}_h P_h - I) D_h^{-1} \right|_{-1 \leftarrow 0} \leq Ch$. ■

Übungsaufgabe 9.2.28 Ist $\Omega \subset \mathbb{R}^n$ ($n \leq 3$) beschränkt und L_h H_h^2 -regulär, so ist L_h bezüglich der Zeilen-summennorm stabil: $\|L_h^{-1}\|_\infty \leq C$. Hinweis: Man verwende $C^{-1}|u_h|_0 \leq \|u_h\|_\infty \leq C|u_h|_2$.

Für allgemeine Lipschitz-Gebiete Ω wie zum Beispiel das L-Gebiet aus Beispiel 2.1.4 kann L (L_h) nicht H^2 -regulär (H_h^2 -regulär) sein. Satz 9.1.25 garantiert jedoch die $H_0^{1+s}(\Omega)$ -Regularität für $s \in [0, 1/2)$. Für L_h kann in analoger Weise die H_h^{1+s} -Regularität definiert und bewiesen werden (vgl. Hackbusch [37]).

10

Spezielle Differentialgleichungen

Haben die Randwertaufgaben spezielle Eigenschaften, braucht man häufig auch spezielle Diskretisierungen. Wir geben hierzu zwei Beispiele in den Abschnitten 10.1 und 10.2. Für das erste Beispiel ist die Variationsformulierung von entscheidender Bedeutung.

10.1 Differentialgleichungen mit unstetigen Koeffizienten

10.1.1 Formulierung

Die selbstadjungierte Differentialgleichung

$$\sum_{i,j=1}^n \frac{\partial}{\partial x_i} \left(a_{ij}(\mathbf{x}) \frac{\partial}{\partial x_j} u \right) + a(\mathbf{x})u(\mathbf{x}) = f \quad \text{in } \Omega \quad (10.1.1a)$$

(vgl. (5.1.17)) tritt häufig in der Physik auf. Sie lässt sich auch als

$$\operatorname{div}(\mathbf{A}(\mathbf{x}) \operatorname{grad} u) + au = f \quad \text{mit } \mathbf{A}(\mathbf{x}) := (a_{ij}(\mathbf{x}))_{i,j=1,\dots,n}$$

schreiben, sodass für $a = 0$ und $f = 0$ die Erhaltungform $\operatorname{div} \phi = 0$ für $\phi := \mathbf{A} \operatorname{grad} u$ resultiert.

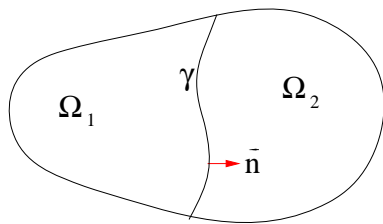


Abb. 10.1.1. innerer Rand
 $\gamma = \Gamma_1 \cap \Gamma_2$

In den physikalischen Anwendungen sind die Koeffizienten a_{ij} im Allgemeinen Materialkonstanten. Die Funktionen a_{ij} können variabel sein, wenn die Materialzusammensetzung ortsabhängig ist. Sobald aber verschiedene Materialien aneinanderstoßen, sind die Koeffizienten $a_{ij}(\mathbf{x})$ an der Berührungsgrenze γ (vgl. Abbildung 10.1.1) unstetig.

Die Gleichung (10.1.1a) lässt sich nur dann im klassischen Sinne verstehen, wenn $a_{ij} \in C^1(\Omega)$. Für unstetige a_{ij} ist deshalb die Variationsformulierung anzuwenden. Ergänzt man Gleichung (10.1.1a) durch die Dirichlet-Bedingung

$$u = 0 \quad \text{auf } \Gamma, \quad (10.1.1b)$$

so lautet die schwache Formulierung:

$$\text{suche } u \in H_0^1(\Omega) \quad \text{mit} \quad a(u, v) = (f, v)_{L^2(\Omega)} \quad \text{für alle } v \in H_0^1(\Omega), \quad (10.1.2a)$$

wobei

$$a(u, v) := \int_{\Omega} \left[a(\mathbf{x})u(\mathbf{x})v(\mathbf{x}) - \sum_{i,j=1}^n a_{ij}(\mathbf{x}) \frac{\partial u}{\partial x_j} \frac{\partial v}{\partial x_i} \right] d\mathbf{x}. \quad (10.1.2b)$$

Man beachte, dass die Gleichungen (10.1.2a,b) anders als (10.1.1a) für beliebige $a_{ij} \in L^\infty(\Omega)$ definiert sind.

Im Weiteren nehmen wir die Situation von Abbildung 10.1.1 an: Ω ist durch die Trennlinie γ in die Teilgebiete Ω_1 und Ω_2 zerlegt. Die Koeffizienten a_{ij} seien stückweise glatt: $a_{ij} \in C^1(\Omega_k)$ für $k = 1, 2$. Entlang γ können die Koeffizienten unstetig sein, sodass die einseitigen Grenzwerte

$$a_{ij}^{(1)}(\mathbf{x}) := \lim_{\Omega_1 \ni \mathbf{y} \rightarrow \mathbf{x}} a_{ij}(\mathbf{y}), \quad a_{ij}^{(2)}(\mathbf{x}) := \lim_{\Omega_2 \ni \mathbf{y} \rightarrow \mathbf{x}} a_{ij}(\mathbf{y}) \quad (\mathbf{x} \in \gamma)$$

verschieden sein können, Ferner setzen wir voraus, dass die Lösung u stetig ist: $u \in C^0(\overline{\Omega})$, aber nur stückweise glatt ist: $u \in C^1(\Omega_1)$, $u \in C^1(\Omega_2)$. Die einseitigen Grenzwerte der Ableitungen sind durch $u_{x_j}^{(1)}(\mathbf{x})$ und $u_{x_j}^{(2)}(\mathbf{x})$ ($\mathbf{x} \in \gamma$) gekennzeichnet. Unter diesen Voraussetzungen liefert partielle Integration von $-\int_{\Omega_1} a_{ij} v_{x_i} u_{x_j} d\mathbf{x}$ das Resultat $\int_{\Omega_1} v(a u_{x_j})_{x_i} d\mathbf{x} - \int_{\gamma} v a_{ij}^{(1)} u_{x_j}^{(1)} n_i d\Gamma$. Da $\vec{n} = (n_1, n_2, \dots)$ aus Abbildung 10.1.1 die äußere Normale bezüglich Ω_1 aber die innere Normale bezüglich Ω_2 ist, ergibt sich in Ω_2 für die gleiche Richtung \vec{n} , dass $-\int_{\Omega_2} a_{ij} v_{x_i} u_{x_j} d\mathbf{x} = \int_{\Omega_2} v(a_{ij} u_{x_j})_{x_i} d\mathbf{x} + \int_{\gamma} v a_{ij}^{(2)} u_{x_j}^{(2)} n_i d\Gamma$. Zusammen erhält man

$$a(u, v) = \int_{\Omega} \left[a(\mathbf{x}) u(\mathbf{x}) - \sum_{i,j=1}^n \frac{\partial}{\partial x_i} \left(a_{ij}(\mathbf{x}) \frac{\partial u}{\partial x_j} \right) \right] v(\mathbf{x}) d\mathbf{x} + \int_{\gamma} \sum_{i,j=1}^n \left[a_{ij}^{(2)} u_{x_j}^{(2)} - a_{ij}^{(1)} u_{x_j}^{(1)} \right] n_i v d\Gamma.$$

Aus der Variationsgleichung (10.1.2a) folgt daher neben der Differentialgleichung (10.1.1a) auch die Übergangsbedingung

$$\sum_{i,j=1}^n a_{ij}^{(1)} u_{x_j}^{(1)} n_i = \sum_{i,j=1}^n a_{ij}^{(2)} u_{x_j}^{(2)} n_i \quad \text{auf } \gamma. \quad (10.1.3)$$

Damit ist das folgende Lemma bewiesen:

Lemma 10.1.1 *Sei $a_{ij} \in C^1(\Omega_k)$, $k = 1, 2$. Ist die schwache Lösung u von (10.1.2a) in Ω stetig und stückweise in Ω_1 und Ω_2 differenzierbar, so ist sie klassische Lösung der Differentialgleichung (10.1.1a) in $\Omega_1 \cup \Omega_2 = \Omega \setminus \gamma$. Außerdem erfüllt sie neben der Randbedingung (10.1.1b) auf Γ die Übergangsbedingung (10.1.3) auf dem inneren Rand γ .*

Zusatz 10.1.2 *Sind die Koeffizienten unstetig, so gehört die Lösung u der Gleichung (10.1.2a) im Allgemeinen nicht zu $C^2(\Omega)$, sondern hat unstetige Ableitungen in γ . Lediglich die Tangentialableitung entlang γ kann stetig sein.*

Beispiel 10.1.3 *Sei $\xi \in (0, 1)$. Der Koeffizient von $a(u, v) := \int_0^1 a(x) u' v' dx$ sei durch $a(x) = 1$ in $(0, \xi)$ und $a(x) = 2$ in $(\xi, 1)$ gegeben. Für dieses eindimensionale Beispiel ersetzt der Punkt ξ die Kurve γ . Die Lösung der Aufgabe (10.1.2a) mit $f(x) = 1$ lautet*

$$u(x) = \frac{1}{2} \left[\frac{1 + \xi^2}{1 + \xi} x - x^2 \right] \quad \text{in } \Omega_1 = (0, \xi), \quad u(x) = \frac{1}{4} \left[\frac{1 + 2\xi - \xi^2}{1 + \xi} (1 - x) - (1 - x)^2 \right] \quad \text{in } \Omega_2 = (\xi, 1).$$

Sie erfüllt $-au'' = 1$ in $(0, \xi) \cup (\xi, 1)$, $u(0) = u(1) = 0$ und die Übergangsbedingung $1 \cdot u'(\xi - 0) = 2 \cdot u'(\xi + 0)$.

Wie anfangs erwähnt, lässt sich die Differentialgleichung in der Form $\operatorname{div} \phi = f$ in $\Omega_1 \cup \Omega_2$ schreiben, wobei $\phi := \mathbf{A}(\mathbf{x}) \operatorname{grad} u$ in Ω und $u = 0$ auf Γ . Die Übergangsbedingung (10.1.3) besagt, dass $\langle \phi, \vec{n} \rangle$ in γ stetig ist. Da auch $\langle \phi, \vec{t} \rangle$ für jede Tangentialrichtung t stetig ist, ist ϕ in γ und damit in Ω stetig.

Die Regularitätsbeweise des Abschnitt 9.1 lassen sich auf den vorliegenden Fall übertragen. Die Glattheitsvoraussetzungen an die Koeffizienten von $a(\cdot, \cdot)$ sind jeweils stückweise in Ω_1 und Ω_2 zu fordern, ferner muss die Trennlinie (-hyperfläche) γ hinreichend glatt sein. Dann ergibt sich die stückweise Regularitätsaussage $u \in H^{m+s}(\Omega_1)$, $u \in H^{m+s}(\Omega_2)$ anstelle von $u \in H^{m+s}(\Omega)$. Zusammen mit der Übergangsbedingung (10.1.3) erhält man $\phi = \mathbf{A}(\mathbf{x}) \operatorname{grad} u \in H^{m+s-1}(\Omega)$ im Gesamtgebiet.

10.1.2 Diskretisierung

Bei einer Diskretisierung durch finite Elemente stößt man auf folgende Schwierigkeiten:

Bemerkung 10.1.4 *Lineare oder bilineare Elemente ergeben eine Finite-Element-Lösung mit der Fehlerabschätzung $|u - u^h|_1 = \mathcal{O}(h^{1/2})$. Die $L^2(\Omega)$ -Fehlerschranke ist im Allgemeinen nicht besser als $|u - u^h|_0 = \mathcal{O}(h)$.*

Beweis. $\Omega_\gamma := \cup\{t \in \tau : \tau \cap \gamma \neq \emptyset\}$ besteht aus allen finiten Elementen, die von γ getroffen werden. Da u unstetige Ableitungen auf γ besitzt, kann man auf Ω_γ keine bessere Abschätzung als $\nabla u - \nabla v = \mathcal{O}(1)$ ($v \in V_h$) erwarten. Die Fläche von Ω_γ beträgt $\mathcal{O}(h)$, sodass sich $|u - v|_1 = \mathcal{O}(h^{1/2})$ ergibt. Zur $|u - u^h|_0$ -Abschätzung vergleiche man das folgende Beispiel. ■

Beispiel 10.1.5 *In Beispiel 10.1.3 wähle man $\xi = (1 + h)/2$ und diskretisiere durch stückweise lineare Elemente der Intervalllänge h . Dann ergibt sich der Fehler in $x = 1/2$ zu $u(x) - u^h(x) = \alpha h + \mathcal{O}(h^2)$ mit $\alpha \approx 0,00463$. Da $u - u^h$ in $(0, 1/2)$ linear verläuft, erhält man den Fehler $\mathcal{O}(h)$ nicht nur für die L^2 -Norm $|u - u^h|_0$, sondern auch für die Maximumnorm $|u - u^h|_\infty$ und die L^1 -Norm $\|u - u^h\|_{L^1(0,1)}$.*

Die üblichen Fehlerschranken $|u - u^h|_1 = \mathcal{O}(h)$ und $|u - u^h|_0 = \mathcal{O}(h^2)$ lassen sich jedoch erreichen. Dazu hat man die Geometrie der Triangulation der Kurve γ anzupassen. Ist die Trennlinie γ stückweise linear, muss man die Triangulation so wählen, dass γ mit (inneren) Dreiecksseiten übereinstimmt. Für die Fehlerabschätzung in Abschnitt 8.4.3 wird nur die Glattheit der Beschränkungen von u auf die finiten Elemente $t \in \tau$ benötigt. Stückweise Glattheit $u \in H^{m+s}(\Omega_k)$ ($k = 1, 2$) und $t \subset \Omega_k$ für genau ein k reicht daher aus, um die gleichen Fehlerabschätzungen zu beweisen. Ist γ eine gekrümmte Kurve, lässt sie sich durch isoparametrische Elemente approximieren (vgl. §8.5.3).

Die analoge Aussage gilt für das Differenzenschema (5.1.18).

Beispiel 10.1.6 *Sei ξ aus Beispiel 10.1.3 ein Gitterpunkt, d.h. $\xi/h \in \mathbb{N}$. Dann ist das Differenzenschema (5.1.18), das hier die Gestalt*

$$\frac{1}{h^2} \left[a \left(\left(\nu + \frac{1}{2} \right) h \right) (u_{\nu+1} - u_\nu) - a \left(\left(\nu - \frac{1}{2} \right) h \right) (u_\nu - u_{\nu-1}) \right] = 1 \quad \left(1 \leq \nu \leq \frac{1}{h} - 1 \right)$$

annimmt, für die Gleichung aus Beispiel 10.1.3 geeignet. Im Allgemeinen beträgt der Fehler $\mathcal{O}(h^2)$. Da hier die Lösung der Differentialgleichung stückweise quadratisch ist, wird sie durch die Differenzenlösung exakt wiedergegeben. Ist dagegen ξ kein Gitterpunkt, beträgt der Fehler $\mathcal{O}(h)$.

Im zweidimensionalen Fall erhält man $\mathcal{O}(h^2)$ -Differenzenlösungen, wenn γ mit den Gitterlinien übereinstimmt. Andernfalls verschlechtert sich der Fehler auf $\mathcal{O}(h)$. Eine weitere Möglichkeit der Differenzenapproximation besteht darin, die Differentialgleichungen in Ω_1 und Ω_2 getrennt zu approximieren und für die unbekannten Werte auf γ die Übergangsbedingung (10.1.3) zu diskretisieren.

10.2 Ein singular gestörtes Problem

10.2.1 Die Konvektionsdiffusionsgleichung

Im Folgenden betrachten wir das Randwertproblem

$$-\varepsilon \Delta u + \sum_{i=1}^n c_i \frac{\partial u}{\partial x_i} = f \quad \text{in } \Omega, \quad u = 0 \quad \text{auf } \Gamma \quad (10.2.1)$$

für $\varepsilon > 0$. Den Hauptteil $-\varepsilon \Delta u$ bezeichnet man als *Diffusionsterm* und $\sum c_i u_{x_j} = \langle \mathbf{c}, \nabla u \rangle$ als *Konvektionsanteil*. Die entsprechende Variationsformulierung ist

$$\int_{\Omega} \left[\varepsilon \langle \nabla u, \nabla v \rangle + \sum_{i=1}^n c_i \frac{\partial u}{\partial x_i} v \right] dx = \int_{\Omega} f v dx. \quad (10.2.1')$$

Gleichung (10.2.1) ist für alle $\varepsilon > 0$ elliptisch. Die eindeutige Lösbarkeit folgt aus Satz 5.1.11 (vgl. auch Übungsaufgabe 10.2.1). Die Lösung sei mit u_ε bezeichnet.

Im Folgenden soll analysiert werden, was passiert, wenn ε klein wird, d.h. wenn der Konvektionsterm dominiert.

Übungsaufgabe 10.2.1 Die Koeffizienten c_i seien konstant. Gleichung (10.2.1) lässt sich durch die Transformation $v(\mathbf{x}) := u(\mathbf{x}) \exp(-\sum c_i x_i / 2\varepsilon) = u(\mathbf{x}) \exp(-\langle \mathbf{c}, \mathbf{x} \rangle / 2\varepsilon)$ in die symmetrische $H_0^1(\Omega)$ -elliptische Gleichung (10.2.1') umwandeln:

$$-\varepsilon \Delta v + \left(\sum_{i=1}^n \frac{c_i^2}{4\varepsilon} \right) v = f \exp(-\langle \mathbf{c}, \mathbf{x} \rangle / 2\varepsilon) \quad \text{in } \Omega, \quad v = 0 \quad \text{auf } \Gamma. \quad (10.2.1'')$$

Zur Konvergenz von u_ε für $\varepsilon \rightarrow 0$ ergibt sich zunächst ein negatives Resultat.

Bemerkung 10.2.2 u_ε kann nicht in der Norm $\|\cdot\|_{H^1(\Omega)}$ konvergieren.

Beweis. Andernfalls könnte man in Gleichung (10.2.1') den Grenzprozess $\varepsilon \rightarrow 0$ durchführen und erhielte für $u_0 := \lim u_\varepsilon$ die Variationsgleichung $\int_\Omega \sum_{i=1}^n c_i \frac{\partial u_0}{\partial x_i} v d\mathbf{x} = \int_\Omega f v d\mathbf{x}$ und die Differentialgleichung

$$c_i \frac{\partial u_0}{\partial x_i} = f \quad \text{in } \Omega \quad (10.2.2)$$

erster Ordnung. Diese Gleichung vom hyperbolischen Typ ist nicht aber mit der Randbedingung $u = 0$ auf Γ verträglich (vgl. §1.4), d.h. Gleichung (10.2.2) hat im Allgemeinen keine Lösung mit $u_0 = 0$ auf Γ . ■

Gleichung (10.2.2) heißt die “reduzierte Gleichung”. Die Gleichungen (10.2.1) und (10.2.2) unterscheiden sich um den “Störungsterm” $-\varepsilon \Delta u$. Da Gleichung (10.2.1) und Gleichung (10.2.2) von *verschiedenem Typ* sind, spricht man von einer “singulären” Störung. Ein anderes Beispiel einer singulär gestörten Gleichung ist die *Reaktions-Diffusionsgleichung* $-\varepsilon \Delta u + au = f$. Ein Übergang vom elliptischen zum parabolischen Typ liegt bei $-\varepsilon u_{xx} - u_{yy} + u_x = f$ vor. Auch $\varepsilon \Delta^2 u - \Delta u = f$ ist singulär gestört, da diese Gleichungen für $\varepsilon > 0$ und $\varepsilon = 0$ zwar beide elliptisch sind, aber unterschiedlich viele Randbedingungen benötigen.

Das folgende Beispiel wird zeigen, dass der punktweise Limes $u_0 = \lim u_\varepsilon$ in Ω existiert und Gleichung (10.2.2) erfüllt, dass aber die Randbedingung $u_0 = 0$ nur auf einem Teil Γ_0 des Randes Γ gilt.

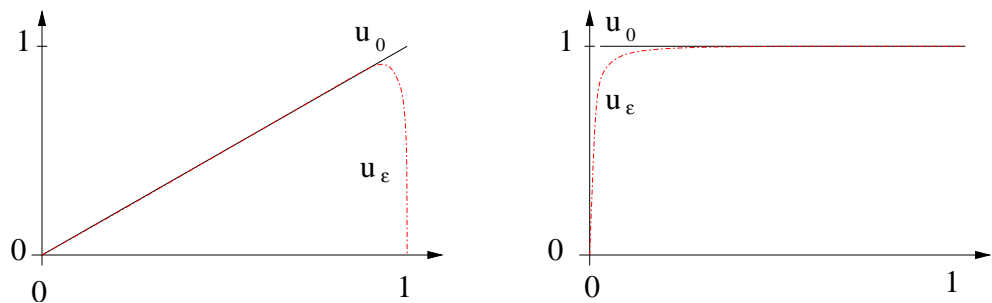


Abb. 10.2.1. a) Lösung aus Beispiel 10.2.3a und b) Lösung aus Beispiel 10.2.3b

Beispiel 10.2.3 a) Die Lösung der gewöhnlichen Randwertaufgabe

$$-\varepsilon u'' + u' = 1 \quad \text{in } \Omega := (0, 1), \quad u(0) = u(1) = 0, \quad (10.2.3a)$$

lautet $u_\varepsilon(x) = x - (e^{x/\varepsilon} - 1)/(e^{1/\varepsilon} - 1)$. In $[0, 1]$ konvergiert $u_\varepsilon(x)$ punktweise gegen $u_0(x) = x$. Diese Funktion erfüllt die reduzierte Gleichung (10.2.2): $u' = 1$ und die linke Randbedingung $u_0(0) = 0$, nicht jedoch rechte Randbedingung $u_0(1) = 0$.

b) Die Lösung von

$$-\varepsilon u'' - u' = 0 \quad \text{in } \Omega := (0, 1), \quad u(0) = 0, \quad u(1) = 1, \quad (10.2.3b)$$

lautet $u_\varepsilon(x) = (1 - e^{-x/\varepsilon})/(1 - e^{-1/\varepsilon})$. $u_0(x) := \lim_{\varepsilon \rightarrow 0} u_\varepsilon(x) = 1$ erfüllt Gleichung (10.2.2): $-u' = 0$ und $u_0(1) = 1$, nicht jedoch die linke Randbedingung bei $x = 0$.

Welche Randbedingungen erfüllt werden, hängt in den Gleichungen (10.2.3a,b) vom Vorzeichen des Konvektionsterms $\pm u'$ ab. Im mehrdimensionalen Fall entscheidet hierüber die Richtung des Vektors $\mathbf{c} = (c_1, \dots, c_n)$.

In den Abbildungen 10.2.1a,b sind die Lösungen aus Beispiel 10.2.3 skizziert. Im Inneren liegt u_ε nahe an u_0 , nur in der Umgebung von $x = 1$ (Abbildung 10.2.1a) bzw. $x = 0$ (Abbildung 10.2.1b) weicht u_ε von u_0 ab, um die zweite Randbedingung zu erfüllen. Diese Umgebungen, in denen u_ε Ableitungen der Größenordnung $\mathcal{O}(1/\varepsilon)$ erreicht, nennt man die *Randschicht*. Für Beispiel 10.2.3 beträgt die Randschichtbreite $\mathcal{O}(\varepsilon)$.

ε	x	0	1/32	2/32	3/32	4/32	...	30/32	31/32	1
0.5	1	0.93	0.87	0.81	0.75	...	0.04	0.02	0	
0.1	1	0.97	0.95	0.92	0.89	...	0.22	0.13	0	
0.05	1	0.99	0.97	0.96	0.94	...	0.47	0.31	0	
0.02	1	0.99	0.99	0.98	0.98	...	0.82	0.73	0	
0.01	1	1.00	0.99	0.99	0.99	...	0.87	1.11	0	
1_{10-3}	1	1.03	0.99	1.04	0.99	...	0.23	1.89	0	
1_{10-4}	1	4.95	0.95	5.00	0.90	...	0.08	5.88	0	
1_{10-5}	1	48.6	0.94	48.7	0.88	...	0.06	49.6	0	
1_{10-7}	1	4859.5	0.94	4859.6	0.88	...	0.06	4860.4	0	

Tabelle 10.2.1. Werte $u_{\varepsilon,h}(x, 1/2)$ der Lösung von Gleichung (10.2.4)

Übungsaufgabe 10.2.4 Das Intervall $[1-\xi, 1]$, in dem die Funktion $(e^{x/\varepsilon}-1)/(e^{\xi/\varepsilon}-1)$ den Wert $0 < \eta < 1$ überschreitet, hat die Breite $\xi = \mathcal{O}(\varepsilon |\log \eta|)$.

Eine ausführliche Analyse der singular gestörten Probleme findet man z.B. in der Monographie Roos-Stynes-Tobiska [71] (vgl. auch Großmann-Roos [36, §7]).

10.2.2 Stabile Differenzenschemata

Der Spezialfall der Gleichung (10.2.1) für $\mathbf{c} = (1, 0)$ lautet

$$-\varepsilon \Delta u + u_x = f \quad \text{in } \Omega, \quad u = \varphi \quad \text{auf } \Gamma. \quad (10.2.4)$$

Die symmetrische Differenzenformel (5.1.10) für Gleichung (10.2.4) lautet

$$L_h = h^{-2} \begin{bmatrix} -\varepsilon & -\varepsilon \\ -\varepsilon - h/2 & 4\varepsilon & -\varepsilon - h/2 \\ -\varepsilon & -\varepsilon \end{bmatrix}. \quad (10.2.5)$$

Für festes ε ist L_h von der Konsistenzordnung 2. Aus Korollar 5.1.19 folgt die folgende Bemerkung.

Bemerkung 10.2.5 Sobald $h < 2\varepsilon$, führt das Differenzenschema (10.2.5) zu einer M-Matrix.

Die M-Matrix-Eigenschaft wurde in Abschnitt 5.1.4 benutzt, um die Lösbarkeit der diskreten Gleichung nachzuweisen. Es zeigt sich, dass die Differenzengleichungen auch für $h \geq 2\varepsilon$ lösbar sind. Aber mit steigendem Verhältnis h/ε entwickelt sich eine Instabilität, die in Tabelle 10.2.1 veranschaulicht wird.

Tabelle 10.2.1 enthält die Werte $u_{\varepsilon,h}(x, 1/2)$ der Differenzlösung $u_{\varepsilon,h}$ zu Gleichung (10.2.4) im Einheitsquadrat $\Omega = (0, 1) \times (0, 1)$ mit $f = 0$ und $\varphi(x, y) = (1-x)\sin(\pi y)$. Die Gitterweite ist $h = 1/32$. Solange $\varepsilon > h/2 = 1/64 = 0.015625$, ist L_h eine (irreduzibel diagonaldominante) M-Matrix. Auf Grund des diskreten Maximumprinzips (vgl. Bemerkung 4.4.4) liegen die Werte zwischen $\min \varphi = 0$ und $\max \varphi = 1$. Die Lösung der reduzierten Gleichung (10.2.2) ist $u_0(x, y) = \lim_{\varepsilon \rightarrow 0} u_\varepsilon(x, y) = \sin \pi y$. Dementsprechend nähern sich die Werte $u_{\varepsilon,h}(x, 1/2)$ aus dem ersten Teil der Tabelle 10.2.1 dem Grenzwert $\sin \pi/2 = 1$.

Im zweiten Teil der Tabelle ist $\varepsilon < h/2$. L_h ist daher keine M-Matrix. Man beobachtet $u_{\varepsilon,h}(31/32, 1/2) > 1$, d.h. auch das Maximumprinzip gilt nicht mehr. Außerdem erkennt man, dass $u_{\varepsilon,h}(x, 1/2)$ für gradzahlige Vielfache $x = \nu h$ gegen $1-x$ konvergiert, also nicht gegen $u_0(x, y) = \sin \pi y$. Dazwischen entwickeln sich für ungradzahlige Vielfache $x = \nu h$ Oszillationen der Höhe $\mathcal{O}(h^2/2\varepsilon)$.

Die sich ergebende Problematik ist der bei Anfangswertaufgaben für steife (gewöhnliche) Differentialgleichungen ähnlich: Hält man ε fest und lässt h gegen null streben, so gelten alle Konvergenzaussagen des Abschnitt 5.1.4. Ist aber ε sehr klein, lässt sich die Bedingung $h < 2\varepsilon$, ohne die man keine vernünftige Lösung erhält, praktisch nicht erfüllen.

Ein Ausweg wurde bereits in Abschnitt 5.1.4 beschrieben: Man muss den Konvektionsterm $\sum_{i=1}^n c_i \frac{\partial u}{\partial x_i}$ (hier mit $n = 2$) durch geeignete *einseitige* Differenzen (5.1.14) approximieren:

$$L_h = h^{-2} \begin{bmatrix} -\varepsilon + hc_2^- & & \\ -\varepsilon - hc_1^+ & 4\varepsilon + h|c_1| + h|c_2| & -\varepsilon + hc_1^- \\ & -\varepsilon - hc_2^+ & \end{bmatrix}, \quad \text{wobei } \begin{cases} c_i^+ := \max\{0, c_i\}, \\ c_i^- := \min\{0, c_i\}. \end{cases} \quad (10.2.6)$$

Im Falle von $c_1, c_2 \geq 0$, d.h. $L_h = h^{-2} \begin{bmatrix} & -\varepsilon & \\ -\varepsilon - hc_1 & 4\varepsilon + hc_1 + hc_2 & -\varepsilon \\ & -\varepsilon - hc_2 & \end{bmatrix}$, treten beispielsweise nur Rückwärtsdifferenzen auf.

Bemerkung 10.2.6 Diskretisiert man Gleichung (10.2.1) mit Hilfe von (10.2.6), so erhält man für alle $\varepsilon > 0$ und $h > 0$ eine M-Matrix L_h . Für festes ε hat das Schema die Konsistenzordnung 1.

Übungsaufgabe 10.2.7 Die (10.2.6) entsprechende Diskretisierung der Gleichung (10.2.3b) lautet $L_h = h^{-2} \begin{bmatrix} -\varepsilon & 2\varepsilon + h & -\varepsilon - h \end{bmatrix}$ und ergibt die diskrete Lösung $u_h(x) = [1 - (1 + h/\varepsilon)^{-x/h}] / [1 - (1 + h/\varepsilon)^{-1/h}]$.

Wendet man den Differenzenoperator (10.2.6) auf eine glatte Funktion an, ergibt die Taylor-Entwicklung

$$L_h u = Lu - [|c_1| u_{xx} + |c_2| u_{yy}] h/2 + \mathcal{O}(h^2). \quad (10.2.7)$$

Den $\mathcal{O}(h)$ -Term $(h/2)(|c_1| u_{xx} + |c_2| u_{yy})$ nennt man *numerische Viskosität* (oder *numerische Elliptizität*), da er den Hauptteil verstärkt. Man beachte die Abhängigkeit von der "Stromlinienrichtung" $\mathbf{c} = (c_1, c_2)$. Falls beispielsweise $c_2 = 0$ (Konvektion nur in x -Richtung), so ist die numerische Viskosität eine zweite Ableitung nur in x -Richtung.

Ein zweiter Ausweg besteht darin, den Parameter ε bei der Diskretisierung durch $\varepsilon_h \geq \varepsilon$ zu ersetzen. Wählt man

$$\varepsilon_h := \max \left\{ \varepsilon, \frac{h}{2} |c_1|, \frac{h}{2} |c_2| \right\} \quad \text{oder} \quad \varepsilon_h := \varepsilon + \frac{h}{2} \|\mathbf{c}\|_\infty, \quad (10.2.8)$$

führt das symmetrische Differenzenverfahren

$$L_h = \varepsilon_h h^{-2} \begin{bmatrix} & -1 & \\ -1 & 4 & -1 \\ & -1 & \end{bmatrix} + \frac{1}{2} h^{-1} \begin{bmatrix} & c_2 & \\ -c_1 & 0 & c_1 \\ & -c_2 & \end{bmatrix} \quad (10.2.9)$$

zu einer M-Matrix. Zwar wird der Konvektionsterm von zweiter Konsistenzordnung diskretisiert, jedoch beträgt der Fehler des Diffusionsterms $\mathcal{O}(\varepsilon_h - \varepsilon)$, der für den praktisch relevanten Fall $h > 2\varepsilon / \|\mathbf{c}\|_\infty$ mit $\mathcal{O}(h)$ übereinstimmt. Anstelle von (10.2.7) erhält man

$$L_h u = Lu - (\varepsilon_h - \varepsilon) \Delta u + \mathcal{O}(h^2). \quad (10.2.10)$$

Der Differenzbetrag $(\varepsilon_h - \varepsilon) \Delta u$ heißt *künstliche Viskosität*. Anders als in (10.2.7) verschlechtert man die Konsistenz gleichmäßig in allen Richtungen.

Im eindimensionalen Fall unterscheiden sich die Ansätze mit numerischer und künstlicher Viskosität nicht:

Bemerkung 10.2.8 Wählt man im eindimensionalen Fall ε_h gemäß der zweiten Alternative in (10.2.8), so stimmen die Differenzenformeln (10.2.6) und (10.2.9) überein.

10.2.3 Finite Elemente

Die im vorigen Abschnitt beschriebene Problematik ist nicht auf Differenzenverfahren beschränkt. Zuerst wird das Standard-Galerkin-Verfahren und seine Problematik gezeigt. In §10.2.3.2 wird eine an das Problem angepasste Variante des Galerkin-Verfahrens, das Stromlinien-Diffusionsverfahren vorgestellt.

10.2.3.1 Standard-Galerkin-Verfahren

Die schwache Formulierung der Konvektionsdiffusionsgleichung lautet

$$\text{suche } u \in H_0^1(\Omega) \text{ mit } a(u, v) = f(v) \text{ für alle } v \in H_0^1(\Omega), \quad \text{wobei} \quad (10.2.11a)$$

$$a(u, v) := \int_\Omega \{ \varepsilon \langle \nabla u(\mathbf{x}), \nabla v(\mathbf{x}) \rangle + \langle \mathbf{c}(\mathbf{x}), \nabla u(\mathbf{x}) \rangle v(\mathbf{x}) + auv \} d\mathbf{x}, \quad f(v) = \int_\Omega f(\mathbf{x}) v(\mathbf{x}) d\mathbf{x}, \quad (10.2.11b)$$

und $\mathbf{c} = (c_i)_{i=1, \dots, n}$. Im ersten Teil beschränken wir uns auf $a = 0$ in (10.2.11b).

Übungsaufgabe 10.2.9 Man zeige: a) Lineare finite Elemente über einer Quadratgittertriangulation angewandt auf Gleichung (10.2.4) ergeben eine Diskretisierung, die mit dem Differenzenverfahren

$$L_h = \varepsilon \begin{bmatrix} & -1 & \\ -1 & 4 & -1 \\ & -1 & \end{bmatrix} + h \begin{bmatrix} 0 & -1/6 & 1/6 \\ -1/3 & 0 & 1/3 \\ -1/6 & 1/6 & 0 \end{bmatrix} \quad (10.2.12a)$$

identisch ist (vgl. Übungsaufgabe 8.3.13).

b) Für bilineare Elemente (vgl. Übungsaufgabe 8.3.17) erhält man

$$L_h = \frac{\varepsilon}{3} \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} + \frac{h}{12} \begin{bmatrix} -1 & 0 & +1 \\ -4 & 0 & +4 \\ -1 & 0 & +1 \end{bmatrix}. \quad (10.2.12b)$$

c) Eindimensionale lineare Elemente für $-\varepsilon u'' + u' = f$ führen zur zentralen Differenzenformel

$$L_h = \varepsilon h^{-1} [-1 \quad 2 \quad -1] + [-1 \quad 0 \quad 1]. \quad (10.2.12c)$$

Die Übungsaufgaben 10.2.9b,c zeigen, dass Finite-Element-Verfahren den zentralen Differenzenformeln entsprechen und damit ebenfalls zur Instabilität führen können.

Dem Verfahren der künstlichen Viskosität entspricht die Finite-Element-Lösung der Gleichung $-\varepsilon_h \Delta u + \langle \mathbf{c}, \text{grad } u \rangle = f$ mit geeignetem ε_h . Wie in Übungsaufgabe 10.2.9b beweist man

Bemerkung 10.2.10 Setzt man $\varepsilon_h := \max\{\varepsilon, |c_1| h, |c_2| h\}$ und wendet bilineare Elemente an, so führt die Diskretisierung der Gleichung (10.2.1) zu einer M-Matrix.

Dagegen hat die Matrix (10.2.12a) für alle $\varepsilon > 0$ unterschiedliche Vorzeichen in der Nebendiagonale, sodass sich für kein ε_h eine M-Matrix erreichen lässt.

Das Analogon der einseitigen Differenzen ist schwieriger zu konstruieren. Ein Zugang besteht in der Kombination einer Finite-Element-Methode für den Diffusionsterm mit einem (einseitigen) Differenzenverfahren für den Konvektionsterm (vgl. Thomasset [81, §2.4]).

Eine zweite Möglichkeit ist die Verallgemeinerung des Galerkin-Verfahrens zum Petrov-Galerkin-Verfahren, bei dem die diskrete Lösung des allgemeinen Problems (8.1.1) durch Aufgabe (10.2.13) definiert wird:

$$\text{suche } u \in V_h, \text{ sodass } a(u, v) = f(v) \quad \text{für alle } v \in W_h \quad (10.2.13)$$

(vgl. Fletcher [30, §7.2], Thomasset [81, §2.2]). Dabei ist

$$\dim V_h = \dim W_h, \quad \text{aber im Allgemeinen } V_h \neq W_h.$$

10.2.3.2 Stromlinien-Diffusionsverfahren

Wir folgen im Wesentlichen der Darstellung in [71, §III.3.2.1] und gehen von der schwachen Formulierung $a(u, v) = f(v)$ aus (10.2.11a,b) aus. Die (schwache) Lösung u erfüllt die Gleichung $-\varepsilon \Delta u + \langle \mathbf{c}, \nabla u \rangle + cu = f$ in $H^{-1}(\Omega)$. Wenn wir $f \in L^2(\Omega)$ voraussetzen, können beide Seiten der Gleichung mit dem Ausdruck $\langle \mathbf{c}, \nabla v \rangle \in L^2(\Omega)$ für eine Testfunktion $v \in H_0^1(\Omega)$ und einer Gewichtsfunktion ω multipliziert werden:

$$\int_{\Omega} \omega(\mathbf{x}) (-\varepsilon \Delta u + \langle \mathbf{c}, \nabla u \rangle + cu) \langle \mathbf{c}, \nabla v \rangle \, d\mathbf{x} = \int_{\Omega} f(\mathbf{x}) \langle \mathbf{c}, \nabla v \rangle \, d\mathbf{x}.$$

Addiert man diese Gleichung, die allerdings nur für die Lösung von (10.2.11a) gilt, zu der ursprünglichen Variationsformulierung (10.2.11a), erhält man

$$a^{SD}(u, v) = f^{SD}(v) \quad \text{für alle } v \in H_0^1(\Omega) \quad \text{mit} \quad (10.2.14a)$$

$$a^{SD}(u, v) := \int_{\Omega} \{ \varepsilon \langle \nabla u, \nabla v \rangle + \langle \mathbf{c}, \nabla u \rangle v + auv + \omega(\mathbf{x}) (-\varepsilon \Delta u + \langle \mathbf{c}, \nabla u \rangle + cu) \langle \mathbf{c}, \nabla v \rangle \} \, d\mathbf{x}, \quad (10.2.14b)$$

$$f^{SD}(v) := \int_{\Omega} f(\mathbf{x}) (v + \langle \mathbf{c}, \nabla v \rangle) \, d\mathbf{x}. \quad (10.2.14c)$$

Trotz der Änderung der schwachen Formulierung gilt: Die Lösung von (10.2.11a) ist auch Lösung von (10.2.14a).

Bei der Diskretisierung mit finiten Elementen tritt ein Problem auf. Da im Allgemeinen nur $V_h \subset V = H_0^1(\Omega)$, aber nicht $V_h \subset H_0^1(\Omega) \cap H^2(\Omega)$ gilt, gehört Δu^h im Allgemeinen nicht zu $L^2(\Omega)$. Im Inneren jedes Dreiecks T der Finite-Element-Triangulierung \mathcal{T} ist $u^h|_T$ ein Polynom, sodass $\Delta u^h|_T$ wiederum ein Polynom und somit beliebig glatt ist. Über die Dreiecksränder springt die Ableitung (wenn nicht $V_h \subset H^2(\Omega)$), sodass die zweite Ableitung Δu^h eine Distribution mit Träger in $\bigcup_{T \in \mathcal{T}} \partial T$ ist. Die Diskretisierung, die nun vorgestellt werden soll, ignoriert den Distributionsbeitrag in Δu^h und berücksichtigt nur die Integrale mit $\Delta u^h|_T$. Die Gewichtsfunktion wird konstant pro Dreieck gewählt, sodass ein individueller Faktor ω_T auftritt. Gesucht ist eine Lösung $u^h \in V_h$ von

$$a_h^{SD}(u^h, v) = f_h^{SD}(v) \quad \text{für alle } v \in V_h \quad \text{mit} \quad (10.2.15a)$$

$$a_h^{SD}(u^h, v) := \int_{\Omega} \{ \varepsilon \langle \nabla u^h, \nabla v \rangle + \langle \mathbf{c}, \nabla u^h \rangle v + a u^h v \} \, d\mathbf{x} \quad (10.2.15b)$$

$$+ \sum_{T \in \mathcal{T}} \omega_T \int_T (\varepsilon \Delta u^h + \langle \mathbf{c}, \nabla u^h \rangle + a u^h) |_T \langle \mathbf{c}, \nabla v \rangle \, d\mathbf{x}$$

$$f_h^{SD}(v) := f_h^{SD}(v) = \int_{\Omega} f(\mathbf{x}) (v + \langle \mathbf{c}, \nabla v \rangle) \, d\mathbf{x}. \quad (10.2.15c)$$

Wegen der speziellen Behandlung des Δu^h -Terms gilt – anders als im kontinuierlichen Fall – *nicht*, dass die Finite-Element-Lösung von (10.2.11a) auch eine Lösung von (10.2.15a) wäre. Im Gegenteil soll die Diskretisierung (10.2.15a) stabiler ausfallen, wie noch zu prüfen ist.

Bemerkung 10.2.11 *Im einfachsten Fall der stückweise linearen finiten Elemente gilt $\Delta u^h = 0$, sodass die Summe in (10.2.15b) verschwindet.*

Die Diskretisierung (10.2.15a) nennt man die *Stromlinien-Diffusionsmethode*, da sie der Stabilisierung durch einseitige Differenzen im Falle der Differenzenverfahren entspricht. Um den letztgenannten Effekt zu sehen, nehmen wir zur Vereinfachung an, dass das Finite-Element-Gitter die regelmäßige Quadratgittertriangulierung aus Abbildung 8.3.3 ist und stückweise lineare Elemente gewählt werden. Zudem habe der Vektor nur eine Komponente in der Koordinatenrichtung x : $\langle \mathbf{c}, \nabla u^h \rangle = c_1 \partial u^h / \partial x$, wobei c_1 konstant sei. Wegen Bemerkung 10.2.11 entfällt der Δu^h -Term. Sei T ein Dreieck mit den Eckpunkten (x, y) , $(x + h, y)$, $(x, y + h)$. Da u^h und v auf T linear sind, sind die Ableitungen konstant und stimmen mit den Differenzenquotienten überein:

$$\int_T \langle \mathbf{c}, \nabla u^h \rangle \langle \mathbf{c}, \nabla v \rangle \, d\mathbf{x} = \int_T c_1^2 \frac{\partial u^h}{\partial x} \frac{\partial v}{\partial x} \, dxdy = c_1^2 \frac{h^2}{2} \partial_{x,h}^+ u^h(x, y) \partial_{x,h}^+ v(x, y).$$

Für das anliegende Dreieck mit den Ecken (x, y) , $(x, y + h)$, $(x + h, y + h)$ gilt entsprechend $\int_T \dots = c_1^2 \mu(T) \partial_{x,h}^+ u^h(x + h, y) \partial_{x,h}^+ v(x + h, y)$. In der Finite-Element-Matrix \mathbf{L} ergibt sich damit ein zusätzlicher Anteil, der dem Differenzenstern

$$\frac{c_1^2}{2} \begin{bmatrix} 0 & 0 & 0 \\ -1 & 2 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

entspricht. Es wird also eine zweite Differenz (Diffusionsterm) nur in Richtung der Stromlinien (das ist die Richtung $\mathbf{c} = (c_1, 0)$) addiert, wie es auch bei der einseitigen Differenz (10.2.6) (für $c_2 = 0$) der Fall ist.

Die Stabilität der Stromlinien-Diffusionsmethode wird durch eine Ungleichung bewiesen, die der V-Elliptizität nahekommt. Hierzu werden einige Voraussetzungen benötigt. Für die Koeffizienten \mathbf{c} und a aus (10.2.11b) gelte

$$a - \frac{1}{2} \operatorname{div} \mathbf{c} \geq c_0 > 0 \quad \text{in } \Omega. \quad (10.2.16a)$$

Im Falle konstanter Koeffizienten ist diese Bedingung äquivalent zu $a > 0$. Die Bedeutung der Ungleichung (10.2.16a) ergibt sich aus

Lemma 10.2.12 *Es gelte (10.2.16a). Dann gilt für alle $u \in H_0^1(\Omega)$ die Abschätzung*

$$\int_{\Omega} \{ a u + \langle \mathbf{c}, \nabla u \rangle \} u \, d\mathbf{x} \geq c_0 \|u\|_{L^2(\Omega)}^2.$$

Beweis. Aus $\int_{\Omega} c_i \frac{\partial u}{\partial x_i} u d\mathbf{x} = \frac{1}{2} \int_{\Omega} c_i \frac{\partial u^2}{\partial x_i} d\mathbf{x} = -\frac{1}{2} \int_{\Omega} u^2 \frac{\partial c_i}{\partial x_i} d\mathbf{x}$ (wegen $u^2 = 0$ auf $\partial\Omega$) folgt die Behauptung. ■

Eine Abschätzung von a nach oben wird für jedes Dreieck formuliert:

$$c_T := \max_{\mathbf{x} \in T} |a(\mathbf{x})|. \quad (10.2.16b)$$

Wir definieren

$$h_T := \text{Durchmesser des Dreiecks } T. \quad (10.2.16c)$$

Auf T ist $v \in V_h$ ein Polynom festen Grades. Hierfür gilt die inverse Abschätzung

$$\|\Delta v\|_{L^2(T)} \leq \frac{\mu}{h_T} \|\nabla v\|_{L^2(T)} \quad \text{für alle } T \in \mathcal{T}_h \text{ und alle } v \in V_h. \quad (10.2.16d)$$

Dabei ist $\|\nabla v\|_{L^2(T)}^2 = \sum_{i=1}^n \|\partial v / \partial x_i\|_{L^2(T)}^2$. Wegen Bemerkung 10.2.11 gilt (10.2.16d) für stückweise lineare Elemente mit $\mu = 0$.

Im Zusammenhang mit der Bilinearform $a_h^{SD}(\cdot, \cdot)$ spielt

$$\|v\|_{SD} := \sqrt{\varepsilon \|\nabla v\|_{L^2(\Omega)}^2 + c_0 \|v\|_{L^2(\Omega)}^2 + \sum_{T \in \mathcal{T}} \omega_T \|\langle \mathbf{c}, \nabla v \rangle\|_{L^2(T)}^2}$$

die Rolle der Energienorm.

Satz 10.2.13 (Stabilität) *Es gelte (10.2.16a-d) mit den dort definierten Konstanten. Die Gewicht ω_T mögen*

$$0 < \omega_T \leq \min \left\{ \frac{c_0}{c_T^2}, \frac{h_T^2}{\varepsilon \mu} \right\} \quad \text{für alle } T \in \mathcal{T}_h \quad (10.2.17)$$

erfüllen. Dann genügt die Bilinearform $a_h^{SD}(\cdot, \cdot) : V_h \times V_h \rightarrow \mathbb{R}$ der Ungleichung

$$a_h^{SD}(v, v) \geq \frac{1}{2} \|v\|_{SD}^2 \quad \text{für alle } v \in V_h. \quad (10.2.18)$$

Beweis. Mit Lemma 10.2.12 erhält man

$$a_h^{SD}(v, v) \geq \varepsilon \|\nabla v\|_{L^2(\Omega)}^2 + c_0 \|v\|_{L^2(\Omega)}^2 + \sum_{T \in \mathcal{T}} \omega_T \|\langle \mathbf{c}, \nabla v \rangle\|_{L^2(T)}^2 + \sum_{T \in \mathcal{T}} \omega_T \int_T (\varepsilon \Delta v + av) |_T \langle \mathbf{c}, \nabla v \rangle d\mathbf{x}.$$

Mit der Ungleichung $\alpha\beta \leq \frac{1}{2}\alpha^2 + \frac{1}{2}\beta^2$ schätzt man den letzten Term ab:

$$\begin{aligned} \left| \sum_{T \in \mathcal{T}} \omega_T \int_T (\varepsilon \Delta v + av) |_T \langle \mathbf{c}, \nabla v \rangle d\mathbf{x} \right| &\leq \sum_{T \in \mathcal{T}} \omega_T \left[\frac{1}{2} \|\varepsilon \Delta v + av\|_{L^2(T)}^2 + \frac{1}{2} \|\langle \mathbf{c}, \nabla v \rangle\|_{L^2(T)}^2 \right] \\ &\leq \sum_{T \in \mathcal{T}} \omega_T \left[\|\varepsilon \Delta v\|_{L^2(T)}^2 + \|av\|_{L^2(T)}^2 + \frac{1}{2} \|\langle \mathbf{c}, \nabla v \rangle\|_{L^2(T)}^2 \right] \\ &\leq \sum_{T \in \mathcal{T}} \omega_T \left[\varepsilon^2 \|\Delta v\|_{L^2(T)}^2 + c_T^2 \|v\|_{L^2(T)}^2 + \frac{1}{2} \|\langle \mathbf{c}, \nabla v \rangle\|_{L^2(T)}^2 \right]. \end{aligned}$$

Die Kombination beider Ungleichungen die gewünschte Abschätzung. ■

Für $\mu = 0$ ist $\frac{h_T^2}{\varepsilon \mu} = \infty$ und damit $\min \left\{ \frac{c_0}{c_T^2}, \frac{h_T^2}{\varepsilon \mu} \right\} = \frac{c_0}{c_T^2}$ in (10.2.17).

Das folgende Lemma beschreibt die Konsistenz der Stromlinien-Diffusionsmethode.

Lemma 10.2.14 *Die Voraussetzungen von Satz 10.2.13 und (10.2.17) seien erfüllt. Im Falle von $\mu = 0$ sei $\varepsilon \omega_T \leq Ch_T^2$ explizit vorausgesetzt¹. V_h enthalte stückweise alle Polynomen vom Grad $\leq k$ ($k \geq 1$). Die Lösung u von (10.2.1') gehöre zu $H^{k+1}(\Omega)$. $R_h : H^{k+1}(\Omega) \cap H_0^1(\Omega) \subset V \rightarrow V_h$ sei die Beschränkung auf die Knotenpunkte, d.h. $R_h u \in V_h$ ist die Interpolierende. Dann lautet der Konsistenzfehler*

$$\|R_h u - u^h\|_{SD} \leq Ch^k \sqrt{\sum_{T \in \mathcal{T}} [\varepsilon + \omega_T + h_T^2/\omega_T + h_T^2] \|u\|_{H^{k+1}(T)}^2}. \quad (10.2.19)$$

¹Andernfalls folgt die Ungleichung aus (10.2.17).

Beweis. 1) $R_h u - u^h$ eingesetzt in (10.2.18) liefert die erste Ungleichung in

$$\frac{1}{2} \|R_h u - u^h\|_{SD}^2 \leq a_h^{SD}(R_h u - u^h, R_h u - u^h) = a_h^{SD}(R_h u - u, R_h u - u^h). \quad (10.2.20a)$$

Die zweite Gleichheit beruht auf der Projektionseigenschaft $a_h^{SD}(u^h - u, v) = 0$ für alle $v \in V_h$, die sich aus der Kombination von (10.2.14a) und (10.2.15a) ergibt.

2) In jedem Dreieck $T \in \mathcal{T}$ gilt $\|R_h u - u\|_{H^1(T)} \leq Ch^k \|u\|_{H^{k+1}(T)}$. Stückweise Abschätzung der Terme aus $a_h^{SD}(R_h u - u, R_h u - u^h)$ zusammen mit der Schwarzschen Ungleichung liefert

$$a) \quad \int_{\Omega} \varepsilon \langle \nabla (R_h u - u), \nabla (R_h u - u^h) \rangle d\mathbf{x} \leq \sqrt{\varepsilon} h^k |u|_{k+1} \|R_h u - u^h\|_{SD}, \quad (10.2.20b)$$

$$\begin{aligned} b) \quad & \int_{\Omega} [\langle \mathbf{c}, \nabla (R_h u - u) \rangle (R_h u - u^h) + a (R_h u - u) (R_h u - u^h)] d\mathbf{x} \\ &= \int_{\Omega} [(a - \operatorname{div} \mathbf{c}) (R_h u - u) (R_h u - u^h) - (R_h u - u) \langle \mathbf{c}, \nabla (R_h u - u^h) \rangle] d\mathbf{x} \\ &\leq \left\{ C \sqrt{\sum_{T \in \mathcal{T}} \|u\|_{L^2(T)}^2} + \sqrt{\sum_{T \in \mathcal{T}} \frac{1}{\omega_T} \|u\|_{L^2(T)}^2} \right\} \|R_h u - u^h\|_{SD} \\ &\leq C' h^k \sqrt{\sum_{T \in \mathcal{T}} h_T^2 (1 + 1/\omega_T) \|u\|_{H^{k+1}(T)}^2} \|R_h u - u^h\|_{SD}, \end{aligned} \quad (10.2.20c)$$

$$\begin{aligned} c) \quad & \left| \sum_{T \in \mathcal{T}} \omega_T \int_T [-\varepsilon \Delta (R_h u - u) + \langle \mathbf{c}, \nabla (R_h u - u) \rangle + a (R_h u - u)] \langle \mathbf{c}, \nabla (R_h u - u^h) \rangle d\mathbf{x} \right| \\ &\leq C \sum_{T \in \mathcal{T}} \sqrt{\omega_T} [\varepsilon h_T^{k-1} + h_T^k + h_T^{k+1}] \|u\|_{H^{k+1}(T)} \sqrt{\omega_T} \|\langle \mathbf{c}, \nabla (R_h u - u^h) \rangle\|_{L^2(T)} \leq \text{mit } \varepsilon \omega_T \leq Ch_T^2 \\ &\leq C' \sqrt{\sum_{T \in \mathcal{T}} (\varepsilon + \omega_T) h_T^{2k} \|u\|_{H^{k+1}(T)}^2} \|R_h u - u^h\|_{SD}. \end{aligned} \quad (10.2.20d)$$

Die Kombination der Abschätzungen (10.2.20a-d) nach Kürzen mit $\|R_h u - u^h\|_{SD}$ zeigt (10.2.19). ■

Die bewiesene Ungleichung (10.2.19) legt nahe, die Gewichte so zu wählen, dass die drei Größen ε , ω_T , h_T^2/ω_T von ähnlicher Größenordnung sind, wobei allerdings die Ungleichungen (10.2.17) zu beachten sind. Dies führt auf

$$\omega_T = \begin{cases} \delta_0 h_T & \text{falls } \|\mathbf{c}\|_{L^\infty(T)} h_T > 2\varepsilon \quad (\text{überwiegende Konvektion}), \\ \delta_0 h_T^2/\varepsilon & \text{falls } \|\mathbf{c}\|_{L^\infty(T)} h_T \leq 2\varepsilon \quad (\text{überwiegende Diffusion}). \end{cases} \quad (10.2.21)$$

Stabilität zusammen mit der Konsistenz des Verfahrens liefert das folgende Konvergenzresultat (vgl. [71, Theorem 3.30] für den Beweis).

Satz 10.2.15 (Konvergenz) *Die Voraussetzungen des Lemmas 10.2.14 und (10.2.21) seien erfüllt. Dann genügt die Lösung u^h der Stromlinien-Diffusionsmethode (10.2.15a) der Fehlerabschätzung*

$$\|u - u^h\|_{SD} \leq C (\sqrt{\varepsilon} + \sqrt{h}) h^k |u|_{k+1}.$$

Beweis. Die Kombination von $\|R_h u - u^h\|_{SD} \leq C(\sqrt{\varepsilon} + \sqrt{h}) h^k |u|_{k+1}$ aus (10.2.19) und der aus $\|R_h u - u\|_{H^1(T)} \leq Ch^k \|u\|_{H^{k+1}(T)}$ folgenden Ungleichung $\|R_h u - u\|_{SD} \leq C(\sqrt{\varepsilon} + \sqrt{h}) h^k |u|_{k+1}$ zeigt die Behauptung. ■

Man vergleiche auch Knabner-Angermann [52, §9].

11

Eigenwertprobleme elliptischer Operatoren

11.1 Formulierung der Eigenwertprobleme

Die klassische Formulierung eines Eigenwertproblems lautet

$$Le = \lambda e \quad \text{in } \Omega, \quad B_j e = 0 \quad \text{auf } \Gamma \quad (j = 1, \dots, m). \quad (11.1.1)$$

Dabei ist L ein elliptischer Differentialoperator der Ordnung $2m$, und B_j sind Randoperatoren. Eine Lösung e von (11.1.1) heißt *Eigenfunktion*, falls $e \neq 0$. In diesem Fall ist λ der (zu e gehörende) Eigenwert.

Wie in Kapitel 7 kann man die klassische Darstellung (11.1.1) durch eine Variationsformulierung ersetzen, wobei eine geeignete Bilinearform $a(\cdot, \cdot) : V \times V \rightarrow \mathbb{R}$ an die Stelle von $\{L, B_j\}$ tritt :

$$\text{suche } e \in V, \text{ sodass } a(e, v) = \lambda(e, v)_0 \quad \text{für alle } v \in V. \quad (11.1.2a)$$

$(u, v)_0 = \int_{\Omega} uv dx$ ist das $L^2(\Omega)$ -Skalarprodukt. Eigentlich hat man $(\cdot, \cdot)_0$ durch $(\cdot, \cdot)_U$ zu ersetzen, wobei U der Hilbert-Raum des Gelfand-Dreiers $V \subset U \subset V'$ ist (vgl. Abschnitt 6.3.3). Wir beschränken uns hier aber auf den Standardfall $U = L^2(\Omega)$.

Als *adjungiertes Eigenwertproblem* bezeichnet man die Aufgabe

$$\text{suche } e^* \in V, \text{ sodass } a(v, e^*) = \bar{\lambda}(v, e^*)_0 \quad \text{für alle } v \in V. \quad (11.1.2b)$$

Definition 11.1.1 Sei $\lambda \in \mathbb{C}$. Mit $E(\lambda)$ wird der Unterraum aller $e \in V$ bezeichnet, die Gleichung (11.1.1) bzw. (11.1.2a) erfüllen. $E(\lambda)$ heißt der Eigenraum zu λ . Mit $E^*(\lambda)$ bezeichnet man den entsprechenden Eigenraum von Gleichung (11.1.2b). λ heißt Eigenwert, falls $\dim E(\lambda) \geq 1$.

Die Sätze 6.5.15 und 7.2.14 enthalten bereits die folgenden Aussagen:

Satz 11.1.2 $V \subset L^2(\Omega)$ sei stetig, dicht und kompakt eingebettet [z.B. sei $V = H_0^m(\Omega)$ mit beschränktem Ω]. $a(\cdot, \cdot)$ sei V -koerziv. Dann haben die Aufgaben (11.1.2a, b) abzählbar viele Eigenwerte $\lambda \in \mathbb{C}$, die sich nur in ∞ häufen. Für alle $\lambda \in \mathbb{C}$ gilt

$$\dim E(\lambda) = \dim E^*(\lambda) < \infty. \quad (11.1.3)$$

Übungsaufgabe 11.1.3 Zusätzlich zu den Voraussetzungen des Satzes 11.1.2 sei $a(\cdot, \cdot)$ symmetrisch. Man zeige: Alle Eigenwerte sind reell, und die Aufgaben (11.1.2a), (11.1.2b) sind identisch, sodass $E(\lambda) = E^*(\lambda)$.

Für gewöhnliche Differentialgleichungen zweiter Ordnung (d.h. $\Omega \subset \mathbb{R}^1$, $m = 1$) weiß man, dass alle Eigenwerte einfach sind: $\dim E(\lambda) = 1$. Diese Aussage ist für partielle Differentialgleichungen falsch, wie das folgende Beispiel zeigt.

Beispiel 11.1.4 (Vielfachheit der Eigenwerte) Die Poisson-Gleichung $-\Delta e = \lambda e$ im Rechteck $(0, a) \times (0, b)$ mit Dirichlet-Randwerten $e = 0$ auf Γ hat die Eigenwerte

$$\lambda = (\nu\pi/a)^2 + (\mu\pi/b)^2 \quad \text{für alle } \nu, \mu \in \mathbb{N}.$$

Die zugehörige Eigenfunktion lautet $e(x, y) = e^{\nu, \mu}(x, y) := \sin(\nu\pi x/a) \sin(\mu\pi y/b)$. Im Falle des Quadrates $a = b$ ergeben sich für $\nu \neq \mu$ Eigenwerte $\lambda = \lambda^{\nu, \mu} = \lambda^{\mu, \nu}$, die mindestens die Vielfachheit 2 haben, da $e^{\nu, \mu}$ und $e^{\mu, \nu}$ linear unabhängige Eigenfunktionen zum gleichen Eigenwert sind. Ein dreifacher Eigenwert liegt z.B. für $a = b$, $\lambda = 50\pi^2/a^2$ vor: $E(\lambda) = \text{span}\{e^{1,7}, e^{7,1}, e^{5,5}\}$.

Die Eigenfunktionen $e \in E(\lambda)$ gehören definitionsgemäß zu V . Die Regularitätsuntersuchungen des Abschnittes 9.1 ergeben sofort höhere Regularität.

Satz 11.1.5 Es seien $V = H_0^m(\Omega)$ oder $V = H^m(\Omega)$. Im zweiten Falle sei $m = 1$. Unter den Voraussetzungen der Sätze 9.1.20 bzw. 9.1.21 gilt $E(\lambda) \subset H^{m+s}(\Omega)$.

Beweis. Mit $a(\cdot, \cdot)$ erfüllt auch $a_\lambda(u, v) := a(u, v) - \lambda(u, v)_0$ die Voraussetzungen. Da $a_\lambda(e, v) = 0$ für $e \in E(\lambda)$, $v \in V$, folgt die Behauptung aus Korollar 9.1.23. ■

Neben der Standardform (11.1.2a) gibt es verallgemeinerte Eigenwertaufgaben. Ein Beispiel ist das *Steklov-Problem*

$$-\Delta e = 0 \quad \text{in } \Omega, \quad \partial e / \partial n = \lambda e \quad \text{auf } \Gamma,$$

dessen Variationsformulierung $e \in H^1(\Omega)$, $\int_\Omega \langle \nabla e, \nabla v \rangle dx = \lambda \int_\Gamma e v d\Gamma$ ($v \in H^1(\Omega)$) lautet. Es lässt sich zeigen, dass alle Eigenwerte reell sind und die Aussagen des Satzes 11.1.2 gelten.

11.2 Finite-Element-Diskretisierung

11.2.1 Diskretisierung

$V_h \subset V$ sei ein (Finite-Element-)Unterraum. Die Ritz-Galerkin- (bzw. Finite-Element-)Diskretisierungen der Eigenwertprobleme (11.1.2a,b) lauten

$$\text{suche } e^h \in V_h, \text{ sodass } a(e^h, v) = \lambda_h(e^h, v)_0 \quad \text{für alle } v \in V_h. \quad (11.2.1a)$$

$$\text{suche } e^{*h} \in V_h, \text{ sodass } a(v, e^{*h}) = \bar{\lambda}_h(v, e^{*h})_0 \quad \text{für alle } v \in V_h. \quad (11.2.1b)$$

Die diskreten Eigenräume $E_h(\lambda_h), E^*(\lambda_h)$ werden von den Lösungen der Aufgaben (11.2.1a) bzw. (11.2.1b) aufgespannt. In Analogie zu Satz 11.1.2 gilt $\dim E_h(\lambda_h) = \dim E^*(\lambda_h)$. Falls $a(\cdot, \cdot)$ symmetrisch ist, gilt $E_h(\lambda_h) = E^*(\lambda_h)$.

Wie in Abschnitt 8.1 lässt sich die Formulierung (11.2.1a,b) in Matrixschreibweise umsetzen.

Bemerkung 11.2.1 \mathbf{e} und \mathbf{e}^* seien die Koeffizientenvektoren zu $e^h = \mathbf{P}\mathbf{e}$ und $e^{*h} = \mathbf{P}\mathbf{e}^*$ (vgl. (8.1.6)). Die Eigenwertaufgaben (11.2.1a,b) sind äquivalent zu

$$\mathbf{L}\mathbf{e} = \lambda_h \mathbf{M}\mathbf{e}, \quad \mathbf{L}^\top \mathbf{e}^* = \bar{\lambda}_h \mathbf{M}\mathbf{e}^*, \quad (11.2.1')$$

wobei die Systemmatrix \mathbf{L} wie in Satz 8.1.3 und die Massematrix \mathbf{M} durch (8.7.7) definiert sind. Da im Allgemeinen $\mathbf{M} \neq \mathbf{I}$, handelt es sich bei (11.2.1') um verallgemeinerte Eigenwertprobleme.

Übungsaufgabe 11.2.2 Man zeige: 1) \mathbf{M} ist positiv definit und besitzt eine Zerlegung $\mathbf{M} = \mathbf{A}^\top \mathbf{A}$ (z.B. $\mathbf{A} = \mathbf{M}^{1/2}$ oder Cholesky-Faktor).

2) Das erste Problem in (11.2.1') lässt sich äquivalent als Standard-Eigenwertproblem $\tilde{\mathbf{L}}\tilde{\mathbf{e}} = \lambda_h \tilde{\mathbf{e}}$ mit $\tilde{\mathbf{L}} := (\mathbf{A}^\top)^{-1} \mathbf{L} \mathbf{A}^{-1}$, $\tilde{\mathbf{e}} = \mathbf{A}\mathbf{e}$ formulieren. Die zweite Aufgabe in (11.2.1') entspricht $\tilde{\mathbf{L}}^\top \tilde{\mathbf{e}}^* = \lambda_h \tilde{\mathbf{e}}^*$ mit $\tilde{\mathbf{e}}^* = \mathbf{A}\mathbf{e}^*$.

Bei der Konvergenzuntersuchung hat man folgende Schwierigkeiten zu beachten:

(i) Eine gleichmäßige Approximation sämtlicher Eigenwerte und -funktionen durch diskrete Eigenwerte und -funktionen ist nicht möglich, da den unendlich vielen Eigenwerten von (11.2.2a) nur endlich viele von

(11.2.1a) gegenüberstehen. Es ist nur möglich, einen festen Eigenwert λ von (11.1.2a) als Häufungspunkt diskreter Eigenwerte $\{\lambda_h : h > 0\}$ zu charakterisieren und Abschätzungen für $|\lambda - \lambda_h|$ aufzustellen.

(ii) Sind λ und λ_h die kontinuierlichen bzw. diskreten Eigenwerte, so braucht keinesfalls $\dim E(\lambda) = \dim E_h(\lambda_h)$ zu gelten. Man beschränkt sich daher gerne auf den Fall einfacher Eigenwerte, wo $\dim E(\lambda) = \dim E_h(\lambda_h) = 1$. Falls $\dim E(\lambda) = k > 1$, muss man damit rechnen, dass der mehrfache Eigenwert λ durch mehrere diskrete Eigenwerte $\lambda_h^{(i)}$ ($i = 1, \dots, k$) approximiert wird, wobei $\dim E(\lambda) = \sum_{i=1}^k \dim E_h(\lambda_h^{(i)})$. Die Fehlerabschätzungen von $|\lambda - \lambda_h^{(i)}|$ sind dann im Allgemeinen schlechter als bei einfachen Eigenwerten. Lediglich für den Mittelwert $\hat{\lambda}_h := \frac{1}{k} \sum_{i=1}^k \lambda_h^{(i)}$ erhält man die üblichen Abschätzungen (vgl. Babuška-Aziz [5, p. 338]).

Übungsaufgabe 11.2.3 Das Eigenwertproblem sei wie in Beispiel 11.1.4 mit $a = b$ und $\lambda = 50\pi^2/a^2$. V_h bestehe aus linearen Elementen über einer Quadratgittertriangulation. Man zeige: Dem angegebenen dreifachen Eigenwert λ entsprechen ein doppelter Eigenwert $\lambda_h^{(1)}$ und ein davon verschiedener einfacher Eigenwert $\lambda_h^{(2)}$ mit $\lim_{h \rightarrow 0} \lambda_h^{(i)} = \lambda$. Hinweis: Die Knotenwerte der diskreten Eigenfunktionen stimmen mit den kontinuierlichen Eigenfunktionen $e^{1,7}$, $e^{7,1}$ und $e^{5,5}$ überein.

11.2.2 Qualitative Konvergenzresultate

In diesem Abschnitt geht es um die Frage, ob $\lambda_h \rightarrow \lambda$ und $e^h \rightarrow e$ für $h \rightarrow 0$. Die Konvergenzgeschwindigkeit wird in Abschnitt 11.2.3 diskutiert werden. Grundvoraussetzung sei im Folgenden:

$$a(\cdot, \cdot) : V \times V \rightarrow \mathbb{C} \quad \text{sei } V\text{-koerziv,} \quad (11.2.2a)$$

$$V \subset L^2(\Omega) \quad \text{sei stetig, dicht und kompakt eingebettet,} \quad (11.2.2b)$$

sodass die Riesz-Schauder-Theorie (Satz 11.1.2) anwendbar ist. Außerdem sei eine Folge von Unterräumen V_{h_i} ($h_i \rightarrow 0$) gegeben, die V zunehmend approximieren (vgl. (8.2.4a)):

$$\lim_{h \rightarrow 0} \inf \{ \|u - v\|_V : v \in V_h \} = 0 \quad \text{für alle } u \in V. \quad (11.2.2c)$$

Wir definieren

$$a_\lambda(\cdot, \cdot) : V \times V \rightarrow \mathbb{C}, \quad a_\lambda(u, v) := a(u, v) - \lambda(u, v)_0, \quad (11.2.3a)$$

$$\omega(\lambda) := \inf_{u \in V, \|u\|_V=1} \sup_{v \in V, \|v\|_V=1} |a_\lambda(u, v)|, \quad (11.2.3b)$$

$$\omega_h(\lambda) := \inf_{u \in V_h, \|u\|_V=1} \sup_{v \in V_h, \|v\|_V=1} |a_\lambda(u, v)|. \quad (11.2.3c)$$

Die Interpretation der Größen $\omega(\lambda)$ und $\omega_h(\lambda)$ ist Gegenstand der folgenden Übungsaufgabe.

Übungsaufgabe 11.2.4 L und L_h seien die zu $a(\cdot, \cdot) : V \times V \rightarrow \mathbb{R}$ und $a(\cdot, \cdot) : V_h \times V_h \rightarrow \mathbb{R}$ gehörenden Operatoren (vgl. (8.1.11a)). Man zeige:

(i) Ist λ kein Eigenwert, so gilt

$$\omega(\lambda) = 1 / \left\| (L - \lambda I)^{-1} \right\|_{V \leftarrow V'}, \quad \omega_h(\lambda) = 1 / \left\| (L_h - \lambda I)^{-1} \right\|_{V_h \leftarrow V'_h} \quad (11.2.4)$$

(vgl. Lemma 6.5.3 und Übungsaufgabe 8.1.17).

(ii) $\omega(\lambda)$ und $\omega_h(\lambda)$ sind stetig in $\lambda \in \mathbb{C}$.

(iii) Ersetzt man in (11.2.3b,c) $a_\lambda(u, v)$ durch $a_\lambda(v, u)$, erhält man die Größen $\omega^*(\lambda)$ und $\omega_h^*(\lambda)$, die dem adjungierten Problem entsprechen. Es gilt $\omega^*(\lambda) = \omega(\lambda)$ und $\omega_h^*(\lambda) = \omega_h(\lambda)$. Hinweis: Lemma 6.5.17.

Mit Hilfe von (11.2.4) und Satz 6.5.15 beweist man den folgenden Zusammenhang zwischen $\omega(\lambda)$, $\omega_h(\lambda)$ und den Eigenwertproblemen.

Bemerkung 11.2.5 Es gelte (11.2.2a,b). λ ist genau dann Eigenwert von (11.1.2a), wenn $\omega(\lambda) = 0$. λ_h ist genau dann Eigenwert von (11.2.1a), wenn $\omega_h(\lambda_h) = 0$.

Lemma 11.2.6 $a(\cdot, \cdot)$ sei V -koerziv (vgl. (11.2.2a)). Dann gibt es ein $\mu \in \mathbb{R}$, sodass $a_\mu(\cdot, \cdot)$ V -elliptisch ist. Es gilt $\omega(\mu) \geq 1/C_E$ und $\omega_h(\mu) \geq 1/C_E$ mit $C_E > 0$ aus Definition 6.5.13.

Lemma 11.2.7 $\Lambda \subset \mathbb{C}$ sei kompakt. Es gelte (11.2.2a-c). Dann gibt es von $\lambda \in \Lambda$ unabhängige Zahlen $C > 0$ und $\eta(h) > 0$ mit $\lim_{h \rightarrow 0} \eta(h) = 0$, sodass

$$\omega_h(\lambda) \geq C\omega(\lambda) - \eta(h), \quad \omega(\lambda) \geq C\omega_h(\lambda) - \eta(h). \quad (11.2.5)$$

Beweis. Die Operatoren $Z = Z(\lambda) : V \rightarrow V$ und $Z_h = Z_h(\lambda) : V \rightarrow V_h$ seien wie folgt definiert:

$$z := Z(\lambda)u \in V \quad \text{sei Lösung von} \quad a_\mu(z, v) = (\lambda - \mu)(u, v)_0 \quad \text{für alle } v \in V, \quad (11.2.6a)$$

$$z^h := Z_h(\lambda)u \in V \quad \text{sei Lösung von} \quad a_\mu(z^h, v) = (\lambda - \mu)(u, v)_0 \quad \text{für alle } v \in V_h, \quad (11.2.6b)$$

wobei μ gemäß Lemma 11.2.6 gewählt ist. Nach Lemma 11.2.6 gibt es ein C_Z mit

$$\|Z(\lambda)\|_{V \leftarrow V'} \leq C_Z, \quad \|Z_h(\lambda)\|_{V \leftarrow V'} \leq C_Z \quad \text{für alle } \lambda \in \Lambda. \quad (11.2.6c)$$

Aus $a_\lambda(u, v) = a(u, v) - \lambda(u, v)_0 = a(u, v) - \mu(u, v)_0 - (\lambda - \mu)(u, v)_0 = a_\mu(u, v) - a_\mu(z, v) = a_\mu(u - z, v)$ ($z := Z(\lambda)u$) und der Definition von $\omega(\lambda)$ folgert man

$$\omega(\lambda) \|u\|_V \leq \sup_{v \in V, \|v\|_V=1} |a_\lambda(u, v)| = \sup_{v \in V, \|v\|_V=1} |a_\mu(u - z, v)| \leq C_S \|u - z\|_V \quad (11.2.6d)$$

mit $C_S := \|L - \mu I\|_{V' \leftarrow V}$. Für $u \in V_h$ schließt man aus $a_\lambda(u, v) = a_\mu(u - z^h, v)$ für $z^h = Z_h(\lambda)u$, Lemma 11.2.6 und (11.2.6d), dass

$$\begin{aligned} \sup_{v \in V_h, \|v\|_V=1} |a_\lambda(u, v)| &= \sup_{v \in V_h, \|v\|_V=1} |a_\lambda(u - z^h, v)| \geq C_E^{-1} \|u - z^h\|_V \geq C_E^{-1} [\|u - z\|_V - \|z - z^h\|_V] \\ &\geq C_E^{-1} [C_S^{-1} \omega(\lambda) - \|Z - Z_h\|_{V \leftarrow V}] \|u\|_V \quad \text{für alle } u \in V_h. \end{aligned}$$

Hieraus folgt der erste Teil von (11.2.5) mit $C = (C_E C_S)^{-1} > 0$ und $\eta(h) = \|Z - Z_h\|_{V \leftarrow V} / C_E$, falls

$$\lim_{h \rightarrow 0} \sup_{\lambda \in \Lambda} \|Z - Z_h\|_{V \leftarrow V} = 0. \quad (11.2.6e)$$

Der Beweis von (11.2.6e) wird indirekt geführt. Die Negation lautet: Es gibt $\varepsilon > 0$, $\lambda_i \in \Lambda$, $h_i \rightarrow 0$ mit $\|Z(\lambda_i) - Z_{h_i}(\lambda_i)\|_{V \leftarrow V} \geq \varepsilon > 0$. Dann existieren $u_i \in V$ mit

$$\|u_i\|_V = 1, \quad \|[Z(\lambda_i) - Z_{h_i}(\lambda_i)] u_i\|_V \geq \varepsilon/2 > 0. \quad (11.2.6f)$$

Wegen (11.2.2b) und der Kompaktheit von Λ gibt es eine Teilfolge $\lambda_j \in \Lambda$, $u_j \in V$ mit $\lim \lambda_j = \lambda^*$ und $\lim u_j = u^* \in V'$ in V' . (11.2.2c) und Satz 8.2.2 zeigen $\|[Z(\lambda^*) - Z_{h_j}(\lambda^*)] u^*\|_V \rightarrow 0$. Zusammen mit (11.2.6c) ergibt sich

$$\begin{aligned} \|[Z(\lambda_j) - Z_{h_j}(\lambda_j)] u_j\|_V &\leq \|[Z(\lambda_j) - Z(\lambda^*)] u_j\|_V + \|[Z_{h_j}(\lambda^*) - Z_{h_j}(\lambda_j)] u_j\|_V \\ &\quad + \|Z(\lambda^*)[u_j - u^*]\|_V + \|Z_{h_j}(\lambda^*)[u_j - u^*]\|_V + \|[Z(\lambda^*) - Z_{h_j}(\lambda^*)] u^*\|_V \\ &\leq 2C |\lambda_j - \lambda^*| + 2C_Z \|u_j - u^*\|_{V'} + \|[Z(\lambda^*) - Z_{h_j}(\lambda^*)] u^*\|_V \rightarrow 0 \end{aligned}$$

im Widerspruch zu (11.2.6f).

Zum Beweis des zweiten Teils von (11.2.5) ersetzt man (11.2.6d) und die nachfolgende Abschätzung durch

$$\omega_h(\lambda) \|u^h\|_V \leq \sup_{v \in V_h, \|v\|_V=1} |a_\mu(u^h - z^h, v)| \leq C_S \|u^h - z^h\|_V \quad \text{für alle } u^h \in V_h,$$

$$\begin{aligned} \text{und} \quad \sup_{v \in V_h, \|v\|_V=1} |a_\lambda(u^h, v)| &= \sup_{v \in V_h, \|v\|_V=1} |a_\lambda(u^h - z, v)| \geq C_E^{-1} \|u^h - z\|_V \\ &\geq C_E^{-1} [C_S^{-1} \omega_h(\lambda) - \|Z(\lambda) - Z_h(\lambda)\|_{V \leftarrow V}] \|u^h\|_V \quad \text{für alle } u^h \in V_h. \end{aligned}$$

Sei $u \in V$ mit $\|u\|_V = 1$ so gewählt, dass $\sup_{v \in V, \|v\|_V=1} |a_\lambda(u, v)| = \inf_{u \in V, \|u\|_V=1} \sup_{v \in V, \|v\|_V=1} |a_\lambda(u, v)| = \omega(\lambda)$.
 Da $\sup |a_\lambda(u - u^h, v)| \leq C_S \|u - u^h\|_V$, folgt für beliebiges $u^h \in V_h$:

$$\omega(\lambda) \geq C\omega_h(\lambda) - \|Z(\lambda) - Z_h(\lambda)\|_{V \leftarrow V} - C_S \|u - u^h\|_V.$$

Aus (11.2.2c) und (11.2.6e) folgt der zweite Teil von (11.2.5). ■

Eine Folgerung aus Lemma 11.2.7 ist der Satz 8.2.8. Falls Problem (8.1.1) für alle $f \in V'$ lösbar ist, kann $\lambda = 0$ kein Eigenwert sein, d.h. $\omega(0) > 0$. Somit folgt für ε_{N_i} aus Satz 8.2.8, dass $\varepsilon_{N_i} \geq \omega_{N_i}(0) \geq \varepsilon := \frac{1}{2}C\omega(\lambda) > 0$ für hinreichend großes i .

Eine zweite Folgerung betrifft die Konvergenz der diskreten Eigenwerte.

Satz 11.2.8 *Es gelte (11.2.2a-c). Sind λ_{h_i} ($i \rightarrow \infty$, $h_i \rightarrow 0$) diskrete Eigenwerte von (11.2.1a) mit $\lambda_{h_i} \rightarrow \lambda_0$, so ist λ_0 ein Eigenwert von (11.1.2a).*

Beweis. Wäre λ_0 kein Eigenwert, so wäre $\omega(\lambda) \geq \eta_0 > 0$ in der ε -Umgebung $K_\varepsilon(\lambda_0)$, da $\omega(\lambda)$ stetig ist (vgl. Übungsaufgabe 11.2.4ii). Es gäbe $h_0 > 0$, sodass $\eta(h) \leq C\eta_0/2$ für alle $h \leq h_0$ (C und $\eta(h)$ aus (11.2.5)). Für alle $\lambda_{h_i} \in K_\varepsilon(\lambda_0)$ mit $h_i \leq h_0$ folgt der Widerspruch

$$0 = \omega_{h_i}(\lambda_{h_i}) \geq C\omega(\lambda_{h_i}) - \eta(h_i) \geq C\eta_0 - \frac{1}{2}C\eta_0 = \frac{1}{2}C\eta_0 > 0$$

aus (11.2.5). ■

Lemma 11.2.9 (Minimumprinzip) *Es gelte (11.2.2a,b). Die Funktionen $\omega(\lambda)$ und $\omega_h(\lambda)$ haben im Inneren von $\Lambda \subset \mathbb{C}$ kein eigentliches, positives Minimum.*

Beweis. L sei der zu $a(\cdot, \cdot)$ gehörende Operator. λ^* mit $\omega(\lambda^*) > 0$ sei beliebig aus dem Inneren von Λ gewählt. Für hinreichend kleines $\varepsilon > 0$ gilt $K_\varepsilon(\lambda^*) \subset \Lambda$ und $\omega(\lambda) > 0$ in $K_\varepsilon(\lambda^*)$. Damit ist $(L - \lambda I)^{-1}$ in $K_\varepsilon(\lambda^*)$ erklärt und holomorph. Für beliebige, aber feste $u, v \in V'$ ist damit $\Omega(\lambda) := \langle (L - \lambda I)^{-1}u, v \rangle_{V \times V'}$, ebenfalls in Λ holomorph. Da Beträge holomorpher Funktionen ihr Maximum auf dem Rand annehmen, ist $|\Omega(\lambda)| \leq \max_{\zeta \in \partial K_\varepsilon(\lambda^*)} |\Omega(\zeta)|$. Für $\|u\|_{V'} = \|v\|_{V'} = 1$ ist $|\Omega(\zeta)| \leq \|(L - \zeta I)^{-1}\|_{V \leftarrow V'}$, sodass

$$|\Omega(\lambda)| \leq \max_{\zeta \in \partial K_\varepsilon(\lambda^*)} \|(L - \zeta I)^{-1}\|_{V \leftarrow V'} = \max_{\zeta \in \partial K_\varepsilon(\lambda^*)} \frac{1}{\omega(\zeta)}.$$

Da $\frac{1}{\omega(\lambda)} = \|(L - \lambda I)^{-1}\|_{V \leftarrow V'}$, das Infimum von $|\Omega(\lambda)|$ über alle $u, v \in V'$ mit $\|u\|_{V'} = \|v\|_{V'} = 1$ ist, folgt

$$\frac{1}{\omega(\lambda)} = \|(L - \lambda I)^{-1}\|_{V \leftarrow V'} \leq \max_{\zeta \in \partial K_\varepsilon(\lambda^*)} \|(L - \zeta I)^{-1}\|_{V \leftarrow V'} = \max_{\zeta \in \partial K_\varepsilon(\lambda^*)} \frac{1}{\omega(\zeta)},$$

d.h. $\omega(\lambda) \geq \min\{\omega(\zeta) : \zeta \in \partial K_\varepsilon(\lambda^*)\}$ (vgl. Übungsaufgabe 11.2.4i). Somit kann $\omega(\lambda)$ in $K_\varepsilon(\lambda^*)$ kein eigentliches Minimum annehmen. Für $\omega_h(\lambda)$ schließt man ebenso. ■

Die Umkehrung des Satzes 11.2.8 enthält der

Satz 11.2.10 *Es gelte (11.2.2a-c). λ_0 sei Eigenwert von (11.1.2a). Dann existieren diskrete Eigenwerte λ_h von (11.2.1a) [für alle h], sodass $\lim_{h \rightarrow 0} \lambda_h = \lambda_0$.*

Beweis. Sei $\varepsilon > 0$ beliebig. Gemäß Satz 11.1.2 ist λ_0 isolierter Eigenwert: $\omega(\lambda) > 0$ für $0 < |\lambda - \lambda_0| \leq \varepsilon$ (ε hinreichend klein). Da $\omega(\lambda)$ stetig und $\partial K_\varepsilon(\lambda_0)$ kompakt, ist $\rho_\varepsilon := \min\{\omega(\lambda) : |\lambda - \lambda_0| = \varepsilon\}$ positiv. Wegen (11.2.5) und $\omega(\lambda_0) = 0$ erhält man für hinreichend kleine h :

$$\omega_h(\lambda) \geq C\omega(\lambda) - \eta(h) \geq C\rho_\varepsilon - \eta(h) > \eta(h)/C \geq \omega_h(\lambda_0) \quad \text{für alle } \lambda \in \partial K_\varepsilon(\lambda_0).$$

Damit muss $\omega_h(\lambda)$ in $K_\varepsilon(\lambda_0)$ ein eigentliches Minimum besitzen. Wegen Lemma 11.2.9 ist der Minimalwert null.

Damit existiert ein $\lambda_h \in K_\varepsilon(\lambda_0)$, das diskreter Eigenwert ist: $\omega_h(\lambda_h) = 0$. ■

Die Konvergenz der Eigenfunktionen erhält man aus

Satz 11.2.11 *Es gelte (11.2.2a-c). $e^h \in E_h(\lambda_h)$ sei eine Folge diskreter Eigenfunktionen mit $\|e^h\|_V = 1$ und $\lim \lambda_h = \lambda_0$. Dann existiert eine Teilfolge e^{h_i} , die in V gegen eine Eigenfunktion $e \in E(\lambda_0)$ konvergiert:*

$$e \in E(\lambda_0), \quad \|e^{h_i} - e\|_V \rightarrow 0 \quad (i \rightarrow \infty), \quad \|e\|_V = 1.$$

Beweis. Die Funktionen e^h sind in V gleichmäßig beschränkt. Da $V \subset L^2(\Omega)$ kompakt eingebettet ist (vgl. (11.2.2b)), existiert eine Teilfolge e^{h_i} , die in $L^2(\Omega)$ gegen ein $e \in L^2(\Omega)$ konvergiert:

$$\|e - e^{h_i}\|_{L^2(\Omega)} \rightarrow 0 \quad (i \rightarrow \infty). \quad (11.2.7a)$$

Wir definieren $z = Z(\lambda_0)e$, $z^{h_i} = Z_{h_i}(\lambda_0)e$ gemäß (11.2.6a,b). Nach Satz 8.2.2 gibt es ein $h_1(\varepsilon) > 0$, sodass

$$\|z - z^{h_i}\|_V \leq \varepsilon/2 \quad \text{für } h_i \leq h_1(\varepsilon). \quad (11.2.7b)$$

e^{h_i} löst

$$a_\mu(e^{h_i}, v) = (\lambda_{h_i} - \mu)(e^{h_i}, v)_0 \quad \text{für alle } v \in V_{h_i}. \quad (11.2.7c)$$

Kombination von $z^{h_i} = Z_{h_i}(\lambda_0)e$ [d.h. (11.2.6b) für $\lambda = \lambda_0$] und (11.2.7c) liefert

$$a_\mu(z^{h_i} - e^{h_i}, v) = F_i(v) := (\lambda_0 - \mu)(e - e^{h_i}, v)_0 - (\lambda_{h_i} - \lambda_0)(e^{h_i}, v)_0 \quad \text{für alle } v \in V_{h_i}. \quad (11.2.7d)$$

Da $\|F_i\|_{V'} \rightarrow 0$ wegen $\lambda_{h_i} \rightarrow \lambda_0$ und (11.2.7a), gibt es ein $h_2(\varepsilon) > 0$, sodass $\|F_i\|_{V'} \leq \varepsilon/[2C_E]$ (C_E aus Lemma 11.2.6) und

$$\|z^{h_i} - e^{h_i}\|_V \leq \varepsilon/2 \quad \text{für } h_i \leq h_2(\varepsilon). \quad (11.2.7e)$$

(11.2.7b) und (11.2.7e) zeigen $\|z - e^{h_i}\|_V \leq \varepsilon$ für $h_i \leq \min\{h_1(\varepsilon), h_2(\varepsilon)\}$, also $\lim_{i \rightarrow \infty} e^{h_i} = z$ in V . Damit gilt aber auch $\lim e^{h_i} = z$ in $L^2(\Omega) \subset V$. (11.2.7a) beweist $z = e \in V$, sodass $e = z = Z(\lambda_0)e$ zu $a(e, v) = \lambda_0(e, v)_0$ wird. Somit ist $e = \lim e^{h_i}$ Eigenfunktion von (11.1.2a). Insbesondere ist $\|e\|_V = \lim \|e^{h_i}\|_V = 1$. ■

Übungsaufgabe 11.2.12 *Es gelte (11.2.2a-c). $\lambda_h, e^h, \lambda_0$ und e seien wie in Satz 11.2.11. Man zeige:*
a) *Wenn $\dim E(\lambda_0) = 1$, so auch $\lim_{h \rightarrow 0} \dim E_h(\lambda_h) = 1$.*
b) *Es sei $\dim E(\lambda_0) = 1$. Dann gilt $\lim \hat{e}^h = e$ in V für $\hat{e}^h := e^h / (e^h, e)_V$, falls $|(e^h, e)_V| \geq 1/2$, und $\hat{e}^h := e^h$ sonst.*

11.2.3 Quantitative Konvergenzresultate

Die geometrische und algebraische Vielfachheit eines Eigenwertes λ_0 von (11.1.2a) stimmen überein, falls¹

$$\dim \text{Kern}(L - \lambda_0 I) = \dim \text{Kern}((L - \lambda_0 I)^2). \quad (11.2.8)$$

Lemma 11.2.13 *Es gelte (11.2.2a,b) und $\dim E(\lambda_0) = 1$. Dann ist (11.2.8) äquivalent zu $(e, e^*)_0 \neq 0$ für $e \in E(\lambda_0) \setminus \{0\}$, $0 \neq e^* \in E^*(\lambda_0) \setminus \{0\}$.*

Beweis. Es ist $\text{Kern}(L - \lambda_0 I) = E(\lambda_0) = \text{span}\{e\}$. $\dim \text{Kern}(L - \lambda_0 I)^2 > 1$ gilt genau dann, wenn eine Lösung $v \in V$ zu $(L - \lambda_0 I)v = e$ existiert. Nach Satz 6.5.15c hat diese Gleichung genau dann eine Lösung, wenn $(e, e^*)_0 = 0$. Damit ist (11.2.8) zu $(e, e^*)_0 \neq 0$ äquivalent. ■

Seien $E(\lambda_0) = \text{span}\{e\}$ und $E^*(\lambda_0) = \text{span}\{e^*\}$. Unter der Voraussetzung (11.2.8) können e und e^* so normiert werden, dass

$$(e, e^*)_0 = 1. \quad (11.2.8')$$

¹Die Vielfachheiten stimmen nicht überein, wenn die Jordan-Normalform von $L - \lambda_0 I$ einen echten $k \times k$ -Jordan-Block $J = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1 \\ & & & 0 \end{bmatrix}$ enthält, d.h. wenn $k \geq 2$. Da $\dim \text{Kern}(J) = 1 < \dim \text{Kern}(J^2) = 2$, folgt die Aussage.

Wir definieren $\hat{V} := \{v \in V : (v, e^*)_0 = 0\}$, $\hat{V}' = \{v' \in V' : (v', e^*)_0 = 0\}$. $\|\cdot\|_{\hat{V}'}$ sei die Dualnorm zu $\|\cdot\|_{\hat{V}} = \|\cdot\|_V$. Zum Problem (11.2.9):

$$\text{zu } f \in \hat{V}' \text{ suche } u \in \hat{V} \text{ mit } a_\lambda(u, v) = (f, v)_0 \text{ für alle } v \in \hat{V}, \quad (11.2.9)$$

definiert man die (11.2.3b) entsprechende Größe

$$\hat{\omega}(\lambda) := \inf_{u \in \hat{V}, \|u\|_V=1} \sup_{v \in \hat{V}, \|v\|_V=1} |a_\lambda(u, v)|. \quad (11.2.10)$$

Lemma 11.2.14 *Es gelte (11.2.2a,b), (11.2.8) und $\dim E(\lambda_0) = 1$. Dann gibt es ein $\varepsilon > 0$, sodass $\hat{\omega}(\lambda) \geq C > 0$ für alle $|\lambda - \lambda_0| \leq \varepsilon$. Aufgabe (11.2.9) hat genau dann, wenn $\hat{\omega}(\lambda) > 0$, eine Lösung $u \in \hat{V}$. Diese erfüllt*

$$\|u\|_V \leq \|f\|_{V'} / \hat{\omega}(\lambda).$$

Beweis. $\hat{L} : \hat{V} \rightarrow \hat{V}'$ sei der zu $a(\cdot, \cdot) : \hat{V} \times \hat{V} \rightarrow \mathbb{C}$ gehörende Operator. Für $0 < |\lambda - \lambda_0| < \varepsilon$ (ε hinreichend klein) hat $(L - \lambda I)u = f$ eine eindeutige Lösung $u \in V$. Aus $f \in \hat{V}'$ folgt

$$0 = (f, e^*)_0 = ([L - \lambda I]u, e^*)_0 = (u, [L - \lambda I]^* e^*)_0 = (\bar{\lambda}_0 - \bar{\lambda})(u, e^*)_0,$$

d.h. $u \in \hat{V}$. Somit existiert $(\hat{L} - \lambda I)^{-1} : \hat{V}' \rightarrow \hat{V}$ als Beschränkung von $(L - \lambda I)^{-1}$ auf $\hat{V}' \subset V'$. Für $\lambda = \lambda_0$ hat Aufgabe (11.2.9) gemäß Satz 6.5.15c eine eindeutige Lösung. Damit existiert $(\hat{L} - \lambda I)^{-1}$ für alle $\lambda \in \underline{K}_\varepsilon(\lambda_0)$. Nach Bemerkung 11.2.5 (mit \hat{V} statt V) muss $\hat{\omega}(\lambda)$ in $\underline{K}_\varepsilon(\lambda_0)$ positiv sein. Die Stetigkeit von $\hat{\omega}(\lambda)$ beweist $\hat{\omega}(\lambda) \geq C > 0$. In Analogie zu (11.2.4) hat man $\|u\|_V = \|u\|_{\hat{V}} \leq \|f\|_{\hat{V}'} / \hat{\omega}(\lambda)$. Die Abschätzung durch $\|f\|_{V'} / \hat{\omega}(\lambda)$ ergibt sich aus der folgenden Übungsaufgabe. ■

Übungsaufgabe 11.2.15 *Man zeige: $\|f\|_{\hat{V}'} \leq \|f\|_{V'}$ für alle $f \in \hat{V}'$.*

Lemma 11.2.16 *Es gelte (11.2.2a-c), $\dim E(\lambda_0) = 1$ und (11.2.8). Seien λ_h diskrete Eigenwerte mit $\lim \lambda_h = \lambda_0$. Nach Übungsaufgabe 11.2.12b gibt es*

$$e^h \in E_h(\lambda_h) \text{ und } e^{*h} \in E_h^*(\lambda_h) \text{ mit } e^h \rightarrow e \in E(\lambda_0), e^{*h} \rightarrow e^* \in E^*(\lambda_0), (e, e^*)_0 = 1.$$

Damit lassen sich der Raum $\hat{V}_h := \{v^h \in V_h : (v^h, e^h)_0 = 0\}$ und die Größe

$$\hat{\omega}_h(\lambda) := \inf_{u \in \hat{V}_h, \|u\|_V=1} \sup_{v \in \hat{V}_h, \|v\|_V=1} |a_\lambda(u, v)|$$

konstruieren. Dann gibt es ein von h und $\lambda \in \mathbb{C}$ unabhängiges $C > 0$, sodass $\hat{\omega}_h(\lambda) \geq C\omega_h(\lambda)$. Für hinreichend kleine $\varepsilon > 0$ und h ist $\hat{\omega}_h(\lambda) \geq \eta > 0$ für alle $|\lambda - \lambda_0| \leq \varepsilon$.

Beweis. 1) 1. Behauptung: Es gibt $h_0 > 0$ und $\hat{C} > 0$, sodass

$$\min \{ \|v + \alpha e^{*h}\|_V : \alpha \in \mathbb{C} \} \geq \hat{C} \|v\|_V \quad \text{für alle } v \in \hat{V}_h \text{ und alle } 0 < h \leq h_0. \quad (*)$$

Der Beweis wird indirekt geführt. Die Negation lautet: Es gibt eine Folge $\alpha_i \in \mathbb{C}$, $h_i \rightarrow 0$, $v_i \in \hat{V}_{h_i}$ mit $\|v_i\|_V = 1$ und $\|v_i + \alpha_i e^{*h_i}\|_V \rightarrow 0$. Damit existieren Teilfolgen mit $\alpha_i \rightarrow \alpha^*$, $v_i \rightarrow v^*$ in $L^2(\Omega)$. Offenbar muss $w_i := v_i + \alpha_i e^{*h_i}$ den Grenzwert $w^* = \lim w_i = v^* + \alpha^* e^*$ in $L^2(\Omega)$ besitzen. Da

$$\|w^*\|_{L^2(\Omega)} = \lim \|w_i\|_{L^2(\Omega)} \leq C \lim \|w_i\|_V = 0,$$

folgt $w^* = 0$, also $v^* = -\alpha^* e^*$. Aus $0 = \lim (v_i, e^{*h_i})_0 = (v^*, e^*)_0 = -\alpha^* \|e^*\|_0^2$ schließt man $\alpha^* = 0$. Damit ergibt sich der Widerspruch aus $1 = \lim \|v_i\|_V \leq \limsup \|w_i\|_V + \limsup \|\alpha_i e^{*h_i}\|_V = \lim \|w_i\|_V = 0$.

2) 2. Behauptung: $\hat{\omega}_h(\lambda) \geq \hat{C}\omega_h(\lambda)$ mit \hat{C} aus 1). Dies folgt aus

$$\begin{aligned} \hat{\omega}_h(\lambda) & \stackrel{(11.2.10)}{=} \inf_{u \in \hat{V}_h \setminus \{0\}} \sup_{v \in \hat{V}_h \setminus \{0\}} \frac{|a_\lambda(u, v)|}{\|u\|_V \|v\|_V} \\ & \geq \hat{C} \inf_{u \in \hat{V}_h \setminus \{0\}} \sup_{v \in \hat{V}_h \setminus \{0\}} \max_{\alpha \in \mathbb{C}} \frac{|a_\lambda(u, v + \alpha e^{*h})|}{\|u\|_V \|v + \alpha e^{*h}\|_V} \\ & \stackrel{(*) \text{ und } a_\lambda(u, v) = a_\lambda(u, v + \alpha e^{*h}) \text{ für } u \in \hat{V}_h}{=} \hat{C} \inf_{u \in \hat{V}_h \setminus \{0\}} \sup_{w \in V_h \setminus \{0\}} \frac{|a_\lambda(u, w)|}{\|u\|_V \|w\|_V} \geq \hat{C} \inf_{u \in V_h \setminus \{0\}} \sup_{w \in V_h \setminus \{0\}} \frac{|a_\lambda(u, w)|}{\|u\|_V \|w\|_V} = \hat{C}\omega_h(\lambda). \end{aligned}$$

3) $\varepsilon > 0$ sei so gewählt, dass λ_0 einziger Eigenwert in $\overline{K_\varepsilon(\lambda_0)}$ ist. Für hinreichend kleines h ist λ_h einziger diskreter Eigenwert in $\overline{K_\varepsilon(\lambda_0)}$. Im Beweis des Satzes 10 wurde bereits $\omega_h(\lambda) \geq \eta' > 0$ für $\lambda \in \partial K(\lambda_0)$, $h \leq h_0(\varepsilon)$ benutzt. Aus Teil 2) folgt $\hat{\omega}_h(\lambda) \geq \eta := \eta' C > 0$ für $\lambda \in \partial K_\varepsilon(\lambda_0)$. Nach Übungsaufgabe 11.2.12a ist $a_\lambda(u, v) = (f, v)_0$ ($v \in V_h$) für jedes $f \in V_h$ und alle $\lambda \in K_\varepsilon(\lambda_0)$ lösbar, sodass $\hat{\omega}_h(\lambda) = 0$ ausgeschlossen ist. Lemma 11.2.9 zeigt $\hat{\omega}_h(\lambda) \geq \eta > 0$ in $K_\varepsilon(\lambda_0)$. ■

Übungsaufgabe 11.2.17 Es gelte (11.2.2a,b). Die Funktionen $u, v \in V$ mögen $(u, v)_0 \neq 0$ erfüllen. $d(\cdot, V_h)$ sei wie in (8.2.2) definiert. Man zeige: Ist $d(u, V_h)$ hinreichend klein, so gibt es ein $u^h \in V_h$ mit

$$\|u^h - u\|_V \leq 2d(u, V_h) \quad \text{mit der Normierung} \quad (u^h - u, v)_0 = 0.$$

Lemma 11.2.18 Es gelte (11.2.2a-c). λ_0 sei Eigenwert mit den Eigenschaften (11.2.8) und $\dim E(\lambda_0) = 1$. Für hinreichend kleines h gibt es $e^h \in E_h(\lambda_h)$ mit

$$\|e^h - e\|_V \leq C[|\lambda_0 - \lambda_h| + d(e, V_h)].$$

Beweis. $z^h := Z_h(\lambda_0)e$ sei die Lösung von (11.2.6b). Da $e = Z(\lambda_0)e$, hat man

$$\|e - z^h\|_V \leq C_1 d(e, V_h). \quad (11.2.11a)$$

Sei μ der Wert aus Lemma 11.2.6. Für alle $v \in V_h$ ist

$$\begin{aligned} a_\mu(z^h - e^h, v) &= a_\mu(z^h, v) - a_\mu(e^h, v) \stackrel{(11.2.6b)}{=} (\lambda_0 - \mu)(e, v)_0 - a_\mu(e^h, v) \stackrel{(11.2.1a)}{=} \\ &= (\lambda_0 - \mu)(e, v)_0 - (\lambda_h - \mu)(e^h, v)_0 = (\lambda_0 - \lambda_h)(e, v)_0 + (\lambda_h - \mu)(e - z^h, v)_0 + (\lambda_h - \mu)(z^h - e^h, v)_0, \end{aligned}$$

sodass

$$a_{\lambda_h}(z^h - e^h, v) = a_\mu(z^h - e^h, v) - (\lambda_h - \mu)(z^h - e^h, v)_0 = (\lambda_0 - \lambda_h)(e, v)_0 + (\lambda_h - \mu)(e - z^h, v)_0. \quad (11.2.11b)$$

Wegen Lemma 11.2.13 und $e^h \rightarrow e$, $e^{*h} \rightarrow e^*$ (vgl. Lemma 11.2.16) gilt $|(e, e^{*h})_0|, |(e^h, e^{*h})_0| \geq \eta > 0$ für hinreichend kleines h . Damit kann e^h so skaliert werden, dass $(z^h - e^h, e^{*h})_0 = 0$. Folglich entspricht (11.2.11b) der Aufgabe (11.2.9) mit \tilde{V}_h statt \tilde{V} , $z^h - e^h$ statt u und $(\lambda_0 - \lambda_h)e + (\lambda_h - \mu)(e - z^h)$ statt f . Lemma 11.2.14 beweist

$$\|z^h - e^h\|_V \leq \hat{\omega}_h(\lambda_h)^{-1} C[|\lambda_0 - \lambda_h| + \|e - z^h\|_V] \stackrel{(11.2.11a)}{\leq} C'[|\lambda_0 - \lambda_h| + d(e, V_h)].$$

Mit $\|e^h - e\|_V \leq \|e - z^h\|_V + \|z^h - e^h\|_V$ und (11.2.11a) erhält man die Behauptung. ■

Satz 11.2.19 Es gelte (11.2.2a-c). λ_0 sei Eigenwert mit den Eigenschaften (11.2.8) und $\dim E(\lambda_0) = 1$. Seien $e \in E(\lambda_0)$, $e^* \in E^*(\lambda_0)$ mit $\|e\|_V = 1$, $(e, e^*)_0 = 1$. Dann gibt es diskrete Eigenwerte λ_h mit

$$|\lambda_0 - \lambda_h| \leq C d(e, V_h) d(e^*, V_h). \quad (11.2.12)$$

Beweis. u^h sei gemäß Übungsaufgabe 11.2.17 so gewählt, dass $\|e^* - u^h\|_V \leq 2d(e^*, V_h)$, $(e^* - u^h, e)_0 = 0$. Diskrete Eigenwerte $\lambda_h \rightarrow \lambda_0$ existieren nach Satz 11.2.10. Aus

$$\begin{aligned} 0 &= a_{\lambda_0}(e^h, e^*) = a_{\lambda_h}(e^h, e^*) - (\lambda_0 - \lambda_h)(e^h, e^*)_0 \\ &= a_{\lambda_h}(e^h, e^* - u^h) - (\lambda_0 - \lambda_h)(e^h, e^*)_0 \quad (\text{wegen } a_{\lambda_h}(e^h, v^h) = 0 \text{ für alle } v^h \in V_h) \\ &= a_{\lambda_h}(e^h - e, e^* - u^h) - (\lambda_0 - \lambda_h)[(e^h, e^*)_0 - (e, e^* - u^h)_0] \quad \left(\begin{array}{l} \text{für alle } v \in V \text{ gilt } 0 = a_\lambda(e, v) \\ = a_{\lambda_h}(e, v) + (\lambda_h - \lambda_0)(e, v)_0 \end{array} \right) \\ &= a_{\lambda_h}(e^h - e, e^* - u^h) - (\lambda_0 - \lambda_h) \left[\underbrace{(e, e^*)_0}_{=1} + (e^h - e, e^*)_0 - \underbrace{(e, e^* - u^h)_0}_{=0 \text{ nach Voraussetzung}} \right] \end{aligned}$$

folgt $|\lambda_0 - \lambda_h| \leq C\|e^h - e\|_V\|e^* - u^h\|_V + |\lambda_0 - \lambda_h| \|e^h - e\|_V\|e^*\|_V \leq C'\|e^h - e\|_V [\|e^* - u^h\|_V + |\lambda_0 - \lambda_h|]$. Nach Lemma 11.2.18 gibt es ein $e^h \in E_h(\lambda_h)$, sodass

$$|\lambda_0 - \lambda_h| \leq C' C'' [|\lambda_0 - \lambda_h| + d(e, V_h)] [|\lambda_0 - \lambda_h| + 2d(e^*, V_h)].$$

Hieraus erhält man (11.2.12) mit $C = 3C' C''$ für hinreichend kleines h , da $|\lambda_0 - \lambda_h| \rightarrow 0$, $d(e, V_h) \rightarrow 0$, $d(e^*, V_h) \rightarrow 0$. ■

Satz 11.2.20 *Unter den Voraussetzungen des Satzes 11.2.19 gibt es zu $e \in E(\lambda_0)$ und $e^* \in E^*(\lambda_0)$ diskrete Eigenfunktionen $e^h \in E_h(\lambda_h)$, $e^{*h} \in E^*(\lambda_h)$ mit*

$$\|e^h - e\|_V \leq Cd(e, V_h), \quad \|e^{*h} - e^*\|_V \leq Cd(e^*, V_h) \quad (11.2.13)$$

Beweis. Man setze (11.2.12) mit $d(e^*, V_h) \leq \text{const}$ in Lemma 11.2.18 ein. Die zweite Abschätzung in (11.2.13) ergibt sich analog. ■

Sei im Weiteren $V \subset H^1(\Omega)$. Satz 11.1.5 beweist unter den dortigen Voraussetzungen

$$E(\lambda_0) \subset H^{1+s}(\Omega), \quad E^*(\lambda_0) \subset H^{1+s}(\Omega). \quad (11.2.14a)$$

Ferner gelte (11.2.14b) (vgl. (8.4.12)):

$$d(u, V_h) \leq Ch^s \|u\|_{H^{1+s}(\Omega)} \quad \text{für alle } u \in E(\lambda_0) \cup E^*(\lambda_0). \quad (11.2.14b)$$

Folgerung 11.2.21 *Es gelte (11.2.2a), (11.2.2c) und (11.2.14a,b). λ_0 sei Eigenwert mit den Eigenschaften (11.2.8) und $\dim E(\lambda_0) = 1$. Dann gibt es λ_h , $e^h \in E_h(\lambda_h)$ und $e^{*h} \in E^*(\lambda_h)$, sodass*

$$|\lambda_0 - \lambda_h| \leq Ch^{2s}, \quad \|e^h - e\|_V \leq Ch^s, \quad \|e^{*h} - e^*\|_V \leq Ch^s. \quad (11.2.15)$$

Gelegentlich können Eigenfunktionen bessere Regularität besitzen, als für übliche Randwertaufgaben bewiesen werden kann. Sei zum Beispiel $-\Delta e = \lambda e$ im Quadrat $\Omega = (0, 1) \times (0, 1)$ und $e = 0$ auf Γ . Zunächst ergibt Satz 11.1.5 $e \in H^2(\Omega) \cap H_0^1(\Omega)$, also $e \in C^0(\overline{\Omega})$ (vgl. Satz 6.2.31). Damit gilt $e = 0$ auch in den Ecken von Ω . Nach Beispiel 9.1.29 folgt $e \in H^s(\Omega)$ für $s < 4$.

Wie in Abschnitt 8.4.4 erhält man bessere Fehlerabschätzungen für $e - e^h$ in der L^2 -Norm. Der Beweis des folgenden Satzes wird anschließend an Korollar 11.2.29 nachgeholt werden.

Satz 11.2.22 *Es gelte (11.2.2a-c), (11.2.8), $\dim E(\lambda_0) = 1$ und (11.2.14a,b) mit $s = 1$. $a(\cdot, \cdot)$ und $a^*(\cdot, \cdot)$ seien H^2 -regulär, d.h. für $f \in L^2(\Omega)$ haben $a_\mu(u, v) = (f, v)_0$ und $a_\mu(v, u^*) = (v, f)_0$ ($v \in V_h$, μ aus Lemma 11.2.6) Lösungen $u, u^* \in H^2(\Omega)$. Seien $e \in E(\lambda_0)$ und $e^* \in E^*(\lambda_0)$. Dann gibt es $\lambda_h, e^h \in E_h(\lambda_h)$, $e^{*h} \in E^*(\lambda_h)$ mit*

$$\|e^h - e\|_{L^2(\Omega)} \leq Ch^2, \quad \|e^{*h} - e^*\|_{L^2(\Omega)} \leq C'h^2.$$

Gilt (11.2.14a,b) außerdem mit einem $s > 1$, darf man $C'h^2$ durch $C'h^{1+s}$ ersetzen.

11.2.4 Komplementäre Probleme

In Aufgabe (11.2.9) trat bereits eine singuläre Gleichung auf, die trotzdem lösbar war. Im Folgenden sei λ_0 einziger Eigenwert im Kreis $\overline{K_r(\lambda_0)}$ (gilt stets für hinreichend kleines r). Die Gleichung

$$a_\lambda(u, v) = (f, v)_0 \quad \text{für alle } v \in V \quad (11.2.16a)$$

ist für $\lambda = \lambda_0$ singulär. Für $\lambda \approx \lambda_0$ ist Gleichung (11.2.16a) schlecht konditioniert. Es soll im Weiteren gezeigt werden, dass Gleichung (11.2.16a) wohldefiniert und gut konditioniert ist, falls die rechte Seite f im orthogonalen Komplement von $E^*(\lambda_0)$ liegt:

$$f \perp E^*(\lambda_0) \quad (\text{d.h. } (f, e^*)_0 = 0 \text{ für alle } e^* \in E^*(\lambda_0)). \quad (11.2.16b)$$

Im Falle von $\lambda = \lambda_0$ ist mit u auch $u + e$ ($e \in E(\lambda_0)$) Lösung. Die Eindeutigkeit der Lösung erhält man unter den Bedingungen (11.2.8) und (11.2.16c):

$$u \perp E^*(\lambda_0). \quad (11.2.16c)$$

Bemerkung 11.2.23 *Es gelte (11.2.2a,b) und (11.2.8). λ_0 sei einziger Eigenwert in $\overline{K_r(\lambda_0)}$. Dann hat (11.2.16a,b) für alle $|\lambda - \lambda_0| \leq r$ genau eine Lösung u , die (11.2.16c) erfüllt. Es gibt ein von f und λ unabhängiges C , sodass $\|u\|_V \leq C \|f\|_V$.*

Beweis. Folgt nach Lemma 11.2.14, in dem die Voraussetzung $E(\lambda_0) = 1$ nicht benötigt wird. ■

Die Finite-Element-Diskretisierung der Gleichung (11.2.16a) lautet:

$$\text{suche } u^h \in V_h \text{ mit } a_\lambda(u^h, v) = (f, v)_0 \quad \text{für alle } v \in V_h. \quad (11.2.17)$$

Gleichung (11.2.17) kann auch unter der Voraussetzung (11.2.16b) nicht wohldefiniert zu sein.

Der Einfachheit halber beschränken wir uns im Folgenden auf einfache Eigenwerte: $\dim E(\lambda_0) = 1$. Gleichung (11.2.17) ersetzen wir durch (11.2.18a):

$$\text{suche } u^h \in V_h \text{ mit } a_\lambda(u^h, v) = (f^{(h)}, v)_0 \quad \text{für alle } v \in V_h \quad (11.2.18a)$$

$$\text{wobei } f^{(h)} \perp E_h^*(\lambda_h), \quad (11.2.18b)$$

$$u^h \perp E_h^*(\lambda_h). \quad (11.2.18c)$$

Zur praktischen Behandlung von Nebenbedingungen sei auf Abschnitt 8.3.6 verwiesen.

Übungsaufgabe 11.2.24 Man zeige: Sei $\hat{V}_h = V_h \cap E_h^*(\lambda_n)^\perp$ wie in Lemma 11.2.16. Aufgabe (11.2.18a-c) ist äquivalent zu: Suche $u^h \in \hat{V}_h$ mit $a_\lambda(u^h, v) = (f^{(h)}, v)_0$ für alle $v \in \hat{V}_h$.

Lemma 11.2.16 beweist die

Bemerkung 11.2.25 Es gelte (11.2.2a-c), (11.2.8) und $\dim E(\lambda_0) = 1$. λ_0 sei einziger Eigenwert in $\overline{K_r(\lambda_0)}$. Dann gibt es ein $h_0 > 0$, sodass die Aufgabe (11.2.18a,b) für alle $h \leq h_0$ und alle $\lambda \in \overline{K_r(\lambda_0)}$ eine eindeutige Lösung $u^h = u^h(\lambda)$ besitzt, die die Zusatzbedingung (11.2.18c) erfüllt. Ferner gibt es ein von h , λ und $f^{(h)}$ unabhängiges C , sodass $\|u^h\|_V \leq C\|f^{(h)}\|_V$. Sobald $E_h^*(\lambda_h) \neq E^*(\lambda_0)$, braucht f aus (11.2.16b) die Bedingung (11.2.18b) nicht zu erfüllen. Ist $0 \neq e^{*h} \in E_h^*(\lambda_h)$ bekannt, so lässt sich

$$f^{(h)} := Q_{*h}f := f - \frac{(f, e^{*h})_0}{(e^{*h}, e^{*h})_0} e^{*h} \quad (11.2.19)$$

definieren. $f^{(h)}$ erfüllt (11.2.18b), da Q_{*h} die orthogonale Projektion auf $E_h^*(\lambda_h)^\perp$ darstellt.

Übungsaufgabe 11.2.26 Sei $u \perp E(\lambda_0)$, $\dim E(\lambda_0) = \dim E_h(\lambda_h) = 1$, $\|e_h^*\|_V = 1$ und $(e^h, e^{*h})_0 = 1$. Man zeige:

$$d(u, V_h \cap E_h^*(\lambda_n)^\perp) = \inf_{v^h \perp E_h^*(\lambda_n)} \|u - v^h\|_V \leq C \left[d(u, V_h) + \|u\|_V \inf_{e^* \in E^*(\lambda_0)} \|e^{*h} - e^*\|_V \right].$$

Satz 11.2.27 Es gelte (11.2.2a-c), (11.2.8) und $\dim E(\lambda_0) = \dim E_h(\lambda_h) = 1$. λ_0 sei einziger Eigenwert in $\overline{K_r(\lambda_0)}$. h sei hinreichend klein, sodass (gemäß Bemerkung 11.2.25) die Aufgabe (11.2.18a-c) lösbar ist. Für die Lösungen u und u^h von (11.2.16a-c) und (11.2.18a-c) gilt die Fehlerabschätzung

$$\|u^h - u\|_V \leq C \left[d(u, V_h) + \|f^{(h)} - f\|_{V'} + \|f\|_{V'} \inf_{e^* \in E^*(\lambda_0)} \|e^{*h} - e^*\|_V \right] \quad (11.2.20)$$

mit C unabhängig von f , $f^{(h)}$ und h .

Beweis. Man wiederhole den Beweis des Satzes 8.2.1 für $a_\lambda(\cdot, \cdot)$ statt $a(\cdot, \cdot)$. Dabei muss $w \in V_h$ mit $w \perp E_h^*(\lambda_h)$ gewählt werden. Außerdem wird (8.2.3) zu

$$a_\lambda(u^h - u, v) = (f^{(h)} - f, v)_0 \quad \text{für alle } v \in V_h.$$

ε_N stimmt mit $\hat{\omega}_h(\lambda) \geq \eta > 0$ ($\lambda \in \overline{K_r(\lambda_0)}$) überein (vgl. Lemma 11.2.16). $\|u - w\|_V$ wird mit Hilfe von Übungsaufgabe 11.2.26 abgeschätzt, wobei $\|u\|_V \leq C\|f\|_{V'}$ hinzukommt. ■

Zusatz 11.2.28 Ist $f^{(h)}$ durch (11.2.19) definiert, so wird Ungleichung (11.2.20) zu

$$\|u^h - u\|_V \leq C' \left[d(u, V_h) + \|f\|_{V'} \inf_{e^* \in E^*(\lambda_0)} \|e^{*h} - e^*\|_V \right]. \quad (11.2.21a)$$

Beweis. $\|f^{(h)} - f\|_{V'} \leq C |(f, e^{*h})_0| = C |(f, e^* - e^{*h})_0| \leq C \|f\|_{V'} \|e^* - e^{*h}\|_{V'}$. ■

Zusatz 11.2.29 *Gelten zusätzlich die Voraussetzungen $u \in H^{1+s}(\Omega)$, (11.2.14a) und $d(u, V_h) \leq Ch^s |u|_{1+s}$ für $u \in H^{1+s}(\Omega)$, so ergibt (11.2.21a) die Abschätzung*

$$\|u^h - u\|_V \leq Ch^s |u|_{H^{1+s}(\Omega)}. \quad (11.2.21b)$$

Nachzutragen bleibt der

Beweis des Satzes 11.2.22. Zu $e \in E(\lambda_0)$ gibt es $e^h \in E_h(\lambda_h)$ mit $f := e - e^h \perp E(\lambda_0)$ und $|f|_1 = \|f\|_V \leq Ch^s = Ch$. Nach Bemerkung 11.2.25 hat die Aufgabe $a_{\lambda_0}(v, w) = (v, f)_0$ für alle $v \in V$ eine Lösung $w \perp E^*(\lambda_0)$. Die Regularitätsannahme liefert $w \in H^2(\Omega)$ und $|w|_2 \leq C |f|_0$, sodass $w^h \in V_h$ mit $w^h \perp E_h^*(\lambda_h)$, $|w - w^h|_1 \leq Ch |w|_2 \leq C'h |f|_0$ existiert. Die Größe

$$a_{\lambda_0}(f, w^h) = a_{\lambda_0}(e, w^h) - a_{\lambda_0}(e^h, w^h) = 0 - a_{\lambda_0}(e^h, w^h) = (\lambda_0 - \lambda_h) (e^h, w^h)_0 - a_{\lambda_h}(e^h, w^h) = (\lambda_0 - \lambda_h) (e^h, w^h)_0$$

lässt sich durch $Ch^2 |w^h|_0 |e^h|_0 \leq C'h^2 |f|_0$ abschätzen (vgl. (11.2.15)). Aus

$$|f|_0^2 = (f, f)_0 = a_{\lambda_0}(f, w) = a_{\lambda_0}(f, w - w^h) + (\lambda_0 - \lambda_h) (e^h, w^h)_0 \leq C [C'h |f|_1 |f|_0 + h^2 |f|_0^2]$$

und $|f|_1 \leq Ch$ schließt man $|f|_0 < C'h^2$. Ebenso verfährt man für $|e^* - e^{*h}|_0$. ■

Übungsaufgabe 11.2.30 *Man formuliere die Voraussetzungen und den Beweis für die Fehlerabschätzung $\|u - u^h\|_{L^2(\Omega)} \leq Ch^2 |u|_2$ (u, u^h aus Zusatz 11.2.28). Hinweis: Man zerlege $u - u^h$ in $f_1 + f_2$ mit $f_1 \perp E^*(\lambda_0)$ und $f_2 \in E(\lambda_0)$.*

11.3 Diskretisierung durch Differenzenverfahren

Im Folgenden beschränken wir uns auf den Fall eines Differenzenoperators der Ordnung $2m = 2$. Die Differentialgleichung $Lu = f$ mit homogener Dirichlet-Randbedingung wird wie in den Kapiteln 4 und 5 durch die Differenzengleichung $L_h u_h = f_h$ ersetzt. Die Eigenwertgleichungen $Le = \lambda e$, $L^* e^* = \lambda e^*$ werden durch

$$L_h e_h = \lambda_h e_h, \quad L_h^* e_h^* = \bar{\lambda}_h e_h^* \quad (11.3.1)$$

diskretisiert. L_h^* ist die transponierte und komplex konjugierte Matrix zu L_h .

Die generellen Voraussetzungen der folgenden Analyse sind:

$$V = H_0^1(\Omega), \quad \Omega \in C^{0,1} \text{ beschränkt}, \quad (11.3.2a)$$

$$a(u, v) = (Lu, v)_0 \quad \text{ist } H_0^1(\Omega)\text{-koerziv}, \quad (11.3.2b)$$

$$|L_h R_h - \tilde{R}_h L|_{-1 \leftarrow -2} \leq Ch \quad (\text{Konsistenzbedingung}). \quad (11.3.2c)$$

Bedingung (11.3.2c) ist in Abschnitt 9.2.2 diskutiert worden. Weiterhin sei angenommen, dass L_h H_h^1 -koerziv ist. Für geeignetes $\mu \in \mathbb{R}$ ist daher

$$L_{\mu, h} := L_h - \mu I \quad (I: \text{Einheitsmatrix})$$

H_h^1 -regulär:

$$(L_{\mu, h} v_h, v_h)_0 \geq C_E |v_h|_1^2 \quad \text{für alle } v_h. \quad (11.3.2d)$$

Ferner sei

$$L_\mu := L - \mu I \quad H^2(\Omega)\text{-regulär}, \quad (11.3.2e)$$

d.h. $|L_\mu^{-1}|_{2 \leftarrow 0} \leq C$. Die Beschränktheit von L und L_h lauten

$$|L|_{-1 \leftarrow -1} \leq C, \quad |L_h|_{-1 \leftarrow -1} \leq C. \quad (11.3.2f)$$

In (9.2.26a,b) wurden Prolongationen $\hat{P}_h : L_h^2 \rightarrow L^2(\mathbb{R}^2)$ und $P_h : L_h^2 \rightarrow L^2(\Omega)$ eingeführt. Jetzt wird eine Abbildung $P_h : H_h^1 \rightarrow H_0^1(\Omega)$ benötigt:

$$\bar{u}(\mathbf{x}) := \begin{cases} \hat{P}_h u_h(\mathbf{x}) & \text{falls } K_{h/2}(\mathbf{x}) \subset \Omega, \\ 0 & \text{sonst,} \end{cases}$$

$$P_h u_h(\mathbf{x}) := (\sigma_h^x \sigma_h^y \bar{u})(\mathbf{x}) \quad (\mathbf{x} \in \Omega),$$

wobei $K_{h/2}(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^2 : \|\mathbf{x} - \mathbf{y}\|_\infty < h/2\}$, \hat{P}_h gemäß (9.2.26b) und $\sigma_h^x \sigma_h^y$ gemäß (9.2.11) definiert sind. Man prüft nach, dass $P_h u_h \in H_0^1(\Omega)$ und

$$|P_h|_{1 \leftarrow 1} \leq C. \quad (11.3.2g)$$

R_h und \tilde{R}_h seien wie in (9.2.12a,b) definiert. Sie erfüllen

$$|R_h|_{0 \leftarrow 0} \leq C, \quad |R_h|_{1 \leftarrow 1} \leq C, \quad |\tilde{R}_h|_{0 \leftarrow 0} \leq C. \quad (11.3.2h)$$

Übungsaufgabe 11.3.1 Man zeige:

$$|R_h - \tilde{R}_h|_{0 \leftarrow 1} \leq Ch, \quad |I - R_h P_h|_{0 \leftarrow 1} \leq Ch, \quad |P_h^* - R_h|_{0 \leftarrow 1} \leq Ch, \quad (11.3.2i)$$

$$|I - P_h^* P_h|_{0 \leftarrow 1} \leq Ch, \quad |I - \tilde{R}_h P_h|_{0 \leftarrow 1} \leq Ch, \quad |I - \tilde{R}_h^* R_h|_{0 \leftarrow 1} \leq Ch \quad (11.3.2j)$$

Die erste Ungleichung in (11.3.2j) ist äquivalent zu

$$|(P_h u_h, P_h v_h)_0 - (u_h, v_h)_0| \leq Ch |u_h|_0 |v_h|_1. \quad (11.3.2j^*)$$

Hinweis: Übungsaufgabe 9.2.12 und Lemma 9.2.15.

Lemma 11.3.2 Es gelte (11.3.2a,c,g,h,i). a) Es gilt

$$\lim_{h \rightarrow 0} |u - P_h R_h u|_1 = \lim_{h \rightarrow 0} |u - \tilde{R}_h^* R_h u|_1 = 0 \quad \text{für alle } u \in H_0^1(\Omega), \quad (11.3.2k)$$

$$\lim_{h \rightarrow 0} |(\tilde{R}_h - R_h)u|_0 = 0 \quad \text{für alle } u \in L^2(\Omega), \quad (11.3.2l)$$

$$\lim_{h \rightarrow 0} |[L_{h,\lambda} R_h - \tilde{R}_h L_\lambda]u|_{-1} = 0 \quad \text{für alle } u \in H_0^1(\Omega), \lambda \in \mathbb{C}. \quad (11.3.2m)$$

b) Ist $u \in H_0^1(\Omega)$ und $\lim_{h \rightarrow 0} |R_h u|_0 = 0$, so gilt $u = 0$.

Beweis. a) Der Nachweis von (2k,l,m) folgt demselben Muster, das anhand von (2l) vorgeführt wird. Zu $\varepsilon > 0$ ist $|(\tilde{R}_h - R_h)u| \leq \varepsilon$ für $h \leq h(\varepsilon)$ zu zeigen. Da $H_0^1(\Omega)$ dicht in $L^2(\Omega)$ liegt, gibt es $\tilde{u} \in H_0^1(\Omega)$ mit $|u - \tilde{u}|_0 \leq \varepsilon/[2|\tilde{R}_h - R_h|_{0 \leftarrow 0}]$, sodass $|(\tilde{R}_h - R_h)(u - \tilde{u})|_0 \leq \varepsilon/2$. Wegen (11.3.2i) folgt die Abschätzung $|(\tilde{R}_h - R_h)\tilde{u}|_0 \leq Ch|\tilde{u}|_1 \leq \varepsilon/2$ für $h \leq h(\varepsilon) := \varepsilon/[2C|\tilde{u}|_1]$. Zusammen ergibt sich $|(\tilde{R}_h - R_h)u|_0 \leq \varepsilon$.

b) Aus (11.3.2k) schließt man $0 = \lim_{h \rightarrow 0} (R_h u, P_h^* u)_0 = \lim_{h \rightarrow 0} (P_h R_h u, u)_0 = (u, u)_0$, also $u = 0$. ■

Übungsaufgabe 11.3.3 Es sei $\Lambda_h := I - \partial_x^+ \partial_x^- - \partial_y^+ \partial_y^-$ und $\Lambda := I - \Delta$. Man zeige:

$$\begin{aligned} |u|_1 &= (\Lambda u, u)_0, & |u_h|_1 &= (\Lambda_h u_h, u_h)_0, \\ |L_\lambda u|_{-1}^2 &= (v, L_\lambda u)_0 & \text{für } u \in H_0^1(\Omega) \text{ und } v &= \Lambda^{-1} L_\lambda u \in H_0^1(\Omega), \\ |L_{\lambda,h} u_h|_{-1}^2 &= (v_h, L_{\lambda,h} u_h)_0 & \text{für } u_h \in H_h^1 \text{ und } v_h &= \Lambda_h^{-1} L_{\lambda,h} u_h \in H_h^1, \\ \lim_{h \rightarrow 0} |(\Lambda_h R_h - \tilde{R}_h \Lambda)u|_{-1} &= 0 & \text{für alle } u \in H_0^1(\Omega). \end{aligned}$$

Die folgende Analyse ist auf die Eigenschaften der vorliegenden Differenzenverfahren zugeschnitten. Eine abstraktere Theorie, die sowohl für finite Elemente wie auch Differenzenverfahren anwendbar ist, sollte hier vermieden werden. Man findet derartige Zugänge bei Stummel [79] und Chatelin [21].

Die Größe $\omega_h(\lambda)$ wird jetzt durch

$$\omega_h(\lambda) := \inf_{|u_h|_1=1} \sup_{|v_h|_1=1} |(L_{\lambda,h} u_h, v_h)_0| = \inf_{|u_h|_1=1} |L_{\lambda,h} u_h|_{-1} \quad (11.3.3)$$

definiert. Wie in Übungsaufgabe 11.2.4 gilt

$$\omega_h(\lambda) = \begin{cases} 0 & \text{falls } \lambda \text{ Eigenwert von } L_h, \\ 1/|L_{\lambda,h}|_{1 \leftarrow -1} & \text{sonst.} \end{cases} \quad (11.3.3')$$

Das Analogon zu Lemma 11.2.7 lautet

Lemma 11.3.4 $K \subset \mathbb{C}$ sei kompakt. Es gelte (11.3.2a,b,c,d,f). Dann gibt es von $\lambda \in K$ unabhängige Größen $C > 0$ und $\eta(h) \rightarrow 0$ ($h \rightarrow 0$), sodass

$$\omega_h(\lambda) \geq C\omega(\lambda) - \eta(h), \quad \omega(\lambda) \geq C\omega_h(\lambda) - \eta(h) \quad \text{für alle } \lambda \in K, h > 0. \quad (11.3.4)$$

Beweis. 1) Da K kompakt, $\omega(\lambda)$ stetig und $\omega_h(\lambda)$ gleichgradig stetig in λ sind, genügt es, die Ungleichungen $\lim_{h \rightarrow 0} \omega_h(\lambda) \geq C\omega(\lambda)$, $\omega(\lambda) \geq C \overline{\lim}_{h \rightarrow 0} \omega_h(\lambda)$ für alle $\lambda \in K$ mit $C > 0$ zu zeigen.

2) Zu $\lambda \in K$ und u_h mit $|u_h|_1 = 1$ und $|L_{\lambda,h}u_h|_{-1} = \omega_h(\lambda)$ definiere man

$$u := P_h u_h, \quad z_h := (\lambda - \mu) L_{\mu,h}^{-1} u_h, \quad z := (\lambda - \mu) L_{\mu}^{-1} u_h \quad \text{mit } \mu \text{ aus (11.3.2e).}$$

Ohne Beschränkung der Allgemeinheit kann $\mu \notin K$ vorausgesetzt werden. Es ist

$$\begin{aligned} |u - z|_1 &= |P_h(u_h - z_h) + P_h z_h - z|_1 \leq |P_h|_{1 \leftarrow 1} |u_h - z_h|_1 + |P_h z_h - z|_1, \\ |P_h z_h - z|_1 &= |P_h[z_h - R_h z] - [I - P_h R_h]z|_1 \\ &\leq |P_h|_{1 \leftarrow 1} |\lambda - \mu| \left| L_{\mu,h}^{-1} [\tilde{R}_h L_{\mu} - L_{\mu,h} R_h] L_{\mu}^{-1} u + L_{\mu,h}^{-1} [I - \tilde{R}_h P_h] u_h \right|_1 \\ &\quad + |[I - P_h R_h]z|_1 \rightarrow 0 \quad \text{für } h \rightarrow 0 \end{aligned}$$

(vgl. (11.3.2g,j,m)), sodass

$$|u_h - z_h|_1 \geq C_1 |u - z|_1 - o(1), \quad C_1 > 0. \quad (11.3.5a)$$

Wie in (11.2.6d) erhält man

$$|u - z|_1 \geq C_2 \omega(\lambda) |u|_1, \quad C_2 > 0. \quad (11.3.5b)$$

Aus

$$\begin{aligned} \omega_h(\lambda) &= |L_{\lambda,h}u_h|_{-1} = |L_{\mu,h}u_h + (\mu - \lambda)u_h|_{-1} \geq (L_{\mu,h}u_h + (\mu - \lambda)u_h, u_h)_0 \\ &\geq -|\mu - \lambda| |u_h|_0^2 + (L_{\mu,h}u_h, u_h)_0 \geq -C_{\mu} |u_h|_0^2 + C_E |u_h|_1^2 = -C_{\mu} |u_h|_0^2 + C_E \end{aligned}$$

mit $C_{\mu} := \max\{|\mu - \lambda| : \lambda \in K\} > 0$ folgt

$$|u_h|_0^2 \geq C_{\mu}^{-1} [C_E - \omega_h(\lambda)].$$

Entweder ist $\omega_h(\lambda) \geq C_E/2$, woraus sich die Behauptung direkt ergibt, oder $\omega_h(\lambda) \leq C_E/2$ liefert

$$|u_h|_0 \geq C_0 = C_0 |u_h|_1 \quad \text{mit } C_0 := \sqrt{C_E/(2C_{\mu})}. \quad (11.3.5c)$$

Wir wollen zeigen, dass es $h_0 > 0$ und $C_P = C_P(C_0)$ gibt, sodass

$$|u_h|_1 \leq C_P |P_h u_h|_1 \quad \text{für alle } u_h \text{ mit } |u_h|_0 \geq C_0 |u_h|_1 \text{ und } h \leq h_0. \quad (11.3.5d)$$

Die Negation von (11.3.5d) lautet: Es gibt u_h mit $|u_h|_1 = 1$, $|u_h|_0 \geq C_0$ und $|P_h u_h|_1 \rightarrow 0$ ($h \rightarrow 0$). Aus $|R_h P_h u_h|_0 \leq |R_h P_h u_h|_1 \leq C |P_h u_h|_1 \rightarrow 0$ und $|u_h - R_h P_h u_h|_0 \leq |I - R_h P_h|_{0 \leftarrow 1} |u_h|_1 \leq Ch \rightarrow 0$ folgt $|u_h|_0 \rightarrow 0$ im Widerspruch zu $|u_h|_0 \geq C_0$. Also gilt (11.3.5d).

Zusammen mit $L_{\mu,h}(u_h - z_h) = L_{\lambda,h}u_h$ ergeben (11.3.2d) und (11.3.5a,b,d) die erste Ungleichung in (11.3.4):

$$\begin{aligned} \omega_h(\lambda) &= |L_{\lambda,h}u_h|_{-1} = |L_{\mu,h}(u_h - z_h)|_{-1} \geq C_E |u_h - z_h|_1 \geq C_E C_1 |u - z|_1 - o(1) \\ &\geq C_E C_1 C_2 \omega(\lambda) |u|_1 - o(1) \geq C \omega(\lambda) - o(1) \quad \text{mit } C := C_E C_1 C_2 / C_P > 0. \end{aligned}$$

3) Sei $\varepsilon > 0$ beliebig. $u \in H_0^1(\Omega)$ mit $|u_h|_1 = 1$ kann so gewählt werden, dass $\omega(\lambda) \geq |L_{\lambda}u|_{-1} - \varepsilon$. Man setze $u_h := R_h u$. Gemäß Übungsaufgabe 11.3.3 gilt

$$\begin{aligned} |L_{\lambda}u|_{-1}^2 &= (v, L_{\lambda}u)_0 \quad \text{für } v := \Lambda^{-1} L_{\lambda}u \in H_0^1(\Omega), \quad \text{wobei } \Lambda = I - \Delta, \\ |L_{\lambda,h}u_h|_{-1}^2 &= (v_h, L_{\lambda,h}u_h)_0 \quad \text{für } v_h := \Lambda_h^{-1} L_{\lambda,h}u_h. \end{aligned}$$

Aus

$$\begin{aligned}
R_h v - v_h &= \Lambda_h^{-1} [\Lambda_h R_h - \tilde{R}_h \Lambda] \Lambda^{-1} L_\lambda u + \Lambda_h^{-1} [\tilde{R}_h L_\lambda - L_{\lambda,h} R_h] u \rightarrow 0 & \text{für } h \rightarrow 0, \\
\tilde{R}_h L_\lambda u - L_{\lambda,h} u_h &= [\tilde{R}_h L_\lambda - L_{\lambda,h} R_h] u \rightarrow 0 & \text{für } h \rightarrow 0, \\
(v, L_\lambda u)_0 - (R_h v, \tilde{R}_h L_\lambda u)_0 &= \left([I - \tilde{R}_h^* R_h] v, L_\lambda u \right)_0 \rightarrow 0 & \text{für } h \rightarrow 0
\end{aligned}$$

(vgl. (11.3.2m), (11.3.2k)) schließt man $|L_{\lambda,h} u_h|_{-1} \rightarrow |L_\lambda u|_{-1}$ und

$$\omega(\lambda) \geq |L_\lambda u|_{-1} - \varepsilon \geq |L_{\lambda,h} u_h|_{-1} - \varepsilon - o(1) \geq \omega_h(\lambda) - \varepsilon - o(1) \quad (h \rightarrow 0)$$

für jedes $\varepsilon > 0$. Somit ist $\omega(\lambda) \geq \overline{\lim}_{h \rightarrow 0} \omega_h(\lambda)$ bewiesen. ■

Folgerung 11.3.5 *Unter den Voraussetzungen des Lemmas 11.3.4 gilt: Existiert L_λ^{-1} für alle $\lambda \in K$, so gibt es ein $h_0 > 0$ derart, dass $L_{\lambda,h}$ für alle $\lambda \in K$ und $h \leq h_0$ H_0^1 -regulär ist:*

$$\sup \left\{ |L_{\lambda,h}^{-1}|_{1 \leftarrow -1} : \lambda \in K, h \leq h_0 \right\} \leq C.$$

Beweis. Nach Voraussetzung ist $\omega(\lambda) > 0$ in K , also $\max\{\omega(\lambda) : \lambda \in K\} =: \eta > 0$. Man wähle h_0 gemäß Lemma 11.3.4 so, dass $\omega_h(\lambda) \geq C\omega(\lambda) - \frac{1}{2}C\eta \geq \frac{1}{2}C\eta$ für $h \leq h_0$. Dann ist $|L_{\lambda,h}^{-1}|_{1 \leftarrow -1} \leq 2/(C\eta)$ für alle $\lambda \in K, h \leq h_0$. ■

Der Beweis des Satzes 11.2.8 lässt sich ohne Änderung übertragen und ergibt den

Satz 11.3.6 *Es gelte (11.3.2a,b,c,d,f). Sind λ_{h_i} ($h_i \rightarrow 0$) diskrete Eigenwerte der Aufgabe (11.3.1) mit $\lambda_{h_i} \rightarrow \lambda_0$, so ist λ_0 ein Eigenwert von (11.1.2a).*

Auch Lemma 11.2.9 und Satz 11.2.10 lassen sich ohne Änderung übertragen:

Satz 11.3.7 *Es gelte (11.3.2a,b,c,d,f). λ_0 sei Eigenwert von (11.1.2a). Dann existieren diskrete Eigenwerte λ_h von (11.3.1) [für alle h], sodass $\lim_{h \rightarrow 0} \lambda_h = \lambda_0$.*

Satz 11.3.8 *Es gelte (11.3.2a,b,c,d,f). e_h seien diskrete Eigenfunktionen mit $|e_h|_1 = 1$ zu λ_h , wobei $\lambda_h \rightarrow \lambda_0$ ($h \rightarrow 0$). Dann existiert eine Teilfolge e_{h_i} , sodass $P_{h_i} e_{h_i}$ in $H_0^1(\Omega)$ gegen eine Eigenfunktion $0 \neq e \in E(\lambda_0)$ konvergiert. Ferner gilt $|e_{h_i} - R_{h_i} e|_1 \rightarrow 0$.*

Beweis. Wegen $|P_h e_h|_1 \leq C$ (vgl. (11.3.2g)) sind die Funktionen $e^h := P_h e_h$ gleichmäßig beschränkt. $H_0^1(\Omega)$ ist kompakt in $L^2(\Omega)$ eingebettet (vgl. (11.3.2a) und Satz 6.4.8a), sodass eine Teilfolge e^{h_i} in $L^2(\Omega)$ gegen ein $e \in L^2(\Omega)$ konvergiert: $|e^{h_i} - e|_0 \rightarrow 0$. Insbesondere gilt auch

$$|R_h e - e_h|_0 \leq |R_h(e - e_h) - (R_h P_h - I)e_h|_0 \rightarrow 0 \quad (h = h_i \rightarrow 0).$$

Abschätzung (11.3.2c) liefert

$$|R_h z - z_h|_0 \leq |R_h z - z_h|_1 \leq Ch |e|_0 \rightarrow 0 \quad \text{für } z := (\lambda_0 - \mu) L_\lambda^{-1} e, \quad z_h := (\lambda_0 - \mu) L_{\lambda,h}^{-1} \tilde{R}_h e.$$

Aus $L_{\mu,h}(z_h - e_h) = (\lambda_0 - \mu)(\tilde{R}_h e - e_h) + (\lambda_h - \lambda)e_h \rightarrow 0$ in H_h^{-1} folgt $|z_h - e_h|_1 \rightarrow 0$, sodass $|R_h(e - z)|_0 \rightarrow 0$ ($h = h_i \rightarrow 0$) resultiert. Lemma 11.3.2b zeigt $e = z \in H_0^1(\Omega)$, und $e = 0$ ist wegen $|\lim R_{h_i} e|_1 = |\lim e_{h_i}|_1 = 1$ ausgeschlossen. ■

Satz 11.3.9 *Es gelte (11.3.2a,h,c,d,f). e_h^* sei Lösung der diskreten Eigenwertaufgabe $L_h^* e_h^* = \bar{\lambda}_h e_h^*$ mit $|e_h^*|_1 = 1$, $\lim_{h \rightarrow 0} \lambda_h = \lambda_0$. Dann existiert eine Teilfolge $e_{h_i}^*$, sodass $P_{h_i} e_{h_i}^*$ in $H_0^1(\Omega)$ gegen eine Eigenfunktion $0 \neq e^* \in E^*(\lambda_0)$ konvergiert. Ferner gilt $|e_{h_i}^* - R_{h_i} e^*|_1 \rightarrow 0$.*

Beweisskizze. Der Beweis ist nicht analog zu dem des Satzes 11.3.8, da die Konsistenzbedingung (11.3.2a,m) nicht notwendig die entsprechenden Aussagen für die adjungierten Operatoren impliziert. Man hat folgende Schritte durchzuführen:

- 1) $e^{*h} := P_h e_h^* \rightarrow e^*$ konvergiert in $L^2(\Omega)$ für eine Teilfolge $h = h_i \rightarrow 0$.
- 2) Für $z = (\lambda_0 - \mu) L_\mu^{*-1} e^*$ und $z_h := (\lambda_0 - \mu) L_{\mu,h}^{*-1} e_h^*$ gilt

$$z - \tilde{R}_h^* z_h = (\overline{\lambda_0 - \mu}) L_\mu^{*-1} \left[(e^* - R_h^* e_h^*) + (R_h^* L_{\mu,h}^* - L_\mu^* \tilde{R}_h^*) L_{\mu,h}^{*-1} e_h^* \right].$$

Für jedes $v \in L^2(\Omega)$ ergibt sich

$$\left(v, z - \tilde{R}_h^* z_h \right)_0 = (\lambda_0 - \mu) \left\{ \begin{aligned} & (L_\mu^{-1} v, e_h^* - e_h^*)_0 + ([P_h^* - R_h] L_\mu^{-1} v, e_h^*)_0 \\ & + \left(L_{\mu,h}^{-1} [L_{\mu,h} R_h - \tilde{R}_h L_\mu] L_\mu^{-1} v, e_h^* \right)_0 \end{aligned} \right\} \rightarrow 0$$

für $h = h_i \rightarrow 0$ (vgl. (11.3.2i), (11.3.2c)).

3) $|z_h - e_h^*|_1 \rightarrow 0$ schließt man aus $L_{\mu,h}^*(z_h - e_h^*) = (\overline{\lambda_0 - \lambda_h}) e_h^* \rightarrow 0$. Insbesondere gilt $(v, \tilde{R}_h^* z_h - R_h^* e_h^*)_0 \rightarrow 0$ für jedes $v \in L^2(\Omega)$.

4) $(v, \tilde{R}_h^* e_h^* - R_h^* e_h^*)_0 = ([\tilde{R}_h - R_h] v, e_h^*)_0 \rightarrow 0$ für jedes $v \in L^2(\Omega)$ (vgl. (11.3.2l)).

5) $(v, R_h^* e_h^* - e^*)_0 \rightarrow 0$ für $h = h_i \rightarrow 0$.

6) Aus 2) bis 5) folgt $(v, z - e^*)_0 = 0$, also $z = e^* \in E^*(\lambda_0)$, wobei $e^* \neq 0$. ■

Übungsaufgabe 11.3.10 Man übertrage Übungsaufgabe 11.2.12.

In Lemma 11.2.16 wurde $\hat{\omega}_h(\lambda)$ definiert. Jetzt setze man

$$\begin{aligned} \hat{V}_h &:= \{u_h : (u_h, e_h^*)_0 = 0\} = \{e_h^*\}^\perp, & L_h^* e_h^* &= \bar{\lambda}_h e_h^*, & \lambda_h &\rightarrow \lambda_0, \\ \hat{\omega}_h(\lambda) &:= \inf_{0 \neq u_h \in \hat{V}_h} \sup_{0 \neq v_h \in \hat{V}_h} |(L_{\lambda,h} u_h, v_h)_0| / (|u_h|_1 |v_h|_1). \end{aligned} \quad (11.3.6)$$

Grundvoraussetzung für das Weitere ist

$$\dim E_h(\lambda_h) = 1, \quad E_h(\lambda_h) = \text{span}\{e_h\}, \quad E_h^*(\lambda_h) = \text{span}\{e_h^*\}. \quad (11.3.7)$$

Dabei ist

$$E_h(\lambda_h) := \{u_h \in H_h^1 : L_h u_h = \lambda_h u_h\}, \quad E_h^*(\lambda_h) := \{u_h \in H_h^1 : L_h^* u_h = \bar{\lambda}_h u_h\}$$

Aus Übungsaufgabe 11.3.10 geht hervor, dass (11.3.7) für $h \leq h_0$ zutrifft, wenn $\dim E(\lambda_0) = 1$.

Lemma 11.3.11 Es gelte (11.3.2a,b,c,d,f), $\dim E(\lambda_0) = 1$ und (11.2.8). Dann gibt es $h_0 > 0$ und ein von $h \leq h_0$ und $\lambda \in \mathbb{C}$ unabhängiges $C > 0$, sodass $\hat{\omega}_h(\lambda) \geq C \omega_h(\lambda)$ für $h \leq h_0$. Für hinreichend kleines $\varepsilon > 0$ und h ist $\hat{\omega}_h(\lambda) \geq \eta > 0$ für alle $|\lambda - \lambda_0| \leq \varepsilon$.

Beweis. 1) Es gibt ein $C > 0$, sodass

$$|v_h + \alpha e_h^*|_1 \geq |v_h|_1 / C \quad \text{für alle } v_h \in \hat{V}_h, \alpha \in \mathbb{C}, h > 0.$$

Sei zunächst h fixiert. Man überzeugt sich leicht, dass die Quotientenraumnorm $\|v_h\|_1 := \inf\{|v_h + \alpha e_h^*|_1 : \alpha \in \mathbb{C}\}$ und $|\cdot|_1$ zwei Normen auf \hat{V}_h sind. Wegen der Äquivalenz der Normen auf endlichdimensionalen Vektorräumen, ist die Behauptung mit einer eventuell von h abhängigen Konstanten $C = C(h)$ richtig, sodass nur das Verhalten $h \rightarrow 0$ zu untersuchen ist.

Indirekter Beweis: Es sei angenommen, es gäbe eine Folge v_{h_i} mit $h_i \rightarrow 0$, $|v_{h_i}|_1 = 1$, $\alpha_i \in \mathbb{C}$, $w_{h_i} := v_{h_i} + \alpha_i e_{h_i}^*$, $|w_{h_i}|_1 \rightarrow 0$. Für eine Teilfolge von $\{h_i\}$ konvergieren

$$\alpha_i \rightarrow \alpha^*, \quad P_{h_i} v_{h_i} \rightarrow v^* \quad \text{und} \quad P_{h_i} e_{h_i}^* \rightarrow e^* \neq 0 \quad \text{in } L^2(\Omega), \quad P_{h_i} w_{h_i} \rightarrow w^* := v^* + \alpha^* e^* = 0.$$

Aus

$$0 = (v_{h_i}, e_{h_i}^*)_0 = (v_{h_i}, [I - P_{h_i}^* P_{h_i}] e_{h_i}^*)_0 + (P_{h_i} v_{h_i}, P_{h_i} e_{h_i}^*)_0 \rightarrow (v, e^*)_0$$

schließt man $(v^*, e^*)_0 = 0$, $\alpha^* = (w^*, e^*)_0 / (e^*, e^*)_0 = 0$, $v^* = 0$. Der Widerspruch ergibt sich aus $1 = \lim |v_{h_i}|_1 = \lim |w_{h_i}|_1 = 0$.

2) Der weitere Beweisgang ist wie zu Lemma 11.2.16. ■

Lemma 11.3.12 Es gelte (11.3.2a-m), $\dim E(\lambda_0) = 1$ und (11.2.8). $0 \neq e \in E(\lambda_0)$ und $e_h \in E_h(\lambda_h)$ lassen sich so wählen, dass $|R_h e - e_h|_1 \leq C[h + |\lambda_0 - \lambda_h|]$ für alle $h > 0$.

Beweis. Es ist $e = (\lambda_0 - \mu)L_\mu^{-1}e \in H^2(\Omega) \cap H_0^1(\Omega)$. Sei $z_h := (\lambda_0 - \mu)L_{\mu,h}^{-1}\tilde{R}_he$. Die Abschätzung (9.2.22) liefert $|R_he - z_h|_1 \leq Ch|e|_2$. Für hinreichend kleine h ist $(e_h, e_h^*) \neq 0$, sodass e_h so skaliert werden kann, dass $(e_h - z_h, e_h^*)_0 = 0$. Aus

$$L_{\lambda,h}(e_h - z_h) = (\lambda_h - \lambda_0)\tilde{R}_he + (\lambda_h - \mu)(z_h - \tilde{R}_he) = (\lambda_h - \lambda_0)\tilde{R}_he + (\lambda_h - \mu)\left[(z_h - R_he) + (R_h - \tilde{R}_h)e\right]$$

schließt man $|e_h - z_h|_1 \leq C[|\lambda_h - \lambda_0| + h]$ (vgl. Lemma 11.3.11), da $|(R_h - \tilde{R}_h)e|_{-1} = \mathcal{O}(h)$. Hieraus folgt die Behauptung. ■

Lemma 11.3.13 *Unter den Voraussetzungen von Lemma 11.3.12 gilt $|\lambda_h - \lambda_0| \leq Ch$.*

Beweis. Man wähle e_h, e_h^* so, dass $(R_he - e_h, e_h^*)_0 = (e_h, R_he^* - e_h^*)_0 = 0$. Für den Rayleigh-Quotienten $\tilde{\lambda}_h := (L_h R_he, R_he^*)_0 / (R_he, R_he^*)_0$ gilt dann

$$|\tilde{\lambda}_h - \lambda_h| \leq C|R_he - e_h|_1 |R_he^* - e_h^*|_1 \leq \varepsilon_h [h + |\lambda_h - \lambda_0|] \quad \text{mit} \quad \varepsilon_h := C|R_he^* - e_h^*|_1.$$

Aus (11.3.2c_j) schließt man

$$(L_h R_he, R_he^*)_0 - (Le, e^*)_0 = \left([L_h R_h - \tilde{R}_h L] e, R_he^* \right)_0 + \left(Le, [\tilde{R}_h^* R_h - I] e^* \right)_0 = \mathcal{O}(h)$$

und $|\tilde{\lambda}_h - \lambda_0| = \mathcal{O}(h)$, sodass $|\lambda_h - \lambda_0| \leq Ch + \varepsilon_h |\lambda_h - \lambda_0|$. Für hinreichend kleines h ist $\varepsilon_h \leq \frac{1}{2}$ (vgl. Satz 11.3.9, Übungsaufgabe 11.3.10), also $|\lambda_h - \lambda_0| \leq 2Ch$. ■

Lemmata 11.3.12 und 11.3.13 ergeben

Satz 11.3.14 *Es gelte (11.3.2a-m), $E(\lambda_0) = \text{span}\{e\}$ und (11.2.8). Für alle $h > 0$ existieren $e_h \in E_h(\lambda_h)$ mit $|R_he - e_h|_1 \leq Ch$.*

Da $|R_he^* - e_h^*|_1 = o(1)$ oder sogar $|R_he^* - e_h^*|_1 = \mathcal{O}(h)$, erwartet man nach Satz 11.2.19, dass $|\lambda_0 - \lambda_h| = o(h)$ bzw. $\mathcal{O}(h^2)$. Diese Abschätzung ist im Allgemeinen falsch, wie das folgende Gegenbeispiel zeigt.

Beispiel 11.3.15 *Die Eigenwertaufgabe $-u'' + u' = \lambda u$ in $(0, 1)$ mit $u(0) = u(1) = 0$ hat die Lösung $u(x) = \exp(x/2) \sin(\pi x)$. Ein Eigenwert ist $\lambda_0 = \pi^2 + 1/4$. Man rechnet nach, dass die Diskretisierung $-\partial^- \partial^+ u + \partial^+ u = \lambda u$ den Eigenwert*

$$\begin{aligned} \lambda_h &= h^{-2} \left[2 - \cos(\pi h) \left(e^{\Lambda' h} + e^{-\Lambda' h} \right) - i \sin(\pi h) \left(e^{\Lambda' h} - e^{-\Lambda' h} \right) \right] + h^{-1} \left[\cos(\pi h) e^{\Lambda' h} - 1 + i \sin(\pi h) e^{\Lambda' h} \right] \\ &= \pi^2 + \frac{1}{4} + \frac{1 - 3\pi^2}{8} h + \mathcal{O}(h^2) \quad \text{mit} \quad \Lambda' := \frac{\log(1 - h)}{2h} \end{aligned}$$

besitzt, sodass $|\lambda_0 - \lambda_h|$ nicht besser als $\mathcal{O}(h)$ ausfällt.

12

Stokes-Gleichungen

12.1 Elliptische Differentialgleichungssysteme

In Beispiel 1.1.11 wurden bereits die *Gleichungen von Stokes* für $\Omega \subset \mathbb{R}^2$ angegeben:

$$-\Delta u_1 + \partial p / \partial x_1 = f_1, \quad (12.1.1a_1)$$

$$-\Delta u_2 + \partial p / \partial x_2 = f_2, \quad (12.1.1a_2)$$

$$-\partial u_1 / \partial x_1 - \partial u_2 / \partial x_2 = 0. \quad (12.1.1b)$$

Im Falle von $\Omega \subset \mathbb{R}^3$ ist eine weitere Gleichung $-\Delta u_3 + \partial p / \partial x_3 = f_3$ hinzuzufügen, und die linke Seite von (12.1.1b) ist durch $-\partial u_3 / \partial x_3$ zu ergänzen. Eine von der Dimension unabhängige Darstellung erreicht man, wenn man (u_1, u_2) bzw. (u_1, u_2, u_3) zum Vektor \mathbf{u} zusammenfasst:

$$-\Delta \mathbf{u} + \nabla p = \mathbf{f} \quad \text{in } \Omega, \quad (12.1.2a)$$

$$-\operatorname{div} \mathbf{u} = 0 \quad \text{in } \Omega, \quad (12.1.2b)$$

Dabei ist div der Divergenzoperator

$$\operatorname{div} \mathbf{u} = \sum_{i=1}^n \frac{\partial u_i}{\partial x_i}.$$

n ist sowohl die Dimension von $\Omega \subset \mathbb{R}^n$ als auch die Anzahl der Komponenten von $\mathbf{u}(\mathbf{x}) \in \mathbb{R}^n$. Von physikalischem Interesse ist nur $n \leq 3$. In der Strömungsmechanik beschreibt die Stokes-Gleichung die Strömung eines inkompressiblen Mediums bei Vernachlässigung der Trägheitsterme. \mathbf{u} beschreibt das Geschwindigkeitsfeld: In $\mathbf{x} \in \mathbb{R}^n$ ist $u_i(\mathbf{x})$ die Geschwindigkeit des Mediums in x_i -Richtung. Die Funktion p bezeichnet den Druck.

Bisher wurden noch keine Randbedingungen formuliert. Wir beschränken uns im Folgenden auf Dirichlet-Randwerte:

$$\mathbf{u} = 0 \quad \text{auf } \Gamma. \quad (12.1.3)$$

Sie besagen, dass die Strömung am Rand ruht. Für p wird keine Randbedingung gegeben. Da mit dem Paar (\mathbf{u}, p) auch $(\mathbf{u}, p + \text{const})$ das Stokes-Problem (12.1.2a,b), (12.1.3) löst, erhält man die

Bemerkung 12.1.1 *Durch die Stokes-Gleichung (12.1.2a,b) und die Randbedingungen (12.1.3) ist p nur bis auf eine Konstante bestimmt.*

Die Stokes-Gleichungen sind als ein Beispiel für ein System von Differentialgleichungen ausgewählt. Es bleibt zu untersuchen, ob Gleichungen (12.1.2a,b) in einem geeignet zu definierendem Sinne elliptisch ist. Zwar sind die Funktionen u_i bei vorgegebenen p Lösungen der elliptischen Poisson-Gleichungen $-\Delta u_i =$

$f_i - \partial p / \partial x_i$, aber für die ebenfalls zu bestimmende Funktion p findet man in (12.1.2a,b) keine im bisherigen Sinne elliptische Gleichung.

Ein allgemeines System von q Differentialgleichungen für q Funktionen u_1, \dots, u_q kann man in der Form

$$\sum_{j=1}^q L_{ij} u_j = f_i \quad \text{in } \Omega \subset \mathbb{R}^n \quad (1 \leq i \leq q) \quad (12.1.4a)$$

mit Differentialoperatoren

$$L_{ij} = \sum_{|\alpha| \leq k_{ij}} c_\alpha D^\alpha \quad (1 \leq i, j \leq q) \quad (12.1.4b)$$

schreiben. Die Gleichungen (12.1.4a) werden zu $Lu = f$ zusammengefasst, wobei L die Matrix (L_{ij}) von Differentialoperatoren ist und $\mathbf{u} = (u_1, \dots, u_q)^\top$, $\mathbf{f} = (f_1, \dots, f_q)^\top$. Die Ordnung des Operators L_{ij} ist höchstens k_{ij} . Zahlen $m_1, \dots, m_q, m'_1, \dots, m'_q$ seien so gewählt, dass

$$k_{ij} = m_i + m'_j \quad (1 \leq i, j \leq q). \quad (12.1.5)$$

Als *Hauptteil* von L_{ij} definiert man $L_{ij}^P := \sum_{|\alpha|=m_i+m'_j} c_\alpha D^\alpha$. Das zu L_{ij} gehörende charakteristische Polynom lautet

$$L_{ij}^P(\xi; \mathbf{x}) := \sum_{|\alpha|=m_i+m'_j} c_\alpha(\mathbf{x}) \xi^\alpha \quad (\xi \in \mathbb{R}^n, \mathbf{x} \in \Omega)$$

und bildet die Matrixfunktion

$$L^P(\xi; \mathbf{x}) = (L_{ij}^P(\xi; \mathbf{x}))_{i,j=1,\dots,q}.$$

Definition 12.1.2 (Agmon-Douglis-Nirenberg [2]) *Es gelte (12.1.5) für m_i, m'_j . Der Differentialoperator $L = (L_{ij})$ heißt elliptisch in $\mathbf{x} \in \Omega$, falls*

$$\det L^P(\xi; \mathbf{x}) \neq 0 \quad \text{für alle } 0 \neq \xi \in \mathbb{R}^n. \quad (12.1.6a)$$

L heißt gleichmäßig elliptisch in Ω , falls ein $\varepsilon > 0$ existiert, sodass

$$|\det L^P(\xi; \mathbf{x})| \geq \varepsilon |\xi|^{2m} \quad \text{für alle } \mathbf{x} \in \Omega, \xi \in \mathbb{R}^n \quad \text{mit} \quad 2m := \sum_{i=1}^q (m_i + m'_i). \quad (12.1.6b)$$

Genauer müsste man L elliptisch hinsichtlich der Indizes m_i, m'_j nennen, da die Definition von L^P von m_i, m'_j abhängt. Zu diesem Problem sowie wegen einer zusätzlichen Bedingung für $\mathbf{x} \in \Gamma$, $q = 2$ vergleiche man die Originalarbeit Agmon-Douglas-Nirenberg [2]. Eine weitere Diskussion dieses Themas findet man bei Cosner [26].

Übungsaufgabe 12.1.3 *Man zeige: a) Die Zahlen m_i, m'_j sind nicht eindeutig. Erfüllen m_i und m'_j die Ungleichung (12.1.5), so auch $m_i - k, m'_j + k$. Die Definition von L_{ij}^P ist von k unabhängig.*

b) Für $q = 1$, d.h. für den Fall einer einzigen Gleichung erhält man aus (12.1.6a) die Definitionen 1.2.3a bzw. (5.1.3a) zurück. (12.1.6b) entspricht (5.1.3a').

c) Für ein System erster Ordnung (d.h. $k_{ij} = 1, m_i = 1, m'_j = 0$) ergibt (12.1.6a) die Definition 1.3.2.

Um die Stokes-Gleichungen in der Form (12.1.4a,b) zu schreiben, setzen wir

$$\begin{aligned} q &:= n + 1, & u &= (u_1, \dots, u_n, p)^\top, & f &= (f_1, \dots, f_n, 0)^\top, \\ L_{ii} &= -\Delta, & L_{iq} &= -L_{qi} = \partial / \partial x_i \quad \text{für } 1 \leq i \leq n, & L_{ij} &= 0 \quad \text{sonst.} \end{aligned}$$

Die Ordnungen sind $k_{ii} = 2, k_{iq} = k_{qi} = 1$ ($i \leq n$), $k_{ij} = 0$ sonst. Die Zahlen

$$m_i = m'_i = \begin{cases} 1 & \text{für } 1 \leq i \leq n = q - 1, \\ 0 & \text{für } i = q \end{cases}$$

erfüllen (12.1.5). L_{ij}^P stimmt mit L_{ij} überein und ist von \mathbf{x} unabhängig:

$$L_{ii}^P(\xi) = -|\xi|^2, \quad L_{iq}^P(\xi) = -L_{qi}^P(\xi) = \xi_i \quad \text{für } i \leq n, \quad L_{ij}^P(\xi) = 0 \quad \text{sonst.}$$

Damit ist

$$|\det L^P(\xi; \mathbf{x})| = \left| \det \begin{bmatrix} -|\xi|^2 & & \xi_1 \\ & \ddots & \\ -\xi_1 & \dots & -\xi_n & 0 \end{bmatrix} \right| = |\xi|^{2m} \quad \text{mit } 2m = 2n,$$

sodass (12.1.6b) mit $\varepsilon = 1$ erfüllt ist.

Übungsaufgabe 12.1.4 (Lamé-Gleichungen) In der Elastizitätstheorie¹ wird die sogenannte Verschiebungsfunktion $\mathbf{u} : \Omega \subset \mathbb{R}^3 \rightarrow \mathbb{R}^3$ durch das System von Lamé beschrieben:

$$\mu \Delta \mathbf{u} + (\lambda + \mu) \operatorname{grad} \operatorname{div} \mathbf{u} = \mathbf{f} \quad \text{in } \Omega, \quad \mathbf{u} = \varphi \quad \text{auf } \Gamma$$

($\mu, \lambda > 0$). Man zeige, dass dieses System von drei Gleichungen gleichmäßig elliptisch ist, wobei $|\det L^P(\xi)| = \mu^2(2\mu + \lambda) |\xi|^6$.

Zur Behandlung der Stokes-Gleichungen werden wir zum nächsten Abschnitt eine Variationsformulierung verwenden. Der Vollständigkeit halber sei auf folgende Umformung hingewiesen.

Bemerkung 12.1.5 Es sei $n = 2$ und somit $\mathbf{u} = (u_1, u_2)$. Wegen $\operatorname{div} \mathbf{u} = 0$ gibt es eine sogenannte Stromfunktion Φ mit $u_1 = \partial \Phi / \partial x_2$, $u_2 = -\partial \Phi / \partial x_1$. Einsetzen in Gleichungen (12.1.1a_{1,2}) ergibt die biharmonische Gleichung $\Delta^2 \Phi = \partial f_2 / \partial x_1 - \partial f_1 / \partial x_2$. Die Randbedingung (12.1.3) bedeutet $\nabla \Phi = 0$ auf Γ . Dies ist äquivalent zu $\partial \Phi / \partial n = 0$ und $\partial \Phi / \partial t = 0$ auf Γ , wobei $\partial / \partial t$ die Tangentialableitung sei. $\partial \Phi / \partial t = 0$ impliziert $\Phi = \text{const}$ auf Γ . Da die Konstante beliebig gewählt werden kann, setzt man $\Phi = \partial \Phi / \partial n = 0$ auf Γ .

12.2 Variationsformulierung

12.2.1 Schwache Formulierung der Stokes-Gleichungen

Da $\mathbf{u} = (u_1, \dots, u_n)$ eine vektorwertige Funktion ist, führen wir

$$\mathbf{H}_0^1(\Omega) := H_0^1(\Omega) \times H_0^1(\Omega) \times \dots \times H_0^1(\Omega) \quad (n\text{-faches Produkt})$$

ein. Eine entsprechende Definition gilt für $\mathbf{H}^{-1}(\Omega)$, $\mathbf{H}^2(\Omega)$ usw. Die zu $\mathbf{H}_0^1(\Omega)$ gehörende Norm wird weiterhin mit $|\cdot|_1$ bezeichnet.

Nach Bemerkung 12.1.1 ist die Druckkomponente p des Stokes-Problems nicht eindeutig bestimmt. Um die Konstante in $p = \tilde{p} + \text{const}$ eindeutig festzulegen, normieren wir p durch die Forderung $\int_{\Omega} p \, d\mathbf{x} = 0$. Deshalb wird im Folgenden p stets zum Unterraum $L_0^2(\Omega) \subset L^2(\Omega)$ gehören:

$$L_0^2(\Omega) := \left\{ p \in L^2(\Omega) : \int_{\Omega} p(\mathbf{x}) \, d\mathbf{x} = 0 \right\}.$$

Zur Herleitung der schwachen Formulierung gehen wir wie in Abschnitt 7.1 vor und nehmen an, dass \mathbf{u} und p klassische Lösungen des Stokes-Problems (12.1.2a,b) sind. Multiplikation der i -ten Gleichung in (12.1.2a) mit $v_i \in C_0^\infty(\Omega)$ und anschließende Integration liefert

$$\begin{aligned} \int_{\Omega} f_i(\mathbf{x}) v_i(\mathbf{x}) \, d\mathbf{x} &= \int_{\Omega} [-\Delta u_i(\mathbf{x}) + \partial p(\mathbf{x}) / \partial x_i] v_i(\mathbf{x}) \, d\mathbf{x} \\ &= \int_{\Omega} [\langle \nabla u_i(\mathbf{x}), \nabla v_i(\mathbf{x}) \rangle - p(\mathbf{x}) \partial v_i(\mathbf{x}) / \partial x_i] \, d\mathbf{x} \quad \text{für } v_i \in C_0^\infty(\Omega) \end{aligned} \quad (12.2.1a)$$

¹Eine detaillierte Beschreibung der verschiedenen Differentialgleichungen in der Elastizitätstheorie findet man in Braess [12, §VI.3].

mit $\mathbf{v} = (v_i)_{i=1,\dots,n}$. Summation über i ergibt

$$\int_{\Omega} [\langle \nabla \mathbf{u}(\mathbf{x}), \nabla \mathbf{v}(\mathbf{x}) \rangle - p(\mathbf{x}) \operatorname{div} \mathbf{v}(\mathbf{x})] \, d\mathbf{x} = \int_{\Omega} \langle \mathbf{f}(\mathbf{x}), \mathbf{v}(\mathbf{x}) \rangle \, d\mathbf{x}, \quad (12.2.1a')$$

wobei die Abkürzung

$$\langle \nabla \mathbf{u}, \nabla \mathbf{v} \rangle := \sum_{i=1}^n \langle \nabla u_i, \nabla v_i \rangle := \sum_{i,j=1}^n \frac{\partial u_i}{\partial x_j} \frac{\partial v_i}{\partial x_j}$$

verwendet wird. Gleichung (12.1.2b) wird mit einem $q \in L_0^2(\Omega)$ multipliziert und integriert:

$$-\int_{\Omega} q(\mathbf{x}) \operatorname{div} \mathbf{u}(\mathbf{x}) \, d\mathbf{x} = 0 \quad \text{für alle } q \in L_0^2(\Omega). \quad (12.2.1b)$$

Mit den Bilinearformen

$$a(\mathbf{u}, \mathbf{v}) := \int_{\Omega} [\langle \nabla \mathbf{u}(\mathbf{x}), \nabla \mathbf{v}(\mathbf{x}) \rangle - p(\mathbf{x}) \operatorname{div} \mathbf{v}(\mathbf{x})] \, d\mathbf{x} \quad \text{für } \mathbf{u}, \mathbf{v} \in \mathbf{H}_0^1(\Omega), \quad (12.2.2a)$$

$$b(p, \mathbf{v}) := -\int_{\Omega} p(\mathbf{x}) \operatorname{div} \mathbf{v}(\mathbf{x}) \, d\mathbf{x} \quad \text{für } p \in L_0^2(\Omega), \quad \mathbf{v} \in \mathbf{H}_0^1(\Omega) \quad (12.2.2b)$$

erhalten wir die schwache Formulierung des Stokes-Problems durch (12.2.3a-c):

$$\text{suche } \mathbf{u} \in \mathbf{H}_0^1(\Omega) \quad \text{und} \quad p \in L_0^2(\Omega), \quad \text{sodass} \quad (12.2.3a)$$

$$a(\mathbf{u}, \mathbf{v}) + b(p, \mathbf{v}) = f(\mathbf{v}) := \int_{\Omega} \langle \mathbf{f}(\mathbf{x}), \mathbf{v}(\mathbf{x}) \rangle \, d\mathbf{x} \quad \text{für alle } \mathbf{v} \in \mathbf{H}_0^1(\Omega), \quad (12.2.3b)$$

$$b(q, \mathbf{u}) = 0 \quad \text{für alle } q \in L_0^2(\Omega). \quad (12.2.3c)$$

In (12.2.3b) ist zunächst “ $v \in H_0^1(\Omega)$ ” durch “ $v \in C_0^\infty(\Omega)$ ” zu ersetzen. Da beide Seiten von (12.2.3b) stetig von $v \in H_0^1(\Omega)$ abhängen und $C_0^\infty(\Omega)$ dicht in $H_0^1(\Omega)$ liegt, folgt (12.2.3b) für alle $v \in H_0^1(\Omega)$.

Bemerkung 12.2.1 Eine klassische Lösung $u \in C^2(\overline{\Omega}) \cap H_0^1(\Omega)$, $p \in C^1(\Omega) \cap L_0^2(\Omega)$ des Stokes-Problems (12.1.2a,b), (12.1.3) ist auch schwache Lösung, d.h. Lösung von (12.2.3a-c). Hat umgekehrt (12.2.3a-c) eine Lösung mit $u \in C^2(\overline{\Omega})$, $p \in C^1(\overline{\Omega})$, so ist sie auch klassische Lösung der Randwertaufgabe (12.1.2a,b), (12.1.3).

Beweis. a) Die vorangehenden Überlegungen beweisen den ersten Teil.

b) Gleichung (12.2.3c) impliziert $\operatorname{div} u = 0$. Sei $i \in \{1, \dots, n\}$. In Gleichung (12.2.3b) kann man v mit $v_i \in C_0^\infty(\Omega)$, $v_j = 0$ für $j \neq i$ wählen. Partielle Integration liefert (12.2.1a) zurück und damit die i -te Gleichung aus (12.1.2a). ■

12.2.2 Sattelpunktprobleme

Die Situation in (12.2.3a-c) ist ein Spezialfall des folgenden Problems. Die Räume $H_0^1(\Omega)$ und $L_0^2(\Omega)$ in (12.2.3a-c) ersetzen wir durch zwei Hilbert-Räume V und W . Es seien

$$a(\cdot, \cdot) : V \times V \rightarrow \mathbb{R} \quad \text{eine stetige Bilinearform auf } V \times V, \quad (12.2.4a)$$

$$b(\cdot, \cdot) : W \times V \rightarrow \mathbb{R} \quad \text{eine stetige Bilinearform auf } W \times V, \quad (12.2.4b)$$

$$f_1 \in V', \quad f_2 \in W'. \quad (12.2.4c)$$

In Verallgemeinerung von (6.5.1) heißt $b(\cdot, \cdot) : W \times V \rightarrow \mathbb{R}$ stetig (oder beschränkt), wenn ein $C_b \in \mathbb{R}$ existiert, sodass

$$|b(w, v)| \leq C_b \|w\|_W \|v\|_V \quad \text{für alle } w \in W, v \in V.$$

Gegenstand dieses Kapitels ist das Problem (12.2.5):

$$\text{Suche } v \in V \text{ und } w \in W \text{ mit } \begin{cases} a(v, x) + b(w, x) &= f_1(x) & \text{für alle } x \in V, \\ b(y, v) &= f_2(y) & \text{für alle } y \in W. \end{cases} \quad (12.2.5)$$

Formal lässt sich (12.2.5) in die bisherige Form

$$\text{suche } u \in X \text{ mit } c(u, z) = f(z) \quad \text{für alle } z \in X \quad (12.2.6a)$$

bringen, wenn man setzt:

$$X := V \times W, \quad \left\{ \begin{array}{l} c(u, z) := a(v, x) + b(w, x) + b(y, z) \\ f(z) := f_1(x) + f_2(y) \end{array} \right\} \quad \text{für } u = \begin{pmatrix} v \\ w \end{pmatrix}, \quad z = \begin{pmatrix} x \\ y \end{pmatrix}. \quad (12.2.6b)$$

Übungsaufgabe 12.2.2 Man zeige: a) $c(\cdot, \cdot) : X \times X \rightarrow \mathbb{R}$ ist eine stetige Bilinearform.
b) Die Probleme (12.2.5) und (12.2.6a,b) sind äquivalent.

Dass die Variationsprobleme (12.2.5) bzw. (12.2.6a) anders als in Kapitel 7 behandelt werden müssen, deutet die folgende Bemerkung an.

Bemerkung 12.2.3 Die Bilinearform $c(\cdot, \cdot)$ aus (12.2.6b) kann nicht X -elliptisch sein.

Beweis. Es ist $c(u, u) = 0$ für alle $u = \begin{pmatrix} 0 \\ w \end{pmatrix}$. ■

In Analogie zu (6.5.9) setzen wir

$$J(v, w) := a(v, v) + 2b(w, v) - 2f_1(v) - 2f_2(w),$$

also $J(v, w) = c(u, u) - 2f(u)$ für $u = \begin{pmatrix} v \\ w \end{pmatrix}$. Für $v \in V$ und $w \in W$ ist $J(v, w)$ weder von unten noch von oben beschränkt. Die Lösung v^*, w^* von (12.2.5) minimiert J daher nicht. Unter geeigneten Bedingungen ist v^*, w^* jedoch ein *Sattelpunkt*, wie er durch (12.2.7) charakterisiert ist.

Satz 12.2.4 Es gelte (12.2.4a-c). $a(\cdot, \cdot)$ sei symmetrisch und V -elliptisch. Das Paar $v^* \in V$, $w^* \in W$ ist genau dann eine Lösung der Aufgabe (12.2.5), wenn

$$J(v^*, w) \leq J(v^*, w^*) \leq J(v, w^*) \quad \text{für alle } w \in W, v \in V. \quad (12.2.7)$$

Eine weitere, äquivalente Charakterisierung ist

$$J(v^*, w^*) = \min_{v \in V} J(v, w^*) = \max_{w \in W} \min_{v \in V} J(v, w). \quad (12.2.8)$$

Beweis. 1a) v^*, w^* löse (12.2.5). Wegen der Symmetrie von $a(\cdot, \cdot)$ ist $a(v^* - v, v^* - v) - 2a(v^*, v^* - v) = -a(v^*, v^*) - a(v, v^*) + a(v^*, v) + a(v, v) = a(v, v) - a(v^*, v^*)$. Die eckige Klammer in

$$J(v, w^*) - J(v^*, w^*) = a(v^* - v, v^* - v) + 2[a(v^*, v^* - v) + b(w^*, v^* - v) - f_1(v^* - v)]$$

verschwindet wegen der ersten Gleichung in (12.2.5). Da $a(v^* - v, v^* - v) > 0$ für alle $v^* \neq v \in V$, folgt die zweite Ungleichung in (12.2.7). Wie in Satz 6.5.12 beweist man auch die Umkehrung: Ist $J(v, w^*)$ für $v = v^*$ minimal, so erhält man die erste Gleichung aus (12.2.5) zurück.

1b) Ist v^* Lösung der zweiten Gleichung in (12.2.5), verschwindet

$$J(v^*, w^*) - J(v^*, w) = 2[b(w^* - w, v^*) - f_2(w^* - w)]$$

für alle w und beweist den ersten Teil von (12.2.7) in der verstärkten Form der Gleichheit $J(v^*, w) = J(v^*, w^*)$. Für die Umkehrung definiere man $w_{\pm} = w^* \pm w$. Der erste Teil von (12.2.7) impliziert

$$0 \leq J(v^*, w^*) - J(v^*, w_{\pm}) = \pm 2[b(w, v^*) - f_2(w)]$$

für beide Vorzeichen, sodass $b(w, v^*) = f_2(w)$. Da $w \in W$ beliebig, ist die zweite Gleichung aus (12.2.5) zurückgewonnen.

2a) Wir setzen

$$j(w) := \min_{v \in V} J(v, w).$$

Gemäß Satz 6.5.12 ist $j(w) = J(v_w, w)$, wobei $v_w \in V$ die Lösung der oberen Gleichung in (12.2.5) ist. Sind v_w und $v_{w'}$ die Lösungen zu w und w' , so folgt

$$a(v_w - v_{w'}, x) = F(x) := b(w - w', x) \quad \text{für alle } x \in X.$$

Da $\|F\|_{V'} \leq C_b \|w - w'\|_V$ und $\|v_w - v_{w'}\|_V \leq C' \|F\|_{V'}$, erhält man

$$\|v_w - v_{w'}\|_V \leq C \|w - w'\|_V \quad \text{für alle } w, w' \in W. \quad (12.2.9a)$$

2b) Unter Ausnutzung der Definition von v_w durch die obere Gleichung aus (12.2.5) formen wir um:

$$\begin{aligned} J(v^*, w^*) - J(v_w, w) &= \quad (\text{wegen } a(v^*, v^*) - a(v_w, v_w) = -2a(v_w, v_w - v^*) + a(v_w - v^*, v_w - v^*)) \\ &= [a(v^*, v^*) + 2b(w^*, v^*) - 2f_1(v^*) - 2f_2(w^*)] - [a(v_w, v_w) + 2b(w, v_w) - 2f_1(v_w) - 2f_2(w)] \\ &= 2[f_1(v_w - v^*) - b(w, v_w - v^*) - a(v_w, v_w - v^*)] + a(v_w - v^*, v_w - v^*) + 2[b(w^* - w, v^*) - f_2(w^* - w)] \\ &= a(v_w - v^*, v_w - v^*) + 2[b(w^* - w, v^*) - f_2(w^* - w)]. \end{aligned} \quad (12.2.9b)$$

2c) Sei v^*, w^* eine Lösung von (12.2.5). Wegen der zweiten Gleichung aus (12.2.5) verschwindet die eckige Klammer in (12.2.9b), und es ist

$$J(v^*, w^*) = J(v_w, w) + a(v_w - v^*, v_w - v^*) \geq J(v_w, w) = j(w).$$

Die obere Gleichung in (12.2.5) beweist $v_{w^*} = v^*$, sodass

$$J(v^*, w^*) = j(w^*) = \max_{w \in W} j(w), \quad (12.2.9c)$$

d.h. (12.2.8) gilt.

2d) Sei nun v^*, w^* eine Lösung von (12.2.8). Setzt man in (12.2.9b) $w = w^*, v_w = v_{w^*}$, ergibt sich aus $J(v^*, w^*) = j(w^*)$, dass $v^* = v_{w^*}$. Damit hängt $a(v_w - v^*, v_w - v^*) = a(v_w - v_{w^*}, v_w - v_{w^*})$, quadratisch von $\|w - w^*\|_V$ ab (vgl. (12.2.9a)). Die Variation über $w := w^* - \lambda y$ ($\lambda \in \mathbb{R}, y \in W$ beliebig) liefert

$$0 = \frac{d}{d\lambda} j(w^* - \lambda y)|_{\lambda=0} = 2[b(y, v^*) - f_2(y)],$$

sodass die untere Gleichung in (12.2.5) bewiesen ist. Die obere Gleichung aus (12.2.5) ist bereits mit $v^* = v_{w^*}$ gezeigt. ■

12.2.3 Existenz und Eindeutigkeit der Lösung eines Sattelpunktproblems

Um das Sattelpunktproblem (12.2.5) etwas durchsichtiger zu machen, führen wir die zu den Bilinearformen gehörenden Operatoren ein:

$$A \in L(V, V') \quad \text{mit } a(v, x) = \langle Av, x \rangle_{V' \times V} \quad \text{für alle } v, x \in V, \quad (12.2.10a)$$

$$B \in L(W, V'), \quad B^* \in L(V, W') \quad \text{mit } b(w, x) = \langle Bw, x \rangle_{V' \times V} = \langle w, B^*x \rangle_{W \times W'}, \quad (12.2.10b)$$

$$C \in L(X, X') \quad \text{mit } c(u, z) = \langle Cu, z \rangle_{X' \times X} \quad \text{für alle } u, z \in X, \quad (12.2.10c)$$

Damit hat Aufgabe (12.2.6a,b) die Gestalt $Cu = f$, während sich (12.2.5) als

$$Av + Bw = f_1, \quad (12.2.11a)$$

$$B^*v = f_2 \quad (12.2.11b)$$

schreiben lässt. Nimmt man die Existenz von $A^{-1} \in L(V', V)$ an, kann man (12.2.11a) nach v auflösen:

$$v = A^{-1}(f_1 - Bw) \quad (12.2.12a)$$

und in (12.2.11b) einsetzen:

$$B^*A^{-1}Bw = B^*A^{-1}f_1 - f_2. \quad (12.2.12b)$$

Die Invertierbarkeit von A ist für die Lösbarkeit des Sattelpunktproblems keineswegs notwendig (die genaue Bedingung wird in Satz 12.2.7 diskutiert). Sie vereinfacht jedoch die Analyse und ist im Falle des Stokes-Problems gegeben.

Bemerkung 12.2.5 a) Unter den Voraussetzungen

$$A^{-1} \in L(V', V), \quad B^* A^{-1} B \in L(W', W) \quad (12.2.13)$$

sind das Sattelpunktproblem (12.2.5) bzw. die Gleichungen (12.2.11a,b) eindeutig lösbar.

b) Eine notwendige Bedingung für die Existenz von $(B^* A^{-1} B)^{-1}$ ist

$$B \in L(W, V') \quad \text{ist injektiv.} \quad (12.2.14)$$

Beweis. a) Unter der Voraussetzung $(B^* A^{-1} B)^{-1} \in L(W', W)$ ist (12.2.12b) eindeutig nach w auflösbar, und (12.2.12a) ergibt v .

b) Die Injektivität von $B^* A^{-1} B$ impliziert (12.2.14). ■

Achtung. Im Allgemeinen ist $B : W \rightarrow V'$ nicht bijektiv, sodass eine Darstellung von $(B^* A^{-1} B)^{-1}$ als $B^{-1} A B^{*-1}$ nicht möglich ist.

Anhand der 3×3 -Matrix $C = \begin{bmatrix} A & B \\ B^\top & 0 \end{bmatrix}$ mit $A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$ und $B = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ sieht man, dass ein System der Form (12.2.11a,b) auch bei singulärer Matrix A auflösbar sein kann. Deshalb ist die Voraussetzung $A^{-1} \in L(V', V)$ nicht notwendig. Eine genauere Betrachtung benutzt den Unterraum

$$V_0 := \text{Kern}(B^*) = \{v \in V : b(y, v) = 0 \text{ für alle } y \in W\} \subset V, \quad (12.2.15)$$

der, wie schon vermerkt ist, im Allgemeinen nicht trivial ist. Der Kern einer stetigen Abbildung ist abgeschlossen, sodass sich V nach Lemma 6.1.17 als Summe von Orthogonalräumen darstellen lässt:

$$V = V_0 \oplus V_\perp \quad \text{mit} \quad V_\perp := (V_0)^\perp. \quad (12.2.16a)$$

Übungsaufgabe 12.2.6 Es gelte (12.2.16a). Man zeige: a) Der Dualraum V' lässt sich darstellen als

$$V' = V'_0 \oplus V'_\perp \quad (12.2.16b)$$

wobei

$$V'_0 := \{v' \in V' : v'(v) = 0 \text{ für alle } v \in V_\perp\}, \quad V'_\perp := \{v' \in V' : v'(v) = 0 \text{ für alle } v \in V_0\}. \quad (12.2.16c)$$

Als Norm auf V'_0 und V'_\perp wird $\|\cdot\|_{V'}$ verwendet.

b) Der Riesz-Isomorphismus $J_V : V \rightarrow V'$ bildet V_0 auf V'_0 und V_\perp auf V'_\perp ab.

c) V'_0 und V'_\perp sind Orthogonalräume bezüglich der Dualnorm $\|\cdot\|_{V'}$.

d) Es gilt

$$\|v'\|_{V'}^2 = \|v'_0\|_{V'}^2 + \|v'_\perp\|_{V'}^2, \quad \text{für alle } v' = v'_0 + v'_\perp \in V', \text{ wobei } v'_0 \in V'_0, v'_\perp \in V'_\perp. \quad (12.2.16d)$$

Die Zerlegungen (12.2.16a,b) von V und V' definieren eine Blockzerlegung des Operators A :

$$A = \begin{bmatrix} A_{00} & A_{0\perp} \\ A_{\perp 0} & A_{\perp\perp} \end{bmatrix}, \quad \text{mit} \quad \begin{cases} A_{00} \in L(V_0, V'_0), & A_{0\perp} \in L(V_\perp, V'_0), \\ A_{\perp 0} \in L(V_0, V'_\perp), & A_{\perp\perp} \in L(V_\perp, V'_\perp). \end{cases}$$

Dabei ist zum Beispiel A_{00} wie folgt definiert:

$$A_{00}v_0 = v'_0 \quad \text{für } v_0 \in V_0, \quad \text{falls } Av_0 = v'_0 + v'_\perp, v'_0 \in V'_0, v'_\perp \in V'_\perp.$$

Die entsprechende Zerlegung von B^* in (B_0^*, B_\perp^*) schreiben wir als $(0, B^*)$, da $B_0^* = 0$ nach Definition von V_0 . Umgekehrt gilt $\text{Bild}(B) \subset V'_\perp$, sodass $B = \begin{pmatrix} 0 \\ B \end{pmatrix}$. Aus dem System (12.2.11a,b) wird daher

$$A_{00}v_0 + A_{0\perp}v_\perp = f_{10}, \quad (12.2.17a)$$

$$A_{\perp 0}v_0 + A_{\perp\perp}v_\perp + Bw = f_{1\perp}, \quad (12.2.17b)$$

$$B^*v_\perp = f_2, \quad (12.2.17c)$$

wobei $v = v_0 + v_\perp$, $v_0 \in V_0$, $v_\perp \in V_\perp$, $f_1 = f_{10} + f_{1\perp}$, $f_{10} \in V'_0$, $f_{1\perp} \in V'_\perp$.

Satz 12.2.7 Es gelte (12.2.4a-c). V_0 sei durch (12.2.15) definiert. Notwendig und hinreichend für die eindeutige Lösbarkeit des Sattelpunktpblems (12.2.5) für alle $f_1 \in V'$ ist die Existenz der Inversen

$$A_{00}^{-1} \in L(V'_0, V_0) \quad \text{und} \quad B^{-1} \in L(V'_\perp, W). \quad (12.2.18)$$

Beweis. 1) (12.2.17a-c) stellt ein gestaffeltes Gleichungssystem dar: (12.2.18) impliziert $B^{*-1} = (B^{-1})^* \in L(W', V'_\perp)$, sodass man (12.2.17c) nach $v_\perp = B^{*-1} f_2$ auflösen kann. $v_0 \in V_0$ erhält man aus (12.2.17a): $v_0 = A_{00}^{-1}(f_{10} - A_{0\perp} v_\perp)$. Schließlich ergibt sich w aus (12.2.17b).

2) Um zu zeigen, dass (12.2.18) notwendig ist, wählen wir $f_{10} \in V'_0$ beliebig, $f_{1\perp} = 0$ und $f_2 = 0$. Nach Voraussetzung existiert hierzu eine Lösung $(v_0, v_\perp, w) \in V_0 \times V_\perp \times W$. $B^* v_\perp = 0$ impliziert $v_\perp \in V_0$, sodass $v_\perp = 0$ wegen $V_0 \cap V_\perp = \{0\}$. Somit hat $A_{00} v_0 = f_{10}$ für jedes $f_{10} \in V'_0$ eine eindeutige Lösung $v_0 \in V_0$. Also ist $A_{00} : V_0 \rightarrow V'_0$ bijektiv und beschränkt, und Satz 6.1.13 zeigt, dass $A_{00}^{-1} \in L(V'_0, V_0)$. Wählt man $f_{1\perp} \in V'_\perp$ beliebig und $f_{10} = 0$, $f_2 = 0$, folgert man $v_\perp = 0$ und $v_0 = 0$, sodass $Bw = f_{1\perp}$ eine eindeutige Lösung $w \in W$ besitzt. Wie für A_{00} schließt man, dass auch $B^{-1} \in L(V'_\perp, W)$. ■

Die Formulierung der Bedingungen (12.2.18) mit Hilfe der Bilinearformen ergibt die *Babuška-Brezzi-Bedingungen*:

$$\inf_{v_0 \in V_0 \text{ mit } \|v_0\|_V=1} \sup_{x_0 \in V_0 \text{ mit } \|x_0\|_V=1} |a(v_0, x_0)| \geq \alpha > 0, \quad (12.2.19a)$$

$$\sup_{x_0 \in V_0 \text{ mit } \|x_0\|_V=1} |a(x_0, v_0)| > 0 \quad \text{für alle } 0 \neq v_0 \in V_0, \quad (12.2.19b)$$

$$\inf_{w \in W \text{ mit } \|w\|_W=1} \sup_{x \in V \text{ mit } \|v\|_V=1} |b(w, x)| \geq \beta > 0. \quad (12.2.19c)$$

Übungsaufgabe 12.2.8 Man zeige: (12.2.19a) und (12.2.19c) sind jeweils äquivalent zu

$$\sup_{x_0 \in V_0 \text{ mit } \|x_0\|_V=1} |a(v_0, x_0)| \geq \alpha \|v_0\|_V \quad \text{für alle } v_0 \in V_0, \quad (12.2.19a')$$

$$\sup_{x \in V \text{ mit } \|v\|_V=1} |b(w, x)| \geq \beta \|w\|_W \quad \text{für alle } w \in W. \quad (12.2.19c')$$

Lemma 12.2.9 Es gelte (12.2.4a,b). V_0 sei durch (12.2.15) definiert. Dann sind die Bedingungen (12.2.18) und (12.2.19a-c) äquivalent. Dabei gilt $\|A_{00}^{-1}\|_{V_0 \leftarrow V'_0} \leq 1/\alpha$, $\|B^{-1}\|_{W \leftarrow V'_\perp} \leq 1/\beta$.

Beweis. Wegen (12.2.16b) und $b(w, x) = 0$ für $x \in V_0$ kann (12.2.19c) in der Form

$$\inf_{w \in W \text{ mit } \|w\|_W=1} \sup_{x \in V_\perp \text{ mit } \|v\|_V=1} |b(w, x)| \geq \beta > 0 \quad (12.2.19d)$$

geschrieben werden. Für $0 \neq x \in V_\perp$ gilt $x \notin V_0$, daher nach (12.2.15):

$$\sup_{w \in W \text{ mit } \|w\|_W=1} |b(w, x)| > 0 \quad \text{für alle } 0 \neq x \in V_\perp. \quad (12.2.19e)$$

Wie im Beweis des Lemmas 6.5.3 ergibt sich die Äquivalenz von (12.2.19a,b) mit $A_{00}^{-1} \in L(V'_0, V_0)$ und von (12.2.19d,e) mit $B^{-1} \in L(V'_\perp, W)$. ■

Zusatz 12.2.10 a) Die Bedingung (12.2.19b) erübrigt sich, falls $a(\cdot, \cdot)$ auf $V_0 \times V_0$ symmetrisch ist oder Lemma 6.5.17 zutrifft.

b) Jede der folgenden Bedingungen ist hinreichend für (12.2.19a,b) und damit auch für $A_{00}^{-1} \in L(V'_0, V_0)$:

$$a(\cdot, \cdot) : V_0 \times V_0 \rightarrow \mathbb{R} \text{ ist } V_0\text{-elliptisch:} \quad a(v_0, v_0) \geq \alpha \|v_0\|_V^2 \quad \text{für alle } v_0 \in V_0, \quad (12.2.20a)$$

$$a(\cdot, \cdot) : V \times V \rightarrow \mathbb{R} \text{ ist } V\text{-elliptisch.} \quad (12.2.20b)$$

Beweis. a) Wie in Lemma 6.5.17. b) (12.2.20b) impliziert (12.2.20a). (12.2.20a) ergibt (12.2.19a,b). ■

Übungsaufgabe 12.2.11 Man zeige: Unter der Voraussetzung (12.2.4a-c) ist (12.2.18) auch äquivalent zur Existenz von $C^{-1} \in L(X', X)$ (vgl. (12.2.10c)). Man gebe eine Schranke für $\|C^{-1}\|_{X \leftarrow X'}$ mit Hilfe von $\|A_{00}^{-1}\|_{V_0 \leftarrow V'_0}$, $\|A\|_{V' \leftarrow V'}$ und $\|B^{-1}\|_{W \leftarrow V'_\perp}$ an.

12.2.4 Lösbarkeit und Regularität des Stokes-Problems

Die Bedingungen (12.2.19a,b) (d.h. $A_{00}^{-1} \in L(V'_0, V_0)$) sind für das Stokes-Problem leicht erfüllbar:

Lemma 12.2.12 Ω sei beschränkt. Dann erfüllen die Formen (12.2.2a,b), die das Stokes-Problem beschreiben, die Bedingungen (12.2.4a,b) und (12.2.19a,b).

Beweis. (12.2.4a,b) ist offensichtlich. Nach Beispiel 7.2.10 ist $\int_{\Omega} \langle \nabla u, \nabla v \rangle d\mathbf{x}$ $H_0^1(\Omega)$ -elliptisch. Hieraus ergibt sich die $\mathbf{H}_0^1(\Omega)$ -Elliptizität von $a(\cdot, \cdot)$. Korollar 12.2.10b beweist (12.2.19a,b). ■

Es bleibt die Bedingung (12.2.19c) zu beweisen, die für das Stokes-Problem die Gestalt

$$\sup_{u \in \mathbf{H}_0^1(\Omega) \text{ mit } |u|_1=1} \left| \int_{\Omega} w(\mathbf{x}) \operatorname{div} u(\mathbf{x}) d\mathbf{x} \right| \geq \beta |w|_0 \quad \text{für alle } w \in L_0^2(\Omega) \quad (12.2.21)$$

oder die äquivalente Form

$$\|\nabla w\|_{\mathbf{H}^{-1}(\Omega)} \geq \beta \|w\|_{L^2(\Omega)} \quad \text{für alle } w \in L_0^2(\Omega) \quad (12.2.21')$$

annimmt.

Lemma 12.2.13 Hinreichend und notwendig für (12.2.21) ist, dass für jedes $w \in L_0^2(\Omega)$ ein $u \in \mathbf{H}_0^1(\Omega)$ existiert, sodass

$$w = \operatorname{div} u, \quad |u|_1 \leq \frac{1}{\beta} |w|_0. \quad (12.2.21'')$$

Beweis. 1) Zu $w \in L_0^2(\Omega)$ wähle u mit (12.2.21'') und setze $\tilde{u} := u/|u|_1$. Die linke Seite in (12.2.21) ist $\geq \int_{\Omega} w(\mathbf{x}) \operatorname{div} \tilde{u}(\mathbf{x}) d\mathbf{x} = |w|_0^2 / |u|_1 \geq \beta |w|_0$.

2) Gilt (12.2.21), so folgert man wie in Abschnitt 12.2.3 die Bijektivität von $B^* : V_{\perp} \rightarrow W$ mit $\|B^{*-1}\|_{V_{\perp} \leftarrow W} \leq 1/\beta$. Daher erfüllt $u := B^{*-1}w$ die Bedingung (12.2.21''). ■

Nečas [65] beweist den

Satz 12.2.14 Bedingung (12.2.21) ist erfüllt, falls $\Omega \in C^{0,1}$ beschränkt ist. Unter dieser Voraussetzung hat damit die Stokes-Aufgabe

$$-\Delta u + \nabla p = f, \quad -\operatorname{div} u = g \quad \text{in } \Omega, \quad u = 0 \quad \text{auf } \Gamma \quad (12.2.22)$$

für $f \in \mathbf{H}^{-1}(\Omega)$ und $g \in L_0^2(\Omega)$ eine eindeutige Lösung $(u, p) \in \mathbf{H}_0^1(\Omega) \times L_0^2(\Omega)$ mit

$$|u|_1 + |p|_0 \leq C_{\Omega} [|f|_{-1} + |g|_0]. \quad (12.2.23a)$$

Bemerkung 12.2.15 Unter den Voraussetzungen, dass $n = 2$ und dass $\Omega \in C^2$ ein beschränktes Gebiet ist, kann der Existenzbeweis wie folgt geführt werden.

Beweis. Es ist (12.2.21'') nachzuweisen. Zu $w \in L_0^2(\Omega)$ löse $-\Delta \varphi = w$ in Ω , $\varphi = 0$ auf Γ . Satz 9.1.20 zeigt $\varphi \in H^2(\Omega)$. Da $\nabla \varphi \in \mathbf{H}^1(\Omega)$ und $\vec{n}(x) \in C^1(\Gamma)$, folgt $g := \partial \varphi / \partial n \in H^{1/2}(\Gamma)$ (vgl. Satz 6.2.41a). Aus (3.4.2) schließt man $\int_{\Gamma} g d\Gamma = \int_{\Omega} w d\mathbf{x} = 0$, da $w \in L_0^2(\Omega)$. Integration von g über Γ liefert $G \in H^{3/2}(\Gamma)$ mit $\partial G / \partial t = g$, wobei $\partial / \partial t$ die Tangentialableitung ist. Es existiert eine Funktion $\psi \in H^2(\Omega)$ mit $\psi = G$ und $\partial \psi / \partial n = 0$ auf Γ und $|\Psi|_2 \leq C |G|_{3/2} \leq C' |g|_{1/2} \leq C'' |\varphi|_2 \leq C''' |w|_0$. Wir setzen

$$u_1 := -\varphi_x - \psi_y, \quad u_2 := -\varphi_y + \psi_x.$$

Offenbar gilt $u_1, u_2 \in H^1(\Omega)$. Die Normalenrichtung in $\mathbf{x} \in \Gamma$ sei $\vec{n}(\mathbf{x}) = (n_1(\mathbf{x}), n_2(\mathbf{x}))^T$. Die Tangentialrichtung ist daher $\vec{t}(\mathbf{x}) := (n_2(\mathbf{x}), -n_1(\mathbf{x}))^T$. Für $u = (u_1, u_2)^T$ erhält man

$$\langle u, \vec{n} \rangle = -\varphi_x n_1 - \psi_y n_1 - \varphi_y n_2 + \psi_x n_2 = -\partial \varphi / \partial n + \partial \psi / \partial t = -g + \partial G / \partial t = 0,$$

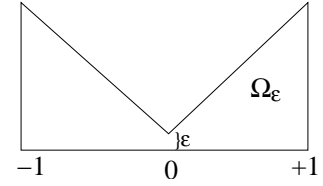
$$\langle u, \vec{t} \rangle = -\varphi_x n_2 - \psi_y n_2 - \varphi_y n_1 + \psi_x n_1 = -\partial \varphi / \partial t - \partial \psi / \partial n = 0,$$

denn $\varphi = 0$ auf Γ impliziert $\partial \varphi / \partial t = 0$, während $\partial \psi / \partial n = 0$ nach Definition gilt. $\langle u, \vec{n} \rangle = \langle u, \vec{t} \rangle = 0$ ergibt $u = 0$ auf Γ , sodass $u = (u_1, u_2)^T \in \mathbf{H}_0^1(\Omega)$ bewiesen ist. Man verifiziert, dass

$$\operatorname{div} u = \partial u_1 / \partial x + \partial u_2 / \partial y = (-\varphi_{xx} - \psi_{yx}) + (-\varphi_{yy} + \psi_{xy}) = -\Delta \varphi = w$$

mit $|u|_1 \leq |\varphi|_2 + |\psi|_2 \leq C |w|_0$. ■

Der vorgeführte Beweis benutzt die H^2 -Regularität des Poisson-Problems und benötigt entsprechende Voraussetzungen an Ω . Auch Satz 12.2.14 setzt $\Omega \in C^{0,1}$ voraus. Da die Poisson-Gleichung $-\Delta u = f$ für jedes Gebiet Ω , das im Kreis $K_R(0)$ oder auch nur im Streifen $\{x \in \mathbb{R}^n : |x_1| < R\}$ enthalten ist, lösbar ist und die Ungleichung $|u|_1 \leq C_R |f|_{-1}$ ergibt, mag man vermuten, dass ein ähnliches Resultat für das Stokes-Problem zutrifft.

Abb. 12.2.1. Gebiet Ω_ε

Beispiel 12.2.16 (Gegenbeispiel) Für $\varepsilon \in (0, 1)$ sei das Gebiet Ω_ε durch $\{(x, y) : -1 < x < 1, 0 < y < \varepsilon + (1 - \varepsilon)|x|\}$ beschrieben (vgl. Abbildung 12.2.1). Alle $\Omega_\varepsilon \in C^{0,1}$ liegen im Rechteck $(-1, 1) \times (0, 1)$. Trotzdem gibt es kein $\beta > 0$, sodass (12.2.21'') für alle $\varepsilon > 0$, $w \in L^2(\Omega_\varepsilon)$ gilt.

Beweis. Wir wählen $w \in L^2_0(\Omega_\varepsilon)$ als $w(x, y) = 1$ für $x > 0$, $w(x, y) = -1$ für $x \leq 0$. Ungleichung (12.2.21'') gelte für Ω_ε mit $\beta_\varepsilon > 0$. $u \in \mathbf{H}^1_0(\Omega_\varepsilon)$ mit $|u|_1 \leq |w|_0 / \beta_\varepsilon$ sei gemäß Lemma 12.2.13 gewählt. Wir setzen $u = 0$ auf \mathbb{R}^2 fort. Für die Beschränkung auf $x = 0$ gilt nach Satz 6.2.29

$$\|u_1(0, \cdot)\|_{L^2(\mathbb{R})} \leq \|u_1(0, \cdot)\|_{H^{1/2}(\mathbb{R})} \leq C |u|_1 \leq C |w|_0 / \beta_\varepsilon \leq 2C / \beta_\varepsilon.$$

Sei $\chi(y) = 1$ für $0 < y < \varepsilon$ und $\chi(y) = 0$ sonst. Da $u_1(0, y) = u_1(0, y)\chi(y)$ und $|\chi|_0 = \sqrt{\varepsilon}$, hat man

$$\left| \int_0^\varepsilon u_1(0, y) dy \right| = \left| \int_{\mathbb{R}} u_1(0, y) \chi(y) dy \right| \leq |u_1(0, \cdot)|_0 |\chi|_0 \leq 2C \sqrt{\varepsilon} / \beta_\varepsilon.$$

Seien $\Omega_\varepsilon^+ = \{(x, y) \in \Omega_\varepsilon : x > 0\}$ und $\gamma := \{(x, y) : x = 0, 0 < y < \varepsilon\} = \partial\Omega_\varepsilon^+ \setminus \partial\Omega_\varepsilon$. Wegen $w = 1$ in Ω_ε^+ und auf Grund der Eigenschaft $\operatorname{div} u = w$ ist

$$\frac{1}{2} \leq \int_{\Omega_\varepsilon^+} |w(\mathbf{x})|^2 d\mathbf{x} = \int_{\Omega_\varepsilon^+} w \operatorname{div} u d\mathbf{x} = \int_{\Omega_\varepsilon^+} \operatorname{div} u d\mathbf{x} = \int_\gamma \langle u, \vec{n} \rangle d\Gamma = - \int_\gamma u_1 d\Gamma = - \int_0^\varepsilon u_1(0, y) dy.$$

Die beiden letzten Ungleichungen ergeben $1/2 \leq 2C \sqrt{\varepsilon} / \beta_\varepsilon$, woraus man schließt, dass β_ε nicht unabhängig von ε nach unten durch ein $\beta_0 > 0$ beschränkt sein kann. ■

Übungsaufgabe 12.2.17 Man konstruiere ein im Streifen $\mathbb{R} \times (0, 1)$ gelegenes Gebiet Ω , in dem die Stokes-Gleichungen nicht lösbar sind. Hinweis: Man setze die Gebiete $\Omega_{1/\nu}$ ($\nu \in \mathbb{N}$) aus Abbildung 12.2.1 aneinander.

Ebenso wie für skalare Differentialgleichungen erhält man höhere Regularität der Stokes-Lösung u, p , falls mehr als $f \in \mathbf{H}^{-1}(\Omega)$ vorausgesetzt wird.

Satz 12.2.18 Ω sei beschränkt und hinreichend glatt. u und p seien die (schwache) Lösung des Stokes-Problems (12.2.22) mit $f \in H^k(\Omega)$, $g \in H^{k+1}(\Omega) \cap L^2_0(\Omega)$ für ein $k \in \mathbb{N}_0$. Dann gilt $u \in H^{k+2}(\Omega) \cap H^1_0(\Omega)$, $p \in H^{k+1}(\Omega) \cap L^2_0(\Omega)$, und es gibt ein nur von Ω abhängiges C , sodass

$$|u|_{k+2} + |p|_{k+1} \leq C [|f|_k + |g|_{k+1}]. \quad (12.2.23b)$$

Beweis. Vgl. Ladyženskaja [55, Kap. III, §5]. ■

In Analogie zu Satz 9.1.26 reicht es aus, die Konvexität von Ω vorauszusetzen, um $u \in H^2(\Omega)$ und $p \in H^1(\Omega)$ aus $f \in L^2(\Omega)$ zu erhalten.

Satz 12.2.19 (Kellogg-Osborn [50]) $\Omega \subset \mathbb{R}^2$ sei ein konvexes Polygongebiet. Ist $f \in L^2(\Omega)$, so besitzt die Stokes-Gleichung (12.1.2a,b) eine eindeutige Lösung $u \in \mathbf{H}^2(\Omega) \cap \mathbf{H}^1_0(\Omega)$, $p \in H^1(\Omega) \cap L^2_0(\Omega)$, die der Abschätzung

$$|u|_2 + |p|_1 \leq C |f|_0$$

genügen. Für das allgemeinere Problem (12.2.22) mit $g \neq 0$ erfüllt die Lösung

$$|u|_2 + |p|_1 \leq C \left[\|f\|_{L^2(\Omega)} + \|g\|_{\mathbf{H}^1_\delta(\Omega)} \right],$$

wenn $f \in L^2(\Omega)$ und $g \in L^2_0(\Omega) \cap H^1_\delta(\Omega)$. Dabei ist $H^1_\delta(\Omega)$ der Unterraum von $H^1(\Omega)$ mit der folgenden (stärkeren) Norm:

$$\|g\|_{H^1_\delta(\Omega)} := \sqrt{\sum_{|\alpha|=1} \|D^\alpha g\|_{L^2(\Omega)}^2 + \|\delta^{-1} g\|_{L^2(\Omega)}^2} \quad \text{mit} \quad \delta(\mathbf{x}) := \min\{|\mathbf{x} - \mathbf{e}| : \mathbf{e} \in \Gamma \text{ Ecken des Polygons } \Omega\}.$$

12.2.5 Eine V_0 -elliptische Variationsformulierung der Stokes-Gleichung

$V_0 \subset \mathbf{H}_0^1(\Omega)$ ist in (12.2.15) durch $V_0 := \{u \in \mathbf{H}_0^1(\Omega) : \operatorname{div} u = 0\}$ definiert worden. Als Kern der Abbildung $B^* = -\operatorname{div} \in L(\mathbf{H}_0^1(\Omega), L_0^2(\Omega))$ ist V_0 ein abgeschlossener Unterraum von $\mathbf{H}_0^1(\Omega)$, also wieder ein Hilbert-Raum zur gleichen Norm $|\cdot|_1$. Im Folgenden untersuchen wir die Aufgabe

$$\text{suche } u \in V_0 \quad \text{mit} \quad a(u, v) = f(v) \quad \text{für alle } v \in V_0, \quad (12.2.24)$$

wobei $a(u, v) := \int_{\Omega} \langle \nabla u(\mathbf{x}), \nabla v(\mathbf{x}) \rangle d\mathbf{x}$. Aufgabe (12.2.24) hat die gleiche Form wie die schwache Formulierung der Poisson-Gleichungen $-\Delta u_i = f_i$ ($i = 1, \dots, n$), nur ist hier $\mathbf{H}_0^1(\Omega)$ durch V_0 ersetzt.

Lemma 12.2.20 Ω sei ein beschränktes Gebiet (oder sei in einer Richtung beschränkt; vgl. Übungsaufgabe 6.2.13b). Die Form $a(\cdot, \cdot)$ ist V_0 -elliptisch. Die Konstante $C_E > 0$ in $a(u, u) \geq C_E |u|_1^2$ hängt nur vom Durchmesser von Ω ab. Insbesondere hat Aufgabe (12.2.24) für $f \in V_0'$ eine eindeutige Lösung $u \in V_0$ mit $|u|_1 \leq C_E^{-1} |f|_{V_0'}$.

Beweis. Die $H_0^1(\Omega)$ -Elliptizität von $a(\cdot, \cdot)$ (vgl. Lemma 12.2.12) überträgt sich auf $V_0 \subset H_0^1(\Omega)$ (vgl. Übungsaufgabe 6.5.6a). Hieraus ergeben sich die übrigen Aussagen (vgl. Satz 6.5.9). ■

Satz 12.2.21 $\Omega \in C^{0,1}$ sei ein beschränktes Gebiet. Es gelte $f \in \mathbf{H}^{-1}(\Omega)$. Dann stimmt die Lösung $u \in V_0$ der Aufgabe (12.2.24) mit der Lösungskomponente u der gemischten Formulierung (12.1.3) überein.

Beweis. V_0 und $V_{\perp} = (V_0)^{\perp}$ seien wie in (12.2.16a). Gemäß Übungsaufgabe 12.2.6 lässt sich $f \in \mathbf{H}^{-1}(\Omega)$ so aufspalten, dass

$$f = f_0 + f_{\perp}, \quad f_0 \in V_0', \quad f_{\perp} \in V', \quad f_0(v_{\perp}) = 0 \quad \text{für } v_{\perp} \in V_{\perp}, \quad f_{\perp}(v) = 0 \quad \text{für } v \in V_0.$$

In (12.2.24) kann $f(v)$ durch $f_0(v)$ ersetzt werden. $u \in V_0 \subset \mathbf{H}_0^1(\Omega)$ ergibt $-\Delta u \in \mathbf{H}^{-1}(\Omega)$. Der zu V_{\perp}' gehörende Teil von $-\Delta u$ ist $g_{\perp} \in V_{\perp}'$ mit

$$g_{\perp}(v_{\perp}) := a(u, v_{\perp}) \quad \text{für alle } v_{\perp} \in V_{\perp} = (V_0)^{\perp}.$$

Satz 12.2.14 beweist die Bedingung (12.2.21'), die die Bijektivität von $B : L_0^2(\Omega) \rightarrow V_{\perp}'$ für $B = \nabla$ nach sich zieht (vgl. Lemma 12.2.13). $p := B^{-1}(f_{\perp} - g_{\perp})$ erfüllt definitionsgemäß

$$b(p, v_{\perp}) = \langle Bp, v_{\perp} \rangle_{\mathbf{H}^{-1}(\Omega) \times \mathbf{H}_0^{-1}(\Omega)} = f_{\perp}(v_{\perp}) - g_{\perp}(v_{\perp}) = f_{\perp}(v_{\perp}) - a(u, v_{\perp}) \quad \text{für alle } v_{\perp} \in V_{\perp}.$$

Da $\operatorname{div} v_0 = 0$ für $v_0 \in V_0$, folgt ferner

$$b(p, v_0) = 0 \quad \text{für alle } v_0 \in V_0.$$

Für beliebiges $v \in \mathbf{H}_0^1(\Omega)$, das in $v = v_0 + v_{\perp}$ mit $v_0 \in V_0$ und $v_{\perp} \in V_{\perp}$ zerlegt sei, hat man wegen (12.2.24):

$$\begin{aligned} b(p, v) &= b(p, v_0) + b(p, v_{\perp}) = f_{\perp}(v_{\perp}) - a(u, v_{\perp}) = f_{\perp}(v_{\perp}) - a(u, v) + a(u, v_0) \\ &= f(v_{\perp}) - a(u, v) + f(v_0) = f(v) - a(u, v). \end{aligned}$$

Da auch $b(w, u) = 0$ für alle $w \in L_0^2(\Omega)$ wegen $u \in V_0$, erfüllen u und p die Variationsformulierung (12.2.3a-c) des Stokes-Problems. ■

Man beachte, dass Aufgabe (12.2.24) für alle beschränkten Gebiete lösbar ist, obwohl Aufgabe (12.2.3a-c) empfindlicher von Ω abhängt (vgl. Beispiel 12.2.16). Satz 12.2.21 zeigt, dass nur die Komponente p eine gebietsabhängige Schranke $|p|_0 \leq C_{\Omega} |f|_{-1}$ besitzt, während $|u|_1 \leq C |f_0|_{-1} \leq C |f|_{-1}$ für alle $\Omega \subset K_R(0)$ gilt.

Die Variationsaufgabe (12.2.24) ist insofern nicht äquivalent zum Stokes-Problem, als zum Beispiel für Ω aus Übungsaufgabe 12.2.17 die Stokes-Gleichungen nicht lösbar sind, wogegen Aufgabe (12.2.24) sehr wohl eine Lösung besitzt.

Die ursprüngliche Formulierung (12.2.3a-c) kann aufgefasst werden als die Gleichung (12.2.24), bei der man die "Nebenbedingung" $\operatorname{div} u = 0$ mit Hilfe der Lagrange-Funktion p einführt (vgl. Abschnitt 8.3.6).

12.3 Gemischte Finite-Element-Methode für das Stokes-Problem

12.3.1 Finite-Element-Diskretisierung eines Sattelpunktproblems

Eine übliche Ritz-Galerkin-Diskretisierung läge vor, wenn man in der Variationsformulierung (12.2.24) den Raum V_0 durch einen endlich-dimensionalen Unterraum $V_h \subset V_0$ ersetzen würde. Dies ist aber nicht so einfach realisierbar.

Übungsaufgabe 12.3.1 Das Quadrat $\Omega = (0, 1) \times (0, 1)$ sei wie in den Abbildungen 8.3.2, 8.3.5a regelmäßig trianguliert bzw. in Gitterquadrate zerlegt. $V_h^{(1)} \subset H_0^1(\Omega)$ sei der Raum der stückweise linearen Dreieckselemente (vgl. (8.3.8)) bzw. der bilinearen Elemente (vgl. (8.3.12b)). Den entsprechenden Unterraum für das Stokes-Problem definiere man als $V_h := \{u = (u_1, u_2) : u_1, u_2 \in V_h^{(1)} \text{ und } \operatorname{div} u = 0\} \subset V_0$. Man zeige: V_h enthält nur die Nullfunktion.

Daher orientiert sich das weitere Vorgehen an der schwachen Formulierung (12.2.3a-c). Der Raum $X = V \times W$ wird durch $X_h = V_h \times W_h$ ersetzt. Das diskrete Problem

$$\text{suche } v^h \in V_h \text{ und } w^h \in W_h \text{ mit } \begin{cases} a(v^h, x) + b(w^h, x) = f_1(x) & \text{für alle } x \in V_h, \\ b(y, v^h) = f_2(y) & \text{für alle } y \in W_h. \end{cases} \quad (12.3.1)$$

nennt man ein *gemischtes Ritz-Galerkin-Problem* [bzw. eine *gemischte Finite-Element-Aufgabe*, wenn V_h und W_h aus finiten Elementen gebildet werden]. Der Formulierung (12.2.6a,b) entspricht die zu (12.3.1) äquivalente Schreibweise:

$$\text{suche } x^h \in X_h \text{ mit } c(x^h, z) = f(z) \text{ für alle } z \in X_h. \quad (12.3.1')$$

Im Falle der Stokes-Gleichungen liefert (12.2.3a-c) die gewünschte Lösung, die die Nebenbedingung $\operatorname{div} u = 0$ erfüllt. Die Finite-Element-Lösung von (12.3.1) genügt dagegen der Bedingung $\operatorname{div} u^h = 0$ im Allgemeinen nicht. Man kann v^h aus (12.3.1) als Lösung einer nichtkonformen Finite-Element-Diskretisierung von (12.2.24) ansehen, wie die folgende Übung zeigt.

Übungsaufgabe 12.3.2 Sei $f_2 = 0$ in der unteren Gleichung von (12.3.1). Ferner sei $V_{0,h} := \{x \in V_h : b(y, x) = 0 \text{ für alle } y \in W\}$ gesetzt. Man zeige: Jede Lösung v^h aus (12.3.1) ist auch Lösung von

$$\text{suche } v^h \in V_{0,h} \text{ mit } a(v^h, x) = f_1(x) \text{ für alle } x \in V_{0,h}. \quad (12.3.2)$$

Da im Allgemeinen $V_{0,h} \not\subset V_0$ (vgl. (12.2.15)), ist (12.3.2) eine nichtkonforme Diskretisierung von (12.2.24).

Seien $\{b_1^V, \dots, b_{N_{V,h}}^V\}$ und $\{b_1^W, \dots, b_{N_{W,h}}^W\}$ geeignete Basen von V_h und W_h , wobei

$$N_{V,h} := \dim V_h, \quad N_{W,h} := \dim W_h. \quad (12.3.3a)$$

Die Koeffizienten von $v \in V_h$ und $w \in W_h$ seien \mathbf{v} und \mathbf{w} :

$$v = P_V \mathbf{v} := \sum_{i=1}^{N_{V,h}} v_i b_i^V, \quad w = P_W \mathbf{w} := \sum_{i=1}^{N_{W,h}} w_i b_i^W. \quad (12.3.3b)$$

Wie in Abschnitt 8.1 beweist man den

Satz 12.3.3 Das Variationsproblem (12.3.1) ist zum Gleichungssystem

$$\begin{bmatrix} \mathbf{A}_h & \mathbf{B}_h \\ \mathbf{B}_h^\top & 0 \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ \mathbf{w} \end{bmatrix} = \begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \end{bmatrix} \quad (12.3.4a)$$

äquivalent, wobei die Matrizen und Vektoren durch

$$\begin{aligned} \mathbf{A}_{h,ij} &:= a(b_j^V, b_i^V), & \mathbf{B}_{h,ik} &:= b(b_k^W, b_i^V) & \text{für } 1 \leq i, j \leq N_{V,h}, 1 \leq k \leq N_{W,h}, \\ \mathbf{f}_{1,i} &:= f_1(b_i^V), & \mathbf{f}_{2,k} &:= f_2(b_k^W), & \text{für } 1 \leq i, j \leq N_{V,h}, 1 \leq k \leq N_{W,h} \end{aligned} \quad (12.3.4b)$$

gegeben sind. Der Zusammenhang zwischen (12.3.1) und (12.3.4a-b) ist durch $v^h = P_V \mathbf{v}$, $w^h = P_W \mathbf{w}$ beschrieben. Für $\mathbf{u} := \begin{pmatrix} \mathbf{v} \\ \mathbf{w} \end{pmatrix}$, $\mathbf{f} := \begin{pmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \end{pmatrix}$ erhält man das Gleichungssystem

$$\mathbf{C}_h \mathbf{u} = \mathbf{f} \quad \text{mit } \mathbf{C}_h := \begin{bmatrix} \mathbf{A}_h & \mathbf{B}_h \\ \mathbf{B}_h^\top & 0 \end{bmatrix}, \quad (12.3.4a')$$

das der Formulierung (12.3.1') entspricht.

12.3.2 Stabilitätsbedingungen

Bei der Wahl der Unterräume V_h und W_h muss man vorsichtig vorgehen, denn auch scheinbar vernünftige Räume führen zu singulären oder aber instabilen Gleichungssystemen (12.3.1). Der erste Fall liegt beim folgenden Beispiel vor.

Beispiel 12.3.4 Das Stokes-Problem sei für das L -Gebiet aus Abbildung 8.3.2 vorgelegt. Für alle drei Komponenten u_1, u_2 und p verwenden wir stückweise lineare Dreieckselemente über der Triangulation \mathcal{T} , die durch das zweite oder dritte Bild der Abbildung 8.3.2 gegeben sei. Dabei gelte $u_1 = u_2 = 0$ auf Γ , während $p \in W_h$ die Nebenbedingung $p \in L_0^2(\Omega)$, d.h. $\int_\Omega p dx = 0$, erfülle. Dann ist das diskrete Variationsproblem (12.3.1) nicht lösbar.

Beweis. Die zweite (dritte) Triangulation aus Abbildung 8.3.2 hat 5 (10) innere Knoten \mathbf{x}^i , die je einen Wert $u_1(\mathbf{x}^i)$ und $u_2(\mathbf{x}^i)$ tragen. Daher gilt $\dim V_h = 2 \cdot 5 = 10$ ($\dim V_h = 20$). Wegen der Nebenbedingung $\int_\Omega p dx = 0$ ist die Dimension von W_h um eins kleiner als die Anzahl der inneren und Randknoten: $\dim W_h = 21 - 1 = 20$ ($\dim W_h = 21$). In beiden Fällen ergibt sich die Behauptung aus dem folgenden Lemma. ■

Lemma 12.3.5 Notwendig für die Lösbarkeit von (12.3.1) ist $N_{V,h} \geq N_{W,h}$, d.h. $\dim V_h \geq \dim W_h$.

Beweis. Gemäß Satz 12.3.3 ist die Lösbarkeit von (12.3.1) äquivalent zur Nichtsingularität der Matrix \mathbf{C}_h aus (12.3.4a'). Elementare Überlegungen zeigen $\text{Rang}(\mathbf{C}_h) \leq N_{V,h} + \text{Rang}(\mathbf{B}_h)$. Da \mathbf{B}_h eine $N_{V,h} \times N_{W,h}$ -Matrix ist, folgt aus $N_{V,h} < N_{W,h}$, dass $\text{Rang}(\mathbf{C}_h) < N_{V,h} + N_{W,h}$ und damit \mathbf{C}_h singulär ist. ■

Eine Erhöhung der Dimension von W_h führt daher nicht zu einer immer besseren Approximation von $w \in W$. Die Wahl von V_h und W_h müssen aufeinander abgestimmt sein. Die Ungleichung $\dim V_h \geq \dim W_h$ entspricht der Forderung, dass B aus (12.2.10b) zwar injektiv, aber nicht notwendig surjektiv zu sein braucht.

Um die erforderlichen Stabilitätsbedingungen zu formulieren, definieren wir

$$V_{0,h} := \{v \in V_h : b(y, v) = 0 \text{ für alle } y \in W_h\} \quad (12.3.5)$$

(vgl. Übungsaufgabe 12.3.2). $V_{0,h}$ ist das diskrete Analogon des Raumes V_0 aus (12.2.15). Die auf Brezzi [18] zurückgehenden Bedingungen lauten:

$$\inf_{v_0 \in V_{0,h}} \sup_{x_0 \in V_{0,h}} \frac{|a(v_0, x_0)|}{\|v_0\|_V \|x_0\|_V} \geq \alpha_h > 0, \quad (12.3.6a)$$

$$\inf_{w \in W_h} \sup_{x \in V_h} \frac{|b(w, x)|}{\|w\|_W \|x\|_V} \geq \beta_h > 0. \quad (12.3.6b)$$

Satz 12.3.6 Es gelte (12.2.4a-c) und $\dim V_h < \infty$. Die Brezzi-Bedingungen (12.3.6a-b) sind hinreichend und notwendig für die Lösbarkeit des diskreten Problems (12.3.1). Die Lösung $u^h = (v^h, w^h) \in V_h \times W_h$ erfüllt

$$\|u^h\|_X := \sqrt{\|v^h\|_V^2 + \|w^h\|_W^2} \leq C_h \|f\|_{X'} := C_h \sqrt{\|f_1\|_{V'}^2 + \|f_2\|_{W'}^2}, \quad (12.3.7)$$

wobei C_h von α_h, β_h und den Schranken C_a und C_b von $|a(v, x)| \leq C_a \|v\|_V \|x\|_V$ und $|b(w, x)| \leq C_b \|w\|_W \|x\|_V$ abhängt. Wenn

$$\alpha_h \geq \underline{\alpha} > 0, \quad \beta_h \geq \underline{\beta} > 0 \quad (12.3.6c)$$

für alle Parameter h einer Folge von Diskretisierungen gilt, heißt die Diskretisierung stabil und C_h bleibt beschränkt: $C_h \leq \overline{C}$ für alle h .

Beweis. Satz 12.2.7 und Lemma 12.2.9 sind anwendbar mit $V_{0,h}$, V_h , W_h anstelle von V_0 , V , W . (12.2.19a,c) stimmen dann mit (12.3.6a,b) überein, während (12.2.19b) wegen $\dim V_{0,h} < \infty$ aus (12.2.19a) folgt (vgl. Übungsaufgabe 6.5.4). ■

Die Bedingung (12.3.6a) ist für das Stokes-Problem trivial:

Übungsaufgabe 12.3.7 *Man zeige: Bedingung (12.3.6a) ist stets mit einer von h unabhängigen Konstante α_h erfüllt, wenn $a(\cdot, \cdot) : V \times V \rightarrow \mathbb{R}$ V -elliptisch ist.*

12.3.3 Stabile Finite-Element-Räume für das Stokes-Problem

12.3.3.1 Stabilitätskriterium

Für das Stokes-Problem muss $V_h \subset H_0^1(\Omega)$, $W_h \subset L_0^2(\Omega)$ gelten. Im beschränkten Gebiet ist $a(\cdot, \cdot)$ $H_0^1(\Omega)$ -elliptisch, sodass (12.3.6a) mit $\alpha_h \geq C_E > 0$ erfüllt ist. Schwieriger ist es, die Bedingungen (12.3.6b,c) nachzuprüfen, die für das Stokes-Problem die Form

$$\sup_{u \in V_h \text{ mit } |u|_1=1} |b(p, u)| \geq \underline{\beta} |p|_0 \quad \text{für alle } p \in W_h \quad (12.3.6d)$$

annehmen, worin $\beta > 0$ von p und h unabhängig sein muss. Einfacher lässt sich die modifizierte Bedingung

$$\sup_{u \in V_h \text{ mit } |u|_0=1} |b(p, u)| \geq \tilde{\beta} |p|_1 \quad \text{für alle } p \in W_h \quad (12.3.8)$$

($\tilde{\beta} > 0$ unabhängig von p und h) nachweisen, bei der die Rollen der $|\cdot|_0$ - und $|\cdot|_1$ -Normen vertauscht sind. Da in (12.3.8) die Norm $|p|_1$ auftritt, erfordert diese Bedingung einen Finite-Element-Raum $W_h \subset H^1(\Omega)$. Damit sind insbesondere stückweise konstante finite Elemente ausgeschlossen.

Zusammen mit üblichen Voraussetzungen ist die Bedingung (12.3.8) hinreichend für die Stabilität.

Satz 12.3.8 $\Omega \in C^{0,1}$ sei beschränkt. Das Poisson-Problem sei H^2 -regulär (hinreichend ist, dass Ω konvex ist; vgl. Satz 9.1.26). V_h erfülle die Approximationseigenschaft

$$\inf_{u^h \in V_h} |u - u^h|_1 \leq C_A h |u|_2 \quad \text{für alle } u \in \mathbf{H}^2(\Omega) \cap \mathbf{H}_0^1(\Omega) \quad (12.3.9a)$$

und die inverse Abschätzung

$$|u^h|_1 \leq C_1 h^{-1} |u^h|_0 \quad \text{für alle } u^h \in V_h. \quad (12.3.9b)$$

Dann ist Bedingung (12.3.8) hinreichend für die Brezzi-Bedingung (12.3.6b,c).

Beweis. 1) Zu gegebenem $p \in W_h$ gibt es $u \in \mathbf{H}_0^1(\Omega)$ mit $|u|_1 = 1$ und $b(p, u) \geq \beta |p|_0$ (vgl. Satz 12.2.14 und (12.2.21)). Gemäß Übungsaufgabe 9.1.28 erfüllt die orthogonale $\mathbf{H}_0^1(\Omega)$ -Projektion u^h von u auf V_h die Bedingungen

$$u = u^h + e, \quad |u^h|_1 \leq |u|_1 = 1, \quad |e|_0 \leq Ch |u|_1 = Ch.$$

Aus $b(p, u^h) = b(p, u) - b(p, e) \geq \beta |p|_0 - b(p, e) \geq \beta |p|_0 - |p|_1 |e|_0 \geq \beta |p|_0 - Ch |p|_1$ und $|u^h|_1 \leq 1$ schließt man

$$\sup_{v^h \in V_h \text{ mit } |v^h|_1=1} |b(p, v^h)| \geq \beta |p|_0 - Ch |p|_1. \quad (12.3.10)$$

2) Wegen (12.3.8) und (12.3.9b) gibt es ein $u^* \in V_h$ mit $|u^*|_0 = 1$ und

$$|b(p, u^*)| \geq \tilde{\beta} |p|_1 = \tilde{\beta} |p|_1 |u^*|_0 \geq (\tilde{\beta}/C_1) h |p|_1 |u^*|_1.$$

Hieraus folgt

$$\sup_{v^h \in V_h \text{ mit } |v^h|_1=1} |b(p, v^h)| \geq \hat{\beta} h |p|_1 \quad \text{mit } \hat{\beta} := \tilde{\beta}/C_1. \quad (12.3.11)$$

Multipliziert man (12.3.10) mit $\hat{\beta}/(C + \hat{\beta})$ und (12.3.11) mit $C/(C + \hat{\beta})$, so lautet die Summe

$$\sup_{v^h \in V_h \text{ mit } |v^h|_1=1} |b(p, v^h)| \geq \underline{\beta} := \hat{\beta} \tilde{\beta} / (C + \hat{\beta}). \quad (12.3.12)$$

Da $\underline{\beta}$ unabhängig von p und h ist, ist (12.3.6d) [d.h. (12.3.6b,c)] bewiesen. ■

12.3.3.2 Finite-Element-Diskretisierungen mit der Blasenfunktion

Sei im Folgenden Ω ein Polygonebiet und \mathcal{T}_h eine zulässige Triangulation. Beispiel 12.3.4 zeigt, dass lineare Dreieckselemente für u und p nicht sinnvoll sind. Wir vergrößern die Dimension von V_h , indem wir zu V_h die sogenannten Blasenfunktionen (englisch “bubble functions” oder “bulb functions”) hinzunehmen.

Auf dem Referenzdreieck $T = \{(\xi, \eta) : \xi, \eta > 0, \xi + \eta < 1\}$ ist die Blasenfunktion durch

$$u(\xi, \eta) := \xi\eta(1 - \xi - \eta) \quad \text{in } T, \quad u = 0 \quad \text{außerhalb,} \quad (12.3.13a)$$

definiert. Der Name leitet sich daraus ab, dass u nur in T positiv ist und auf ∂T sowie außerhalb verschwindet. Die Transformation $\phi : T \rightarrow \tilde{T}$ auf ein allgemeines $\tilde{T} \in \mathcal{T}_h$ (vgl. Übungsaufgabe 8.3.14) ergibt die Darstellung

$$\tilde{u}_{\tilde{T}}(x, y) := u(\phi^{-1}(x, y)) \quad (12.3.13b)$$

der Blasenfunktion auf \tilde{T} .

Übungsaufgabe 12.3.9 \mathcal{T}_h sei eine uniforme Triangulation. Man zeige: Es gibt ein von h unabhängiges $C > 0$, sodass $\int_{\tilde{T}} \tilde{u}_{\tilde{T}}(x, y) dx dy \geq C \int_{\tilde{T}} dx dy$ für alle $\tilde{T} \in \mathcal{T}_h$.

Wir setzen

$$\begin{aligned} V_h^1 &: \text{Linearkombination der linearen Elemente in } H_0^1(\Omega) \text{ und der Blasenfunktionen zu allen } \tilde{T} \in \mathcal{T}_h, \\ V_h &:= V_h^1 \times V_h^1, \quad W_h : \text{lineare Elemente in } L_0^2(\Omega). \end{aligned} \quad (12.3.14)$$

Bezüglich der Nebenbedingung $W_h \subset L_0^2(\Omega)$ sei auf Abschnitt 8.3.6 verwiesen. Da $\tilde{u}_{\tilde{T}} \in H_0^1(\Omega)$, ist $V_h \subset \mathbf{H}_0^1(\Omega)$.

Satz 12.3.10 \mathcal{T}_h sei quasiuniforme Triangulation auf einem beschränkten Polygonebiet Ω . V_h und W_h seien durch (12.3.14) gegeben. Dann ist die Stabilitätsbedingung (12.3.8) erfüllt. Unter den weiteren Voraussetzungen $\Omega \in C^{0,1}$ und “Poisson-Problem $H^2(\Omega)$ -regulär” gilt auch die Brezzi-Bedingung (12.3.6b,c).

Beweis. 1) Sei $p \in W_h$ beliebig. Auf jedem $\tilde{T} \in \mathcal{T}_h$ ist ∇p konstant: $\nabla p = (p_x|_{\tilde{T}}, p_y|_{\tilde{T}})$. Wir setzen

$$v := \sum_{\tilde{T} \in \mathcal{T}_h} \begin{pmatrix} p_x|_{\tilde{T}} \tilde{u}_{\tilde{T}} \\ p_y|_{\tilde{T}} \tilde{u}_{\tilde{T}} \end{pmatrix} \in V_h, \quad \tilde{v} := \frac{1}{|v|_0} v \quad (\tilde{u}_{\tilde{T}} \text{ Blasenfunktion (12.3.13b) auf } \tilde{T}),$$

sodass $|\tilde{v}|_0 = 1$. Übungsaufgabe 12.3.9 liefert

$$b(p, v) = \int_{\Omega} \langle \nabla p, v \rangle dx = \sum_{\tilde{T} \in \mathcal{T}_h} \left(|p_x|_{\tilde{T}}^2 + |p_y|_{\tilde{T}}^2 \right) \int_{\tilde{T}} \tilde{u}_{\tilde{T}} dx dy \geq C \sum_{\tilde{T} \in \mathcal{T}_h} \int_{\tilde{T}} \left(|p_x|_{\tilde{T}}^2 + |p_y|_{\tilde{T}}^2 \right) dx dy = C |\nabla p|_0^2.$$

Da $|\nabla p|_0$ und $|p|_1$ auf dem Unterraum $H^1(\Omega) \cap L_0^2(\Omega)$ äquivalente Normen sind, folgt $|b(p, v)| \geq C' |p|_1 |\nabla p|_0$ und $|b(p, \tilde{v})| \geq C' |p|_1 |\nabla p|_0 / |v|_0$. Ähnlich zeigt man $|v|_0 \leq C'' |\nabla p|_0$ und erhält $|b(p, \tilde{v})| \geq \tilde{\beta} |p|_1$ mit $\tilde{\beta} := C'/C''$ unabhängig von h . Die linke Seite in (12.3.8) ist $\geq |b(p, \tilde{v})|$, sodass (12.3.8) folgt.

2) (12.3.9a) ist für V_h erfüllt (vgl. Satz 8.4.5). Ebenso gilt die inverse Abschätzung (12.3.9b) (vgl. Satz 8.7.5). Daher folgt (12.3.6b,c) aus Satz 12.3.8. ■

Ein allgemeines Resultat zur Stabilisierung durch Blasenfunktionen findet sich bei Brezzi-Pitkäranta [20].

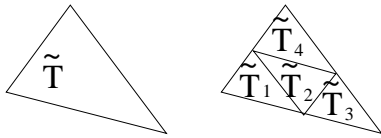
12.3.3.3 Stabile Diskretisierungen mit linearen Elementen in V_h 

Abb. 12.3.1. Triangulationen \mathcal{T}_h und $\mathcal{T}_{h/2}$

Will man ohne Blasenfunktionen auskommen, hat man die Dimension von V_h auf andere Weise zu erhöhen. In diesem Abschnitt werden wir für V_h und W_h zwei verschiedene Triangulationen $\mathcal{T}_{h/2}$ und \mathcal{T}_h betrachten. Indem jedes $\tilde{T} \in \mathcal{T}_h$ wie in Abbildung 12.3.1 durch Halbierung der Seiten in vier ähnliche Dreiecke zerlegt wird, erhält man $\mathcal{T}_{h/2}$. Wir definieren:

$$V_h \subset \mathbf{H}_0^1(\Omega): \text{lineare Elemente zur Triangulation } \mathcal{T}_{h/2}, \quad (12.3.15)$$

$$W_h \subset H^1(\Omega) \cap L_0^2(\Omega): \text{lineare Elemente zur Triangulation } \mathcal{T}_h,$$

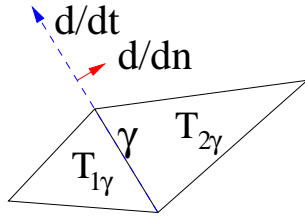
oder

$$\begin{aligned} V_h &\subset \mathbf{H}_0^1(\Omega): \text{quadratische Elemente zur Triangulation } \mathcal{T}_h, \\ W_h &\subset H^1(\Omega) \cap L_0^2(\Omega): \text{lineare Elemente zur Triangulation } \mathcal{T}_h. \end{aligned} \quad (12.3.16)$$

Satz 12.3.11 Satz 12.3.10 gilt sinngemäß für $V_h \times W_h$ aus (12.3.15) oder (12.3.16).

Beweis. 1) $V_h \times W_h$ sei durch (12.3.15) gegeben. Zu jeder inneren Dreiecksseite γ der Triangulation \mathcal{T}_h gibt es zwei Dreiecke $T_{1\gamma}, T_{2\gamma} \in \mathcal{T}_h$ mit $\gamma = \overline{T_{1\gamma}} \cap \overline{T_{2\gamma}}$ (vgl. Abbildung 12.3.2). $\partial/\partial t$ sei die Ableitung in Richtung der Strecke γ , $\partial/\partial n$ sei die dazu senkrechte Richtungsableitung. Es gibt a_γ und b_γ mit

$$a_\gamma^2 + b_\gamma^2 = 1, \quad \frac{\partial}{\partial x} = a_\gamma \frac{\partial}{\partial n} + b_\gamma \frac{\partial}{\partial t}, \quad \frac{\partial}{\partial y} = b_\gamma \frac{\partial}{\partial n} - a_\gamma \frac{\partial}{\partial t}.$$



Im Gegensatz zu $\partial p/\partial n$ ist $\partial p/\partial t$ auf $T_{1\gamma} \cup T_{2\gamma} \cup \gamma$ konstant. Wir bezeichnen diesen Wert mit $p_t|_\gamma$. Der Mittelpunkt \mathbf{x}^γ von γ ist ein Knoten von $\mathcal{T}_{h/2}$. Wir definieren die stückweise lineare Funktion u_γ über $\mathcal{T}_{h/2}$ durch die Knotenwerte

$$u_\gamma(\mathbf{x}^\gamma) = p_t|_\gamma, \quad u_\gamma(\mathbf{x}^j) = 0 \quad \text{in den übrigen Knoten} \quad (12.3.17a)$$

und setzen

Abb. 12.3.2. $\gamma, T_{1\gamma}, T_{2\gamma}$

$$v := \sum_{\gamma} \begin{pmatrix} b_\gamma \\ -a_\gamma \end{pmatrix} p_t|_\gamma u_\gamma \in V_h, \quad \tilde{v} := \frac{1}{|v|_0} v. \quad (12.3.17b)$$

Die Summe \sum_{γ} erstreckt sich über alle inneren Seiten von \mathcal{T}_h . In $T_{1\gamma} \cup T_{2\gamma}$ gilt $\left\langle \nabla p, \begin{pmatrix} b_\gamma \\ -a_\gamma \end{pmatrix} \right\rangle p_t|_\gamma = |p_t|_\gamma|^2$, sodass

$$b(p, v) = \sum_{\gamma} \int_{T_{1\gamma} \cup T_{2\gamma}} \left\langle \nabla p, \begin{pmatrix} b_\gamma \\ -a_\gamma \end{pmatrix} \right\rangle p_t|_\gamma u_\gamma \, d\mathbf{x} = \sum_{\gamma} |p_t|_\gamma|^2 \int_{T_{1\gamma} \cup T_{2\gamma}} u_\gamma \, d\mathbf{x} \geq Ch^2 \sum_{\gamma} |p_t|_\gamma|^2.$$

Sind $\gamma_1, \gamma_2, \gamma_3$ die Seiten von $\tilde{T} \in \mathcal{T}_h$ (\mathcal{T}_h quasiuniform!), so ist $\int_{\tilde{T}} |\nabla p|^2 \, d\mathbf{x} \leq C'h^2 \sum_{i=1}^3 |p_t|_{\gamma_i}|^2$. Hieraus schließt man $b(p, v) \geq C'' |p|_1 |\nabla p|_0$ wie im Beweis des Satzes 12.3.10 und führt den Beweis analog weiter.

2) Im Falle (12.3.16) der quadratischen Elemente hat man die gleichen Knoten wie in 1) (vgl. die Abbildungen 8.3.8a und 12.3.1). Man benutze (12.3.17a,b) zur Definition der quadratischen Funktion $v \in V_h$ und setze den Beweis wie in 1) fort. ■

12.3.3.4 Fehlerabschätzungen

Die Bedingung (8.1.17) ist im Folgenden durch die Stabilitätsbedingung (12.3.6a-c) zu ersetzen. An die Stelle der Approximationseigenschaft (8.4.6') treten die Ungleichungen

$$\inf_{v^h \in V_h} |v - v^h|_1 \leq Ch |v|_2 \quad \text{für alle } v \in \mathbf{H}^2(\Omega) \cap \mathbf{H}_0^1(\Omega), \quad (12.3.18a)$$

$$\inf_{p^h \in W_h} |p - p^h|_1 \leq Ch |p|_1 \quad \text{für alle } p \in H^1(\Omega) \cap L_0^2(\Omega). \quad (12.3.18b)$$

Die Bedingung (12.3.18a) ist äquivalent zur Bedingung (12.3.9a) aus Satz 12.3.8.

Der folgende Satz trifft für allgemeine Sattelpunktprobleme zu und kann auf Satz 8.2.1 zurückgeführt werden.

Satz 12.3.12 $u = \begin{pmatrix} v \\ p \end{pmatrix} \in X = V \times W$ sei Lösung von (12.2.5) bzw. (12.3.6a). Das diskrete Problem (12.3.1) mit $X_h = V_h \times W_h \subset X$ erfülle die Brezzi-Bedingung (12.3.6a-c) und habe die Lösung $u^h = (v^h, w^h)$. Dann gibt es ein von h unabhängiges C , sodass

$$\|u - u^h\|_X \leq C \inf_{x^h \in X_h} \|u - x^h\|_1. \quad (12.3.19)$$

Beweis. Die Brezzi-Bedingung (vgl. Satz 12.3.6) liefert $\|u^h\|_X \leq \bar{C} \|f\|_{X'}$ für alle rechten Seiten $f \in X'$ in (12.2.5), insbesondere für alle $f \in X_h = X'_h$. Die vorherige Ungleichung bedeutet $\|L_h^{-1}\|_{X_h \leftarrow X'_h} \leq \bar{C}$ für den Operator $L_h : X_h \rightarrow X'_h$ der zu $c(\cdot, \cdot) : X_h \times X_h \rightarrow \mathbb{R}$ gehört. Nach Übungsaufgabe 8.1.17 ist $\|L_h^{-1}\|_{X_h \leftarrow X'_h} \leq \bar{C}$ äquivalent zur Bedingung (8.1.17) für $c(\cdot, \cdot) : X \times X \rightarrow \mathbb{R}$ [anstelle von $a(\cdot, \cdot) : V \times V \rightarrow \mathbb{R}$] mit $\varepsilon_N = 1/\bar{C}$. Satz 8.2.1 liefert die Behauptung (12.3.19). ■

Für das Stokes-Problem mit den Bezeichnungen $\binom{u}{p}$, $\binom{u^h}{p^h}$ anstelle von $u = \binom{v}{p}$, $u^h = \binom{v^h}{p^h}$ schreibt sich Ungleichung (12.3.19) wie folgt um:

$$|u - u^h|_1^2 + |p - p^h|_0^2 \leq C^2 \inf \{ |u - v^h|_1^2 + |p - q^h|_0^2 : v^h \in V_h, q^h \in W_h \} \quad (12.3.20a)$$

und

$$|u - u^h|_1 + |p - p^h|_0 \leq \sqrt{2}C \inf \{ |u - v^h|_1 + |p - q^h|_0 : v^h \in V_h, q^h \in W_h \}. \quad (12.3.20b)$$

Satz 12.3.13 Die Stokessche Gleichung (12.1.2a,b) habe eine Lösung mit $u \in \mathbf{H}^2(\Omega) \cap \mathbf{H}_0^1(\Omega)$ und $p \in H^1(\Omega) \cap L_0^2(\Omega)$ (vgl. Satz 12.2.19). Für die Unterräume $V_h \subset \mathbf{H}_0^1(\Omega)$ und $W_h \subset L_0^2(\Omega)$ seien die Brezzi-Bedingung (12.3.6a-c) und die Approximationsbedingungen (12.3.18a,b) erfüllt. Dann genügt die diskrete Lösung u^h, p^h der Abschätzung

$$|u - u^h|_1 + |p - p^h|_0 \leq C'h [|u|_2 + |p|_1]. \quad (12.3.21)$$

Beweis. Man kombiniere die Ungleichungen (12.3.20a,b) und (12.3.18a,b). ■

Mit der gleichen Argumentation wie im zweiten Beweis zu Satz 8.4.11 mit $c(\cdot, \cdot)$ anstelle von $a(\cdot, \cdot)$ beweist man den

Satz 12.3.14 Für jedes $f \in \mathbf{L}^2(\Omega)$, $g \in L_0^2(\Omega) \cap H^1(\Omega)$ habe das Stokes-Problem $-\Delta u + \nabla p = f$, $-\operatorname{div} u = g$ eine Lösung $u \in \mathbf{H}^2(\Omega) \cap \mathbf{H}_0^1(\Omega)$, $p \in H^1(\Omega) \cap L_0^2(\Omega)$ mit $|u|_2 + |p|_1 \leq C[|f|_0 + |g|_1]$. Unter den Voraussetzungen des Satzes 12.3.13 gelten dann die Abschätzungen

$$|u - u^h|_0 + |p - p^h|_{-1} \leq C'h [|u|_1 + |p|_0], \quad (12.3.22a)$$

$$|u - u^h|_0 + |p - p^h|_{-1} \leq C''h^2 [|u|_2 + |p|_1]. \quad (12.3.22b)$$

für die Finite-Element-Lösungen. $|p|_{-1}$ ist dabei die Dualnorm zu $H^1(\Omega) \cap L_0^2(\Omega)$.

Zusatz 12.3.15 Als Kombination von (12.3.22b) und (12.3.21) erhält man

$$|u - u^h|_0 + h|p - p^h|_0 \leq C'h^2 [|u|_2 + |p|_1].$$

Eine ausführliche Darstellung der gemischten Finite-Element-Diskretisierung bietet die Monographie Brezzi-Fortin [19].

Literaturverzeichnis

- [1] R. Adams: *Sobolev spaces*. Academic Press, New York, 1975.
- [2] S. Agmon, A. Douglis und L. Nirenberg: *Estimates near the boundary for solutions of elliptic partial differential equations satisfying general boundary conditions II*. Comm. Pure Appl. Math. **17** (1964), 35–92.
- [3] A. Aziz (Hrsg.): *The mathematical foundation of the finite element method with applications to partial differential equations*. Academic Press, New York, 1972.
- [4] I. Babuška: *The theory of small changes in the domain of existence in the theory of partial differential equations and its applications*. Differential Equations and Their Applications (I. Babuška und M. Zlámal, Hrsg.), Czech. Academy of Sciences, 1963. (auch Acad. Press, New York 1963), pp. 13–26 (Prag, Sept. 1962).
- [5] I. Babuška und A. Aziz: *Survey lectures on the mathematical foundations of the finite element method*. In Aziz [3], pp. 3–363.
- [6] I. Babuška und H. Rosenzweig: *A finite element scheme for domains with corners*. Numer. Math. **20** (1972), 1–21.
- [7] R. E. Bank: *A software package for solving elliptic partial differential equations. Users' guide 6.0*. SIAM, Philadelphia, 1990.
- [8] R. E. Bank, T. F. Dupont und H. Yserentant: *The hierarchical basis multigrid method*. Math. Comp. **52** (1988), 427–458.
- [9] R. Becker und R. Rannacher: *An optimal control approach to a posteriori error estimation in finite element methods*. Acta Numerica 2001, Cambridge University Press, 2001, pp. 1–102.
- [10] H. Blum und K. Dobrowolski: *On finite element methods for elliptic equations on domains with corners*. Computing **28** (1982), 53–63.
- [11] H. Blum und R. Rannacher: *On the boundary value problem of the biharmonic operator on domains with angular corners*. Math. Meth. Appl. Sci. **2** (1980), 556–581.
- [12] D. Braess: *Finite Elemente: Theorie, schnelle Löser und Anwendungen in der Elastizitätstheorie*. 2. Auflage. Springer-Verlag, Berlin, 1997.

- [13] D. Braess und W. Dahmen: *Stability estimates of the mortar finite element method for 3-dimensional problems*. East-West J. Numer. Math. **6** (1998), 249–264.
- [14] J. H. Bramble: *A second order finite element analog of the first biharmonic boundary value problem*. Numer. Math. **9** (1966), 236–249.
- [15] J. H. Bramble und B. Hubbard: *A theorem on error estimation for finite difference analogues of the Dirichlet problem for elliptic equations*. Contributions to Diff. Eq. **2** (1963), 319–340.
- [16] ———: *Approximation of solutions of mixed boundary value problems for Poisson's equation by finite differences*. J. Assoc. Comput. Mach. **12** (1965), 114–123.
- [17] J. H. Bramble und A. H. Schatz: *Higher order local accuracy by averaging in the finite element method*. Math. Comp. **31** (1977), 94–111.
- [18] F. Brezzi: *On the existence, uniqueness and approximation of saddle-point problems arising from Lagrangian multipliers*. RAIRO Anal. Numer. **8** (1974), 129–151.
- [19] F. Brezzi und M. Fortin: *Mixed and hybrid finite element methods*. Springer-Verlag, Berlin, 1991.
- [20] F. Brezzi und J. Pitkäranta: *On the stability of the finite element approximations to the Stokes equations*. In Hackbusch [39], pp. 11–29. (Kiel, Jan. 1984).
- [21] F. Chatelin: *Spectral approximation of linear operators*. Academic Press, New York, 1983.
- [22] P. G. Ciarlet: *The finite element method for elliptic problems*. North-Holland, Amsterdam, 1978.
- [23] P. Clément: *Approximation by finite element functions using local regularization*. RAIRO Anal. Numer. **9**, R-2 (1975), 77–84.
- [24] B. Cockburn, G. Karniadakis und C.-W. Shu (Hrsg.): *Discontinuous Galerkin Methods. Theory, Computation and Applications*. Lecture Notes in Computational Science and Engineering 11. Springer-Verlag, Berlin, 2000. (Newport, Mai 1999).
- [25] L. Collatz: *The Numerical treatment of differential equations*. Springer-Verlag, Berlin, 1966.
- [26] C. Cosner: *On the definition of ellipticity for systems of partial differential equations*. J. Math. Anal. Appl. **158** (1991), 80–93.
- [27] W. Dahmen, B. Faermann, I. G. Graham, W. Hackbusch und S. A. Sauter: *Inverse inequalities on non-quasiuniform meshes and applications to the mortar element method*. Math. Comp. **73** (2004), 1107–1138.
- [28] W. Dörfler: *A convergent adaptive algorithm for Poisson's equation*. SIAM J. Numer. Anal., **33** (1996), 1106–1124.
- [29] K. Eriksson, D. Estep, P. Hansbo und C. Johnson: *Introduction to adaptive methods for differential equations*. Acta Numerica (1995), 105–158.
- [30] C. Fletcher: *Computational Galerkin methods*. Springer-Verlag, New York, 1984.
- [31] F. R. Gantmacher: *Matrizenrechnung, Band I*. Deutscher Verlag der Wissenschaften, Berlin, 1958.
- [32] D. Gilbarg und N. Trudinger: *Elliptic partial differential equations of second order*. Springer-Verlag, Berlin, 1983.
- [33] L. Gladwell und R. Wait (Hrsg.): *A survey of numerical methods for partial differential equations*. Clarendon Press, Oxford, 1979.
- [34] G. Green: *An essay on the application of mathematical analysis in the theories of electricity and magnetism*. Nottingham, 1828.

- [35] P. Grivard: *Elliptic problems in nonsmooth domains*. Pitman, Boston, 1985.
- [36] C. Großmann und H.-G. Roos: *Numerik partieller Differentialgleichungen*. Teubner, Stuttgart, 1992.
- [37] W. Hackbusch: *On the regularity of difference schemes*. Ark. Mat. **19** (1981), 71–95.
- [38] ———: *On the regularity of difference schemes Part II: regularity estimates for linear and nonlinear problems*. Ark. Mat. **21** (1983), 3–28.
- [39] ——— (Hrsg.): *Efficient solvers for elliptic systems*. Notes on Numerical Fluid Mechanics 10. Vieweg-Verlag, Braunschweig, 1984. (Kiel, Jan. 1984).
- [40] ———: *Local defect correction method and domain decomposition techniques*. Comput. Suppl. **5** (1984), 89–113.
- [41] ———: *On first and second order box schemes*. Computing **41** (1989), 277–296.
- [42] ———: *Iterative Lösung großer schwachbesetzter Gleichungssysteme*. 2. Auflage. Teubner, Stuttgart, 1993.
- [43] ———: *Theorie und Numerik elliptischer Differentialgleichungen*. 2. Auflage. Teubner, Stuttgart, 1996.
- [44] ———: *Integralgleichungen. Theorie und Numerik*. 2. Auflage. Teubner, Stuttgart, 1997.
- [45] ———: *Elliptic differential equations. Theory and numerical treatment*. 2. Auflage. SCM 18. Springer-Verlag, Berlin, 2003.
- [46] G. Hellwig: *Partielle Differentialgleichungen*. Teubner, Stuttgart, 1960.
- [47] M. Jung und U. Langer: *Methode der finiten Elemente für Ingenieure*. Teubner, Stuttgart, 2001.
- [48] J. Kadlec: *Über die Regularität der Lösung der Poisson-Gleichung in einem Gebiet, dessen Rand dem eines konvexen Gebietes ähnlich ist*. Czech. Math. J. **14** (1964), 386–393. in Russisch.
- [49] O. D. Kellogg: *Foundations of potential theory*. Springer-Verlag, Berlin, 1929.
- [50] R. Kellogg und J. Osborn: *A regularity result for the Stokes problem in a convex polygon*. J. of Functional Analysis **21** (1976), 397–431.
- [51] C. Kim, R. Lazarov, J. E. Pasciak und P. Vassilevski: *Multiplier spaces for the mortar finite element method in three dimensions*. SIAM J. Numer. Anal. **39** (2001), 519–538.
- [52] P. Knabner und L. Angermann: *Numerik partieller Differentialgleichungen*. Springer-Verlag, Berlin, 2000.
- [53] V. Kondrat'ev: *Boundary value problems for elliptic equations in domains with conical or angular points*. Trans. Moscow Math. Society **16** (1967), 227–313.
- [54] R. Kreß: *Linear integral equations*. Springer-Verlag, Berlin, 1989.
- [55] O. Ladyženskaja: *Funktionalanalytische Untersuchungen der Navier-Stokesschen Gleichungen*. Akademie-Verlag, Berlin, 1965.
- [56] R. Leis: *Vorlesungen über partielle Differentialgleichungen zweiter Ordnung*. B.I.-Hochschultaschenbuch, Mannheim, 1967.
- [57] P. Lesaint und M. Zlámal: *Superconvergence of the gradient of finite element solutions*. RAIRO Anal. Numer. **13** (1979), 139–166.
- [58] J. Lions und E. Magenes: *Non-homogeneous boundary value problems and applications*. Vol. I. Springer-Verlag, Berlin, 1972.

- [59] A. Louis: *Acceleration of convergence for finite element solutions of the Poisson equation*. Numer. Math. **33** (1979), 43–53.
- [60] G. Maeß: *Vorlesungen über numerische Mathematik. I. Lineare Algebra*. Birkhäuser, Basel, 1985.
- [61] G. Marchuk und V. Shaidurov: *Difference methods and their extrapolation*. Springer-Verlag, New York, 1983.
- [62] T. Meis und U. Marcowitz: *Numerische Behandlung partieller Differentialgleichungen*. Springer-Verlag, Berlin, 1978.
- [63] C. Miranda: *Partial differential equations of elliptic type*. Springer-Verlag, Berlin, 1970.
- [64] J. Nečas: *Sur la coercivité des formes sesqui-linéaires elliptiques*. Rev. Roumaine Math. Pures Appl. **9** (1964), 47–69.
- [65] ———: *Equations aux dérivées partielles*. Presse de l'Université de Montréal, 1965.
- [66] J. Nitsche: *Ein Kriterium für die Quasi-Optimalität des Ritzschen Verfahrens*. Numer. Math. **11** (1968), 346–348.
- [67] V. Pereyra, W. Proskurowski und O. B. Widlund: *High order fast Laplace solvers for the Dirichlet problem on general regions*. Math. Comp. **31** (1977), 1–16.
- [68] E. Pflaumann und H. Unger: *Funktionalanalysis I*. Bibliographisches Institut, Mannheim, 1968.
- [69] R. Rannacher: *Approximation of simply supported plates and the Babuška paradox*. ZAMM **59** (1979), T73–T76.
- [70] W. Ritz: *Über eine neue Methode zur Lösung gewisser Variationsprobleme der mathematischen Physik*. J. Reine Angew. Math. **135** (1908), 1–61.
- [71] H.-G. Roos, M. Stynes und L. Tobiska: *Numerical methods for singularly perturbed differential equations: Convection-diffusion and flow problems*. SCM 24. Springer-Verlag, Berlin, 1996.
- [72] S. A. Sauter und C. Schwab: *Randelementmethoden*. Teubner, Stuttgart, 2004.
- [73] A. H. Schatz: *A weak discrete maximum principle and stability in the finite element method in l^∞ on plane polygonal domains*. Math. Comp. **34** (1980), 77–91.
- [74] A. H. Schatz und L. Wahlbin: *Maximum norm estimates in the finite element method on plane polygonal domains. Part 2, refinements*. Math. Comp. **33** (1979), 465–492.
- [75] H. Schwarz: *Methode der finiten Elemente*. Teubner, Stuttgart, 1980.
- [76] G. Shortley und R. Weller: *Numerical solution of Laplace's equation*. J. Appl. Phys. **9** (1938), 334–348.
- [77] J. Stoer: *Einführung in die Numerische Mathematik I*. 5. Auflage. Springer-Verlag, Berlin, 1989.
- [78] G. Strang und G. Fix: *An analysis of the finite element method*. Prentice-Hall, Englewood Cliffs, 1973.
- [79] F. Stummel: *Diskrete Konvergenz linearer Operatoren I*. Math. Ann. **190** (1970), 45–92.
- [80] ———: *The generalized patch test*. SIAM J. Numer. Anal. **16** (1979), 449–471.
- [81] F. Thomasset: *Implementation of finite element methods for Navier-Stokes equations*. Springer-Verlag, New York, 1981.
- [82] V. Thomée: *Discrete interior Schauder estimates for elliptic difference operators*. SIAM J. Numer. Anal. **5** (1968), 626–645.
- [83] ———: *High order local approximations to derivatives in the finite element method*. Math. Comp. **31** (1977), 652–660.

- [84] W. Velte: *Direkte Methoden der Variationsrechnung*. Teubner, Stuttgart, 1976.
- [85] R. Verfürth: *A review of a posteriori error estimation and adaptive mesh-refinement techniques*. J. Wiley und Teubner, Stuttgart, 1996.
- [86] L. Wahlbin: *Superconvergence in Galerkin finite element methods*. Lect. Notes Math. 1605. Springer-Verlag, Berlin, 1995.
- [87] A. J. Wathen: *Realistic eigenvalue bounds for the Galerkin mass matrix*. IMA J. Numer. Anal. **7** (1987), 449–457.
- [88] K. Witsch: *Numerische Quadratur bei Projektionsverfahren*. Numer. Math. **30** (1978), 185–206.
- [89] J. Wloka: *Partielle Differentialgleichungen. Sobolevräume und Randwertaufgaben*. Teubner, Stuttgart, 1982.
- [90] B. Wohlmuth: *Discretization methods and iterative solvers based on domain decomposition*. Habilitationsschrift. Universität Augsburg, 1999 und LNSC 17, Springer-Verlag, Berlin, 2001.
- [91] K. Yosida: *Functional analysis*. Springer-Verlag, Berlin, 1968.
- [92] H. Yserentant: *On the multi-level splitting of finite element spaces*. Numer. Math. **49** (1986), 379–412.
- [93] E. Zeidler (Hrsg.): *Teubner-Taschenbuch der Mathematik*. 2. Auflage. Vol. I. Teubner, Stuttgart, 2003.
- [94] O. Zienkiewicz: *Methode der finiten Elemente*. Carl Hanser Verlag, München, 1984.
- [95] M. Zlámal: *Discretisation and error estimates for elliptic boundary value problems of the fourth order*. SIAM J. Numer. Anal. **4** (1967), 626–639.

Index

- Ableitung
 - klassische, 83
 - konormale, 70
 - Normalen-, 11
 - schwache, 83
 - Tangential-, 70, 104, 111, 174, 190, 217
- Adler-Problem, 114, 131
- Anfangsrandwertvorgabe, 6, 7
- Anfangswertvorgabe, 2
- Anordnung
 - lexikographische, 30
 - Schachbrett-, 31
- B-Splines, 140
- Babuška-Bedingung, 99
- Babuška-Brezzi-Bedingungen, 222
- Babuška-Paradoxon, 75
- Banach-Raum, 80
- Basisfunktionen, *siehe* Finite-Element-Basis-funktionen
- Bedingung von Agmon, 112
- Beltrami-Operator, 10
- BEM, 26
- Bestapproximation, 121
- Bilinearform, **98**, 105–107, 109–115, 117, 118, 123, 124, 133, 136, 139, 162–164, 166, 170, 171, 173, 175, 177, 197, 199, 218, 219
 - adjungierte, **98**, 137, 138
 - beschränkte, 98
 - für die biharmonische Gleichung, 113
 - für die Helmholtz-Gleichung, 106, 110
 - für die Poisson-Gleichung, 107
 - Operator assoziiert zur, 98
 - stetige, 98
 - symmetrische, 98
 - V -elliptische, 99
 - V -koerzive, 101
 - V_h -abhängige, 155
- Blasenfunktion, 229
- Box-Methode, 160
- bubble function, *siehe* Blasenfunktion
- Cauchy-Folge, 80
- Cauchy-Riemann-Differentialgleichungen, 3, 6
- Cholesky-Zerlegung, 200
- DGFEM, 160
- Diagonaldominanz, **34**, 35, 37
 - irreduzible, **34**, 35, 37, 68
- dicht, 81, *siehe* Einbettung
- Differentialgleichung
 - biharmonische, 74–76, 112, 139, 174, 217
 - Cauchy-Riemann-, 3
 - der Ordnung $2m$, 74
 - elliptische, **4–6**, 7–9, 61, **75**, 191, 216
 - gleichmäßig elliptische, **62**, 216
 - hyperbolische, **4**, **5**, 6–8
 - instationäre, 8
 - mit unstetigen Koeffizienten, 189
 - parabolische, **4**, **5**, 7, 8, 192
 - partielle, 1
 - erster Ordnung, 1, 3, 5, 216
 - zweiter Ordnung, 2, 4, 61
 - singulär gestörte, 191
 - stationäre, 8
- Differentialgleichungssystem
 - elliptisches, **5**, 215, 216
 - hyperbolisches, **5**
- Differentialoperator, 61
 - adjungierter, 70

- der Ordnung $2m$, 74
- elliptischer, *siehe* Differentialgleichung
- gleichmäßig elliptischer, 75
- Hauptteil des, *siehe* Hauptteil
- linearer, 5
- Rand-, 70, 110
- Differenz
 - dividierte, 27, 58
 - einseitige, 27, 193
 - linksseitige, 27
 - rechtsseitige, 27
 - Rückwärts-, 27, 41
 - symmetrische, 27, 50, 51
 - Vorwärts-, 27
 - zweite, 28
- Differenzenoperator, 28, 31, 87, 181, 209
- Differenzenstern, *siehe* Stern
- Differenzenverfahren, 66, 71, 176, 193, *siehe* Fünf-,
Neun-, Siebenpunktformel
 - für die biharmonische Gleichung, 76
 - für die Poisson-Gleichung, 27
 - für Eigenwertprobleme, 209
 - Shortley-Weller-, 57
- Diffusion, 191
- direkte Summe, 82
- Dirichlet-Randwertvorgabe, 19, 62, 70, 75
 - homogene, 104, 168
 - inhomogene, 108
- Dirichletsches Integral, 103
- Diskretisierung, *siehe* Differenzenverfahren, finite
Elemente, Galerkin-Verfahren
 - Effizienz der, 146
 - höherer Ordnung, 45
 - Shortley-Weller-, 57
 - stabile, *siehe* Stabilität
- Diskretisierungsfehler, 43
- Divergenzoperator, 215
- Doppelschichtpotential, 26
- Dreiecksungleichung
 - umgekehrte, 79
- Dualabbildung, Dualoperator, 93
- Dualform, 93
- Dualnorm, 93
- Dualraum, 93

- Effizienz einer Diskretisierung, 146
- Eigenfunktionen, 199
 - Konvergenz der diskreten, 203
- Eigenraum, 199
- Eigenwertaufgabe, 97, 101, 107
 - adjungierte, 199
 - elliptische, 8, **199**
 - verallgemeinerte, 6, 200
- Eigenwerte
 - Konvergenz der diskreten, 203
 - Vielfachheit der, 200
- Einbettung
 - dichte, 81
 - kompakte, 96
 - stetige, 81
- Eindeutigkeit der Lösung, 14, 19, 64, 106, 108, 207,
220
- Einfachschichtpotential, 26
- Elementmatrizen, 153
- Elliptizität, numerische, 194
- Existenz einer Lösung, 106, 108
- Extrapolationsverfahren, 45, 60

- Fehlerabschätzung
 - bei Differenzenverfahren, 77
- Fehlerabschätzung
 - a posteriori, 143
 - a priori, 143
 - bei Differenzenverfahren, 43, 47, 59, 60
 - bei Finite-Elemente-Verfahren, 132
 - bei Galerkin-Verfahren, 120
- Fehleräquilibrierung, 147
- Fehlerschätzer, 144
 - asymptotisch exakter, 146
 - effizienter, 146
 - zuverlässiger, 146
- FEM, 124
- finite Elemente
 - bikubische, 140
 - bilineare, 128
 - der Serendipity-Klasse, 130
 - isoparametrische, 142
 - lineare, 124, 126
 - quadratische, 129
- finite Volumen, 160
- Finite-Element-Basisfunktionen
 - hierarchische, 157
 - stückweise lineare, 125, 126
- Finite-Element-Methode, 115
 - adaptive, 147
 - für Eigenwertprobleme, 200
 - für Sattelpunktprobleme, 226
 - gemischte, 226
 - nichtkonforme, 154
- Form
 - Bilinear-, *siehe* Bilinearform
 - Dual-, 93
 - Sesquilinear-, 98
- Formregularität, *siehe* Triangulation
- Fortsetzung, 81
- Fortsetzung(soperator), 88
- Fourier-Transformation, 86
 - inverse, 86
- Fredholmsche Integralgleichung zweiter Art, 26
- Fundamentallösung, *siehe* Singularitätenfunktion

- Fünfpunktformel, 29, 68, 127
- Funktional (lineares), 93, 109
- Galerkin-Verfahren, 26
 - konformes, 115
 - unstetiges, 160
- Gauß-Quadratur, 125
- Gebiet, 9
 - konvexes, 57, 173
 - Normal-, 11
 - unbeschränktes, 17, 105, 106
- Gebietszerlegungsverfahren, 154
- Gelfand-Dreier, 95
- Gitter, 28, 29, 51, 57
 - formreguläres, 135
 - K -, *siehe* K -Gitter
 - uniformes, 135
- Gitterfunktion, 28, 43
- Gitterpunkt
 - benachbarter, 30
 - randferner, 30
 - randnahe, 57
- Gitterverfeinerung, 148
 - adaptive, 148
- Gittervergrößerung, 149
- Gleichungssystem, 28, 30, 48, 116, 132, 227
 - Lösbarkeit des, 48
 - schwach besetztes, 31
- Gradient, 11
- Greensche Formel
 - erste, 11
 - zweite, 11
- Greensche Funktion, 12, **20**, **69**, 103
 - diskrete, 39, 52
 - für die Kugel, **24**
 - zweiter Art, 25
- Grundlösung, **12**, 20
- harmonische Funktion, **10**, 11, 14, 15, 24, 103
- Hauptteil, **5**, 9, 104, 178, **216**
- Helmholtz-Gleichung, 106, 110
- Hermite-Interpolation, 140
- Hesse-Matrix, 62
- hierarchische Basis, 157
- Hilbert-Matrix, 147
- Hilbert-Raum, 82
- Hölder-Stetigkeit, 21, 89
- holomorph, 2, 13, 24
- inf-sup-Bedingung, 98
 - diskrete, 120
- Integralgleichungsmethode, 25, 26
- inverse Abschätzung, 151
- K -Gitter, 135
- Kegelbedingung, gleichmäßige, 97
- klassische Lösung, 19, 21–23, 105, 110, 166, 173, 218
- Kleinste-Quadrate-Minimierung, 160
- Knoten, *siehe* Knotenpunkte
 - hängende, 156
 - innere, 126
 - Rand-, 126
- Knotenpunkte, 124
- Knotenwerte, 124
- Koeffizientenvektor, 116
- Kollokationsmethode, 26
- kompakte Abbildung, 96
- kompakte Menge, 96
- Kondition, 78, 120, 150, 151, 158, 207
- Konormalenableitung, *siehe* Ableitung, konormale
- Konvektion, 191
- Konvergenz, 43
 - der Ordnung k , 43
 - gleichmäßige, 79
 - Super-, 158
- Koordinatentransformation, 5, 24, 62, 65, 73, 86
- L-Gebiet, 10, 24, 90, 157, 188, 227
- Lagrange-Faktoren, 132
- Lamé-Gleichungen, 217
- Laplace-Gleichung, *siehe* Potentialgleichung, 9
- Laplace-Operator, 9, 10
- lexikographische Anordnung, 30
- Lösung
 - klassische, *siehe* klassische Lösung
 - schwache, 105
- M-Matrix, **32**, 38, 46, 49, 58, 60, 66–68, 72, 73, 154, 193–195
- Massematrix, **150**
- Matrix
 - diagonaldominante, *siehe* Diagonaldominanz
 - Hesse-, 62
 - Hilbert-, 147
 - irreduzibel diagonaldominante, *siehe* Diagonaldominanz
 - irreduzible, **32**, 33, 35, 39
 - M-, *siehe* M-Matrix
 - positiv definite, **37**, 61, 68, 76, 118, 200
 - positiv semidefinite, 62
 - schwach besetzte, 31
- Matrixgraph, 32
- Matrixnorm, 36
 - zugeordnete, 36
- Maximumnorm, *siehe* Supremumsnorm, Norm, 33, 36
- Maximumprinzip, **13**, 14, 39, **62**, 74, 154, 193
 - starkes, 63
- Mehrstellenverfahren, 46

- Methode der finiten Elemente, *siehe* Finite-Element-Methode
- Minimierungsproblem, 100, 107, 118, 123, 156
- Minimumprinzip, 203
- Mittelwerteigenschaft, **13**, 14
 - diskrete, 39
 - zweite, 13
- Nebenbedingungen, 131
- Neumann-Randwertvorgabe, 24, 25, 47, 71, 110, 128
- Neunpunktformel, 45, 66
 - kompakte, 45
- nicht sachgemäß gestellte Aufgabe, 7
- Norm, **35**
 - äquivalente, 80
 - Dual-, 93
 - Euklidische, 11, 37
 - Matrix-, *siehe* Matrixnorm
 - Operator-, 80
 - Sobolev-Slobodeckij-, 88
 - Spektral-, 37
 - Supremums-, *siehe* Supremumsnorm
 - Zeilensummen-, 36
- Normalenableitung, **11**, 25
- normales System, 75
- Normalgebiet, 11
- normierter Raum, 79
- Operator, 80
 - adjungierter, 94
 - beschränkter, 80
 - Dual-, 93
 - einer Bilinearform zugeordneter, 98
 - Fortsetzung eines, 81
 - Fortsetzungs-, 88
 - kompakter, 96
 - stetiger, 80
- Operatornorm, 80
- orthogonal, Orthogonalraum, 82
- Partition der Eins, 91
- Patchtest, 155
- Permutationsmatrix, 30
- Petrov-Galerkin-Verfahren, 160
- Plattengleichung, 74
- Poincaré-Friedrichs-Ungleichung, 85
- Poisson-Gleichung, **19**, 27, 107, 174
- Poissonsche Integralformel, **14**, 16
- Polarkoordinaten, 10, 11, 71
- Potentialgleichung, 2–4, 7, 8, **9**, 47, 157
 - diskrete, 39, 41
- präkompakt, 96
- Problem, *siehe* Variationsproblem
- Projektion, 94
 - orthogonale, 94, 117
 - Ritz-, 122, 136, 138
- Prolongation, 116
- Quadratgittertriangulation, 127
- Quasioptimalität, 121
- Quotientenraum, 81
- Randbedingung, *siehe* Dirichlet-Randwertvorgabe, Neumann-Randwertvorgabe
 - dritte, 70
 - erste, 70
 - natürliche, **109**, 128
 - Robin-, 70
 - zweite, 70
- Randdifferentialoperator, *siehe* Differentialoperator
- Randelementmethode, 26
- Randschicht, 192
- Randwertaufgabe, 9
 - der Ordnung $2m$, 112
 - Variationsformulierung einer, 103
- Randwertvorgabe, 2, 6, 7, 9
 - Dirichlet-, *siehe* Dirichlet-Randwertvorgabe
 - Neumann-, *siehe* Neumann-Randwertvorgabe
 - periodische, 71
- Raum
 - Banach-, 80
 - Dual-, 93
 - Hilbert-, 82
 - normierter, 79
 - Orthogonal-, 82
 - Quotienten-, 81
 - Vervollständigung eines, 81
 - vollständiger, 80
- Reaktions-Diffusionsgleichung, 192
- reduzierte Gleichung, 192
- Referenzdreieck (Referenzelement), 127, 129, 133
- Regularität
 - der Differenzengleichungen, 176
- Regularität, **161**
 - H^2 -, 137
 - H^s -, 162
- Restriktion, 43
- Riesz-Isomorphismus, 94
- Riesz-Schauder-Theorie, 97, 101
- Ritz-Galerkin-Verfahren, 115
 - gemischtes, 226
- Ritz-Projektion, 122, 136, 138
- Robin-Randbedingung, 70
- Rückwärtsdifferenz, 27
- sachgemäß gestellte Aufgabe, **16**
- Sattelpunkt, 219
- Sattelpunktproblem, 218

Satz

Riemannscher Abbildungs-, 24
 Rieszscher Darstellungs-, 94
 Sobolevscher Einbettungs-, 90
 Transformations-, 86
 von Aubin-Nitsche, 137
 von Cea, 120
 von der offenen Abbildung, 81
 von Gårding, 107
 von Gerschgorin, 33
 von Harnack, 15
 Schachbrettanordnung, 31
 Schrittweite, 27
 schwache Ableitung, 83
 schwache Formulierung (einer Randwertaufgabe),
 siehe Variationsformulierung
 Schwarzsche Ungleichung, 82
 Sesquilinearform, 98
 Shortley-Weller-Differenzenschema, 57
 Siebenpunktformel, 66
 Singularitätenfunktion, **10**, 75
 diskrete, 53
 Skalarprodukt, 11, 37, 82, 84, 87, 92, 95, 175
 Sobolev-Raum, 82
 Sobolev-Slobodeckij-Norm, 88
 Spektralnorm, 37
 Spektralradius, 34
 Spektrum, 97
 Splines, kubische, 140
 Spur einer Funktion, 88
 Spur einer Matrix, 63
 Stabilität, 43, 69
 Steifigkeitsmatrix, 116, **149**
 Steklov-Problem, 200
 Stern, 31
 stetige Abhängigkeit der Lösung
 gegenüber Gebietsvariation, 17, 75
 von den Koeffizienten, 66
 von den Randwerten, 16, 65
 Stokes-Gleichungen, 4, 74, **215**
 Strang
 erstes Lemma von, 123
 zweites Lemma von, 155
 Stromlinien-Diffusionsverfahren, 195
 Superkonvergenz, 158
 Supremumsnorm, 16, 79, 80
 Systemmatrix, 116, **149**

 Tangentialableitung, *siehe* Ableitung
 Träger (einer Funktion), 83
 Trapezformel, 51
 Trefftz-Verfahren, 156
 Triangulation, 126
 formreguläre, 135, 148
 uniforme, 135

zulässige, 126

Typeneinteilung der partiellen Differentialgleichungen, 1, 4, 5

V-Elliptizität, 99

V-Koerzivität, 101

Variationsformulierung (einer Randwertaufgabe),
103, 115, 161, 189

Variationsproblem, 100
 adjungiertes, 100, 137
 duales oder komplementäres, 157

Vervollständigung, 81

Viskosität
 künstliche, 194
 numerische, 194

Vollständigkeit, 80

Volumenpotential, 26

Vorwärtsdifferenz, 27

Wärmeleitungsgleichung, 3, 6–8

Wellengleichung, 2–4, 6–8

Wilsonsches Rechteck, 154

Zeilensummennorm, 36