

Technische Universität München
Zentrum Mathematik

Zur Approximation der Lösungen elliptischer Systeme partieller Differentialgleichungen mittels Finiter Elemente und \mathcal{H} -Matrizen

Robert Schreittmiller

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. Jurij Suris
Prüfer der Dissertation: 1. Univ.-Prof. Dr. Martin Brokate
2. Univ.-Prof. Dr. Thomas Apel,
Universität der Bundeswehr München
3. Prof. Dr. Dr. h.c. Wolfgang Hackbusch
Christian-Albrechts-Universität zu Kiel
(schriftliche Beurteilung)

Die Dissertation wurde am 27. Juni 2005 bei der Technischen Universität eingereicht und durch die Fakultät für Mathematik am 06. Februar 2006 angenommen.

Zusammenfassung/Abstract

Kern der Untersuchungen bildet der Nachweis, daß für elliptische Systeme von partiellen Differentialgleichungen zweiter Ordnung eine Familie von lokalen Niedrigrangoperatoren existiert, die exponentiell gegen den inversen Differentialoperator konvergiert. Diese Niedrigrangoperatoren werden verwendet, um hierarchische Matrizen (\mathcal{H} -Matrizen) zu konstruieren, die u.a. die inverse Steifigkeitsmatrix approximieren, und um den Approximationsfehler abzuschätzen.

Zur Vorbereitung des Beweises werden die erforderlichen Begriffe und Werkzeuge bereitgestellt und erläutert. Der Beweis verwendet nur die schwache Lösung der Differentialgleichung, nicht jedoch die GREENsche Funktion. Der Vorteil der \mathcal{H} -Matrizen liegt in der problemlosen Handhabung springender Koeffizienten.

Den Abschluß der Arbeit bilden numerische Tests, die die theoretischen Aussagen unterstreichen. Dabei wird die LAMÉ-Gleichung als Beispiel für ein elliptisches System mittels linearer finiter Elemente diskretisiert und die resultierenden linearen Gleichungssysteme mittels der \mathcal{H} -Arithmetik gelöst.

Abstract

The main focus of this work is to prove the existence of a family of local low rank operators, which in case of elliptic systems of partial differential equations of second order converges to the inverse differential operator. These low rank operators will be used to construct hierarchical matrices (\mathcal{H} -matrices) which approximate the inverse of a given matrix (e.g. the stiffness matrix), and to estimate the approximation error.

In the setup to the proof we define and discuss the necessary terms and tools. The proof uses only the weak solution of the differential equation, but not its GREEN's function. The advantage of \mathcal{H} -matrices is the trouble-free handling of L^∞ coefficients.

Finally, we close the work with numerical tests verifying the theoretical results. As an example to elliptic systems we take the LAMÉ equation, discretize it with finite elements, and solve the resulting linear systems of equations using the arithmetics of \mathcal{H} -matrices.

Danksagung

Neben vielen anderen Personen gilt mein Dank Herrn Prof. Dr. MARTIN BROKATE für die interessante und weitsichtige Aufgabenstellung sowie die Freiheit, verschiedene Lösungswege entwickeln zu dürfen und mir beim Beschreiten derselbigen zu helfen.

Dank gebührt aber auch Herrn Prof. Dr. Dr. h.c. WOLFGANG HACKBUSCH, der mich ans MPI nach Leipzig einlud und mich in den Genuß seiner Gastfreundschaft kommen ließ. Dabei stand er mir Rede und Antwort zu vielen Fragen rund um die untersuchte Aufgabe.

Mein besonderer Dank unter seinen Mitarbeitern gilt Herrn Dr. STEFFEN BÖRM, der mir mit seiner herzlichen Art, der immerwährenden Diskussionsfreudigkeit und den vielen fruchtbaren Anregungen zur Seite stand.

Schließlich möchte ich den mir noch unbekanntem Gutachtern danken, daß sie die Zeit und die Mühe für das Lesen der Arbeit aufbringen.

Explizite Erwähnung für ihre moralische Unterstützung haben Sylvia, Christian und Jan sowie in besonderem Maße Harry und Sally verdient.

Symbolverzeichnis

Das Symbolverzeichnis enthält alle Bezeichnungen, die nicht nur lokal in einem Abschnitt verwendet wurden, sondern durch das ganze Dokument dieselbe Bedeutung tragen. Die Seitenzahl gibt die Stelle an, an der eine detaillierte Erklärung zu finden ist. Etwaige Mehrfachverwendungen sind stets so arrangiert, daß Verwechslungen mit der globalen Bedeutung unmöglich sind.

Die Kennzeichnung mit \rightarrow weist stets auf den Zusammenhang mit Differentialgleichungssystemen hin.

Griechische Bezeichner		
Symbol	Bedeutung	Seite
Γ	Rand von Ω	
η	Parameter der Zulässigkeitsbedingung	41
ϑ	Elliptizitätskonstante	6, 13
λ	LAMÉ-Konstante	14
$\lambda_{min}, \lambda_{max}$	kleinster und größter Eigenwert von C	
$\vec{\lambda}_{min}, \vec{\lambda}_{max}$	kleinster und größter Eigenwert von \vec{C}	
μ	Gleitmodul	14
μ_{min}, μ_{max}	kleinster und größter Eigenwert von M und \vec{M}	
κ_C	Verhältnis von λ_{max} zu λ_{min}	6
$\kappa_{\vec{C}}$	Verhältnis von $\vec{\lambda}_{max}$ zu $\vec{\lambda}_{min}$	
$\nu, \nu(x)$	Normalenvektor	
$\varphi, \varphi(x)$	Testfunktion	
ρ_T	Innkreisradius eines (Dreieck-) Elementes T	19
σ	Cluster	
τ	Cluster; einmalig auf S. 19 zulässige Triangulation	
τ_h	zulässige Triangulation der Maschenweite h	S. 19
Ω	Gebiet im \mathbb{R}^n (Definitionsbereich der Differentialgleichungen)	
ω_n	Volumen der n -dim. Einheitskugel	57
Lateinische Bezeichner		
Symbol	Bedeutung	Seite
${}^{(4)}C$		13
$A_{\mathcal{H}}$	\mathcal{H} -Approximation einer Matrix A	
$a(\cdot, \cdot)$	Bilinearform zu L	11

Lateinische Bezeichner (Forts.)

Symbol	Bedeutung	Seite
$\vec{a}(\cdot, \cdot)$	Bilinearform zu \vec{L}	15
$C, C(x)$	Koeffizientenmatrix von L mit den Komponenten c_{ij}	
$\vec{C}, \vec{C}(x)$	Koeffizientenmatrix von \vec{L} mit den Komponenten C_{ij}	13
$\vec{C}_h, \vec{C}_\mathcal{H}$	Hilfsmatrix und ihre \mathcal{H} -Approximierende (Systemfall)	95, 97
$C_{ij}, C_{ij}(x)$	Komponente der Matrix \vec{C}	
c_0, \vec{c}_0	Konstante	62, 70
$c_{(x,y)}$	Konstante, die in der Gleichung (x.y) definiert wird	
c_E	Konstante bei der Elliptizität von Bilinearformen	12, 16
c_{Fak}	Konstante, die bei der Abschätzung von $m!$ auftritt.	56
c_{FEM}	Konstante aus der FE-Fehlerabschätzung	25
$c_{ij}, c_{ij}(x)$	Komponente der Matrix C	
$c_{ijkl}, c_{ijkl}(x)$	Komponente der Matrix C_{ij} bzw. des Tensors ${}^{(4)}C$	
c_J	Konstante	23
c_{lokAF}	Konstante	91
c_m	von m „abhängige“ Konstante	56
$c_{m,n}$	Konstante, bei der m von n abhängt	62
$\tilde{c}_{m,n}$	Konstante, bei der m von n abhängt	57
c_{reg}	Konstante	56
c_S	Konstante bei der Stetigkeit von Bilinearformen	11, 16
C_{sp}	Schwachbesetztheitskonstante	42
c_{TE}	Konstante im Zusammenhang der schwachen TAYLOR-Entwicklung	60
$D^\alpha u$	part. Ableitung $\frac{\partial^{ \alpha } u(x)}{\partial x_1^{\alpha_1} \dots \partial x_n^{\alpha_n}}$	
$D^m u, \nabla^m u, \nabla^m \vec{u}$	Vektor/Tensor aller m -ten Ableitungen	
$F(\cdot), \vec{F}(\cdot)$	Linearform	11, 15
F_h, \vec{F}_h	Lastvektoren	20, 29
f, \vec{f}	rechte Seite der Differentialgleichung	5, 13
$h; h_T$	Schrittweite; längste Kante von T	19
$id = id(x) = x$	Identitätsabbildung	
Id	Einheitsmatrix	
im	Bild	
ker	Kern	
I	Indexmenge	
J	Koeffizientenisomorphismus	20
\vec{J}, \vec{J}_G	Koeffizientenisomorphismus bei Systemen	26
L	elliptischer Differentialoperator	5
L_h	RITZ-GALERKIN-Diskretisierung von L	21
$\vec{L} = (L_{ij})_{i,j}$	elliptischer Matrixdifferentialoperator	13
\vec{L}_h	RITZ-GALERKIN-Diskretisierung von \vec{L}	29
M	Massenmatrix	21
\vec{M}, \vec{M}_G	Massenmatrix beim DG-System	29

Lateinische Bezeichner (Forts.)

Symbol	Bedeutung	Seite
N	Anzahl der Freiheitsgrade (DoF's)	
n	Dimension von Ω	
Q_h, \vec{Q}_h	L^2 -orthogonale Projektion	24, 31
$Q^m u, \vec{Q}^m \vec{u}$	gemittelttes TAYLOR-Polynom von u bzw. \vec{u}	60, 69
q	Anzahl der Gleichungen des DG-Systems	
R_h, \vec{R}_h	RITZ-Projektion	24, 32
R^m, \vec{R}^m	Fehler bei der schwachen TAYLOR-Entwicklung	60, 69
rk	Rang	
\vec{S}	orthogonale Projektion	77
s, \vec{s}, t, \vec{t}	zu σ bzw. τ gehörende Indexteilmenge von I	64, 71
u, \vec{u}	Lösung der Differentialgleichung	5, 12
u_h, \vec{u}_h	RITZ-GALERKIN-Lösung	20, 26
$\vec{u}^{(ext)}$	Fortsetzung von \vec{u}	74
$\mathbf{u}, \mathbf{v}, \mathbf{w}$	(Koeffizienten-) Vektor im \mathbb{R}^N	
$\mathbf{u}_i, \mathbf{v}_i, \mathbf{w}_i$	i -te Komponente von $\mathbf{u}, \mathbf{v}, \mathbf{w}$	
$\vec{\mathbf{u}}, \vec{\mathbf{u}}_i, \mathbf{u}_{ij}$	(Koeffizienten-) Vektor im \mathbb{R}^{qN}	
V, V^q	Test-, Lösungsraum	11, 12
V_h, V_h^q	Ansatzraum	20, 26
$X(D, \mathbb{R}^q)$	Funktionsraum	81, 77

Klammerausdrücke

Symbol	Bedeutung	Seite
$\langle \cdot, \cdot \rangle$	EUKLIDISCHES Skalarprodukt	
$\langle \cdot, \cdot \rangle_h$	skaliertes Skalarprodukt	23
$ u $	EUKLIDISCHE Norm, Betrag	
$ u _{H^m(\Omega)}$	Seminorm in $H^m(\Omega)$	
$ \vec{u} _{H^m(\Omega, \mathbb{R}^n)}$	Seminorm für H^m -Funktionen von Ω nach \mathbb{R}^n	
$\ \cdot\ _h$	skalierte EUKLIDISCHE Norm	23
$\ u\ _{H^m(\Omega)}$	Norm in $H^m(\Omega)$	
$\ \vec{u}\ _{H^m(\Omega, \mathbb{R}^n)}$	Norm für H^m -Funktionen von Ω nach \mathbb{R}^n	
$\ \cdot\ _{V(\Omega, x)}$	V -Norm über Ω , ausgewertet bzgl. x	

Sonstige Bezeichner

Symbol	Bedeutung	Seite
∂_i	partielle Ableitung nach der i -ten Variablen	
$\partial_\nu u, \frac{\partial}{\partial \nu} u$	Normalenableitung	8
\varnothing_M	ČEBYŠEV-Durchmesser der Menge M	40

Inhaltsverzeichnis

Zusammenfassung	i
Symbolverzeichnis	iii
1 Einführung	1
1.1 Zielsetzung und Struktur der Arbeit	1
1.2 Gleichmäßig elliptische Differentialoperatoren	1
1.3 Finite Elemente	2
1.4 \mathcal{H} -Matrizen	3
1.5 Die Existenz der \mathcal{H} -Inversen	3
1.6 Numerik	4
2 Elliptische Differentialgleichungen zweiter Ordnung	5
2.1 Grundbegriffe	5
2.2 Fundamentallösungen	6
2.2.1 Singularitätenfunktionen	6
2.2.2 GREENSche Funktionen	9
2.2.3 Abschließende Bemerkungen zu Fundamentallösungen	10
2.3 Schwache Lösungen	11
2.4 Lineare Systeme von Differentialgleichungen	12
2.4.1 Grundbegriffe	12
2.4.2 Die LAMÉ-Gleichung	14
2.4.3 Fundamentallösungen	15
2.4.4 Schwache Form	15
3 Finite-Element-Methode	19
3.1 RITZ-GALERKIN-Verfahren	20

3.2	Der adjungierte Koeffizientenmorphismus	21
3.3	Koeffizientenmorphismus und FE-Matrizen	22
3.4	Fehler- und Konvergenzanalyse	24
3.5	Finite-Element-Methode bei Systemen von Differentialgleichungen	26
3.5.1	Ansatz- und Testräume	26
3.5.2	Gleichungs- und knotenorientierte Diskretisierung	26
3.5.3	Fehler- und Konvergenzanalyse	31
3.5.4	FE-Formulierung der LAMÉ-Gleichung	34
4	\mathcal{H}-Matrizen	35
4.1	Idee	35
4.1.1	Eine Integralgleichung zur Motivation	35
4.1.2	Zielsetzung und Vorgehensweise	38
4.2	Hierarchische Strukturen	39
4.2.1	\mathcal{H} -Bäume	39
4.2.2	Zulässigkeitsbedingung	40
4.2.3	Produkt von \mathcal{H} -Bäumen	42
4.3	Rk -Matrizen	43
4.3.1	Die Kürzungsmethode	43
4.3.2	Arithmetik	44
4.3.3	Rangadaptivität	45
4.4	\mathcal{H} -Matrizen	45
4.4.1	\mathcal{H} -Approximation von Matrizen	45
4.4.2	Arithmetik	46
4.4.3	\mathcal{H} -Inversion	47
5	Approximationseigenschaft elliptischer Differentialoperatoren	49
5.1	Approximationseigenschaft für skalare Differentialoperatoren	50
5.1.1	Beweisidee	50
5.1.2	Innere Regularität	51
5.1.3	Hilfsaussagen	57
5.1.4	Lokale Niedrigrangapproximation von L^{-1}	60
5.2	Lokale Approximationseigenschaft für Matrixoperatoren I	64
5.2.1	Innere Regularität	64

5.2.2	Lokale Niedrigrangapproximation von \vec{L}^{-1}	69
5.3	Lokale Approximationseigenschaft für Matrixoperatoren II	73
5.3.1	Ein weiteres Regularitätsresultat	73
5.3.2	Funktionalanalytische Betrachtungen	77
5.3.3	\vec{L} -harmonische Approximation von $\vec{u}^{(ext)}$	82
5.3.4	Lokale Niedrigrang- und \mathcal{H} -Approximation von \vec{L}^{-1}	89
5.4	Globale Approximationseigenschaft	91
5.4.1	Steifigkeitsmatrix des inversen Differentialoperators	91
5.4.2	Massen- und Steifigkeitsmatrix	93
5.4.3	\mathcal{H} -Approximation der Hilfsmatrix	97
5.4.4	Gesamtfehler und Rangbestimmung	99
6	Numerische Tests	101
6.1	Algorithmen	101
6.2	Die Testprobleme	103
6.3	Ergebnisevaluierung	105
6.3.1	Geometrie 2 und Parametertyp 3	106
6.3.2	Geometrie 2 und Parametertyp 4	108
6.3.3	Geometrie 3 und Parametertyp 3	110
6.3.4	Geometrie 3 und Parametertyp 4	112
6.3.5	Rangadaptive \mathcal{H} -Invertierung	114
6.4	Praktikabilitätsversuche	116
6.4.1	$P_{\mathcal{H}\text{-Cholesky}}\text{CG}$: adaptiver Rang, Geometrie 3	117
6.4.2	$P_{\mathcal{H}\text{-Cholesky}}\text{CG}$: adaptiver Rang, Geometrie 6	119
6.4.3	$P_{\mathcal{H}\text{-Cholesky}}\text{CG}$: adaptiver Rang, Geometrie 7	121
6.4.4	$P_{\mathcal{H}\text{-Cholesky}}\text{CG}$: adaptiver Rang, Geometrie 8	123
6.5	Resümee	125
A	Mathematische Ergänzungen und Nachträge	127
A.1	Matrixnormen	127
A.2	Äquivalenz von Normen	129
B	Details zu den numerischen Tests	133
B.1	Rechnerarchitektur und Programmatur	133

B.1.1	Angaben zur Hardware und Compiler	133
B.1.2	Verwendete Programme und Bibliotheken	133
B.1.3	Gebiets- und Gittergenerierung	134
B.2	Numerische Resultate	134
B.2.1	\mathcal{H} -Inverse: fester Rang, Geometrie 2, Parametertyp 3	135
B.2.2	\mathcal{H} -Inverse: fester Rang, Geometrie 2, Parametertyp 4	138
B.2.3	\mathcal{H} -Inverse: fester Rang, Geometrie 3, Parametertyp 3	141
B.2.4	\mathcal{H} -Inverse: fester Rang, Geometrie 3, Parametertyp 4	144
B.2.5	\mathcal{H} -Inverse: adaptiver Rang, Geometrie 3, Parametertyp 4	147
B.2.6	$P_{\mathcal{H}\text{-Cholesky}}\text{CG}$: adaptiver Rang, Geometrie 3	151
B.2.7	$P_{\mathcal{H}\text{-Cholesky}}\text{CG}$: adaptiver Rang, Geometrie 6	153
B.2.8	$P_{\mathcal{H}\text{-Cholesky}}\text{CG}$: adaptiver Rang, Geometrie 7	155
B.2.9	$P_{\mathcal{H}\text{-Cholesky}}\text{CG}$: adaptiver Rang, Geometrie 8	157

Kapitel 1

Einführung

Die \mathcal{H} -Matrizen sind eine sehr junge Erfindung von W. HACKBUSCH zur Lösung von linearen Gleichungssystemen mit vollbesetzten Matrizen, wie sie bei der Randelementmethode entstehen, vgl. [Hb98] und [HK99]. In der Vergangenheit haben \mathcal{H} -Matrizen in der Praxis immer wieder demonstriert (vgl. [BH03], [Be04], [Li02]), daß sie auch zur Lösung von linearen Gleichungssystemen im Rahmen der Finiten-Element-Methode effizient eingesetzt werden können.

1.1 Zielsetzung und Struktur der Arbeit

Das Ziel der vorliegenden Arbeit ist, die Existenz einer \mathcal{H} -Approximation der inversen Steifigkeitsmatrix, die wir bei der Finiten-Element-Diskretisierung elliptischer Systeme partieller Differentialgleichungen mit L^∞ -Koeffizienten erhalten, zu beweisen. Hierfür legen wir uns in den Kapiteln 2 bis 4 alle erforderlichen Werkzeuge und Hilfsmittel zurecht und beweisen in Kapitel 5 die Existenz der \mathcal{H} -Inversen auf zwei verschiedenen Wegen. In Kapitel 6 belegen wir die gewonnenen Abschätzungen durch numerische Tests.

1.2 Gleichmäßig elliptische Differentialoperatoren

In Kapitel 2 definieren wir für ein LIPSCHITZ-Gebiet $\Omega \subset \mathbb{R}^n$ gleichmäßig elliptische Differentialoperatoren und stellen die wichtigsten Punkte der schwachen Lösungstheorie zusammen. Anschließend erweitern wir die Begriffe auf q -dimensionale Systeme von partiellen Differentialgleichungen

$$\sum_{j=1}^q L_{ij} u_j = f_i, \quad i = 1, \dots, q,$$

mit $L_{ij} u_j = - \sum_{k,l=1}^n \partial_l (c_{ijkl} \partial_k u_j)$, $c_{ijkl} \in L^\infty(\Omega)$, die nach $\vec{u} = (u_i)_{i=1,\dots,q} : \Omega \rightarrow \mathbb{R}^q$ zu lösen sind. Als Beispiel für ein gleichmäßig elliptisches System von Differenti-

algleichungen diskutieren wir die LAMÉ-Gleichung

$$-(\lambda + \mu) \cdot \nabla (\operatorname{div}(\vec{u})) - \mu \cdot \Delta \vec{u} = \vec{f},$$

die dann auch in Kapitel 6 als Testbeispiel dient.

Wir geben aus mehreren Gründen einen Exkurs über Singularitäten- und GREENsche Funktionen bei. Erstens wurden die \mathcal{H} -Matrizen im Kontext der Integralgleichungen und Randelementmethoden entwickelt und dort sind Kernfunktionen eben sehr wichtig. Zweitens kann unter Verwendung dieser Funktionen der Approximationsbeweis geführt werden, vgl. [BH03]. Drittens ist die GREENsche Funktion Bestandteil des inversen Differentialoperators. Problematisch ist jedoch die Frage nach ihrer Existenz, womit wir unseren Weg motivieren wollen, den Beweis mittels der schwachen Lösung zu führen.

1.3 Finite Elemente

In Kapitel 3 erklären wir die Grundzüge der Finiten-Element-Methode. Wir verwenden quasi-gleichmäßige Triangulierungen mit N inneren Gitterpunkten. Der Ansatzraum V_h^q wird von qN linearen Ansatzfunktionen aufgespannt.

Einen Schwerpunkt setzen wir auf die Darstellung der FE-Matrizen als Verkettung eines sog. Koeffizientenisomorphismus $\vec{J} : \mathbb{R}^{qN} \rightarrow V_h^q$ und des Differentialoperators $\vec{L} = (L_{ij})_{i,j=1,\dots,q}$. Für die RITZ-GALERKIN-Diskretisierung des inversen Differentialoperators \vec{L}^{-1} gilt

$$\left(\vec{L}^{-1}\right)_h = \vec{J}^* \vec{L}^{-1} \vec{J}.$$

Ersetzen wir später \vec{L}^{-1} durch eine Niedrigrangapproximation, sind wir bei der Konstruktion der \mathcal{H} -Inversen der Steifigkeitsmatrix schon fast am Ziel und haben zugleich den Approximationsfehler in der Hand, weil wir Abschätzungen für \vec{L}^{-1} usw. kennen.

Die Struktur der FE-Matrizen hängt von der Assemblierung ab. Um auf Ergebnisse aus [BH03] zurückgreifen zu können und gewisse Abschätzungen zu erleichtern, diskutieren wir die knoten- und die gleichungsorientierte Diskretisierungen. Die resultierenden Massenmatrizen können durch Multiplikation mit einer Permutationsmatrix ineinander überführt werden.

Das Kapitel schließen wir mit einer Fehler- und Konvergenzanalyse. Ist die Lösung \vec{u} des Differentialgleichungssystems $H^{1+\alpha}$ -regulär, $\alpha \in [0, 1]$, und \vec{u}_h die RITZ-GALERKIN-Lösung, so können wir den Diskretisierungsfehler durch die rechte Seite \vec{f} und eine Potenz der Schrittweite h abschätzen,

$$\|\vec{u} - \vec{u}_h\|_{L^2(\Omega, \mathbb{R}^q)} \leq c_{FEM} \cdot h^{2\alpha} \cdot \left\| \vec{f} \right\|_{L^2(\Omega, \mathbb{R}^q)}, \quad c_{FEM} > 0.$$

1.4 \mathcal{H} -Matrizen

In Kapitel 4 motivieren wir anhand einer Integralgleichung die hinter den \mathcal{H} -Matrizen stehende Erkenntnis, daß zwei weit voneinander entfernte Teilgebiete nur wenig miteinander kommunizieren und wir deshalb die korrespondierenden Matrixblöcke ohne große Verluste approximieren können .

Um eine \mathcal{H} -Matrix aufstellen zu können, müssen wir die Indexmenge I partitionieren. Die Indexteilmengen repräsentieren wir dann durch einen \mathcal{H} -Baum. Aus diesem konstruieren wir einen minimal zulässigen \mathcal{H} -Produktbaum, der eine Partitionierung von $I \times I$ angibt. Die Konstruktion erfolgt mittels einer Zulässigkeitsbedingung, die untersucht, ob zwei Indexteilmengen $s, t \subset I$ zulässig sind, genauer

$$t \times s \subset I \times I \text{ ist } \begin{cases} \text{zulässig} & \text{für } \min \{\varnothing_\sigma, \varnothing_\tau\} \leq 2\eta \text{dist}(\sigma, \tau), \\ \text{unzulässig} & \text{sonst,} \end{cases}$$

wobei σ, τ mit s, t assoziierte Teilgebiete von Ω sind und \varnothing_σ den ČEBYŠEV-Radius von σ bezeichnet.

Zulässige Produktmengen $t \times s$ werden nicht mehr weiter unterteilt und die entsprechenden Matrixblöcke durch Matrizen niedrigeren Ranges, den sog. Rk -Matrizen, ersetzt. Der Rang k kann entweder fest vorgegeben werden oder aber er folgt aus dem alternativ vorzugebenden Approximationsfehler (Rangadaptivität).

Unzulässige Produktmengen $t \times s$ werden solange verfeinert, bis sie eine minimale Größe unterschreiten. Der entsprechende Matrixblock kann dann nicht durch eine Rk -Matrix ersetzt werden. Das resultierende Gesamtsystem nennen wir eine \mathcal{H} -Matrix.

Der Schrankensatz für die Spektralnorm stellt fest, daß die Approximierbarkeit einer Matrix A durch eine \mathcal{H} -Matrix $A_{\mathcal{H}}$ äquivalent ist zur Approximierbarkeit auf den zulässigen Blöcken $t \times s$. Diese Aussage bildet die Grundlage für die Untersuchungen in Kapitel 5.

1.5 Die Existenz der \mathcal{H} -Inversen

In Kapitel 5 beweisen wir zuerst die Existenz einer lokalen Niedrigrangapproximation von \vec{L}^{-1} , die auf zulässigen Teilgebieten exponentiell gegen \vec{L}^{-1} konvergiert. Mit den Erkenntnissen aus Kapitel 3 übertragen wir dieses Ergebnis auf die Matrix. Wegen des Schrankensatzes ist damit die \mathcal{H} -Approximierbarkeit der inversen Steifigkeitsmatrix bewiesen.

Für die Herleitung einer lokalen Niedrigrangapproximation von \vec{L}^{-1} beschreiten wir zwei Wege. In den Absätzen 5.1 und 5.2 verwenden wir eine schwache TAYLOR-Entwicklung $\vec{Q}^m \vec{u}$ von \vec{u} der Ordnung m . Wegen $\vec{u} = \vec{L}^{-1} \vec{f}$ definiert das eine Niedrigrangapproximation $\vec{Q}^m \vec{L}^{-1}$ von \vec{L}^{-1} . Dieser Fehler ist das Restglied der TAYLOR-Entwicklung, das im wesentlichen aus den m -ten Ableitungen von \vec{u} besteht und die

können wir im Falle konstanter Koeffizienten durch die rechte Seite abschätzen. Auf einem zulässigen Block ist also die Niedrigrangapproximation von $(\vec{L}^{-1})_h$ durch

$$(\vec{L}^{-1})_{\mathcal{H}} = \vec{J}^* \vec{Q}^m \vec{L}^{-1} \vec{J}$$

gegeben.

Für nichtkonstante Koeffizienten halten wir uns an [BH03], allerdings müssen wir nicht auf die GREENSche Funktion zurückgreifen. Dabei setzen wir \vec{u} in den Außenraum durch 0 fort und approximieren es durch stückweise konstante Funktionen, die auf die Menge der \vec{L} -harmonischen Funktionen projiziert wurden. Somit ist auch die Differenz wieder \vec{L} -harmonisch und kann selbst entsprechend approximiert werden. Auf diese Weise wird eine orthogonale Projektion \vec{S}^m definiert. Diese führt ebenfalls zu einer exponentiell konvergenten Niedrigrangapproximation $\vec{S}^m \vec{L}^{-1}$ von \vec{L}^{-1} .

Daß wir mit dieser in Absatz 5.3 vorgestellten Methode auch L^∞ -Koeffizienten erfassen, liegt daran, daß wir abwechselnd projizieren und approximieren. Beim TAYLOR-Ansatz wurden Approximationen und Projektionen am Stück durchgeführt, was \vec{L} -Harmonizität der m -ten partiellen Ableitungen voraussetzt.

In Absatz 5.4 schließen wir von der nun bewiesenen lokalen \mathcal{H} -Approximierbarkeit auf die globale \mathcal{H} -Approximierbarkeit, d.h. wir zeigen die Existenz einer \mathcal{H} -Matrix $\vec{C}_{\mathcal{H}}$, sodaß

$$\left\| (\vec{L}_h)^{-1} - \vec{C}_{\mathcal{H}} \right\|_2 \leq \varepsilon$$

für ein $\varepsilon > 0$ gilt.

1.6 Numerik

Für die numerischen Tests betrachten wir die LAMÉ-Gleichung mit springenden Koeffizienten und diskretisieren sie mit linearen Finiten Elementen. Die entstehenden Gleichungssysteme werden durch \mathcal{H} -Invertierung und ein mit der \mathcal{H} -Inversen vorkonditioniertes CG-Verfahren gelöst. In beiden Fällen ist die Konvergenzgeschwindigkeit von zweiter Ordnung.

Kapitel 2

Elliptische Differentialgleichungen zweiter Ordnung

2.1 Grundbegriffe

In diesem Abschnitt betrachten wir auf einem LIPSCHITZ-Gebiet $\Omega \subset \mathbb{R}^n$, $n \in \mathbb{N}$, beliebiger Geometrie und Rand $\Gamma := \partial\Omega$ die Randwertaufgabe

$$L[u](x) = f(x) \text{ für } x \in \Omega \quad (2.1)$$

mit verallgemeinerten DIRICHLET-Nullrandwerten

$$u(x) = 0 \text{ für } x \in \Gamma := \partial\Omega \quad (2.2)$$

mit $f : \Omega \rightarrow \mathbb{R}$ gegeben und $u : \bar{\Omega} \rightarrow \mathbb{R}$ gesucht. Weiter ist L ein Differentialoperator zweiter Ordnung der Form

$$Lu = L[u](x) = - \sum_{i,j=1}^n \frac{\partial}{\partial x_j} \left(c_{ij}(x) \cdot \frac{\partial}{\partial x_i} u(x) \right) \quad (2.3)$$

mit den ebenfalls gegebenen reellen Koeffizientenfunktionen $c_{ij} \in L^\infty(\Omega)$. Statt $\frac{\partial}{\partial x_i}$ schreiben wir auch ∂_i . Wir nehmen an, daß $c_{ij}(x) = c_{ji}(x)$ für alle $x \in \Omega$ gilt und fassen die Koeffizienten in der Matrix $C = C(x) = (c_{ij}(x))_{i,j=1,\dots,n}$ zusammen. Seien $\lambda_{min}(x)$ und $\lambda_{max}(x)$ ihr minimaler und maximaler Eigenwert für ein bestimmtes $x \in \Omega$ und

$$\lambda_{min} := \min_{x \in \Omega} \lambda_{min}(x), \quad \lambda_{max} := \max_{x \in \Omega} \lambda_{max}(x).$$

Der Differentialoperator bzw. die Differentialgleichung L heißen **elliptisch im Punkt** $x \in \Omega$, falls $C(x)$ positiv definit ist, d.h.

$$\langle C(x)\xi, \xi \rangle > 0 \quad \forall \xi \in \mathbb{R}^n \setminus \{0\}, \quad (2.4)$$

und **elliptisch in** Ω , falls Bedingung (2.4) für fast alle $x \in \Omega$ erfüllt ist. Für $x \in \Omega$ setzen wir

$$c(x) := \inf \left\{ \sum_{i,j=1}^n a_{ij}(x) \xi_i \xi_j \mid \xi \in \mathbb{R}^n \text{ mit } \|\xi\| = 1 \right\}.$$

Tatsächlich ist $c(x)$ ein Minimum und zugleich der kleinste Eigenwert von $C(x)$, also ist (2.4) äquivalent zu

$$\sum_{i,j=1}^n c_{ij}(x) \xi_i \xi_j \geq c(x) \|\xi\|_2^2 \quad \forall \xi \in \mathbb{R}^n \setminus \{0\}.$$

Wir nennen L **gleichmäßig elliptisch in** Ω , falls

$$\inf \{c(x) \mid x \in \Omega\} > 0,$$

d.h. für alle $x \in \Omega$ existiert ein $\vartheta > 0$, die sog. *Elliptizitätskonstante*, sodaß $c(x) \geq \vartheta$ und damit

$$\langle C(x) \xi, \xi \rangle = \sum_{i,j=1}^n c_{ij}(x) \xi_i \xi_j \geq c(x) \|\xi\|_2^2 \geq \vartheta \|\xi\|_2^2 > 0 \quad \forall \xi \in \mathbb{R}^n \setminus \{0\} \quad (2.5)$$

gilt (insbesondere ist die Wahl $\vartheta = \lambda_{\min}$ möglich). Das Verhältnis von größtem zu kleinstem Eigenwert der Matrix C bezeichnen wir mit

$$\kappa_C(x) := \frac{\lambda_{\max}(x)}{\lambda_{\min}(x)} \quad \text{bzw.} \quad \kappa_C := \frac{\lambda_{\max}}{\lambda_{\min}}. \quad (2.6)$$

2.2 Fundamentallösungen

Wir haben zwei Gründe für eine kurze Auseinandersetzung mit Singularitätenfunktionen. Zum einen finden \mathcal{H} -Matrizen ihren Ursprung in der Lösung linearer Gleichungssysteme, die im Zuge der Randelementmethode, also bei der numerischen Lösung von Integralgleichungen, entstehen. Für Integralgleichungen sind Singularitätenfunktionen von zentraler Bedeutung. Zum anderen besteht eine Möglichkeit, den Beweis der \mathcal{H} -Approximierbarkeit unter Verwendung von Singularitäten- oder GREENSchen Funktionen zu führen. Abschließend wollen wir die Gelegenheit nutzen, um uns klarzumachen, warum wir doch lieber ohne Singularitätenfunktionen auskommen und stattdessen mit schwachen Lösungen arbeiten wollen.

2.2.1 Singularitätenfunktionen

Bei den Definitionen und Begriffsbildungen halten wir uns ausschließlich an die in der HACKBUSCH-Schule gebräuchliche Konvention.

Definition 2.1 Eine Distribution $s : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{\pm\infty\}$ heißt **Singularitätenfunktion**¹ für den Differentialoperator $L = L_x$ aus (2.3), falls für alle $x \in \mathbb{R}^n$ und beliebige, aber feste $y \in \mathbb{R}^n$

$$L[s(x, y)] = \delta_y(x) \quad \text{bzw.} \quad L[s(\cdot, y)] = \delta_y \quad (2.7)$$

im distributionellen Sinne gilt. Ist L rotationssymmetrisch, verlangen wir zusätzlich

$$s(x, y) = s(\|x - y\|). \quad (2.8)$$

Beispiel 2.2 Für den LAPLACE-Operator lautet die Singularitätenfunktion

$$s(x, y) = \begin{cases} -\frac{1}{2\pi} \ln \|x - y\| & \text{für } n = 2, \\ \frac{1}{4\pi} \|x - y\|^{-1} & \text{für } n = 3. \end{cases}$$

Bemerkung 2.3 (a) Im Gegensatz zur GREENSchen Funktion, die wir in Absatz 2.2.2 kennen lernen, ist die Singularitätenfunktion unabhängig vom Definitionsbereich Ω der Differentialgleichung und ihren Randbedingungen: sie bezieht sich einzig und allein auf den Differentialoperator.

- (b) Mit $\delta_y = \delta_y(x) = \delta(y - x)$ bezeichnen wir die DIRACSche δ -Funktion. In der Tat existiert keine klassische Funktion δ_y , sodaß die Forderung (2.7) nur von einer Distribution δ_y erfüllt werden kann, d.h. unsere Lösungen s sind schwach bzw. distributionell.
- (c) Die Singularitäten von s verlaufen meist entlang der Geraden $x = y$. Die Singularitätenfunktion s muß nicht singulär sein — es genügt vollaufen, wenn dies auf ihre (höheren) Ableitungen zutrifft; vielmehr wollen wir, daß die Singularität möglichst schwach ausgeprägt ist. Eine Verstärkung der Singularität erzielen wir durch Differentiation von s (nach x oder y).
- (d) Die in Gleichung (2.8) geforderte Rotationssymmetrie will man nur dann haben, wenn auch der Differentialoperator L rotationssymmetrisch ist wie das z.B. beim LAPLACE-Operator der Fall ist [CZ92, S. 213]. Als Funktion von x erfüllt dann s für $x \neq y$ die homogene Differentialgleichung und hängt nur von der EUKLIDischen Norm $\|x - y\|$ ab, d.h. s ist im Punkt y rotationssymmetrisch, wenn y festgehalten wird und x läuft.

Singularitätenfunktionen kann man im Einzelfall mit Faltungen oder der FOURIER-Transformation konstruieren (siehe z.B. [CZ92, S. 212], [McI00], [Or80]). Im allgemeinen Fall jedoch ist die Herleitung von s eine komplizierte Angelegenheit. Für Operatoren mit konstanten Koeffizienten gewährleistet der Satz von MALGRANGE-EHRENPREIS aus dem Jahre 1955 die Existenz einer Singularitätenfunktion [Ze97, S. 186]; eine allgemeine Darstellung der Singularitätenfunktionen für solche Operatoren findet sich in [SS04, Formel (3.1.3)]. Weitere Literaturhinweise finden sich in [Ha87, Abs. 1.16].

¹Synonyme: Fundamentallösung, Elementarlösung.

Definition 2.4 Sei s die Singularitätenfunktion von L . Als **Grundlösung** von $L[u] = 0$ in Ω bezeichnen wir jede Distribution

$$\gamma : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}, \quad (x, y) \longmapsto s(x, y) + \phi(x, y), \quad (2.9)$$

falls die Funktion ϕ für festes $y \in \Omega$ ebenfalls $L[\phi] = 0$ bedient und in $C^2(\overline{\Omega})$ liegt.

Insbesondere ist $\phi \equiv 0 \in C^\infty(\overline{\Omega})$ mit $L[\phi] = 0$ und somit s selbst eine Grundlösung. Aus diesem Grund und weil die meisten Autoren zwischen s und γ nicht unterscheiden, identifizieren auch wir die beiden und arbeiten fortan nur mit der Singularitätshalfenfunktion. Das hat zur Konsequenz, daß diese eindeutig ist bis auf Lösungen der Gleichung $L[\phi] = 0$.

Die Singularitätenfunktion von L stellt, wie wir gleich sehen werden, einen Teil von L^{-1} und damit von der Lösung der Differentialgleichung dar. Wir betrachten nun wieder die inhomogene Differentialgleichung (2.1) mit den Randdaten (2.2) oder den NEUMANN-Randwerten

$$\frac{\partial}{\partial \nu} u(x) = g_N(x) \quad \text{für } x \in \Gamma, \quad (2.10)$$

wobei

$$\frac{\partial}{\partial \nu} u(x) = \partial_\nu u(x) = \langle \nu(x), \nabla u(x) \rangle \quad (2.11)$$

die **Normalenableitung** von u im Punkt x bezeichnet.

Definition 2.5 Ist s eine Fundamentallösung von L und ϕ integrierbar, so heißt

$$v(x) := \int_{\mathbb{R}^n} s(x, y) \phi(y) dy, \quad x \in \mathbb{R}^n$$

NEWTON- oder **Volumenpotential** von ϕ .

Das Volumenpotential erfüllt $L[v] = \phi$, d.h. für $\phi = f$ ist v Lösung von (2.1):

$$L_x \left[\int_{\Omega} s(x, y) f(y) dy \right] = f(x). \quad (2.12)$$

Konnten wir also bisher nur homogene Aufgaben $L[u] = 0$ lösen, haben wir mit dem Volumenpotential nun die Möglichkeit, für inhomogene Randwertprobleme (2.1) eine Lösung anzugeben [Hb97, S. 357]. Das ist genau die zentrale Eigenschaft der Singularitätenfunktion: kennen wir s , können wir die Differentialgleichung lösen, sogar in dem Fall, daß f eine Distribution ist. Weiterhin gilt [Gr01, S. 101]

$$\int_{\mathbb{R}^n} s(x, y) \cdot L[u](y) dy = \int_{\mathbb{R}^n} s(x, y) \cdot f(y) dy = u(x) = L^{-1}[f](x). \quad (2.13)$$

Damit ist endgültig klar, daß die Singularitätenfunktion ein wesentlicher Bestandteil von L^{-1} ist und wir sind unserem Ziel, eine Darstellung von L^{-1} für den Nachweis der

Approximationseigenschaft zu konstruieren, einen Schritt näher gekommen. Ebenso klar ist uns nun, warum dieses Konzept bei nichtlinearen Gleichungen nicht greift: dort hängt f nämlich von u ab. Was noch fehlt, ist die Berücksichtigung der Randbedingungen.

Setzen wir $f = 0$ in (2.1) und integrieren (2.7) partiell, erhalten wir die Lösung der homogenen Differentialgleichung als das Randintegral

$$u(x) = \int_{\Gamma} \frac{\partial}{\partial \nu} u(y) \cdot s(x, y) - \underbrace{u(y)}_{=0 \text{ nach (2.2)}} \cdot \frac{\partial}{\partial \nu_y} s(x, y) d\Gamma_y \text{ für } x \in \Omega. \quad (2.14)$$

Diese sog. **GREENSche Darstellungsformel** eröffnet einen Zugang, die Differentialgleichung in eine Integralgleichung zu überführen. Gleichung (2.14) ist eine FREDHOLMSche (*Rand-*) *Integralgleichung 3. Art*, d.h. der Integrationsbereich ist fix und die gesuchte Funktion (hier u) tritt innerhalb und außerhalb des Integrals auf und wird dabei noch mit einer Funktion multipliziert. Ein simples Beispiel einer Integralgleichung geben wir in Absatz 4.1.1.

Die Integraldarstellung (2.14) zeigt uns, daß der Wert von u in einem inneren Punkt nur von den Werten von u und $\frac{\partial}{\partial \nu} u$ auf dem Rand abhängt, was aber auch zur Folge hat, daß wir beide Typen von Randbedingungen benötigen. Sind keine NEUMANN-Randbedingungen gegeben, müssen wir entweder ein passendes $\frac{\partial}{\partial \nu} u$ finden — dann sind wir bei der **Integralgleichungsmethode** gelandet — oder die Abhängigkeit von $\frac{\partial}{\partial \nu} u$ eliminieren — diesen Weg beschreiten wir bei der Methode der GREENSchen Funktionen.

2.2.2 GREENSche Funktionen

Das Volumenpotential gibt eine Lösung der inhomogenen Differentialgleichung auf dem ganzen Raum \mathbb{R}^n an. Die GREENSche Darstellungsformel beschreibt die Lösung der homogenen Gleichung auf Ω und berücksichtigt dabei die Randwerte. Die Kombination der beiden versetzt uns in die Lage, für die Lösung des Randwertproblems (2.1, 2.2) die geschlossene Formel

$$u(x) = - \int_{\Omega} s(x, y) \cdot f(y) dy + \int_{\Gamma} s(x, y) \cdot \frac{\partial}{\partial \nu} u(y) d\Gamma_y, \quad x \in \Omega,$$

angeben zu können, vgl. [Hb96, S. 33]. Wir müssen das Randintegral für den Fall, daß $\partial_{\nu} u$ unbekannt ist, eliminieren.

Definition 2.6 *Eine GREENSche Funktionen (erster Art)² $g : \overline{\Omega} \times \overline{\Omega} \longrightarrow \mathbb{R}$ für die DIRICHLET-Randwertaufgabe (2.1, 2.2) ist eine Singularitätenfunktion, die*

$$g(x, y) = 0 \text{ für } x \in \Gamma, y \in \Omega$$

erfüllt.

²Eine GREENSche Funktionen zweiter Art erhalten wir bei der Lösung der NEUMANN-Randwertaufgabe [Hb96, S. 40].

Aus der Definition folgt sofort, daß sich mit der GREENSchen Funktion als Kern eine Lösung der Randwertaufgabe (2.1, 2.2) angeben läßt, nämlich

$$u(x) = L^{-1}[f](x) = K[f](x) = - \int_{\Omega} g(x, y) \cdot f(y) dy, \quad x \in \Omega. \quad (2.15)$$

Im Falle von Nichtnull-DIRICHLET-Randwerten g_D müssen wir zur rechten Seite den Term $\int_{\Gamma} g_D(y) \cdot \frac{\partial}{\partial \nu_y} g(x, y) d\Gamma_y$ addieren.

Die GREENSche Funktion ist also per definitionem eine Distribution, aber wir müssen in ihr etwas stärkeres, glatteres sehen, um den Randbedingungen Sinn zu geben [RR93, Abs. 5.5.2]. Für $x \neq y$ ist die GREENSche Funktion durchaus gutartig, wenn vielleicht auch nicht analytisch. Wir zerlegen die GREENSche Funktion in die Singularitätenfunktion und in die sog. **kompensierende Kernfunktion** g_c [Sh92, S. 142]. Letztere korrigiert s dahingehend, daß die Randbedingung (2.2) erfüllt wird:

$$g(x, y) = s(x, y) + g_c(x, y), \quad x, y \in \bar{\Omega}.$$

Wegen

$$\delta_y(x) = L_x[g(x, y)] = L_x[s(x, y) + g_c(x, y)] = \delta_y(x) + L_x[g_c(x, y)]$$

ist klar, daß g_c

$$L_x[g_c(x, y)] = 0 \text{ für } x, y \in \Omega$$

erfüllt und auf $\Omega \times \Omega$ keine Singularität besitzt. Im Gebietsinnern wird der Charakter der GREENSchen Funktion also im wesentlichen von der Singularitätenfunktion bestimmt. Weiterhin gelten (für $x \in \Omega$)

$$\begin{aligned} \int_{\Omega} (s + g_c) f dy &= u(x) = \int_{\Omega} s f dy + \int_{\Gamma} s \partial_{\nu} u d\Gamma_y, \\ \int_{\Omega} g_c f dy &= \int_{\Gamma} s \partial_{\nu} u d\Gamma_y. \end{aligned}$$

Die Existenz einer GREENSchen Funktion ist eng mit der Lösbarkeit des Randwertproblems verknüpft. Ein Existenz- und Eindeutigkeitsresultat für gleichmäßig elliptische Differentialgleichungen mit L^{∞} -Koeffizienten wird in [GW82] vorgestellt.

Die GREENSche Funktion können wir selten explizit angeben. Weil sie jedoch vor allem für theoretische Ausführungen von Bedeutung ist, müssen wir über Möglichkeiten verfügen, sie in geeigneter Weise approximieren zu können. In Betracht kommt z.B. die Aufspaltung in Singularitätenfunktion und kompensierenden Kern, der dann abzuschätzen ist, vgl. [Gr01, Satz 7.11].

2.2.3 Abschließende Bemerkungen zu Fundamentallösungen

Der Umgang mit Singularitäten- und GREENSchen Funktionen ist nicht zuletzt wegen der Existenzfrage sehr schwierig, die mit der Gutartigkeit der Koeffizienten steht und fällt.

Z.B. treten in der Kontinuumsmechanik die Materialparameter als Koeffizienten des Differentialoperators auf. Wenn diese Koeffizienten nicht konstant sind, wie das bei inhomogenen Materialien der Fall ist, dann ist die Existenz der Grund-/Fundamentallösung nicht automatisch gewährleistet. [Mcl00, S. 200] deutet zwar an, daß Homogenität nicht zwingend erforderlich ist, aber die in Natur und Technik häufiger vorkommende Inhomogenität verkompliziert die Betrachtungen erheblich.

Im Falle von Differentialgleichungssystemen hat die jüngere Forschung entsprechende Ergebnisse hervorgebracht, vgl. Absatz 2.4.3. Jedoch erfordern diese die Einführung neuer Funktionenräume, was wir als weiteren Grund anführen, doch lieber mit schwachen Lösungen zu operieren.

2.3 Schwache Lösungen

Die schwache Formulierung der Randwertaufgabe (2.1, 2.2) ist eine **Variationsgleichung** und lautet

$$a(u, v) = F(v) \quad \forall v \in V := H_0^1(\Omega) \quad (2.16a)$$

mit der zu L gehörenden Bilinearform

$$a(u, v) := \int_{\Omega} \sum_{i,j=1}^n c_{ij} \partial_i u \partial_j v \quad (2.16b)$$

und der Linearform

$$F(v) := (f, v)_{L^2(\Omega)} = \int_{\Omega} f v. \quad (2.16c)$$

Die Lösung $u \in V := H_0^1(\Omega)$ von (2.16a) nennen wir **schwache Lösung** der Randwertaufgabe (2.1, 2.2). Für den Differentialoperator gilt nun $L : V \rightarrow V'$.

Definition 2.7 Die Bilinearform $a : V \times V \rightarrow \mathbb{R}$ heißt **stetig** oder **beschränkt**, falls eine Konstante $c_S > 0$ existiert, sodaß

$$|a(v_1, v_2)| \leq c_S \cdot \|v_1\|_V \cdot \|v_2\|_V \quad \forall v_1, v_2 \in V$$

gilt. Sie heißt $\frac{1}{2}$ **symmetrisch**, falls

$$a(v_1, v_2) = a(v_2, v_1) \quad \forall v_1, v_2 \in V$$

erfüllt ist.

Zwar gibt es den Begriff der elliptischen Bilinearform, den elliptischen Randwertproblemen sind jedoch die koerziven Bilinearformen zugeordnet [Hb96, S. 130].

Definition 2.8 Die Bilinearform a heißt **V -koerziv**, falls sie stetig ist und zwei Konstanten $c_E > 0$ und $c_K \in \mathbb{R}$ existieren mit

$$a(u, u) \geq c_E \|u\|_V^2 - c_K \|u\|_{L^2(\Omega)}^2 \quad \forall u \in V := H_0^1(\Omega).$$

Definition 2.9 Die Bilinearform a heißt **V -elliptisch**, falls sie stetig und positiv definit ist, falls also eine positive reelle Konstante c_E existiert mit

$$a(v, v) \geq c_E \|v\|_V^2 \quad \forall v \in V := H_0^1(\Omega).$$

Wenn wir schon der Vollständigkeit halber auf die Koerzivität eingehen, erwähnen wir auch, daß diese mit der Elliptizität zusammenfällt, wenn L , wie das in der vorliegenden Abhandlung der Fall ist, nur aus dem Hauptteil besteht.

Lemma 2.10 Sei Ω beschränkt und L ein gleichmäßig elliptischer Differentialoperator mit L^∞ -Koeffizienten c_{ij} und nur aus dem Hauptteil bestehend. Dann ist die Bilinearform a H_0^1 -elliptisch.

Beweis: Siehe z.B. [Hb96] oder den Beweis zu Lemma 2.14, das den gleichen Sachverhalt für Systeme von partiellen Differentialgleichungen formuliert. Dort erkennen wir auch, daß die Konstante c_E für skalare wie Systeme von Differentialgleichungen dieselbe ist. \square

Satz 2.11 (LAX-MILGRAM) Sei Ω beschränkt und die Bilinearform a H_0^1 -elliptisch. Bezeichne L^{-1} die Inverse des Differentialoperators L . Dann besitzt die Variationsgleichung (2.16a) genau eine (schwache) Lösung $u \in H_0^1(\Omega)$ mit

$$\begin{aligned} |u|_{H^1(\Omega)} &\leq \frac{1}{c_E} |f|_{H^{-1}(\Omega)} \quad \forall f \in H^{-1}(\Omega), \\ \|L^{-1}\|_{\mathcal{L}(H^{-1}(\Omega), H^1(\Omega))} &\leq \frac{1}{c_E}. \end{aligned}$$

Beweis: Siehe z.B. [Hb96, S. 136]. \square

2.4 Lineare Systeme von Differentialgleichungen

2.4.1 Grundbegriffe

Sei $\Omega \subset \mathbb{R}^n$, $n \in \mathbb{N}$, ein Gebiet beliebiger Geometrie mit Rand $\Gamma := \partial\Omega$ und

$$\vec{u} := (u_j)_{j=1, \dots, q} : \Omega \longrightarrow \mathbb{R}^q, \quad q \in \mathbb{N} \setminus \{1\}.$$

Die Schreibweise $\vec{u} \in V^q := H_0^1(\Omega, \mathbb{R}^q)$ bedeutet, daß jede Komponentenfunktion u_j Element von $H_0^1(\Omega)$ ist. Wir versehen $H^m(\Omega, \mathbb{R}^q)$, $m \in \mathbb{N}$, mit der Norm

$$\|\vec{u}\|_{H^m(\Omega, \mathbb{R}^q)} := \sqrt{\sum_{i=1}^q \sum_{|\alpha| \leq m} \int_{\Omega} |D^\alpha u_i|^2 dx} \quad (2.17a)$$

und der zugehörigen Seminorm

$$|\vec{u}|_{H^m(\Omega, \mathbb{R}^q)} := \sqrt{\sum_{j=1}^q \sum_{|\alpha|=m} \int_{\Omega} |D^\alpha u_j|^2 dx}. \quad (2.17b)$$

Die Norm einer Funktion $\vec{f} = (f_i)_{i=1, \dots, q} \in L^2(\Omega, \mathbb{R}^q)$ lautet dann

$$\|\vec{f}\|_{L^2(\Omega, \mathbb{R}^q)} = \sqrt{\sum_{i=1}^q \|f_i\|_{L^2(\Omega)}^2} = \sqrt{\sum_{i=1}^q \int_{\Omega} f_i^2 dx}.$$

Ein **lineares System von $q \geq 2$ Differentialgleichungen zweiter Ordnung** ist gegeben durch

$$\sum_{j=1}^q L_{ij}[u_j](x) = f_i(x) \quad \text{für } x \in \Omega \text{ und } i = 1, \dots, q, \quad (2.18)$$

oder kurz

$$\vec{L}\vec{u} = \vec{f}$$

in Matrixschreibweise. Der **Matrixdifferentialoperator**

$$\vec{L} : V^q \longrightarrow (V^q)' = H^{-1}(\Omega, \mathbb{R}^q)$$

besteht aus den Differentialoperatoren L_{ij} , die zwar von der Bauart wie in (2.3) sind, also

$$L_{ij}u_j(x) = - \sum_{k,l=1}^n \frac{\partial}{\partial x_l} \left(c_{ijkl}(x) \frac{\partial}{\partial x_k} u_j(x) \right)$$

(wir betrachten wiederum nur den Hauptteil), aber sie müssen nicht elliptisch sein. Die reellen Koeffizientenfunktionen

$$c_{ijkl} \in L^\infty(\Omega), \quad i, j = 1, \dots, q, \quad k, l = 1, \dots, n,$$

bilden die Komponenten eines $q \times q \times n \times n$ -Tensors ${}^{(4)}C$, den wir hier als $nq \times nq$ -Matrix \vec{C} schreiben, d.h. als Matrix, deren Komponenten C_{ij} , $i, j = 1, \dots, q$, wiederum $n \times n$ -Matrizen sind. Wir fordern von den Koeffizienten die Symmetrie

$$c_{ijkl} = c_{jikl} = c_{jilk}, \text{ d.h. } C_{ij}^\top = C_{ij} = C_{ji} := (c_{jikl})_{k,l=1, \dots, n} \quad (2.19)$$

(bei der Elastizität wird selbst im Fall allgemeinsten Materialsymmetrie eine stärkere Symmetrie gefordert).

In Erweiterung von Absatz 2.1 nennen wir \vec{L} **gleichmäßig elliptisch in Ω** , falls für fast³ alle $x \in \Omega$ eine Konstante $\vartheta > 0$ existiert, sodaß

$$\sum_{i,j=1}^q \left(\sum_{k,l=1}^n c_{ijkl}(x) \xi_{k+(i-1)n} \xi_{l+(j-1)n} \right) \geq \vartheta \cdot \|\xi\|_2^2 \quad \forall \xi \in \mathbb{R}^{nq} \setminus \{0\} \quad (2.20)$$

³Diese Einschränkung ergibt sich wegen der L^∞ -Koeffizienten, s. [Hb96, S. 133].

gilt, i.e. die sog. LEGENDRE-Bedingung⁴ [Gi93, S. 8]. Insbesondere ist klar, daß aus der Elliptizität aller L_{ij} die Elliptizität von \vec{L} folgt, aber nicht umgekehrt. Das hat zur Konsequenz, daß wir die Matrizen C_{ij} nicht als positiv definit voraussetzen dürfen. Neben diesem erweiterten Elliptizitätsbegriff existiert noch der etwas abstraktere Begriff nach AGMON-DOUGLIS-NIRENBERG, siehe [Hb96, Definition 12.1.2], [WRL95, S. 376ff].

Wir setzen

$$\xi := \nabla \vec{u}(x) := \begin{pmatrix} \nabla u_1(x) \\ \vdots \\ \nabla u_q(x) \end{pmatrix} \in \mathbb{R}^{nq}.$$

Infolge der gleichmäßigen Elliptizität gilt

$$\langle \vec{C}(x) \nabla \vec{u}(x), \nabla \vec{u}(x) \rangle = \sum_{i,j=1}^q \sum_{k,l=1}^n c_{ijkl}(x) \cdot \partial_k u_i(x) \cdot \partial_l u_j(x) \geq \vartheta \cdot \|\nabla \vec{u}(x)\|_2^2$$

für fast alle $x \in \Omega$. Mit der Integration über ein Teilgebiet $G \subset \Omega$ folgt

$$\int_G \langle \vec{C} \nabla \vec{u}, \nabla \vec{u} \rangle \geq \vartheta \cdot \int_G \langle \nabla \vec{u}, \nabla \vec{u} \rangle. \quad (2.21)$$

2.4.2 Die LAMÉ-Gleichung

Wir kommen nun auf ein spezielles Differentialgleichungssystem, nämlich das der statischen, also zeitunabhängigen, linearen Elastizität zu sprechen. Die Besonderheit hier ist, daß das System quadratisch, d.h. $q = n$ ist.

Wir betrachten den einfachsten Fall und setzen Linearität bei den geometrischen wie physikalischen Eigenschaften voraus. Weitere Vereinfachungen erzielen wir durch die Berücksichtigung von Materialsymmetrien, d.h. das Material weist entlang einer oder mehrerer Achsen bzw. Ebenen identisches Verhalten auf. Bei der Isotropie ist die Richtungsabhängigkeit der physikalischen Eigenschaften aufgehoben, sodaß uns zur Beschreibung des Materials die zwei Parameter

$$\left. \begin{array}{l} \lambda \in]0, \infty[\text{ „LAMÉ-Konstante“} \\ \mu \in]0, \infty[\text{ „Gleitmodul“, „KIRCHHOFF-Konstante“} \end{array} \right\} \quad (2.22)$$

genügen.

⁴Der Vollständigkeit halber sei erwähnt, daß es auch eine LEGENDRE-HADAMARD-Bedingung gibt:

$$\sum_{i,j=1}^q \left(\sum_{k,l=1}^n c_{ijkl}(x) \xi_k \xi_l \right) \eta_i \eta_j \geq \vartheta \cdot \|\xi\|_2^2 \quad \forall \xi \in \mathbb{R}^n \setminus \{0\} \quad \forall \eta \in \mathbb{R}^q \setminus \{0\},$$

vgl. [Gi83, S. 76], [McI00, S. 119]. Die LEGENDRE-Bedingung impliziert die LEGENDRE-HADAMARD-Bedingung vermöge $\xi_{k+(i-1)n} = \xi_k \eta_i$.

Als Differentialgleichungssystem erhalten wir dann die sog. LAMÉ- oder CAUCHY-NAVIER-Gleichung

$$-(\lambda + \mu) \cdot \nabla (\operatorname{div}(\vec{u})) - \mu \cdot \Delta \vec{u} = \vec{f}, \quad (2.23)$$

die unter Berücksichtigung von noch vorzugebenden Randbedingungen nach der Verschiebung $\vec{u} : \Omega \rightarrow \mathbb{R}^q$, $q = n$, zu lösen ist.

2.4.3 Fundamentallösungen

Liegt ein System von Differentialgleichungen vor, wird aus der Singularitätenfunktion eine Matrix, weil bei der Randelementmethode beim Testen von links und rechts jeweils eine vektorwertige Funktion aufmultipliziert wird.

Definition 2.12 Sei Id die $q \times q$ -Einheitsmatrix. Eine Abbildung $\vec{s} : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^{q \times q}$ heißt **Singularitätenmatrix**⁵ für den Differentialoperator $\vec{L} = \vec{L}_x$, falls

$$\vec{L}[\vec{s}(x, y)] = \delta_y(x) \cdot Id \quad (2.24)$$

gilt [Hb97, S. 337]. Im Falle rotationssymmetrischer \vec{L} verlangen wir zusätzlich

$$\vec{s}(x, y) = \vec{s}(\|x - y\|). \quad (2.25)$$

Beispiel 2.13 Für die dreidimensionale Elastizität ist

$$\vec{s}(x, y) = \frac{\lambda + 3\mu}{2(\lambda + 2\mu)} \cdot \left(s(x, y) \cdot Id + \frac{\lambda + \mu}{4\pi(\lambda + 3\mu)} \cdot \frac{(x - y)(x - y)^\top}{|x - y|^3} \right),$$

wobei s die Singularitätenfunktion des LAPLACE-Operators ist.

Ein Existenzsatz für GREENSche Matrizen $\vec{g} : \bar{\Omega} \times \bar{\Omega} \rightarrow \mathbb{R}^{q \times q}$ ist Gegenstand von [DM95, speziell Theorem 1].

2.4.4 Schwache Form

Die schwache Form erhalten wir wie im skalaren Fall durch Multiplikation mit einer Testfunktion \vec{v} und partieller Integration bzgl. x . Sie lautet

$$\vec{a}(\vec{u}, \vec{v}) = \vec{F}(\vec{v}) \quad \forall \vec{v} \in V^q := H_0^1(\Omega, \mathbb{R}^q) \quad (2.26a)$$

mit der zu \vec{L} gehörenden Bilinearform $\vec{a} : V^q \times V^q \rightarrow \mathbb{R}$,

$$\vec{a}(\vec{u}, \vec{v}) := \int_{\Omega} \sum_{i,j=1}^q \sum_{k,l=1}^n c_{ijkl} \partial_k u_j \partial_l v_i \quad (2.26b)$$

⁵Synonyme: Fundamentallösung, Elementarlösung

und der Linearform $\vec{F} : V^q \longrightarrow \mathbb{R}$,

$$\vec{F}(\vec{v}) := \left(\vec{f}, \vec{v} \right)_{L^2(\Omega, \mathbb{R}^q)} = \int_{\Omega} \langle \vec{f}, \vec{v} \rangle = \sum_{i=1}^q \int_{\Omega} f_i v_i. \quad (2.26c)$$

Symmetrie, Stetigkeit, positive Definitheit und V^q -**Koerzivitat** der Bilinearform \vec{a} sind wie bei der mit dem skalaren Differentialoperator L assoziierten Bilinearform a definiert. Speziell nennen wir \vec{a} V^q -**elliptisch**, falls \vec{a} stetig ist und eine Konstante $c_E > 0$ existiert, soda

$$\vec{a}(\vec{v}, \vec{v}) \geq c_E \|\vec{v}\|_{V^q}^2 \quad \forall \vec{v} \in V^q.$$

Lemma 2.14 *Sei Ω beschrankt und \vec{L} ein gleichmaig elliptischer Differentialoperator mit L^∞ -Koeffizienten c_{ijkl} und nur aus dem Hauptteil bestehend. Dann ist die Bilinearform $\vec{a} \in H_0^1(\Omega, \mathbb{R}^q)$ -elliptisch.*

Beweis: Wir mussen zeigen, da \vec{a} stetig und positiv definit ist; die V^q -Elliptizitat folgt dann nach Definition. Seien $\vec{u}, \vec{v} \in V^q$. Dann ist

$$\left| \int c_{ijkl} \partial_k u_j \partial_l v_i \right| \leq \|c_{ijkl}\|_{L^\infty(\Omega)} \cdot \|\partial_k u_j\|_{L^2(\Omega)} \cdot \|\partial_l v_i\|_{L^2(\Omega)}.$$

Da Ω beschrankt ist, existiert eine positive Konstante, soda $\|c_{ijkl}\|_{L^\infty(\Omega)}$ beschrankt ist und damit ist auch \vec{a} beschrankt bzw. stetig. Kommen wir zur positiven Definitheit. Es ist

$$\vec{a}(\vec{v}, \vec{v}) = \int_{\Omega} \sum_{i,j=1}^q \sum_{k,l=1}^n c_{ijkl} \partial_k v_j \partial_l v_i = \int_{\Omega} \sum_{i,j=1}^q \langle \vec{C} \nabla \vec{v}, \nabla \vec{v} \rangle.$$

Nach Voraussetzung ist \vec{L} gleichmaig elliptisch, also gilt fur alle $\xi \in \mathbb{R}^{nq}$

$$\langle \vec{C} \xi, \xi \rangle = \sum_{i,j=1}^q \sum_{k,l=1}^n c_{ijkl}(x) \xi_{k+(i-1)n} \xi_{l+(j-1)n} \stackrel{(2.20)}{\geq} \vartheta \|\xi\|_2^2 = \vartheta \langle \xi, \xi \rangle.$$

Wir identifizieren daher ξ mit $\nabla \vec{v}(x)$ und haben

$$\vec{a}(\vec{v}, \vec{v}) = \int \langle \vec{C} \nabla \vec{v}, \nabla \vec{v} \rangle \geq \vartheta \int \langle \nabla \vec{v}, \nabla \vec{v} \rangle = \vartheta |\vec{v}|_{V^q}^2.$$

Mit der FRIEDRICHS-Ungleichung (A.3) ist $|\vec{v}|_{V^q}^2 \geq c_\Omega^{-1} \|\vec{v}\|_{L^2(\Omega, \mathbb{R}^q)}^2$, also

$$|\vec{v}|_{V^q}^2 = \frac{1}{2} (|\vec{v}|_{V^q}^2 + |\vec{v}|_{V^q}^2) \geq \frac{1}{2} |\vec{v}|_{V^q}^2 + \frac{1}{2c_\Omega} \|\vec{v}\|_{L^2(\Omega, \mathbb{R}^q)}^2 \geq \min \left\{ \frac{1}{2}, \frac{1}{2c_\Omega} \right\} \|\vec{v}\|_{V^q}^2.$$

Somit haben wir gezeigt, da \vec{a} positiv definit ist,

$$\vec{a}(\vec{v}, \vec{v}) \geq c_E \|\vec{v}\|_{V^q}^2, \quad c_E := \frac{\vartheta}{2} \min \left\{ \frac{1}{2}, \frac{1}{2c_\Omega} \right\},$$

und insgesamt V^q -elliptisch ist. **QED**

Satz 2.15 (LAX-MILGRAM) Sei Ω beschränkt und die Bilinearform \vec{a} V^q -elliptisch mit der Elliptizitätskonstanten c_E . Bezeichne \vec{L}^{-1} die Inverse des Differentialoperators \vec{L} . Dann besitzt die Variationsgleichung (2.26a) genau eine (schwache) Lösung $u \in H_0^1(\Omega, \mathbb{R}^q)$ mit

$$|\vec{u}|_{H^1(\Omega, \mathbb{R}^q)} \leq \frac{1}{c_E} \left| \vec{f} \right|_{H^{-1}(\Omega, \mathbb{R}^q)} \quad \forall \vec{f} \in H^{-1}(\Omega, \mathbb{R}^q).$$

Weiterhin gilt

$$\left\| \vec{L}^{-1} \right\|_{\mathcal{L}(H^{-1}(\Omega, \mathbb{R}^q), H^1(\Omega, \mathbb{R}^q))} \leq \frac{1}{c_E}.$$

Da \vec{a} ein Skalarprodukt ist, genügt bereits der RIESZsche Darstellungssatz.

Beispiel 2.16 (LAMÉ-Gleichung) Gehen wir von DIRICHLET-Nullrandbedingungen aus, so lautet die schwache Form von (2.23)

$$\vec{a}(\vec{u}, \vec{v}) = \int_{\Omega} (\lambda + \mu) \cdot \langle \nabla, \vec{u} \rangle \cdot \langle \nabla, \vec{v} \rangle + \int_{\Omega} \mu \cdot \langle \nabla \vec{u}, \nabla \vec{v} \rangle$$

für alle Testfunktionen $\vec{v} \in H_0^1(\Omega, \mathbb{R}^n)$.

Kapitel 3

Finite-Element-Methode

Wir setzen uns mit der Methode der finiten Elemente soweit auseinander, wie das die Fehlerabschätzungen für die \mathcal{H} -Approximationen der Finite-Element-Matrizen erfordern. Der Schwerpunkt liegt dabei auf der Darstellung der Finite-Element-Matrizen durch kontinuierliche Operatoren. In Absatz 3.5 überlegen wir uns z.T. ausführlich die zur Behandlung von Differentialgleichungssystemen notwendigen Modifikationen.

Sowohl für skalare als auch Systeme von partiellen Differentialgleichungen verwenden wir quasi-gleichmäßige Triangulationen. Sei τ eine zulässige Triangulation von Ω , s. [Hb96, Def. 8.3.7]. Bezeichnet h_T die längste Kante eines Elementes $T \in \tau$, so ist

$$h := \max \{h_T | T \in \tau\} \quad (3.1)$$

die längste in τ vorkommende Kante. Diese größte Kantenlänge h ist ein Maß für die Feinheit der Triangulation und direkt proportional zur Anzahl der Gitterpunkte, weshalb wir auch von **Schritt-** oder **Maschenweite** sprechen. Bezeichne ρ_T den **Innkreisradius** von $T \in \tau_h$.

Definition 3.1 Sei $h > 0$ und $\{\tau_h\}$ eine Familie zulässiger Triangulationen. Falls das Verhältnis $\frac{h_T}{\rho_T}$ für alle T aus $\{\tau_h\}$ durch dieselbe, also eine von T unabhängige Konstante beschränkt ist, nennen wir $\{\tau_h\}$ — und damit τ_h — **quasi-uniform** [Br97, S. 108], [Hb96, S. 170]. Falls die stärkere Bedingung

$$\frac{h}{\min \{\rho_T | T \in \tau_h\}} \leq \text{const}$$

erfüllt ist, sprechen wir von einer **uniformen Triangulation**.

Eine uniforme Triangulation mit $h \rightarrow 0$ nennen wir **regulär** [Hb96, S. 170]. Gelegentlich werden aber auch uniforme Triangulationen als regulär bezeichnet [Br97, S. 58].

Aus Gründen der Wahrung der Übersichtlichkeit verwenden wir ausschließlich lineare Ansatzfunktionen, was aber keine allzugroße Einschränkung darstellen sollte.

3.1 Ritz-Galerkin-Verfahren

Die zu lösende elliptische Differentialgleichung zweiter Ordnung (2.1) mit DIRICHLET-Nullrandwerten (2.2) wird in die Variationsgleichung (2.16a) überführt und mit dieser die **Variationsaufgabe**

$$\text{finde } u \in V := H_0^1(\Omega), \text{ soda\ss } a(u, v) = F(v) \quad \forall v \in V \quad (3.2)$$

formuliert. Beim RITZ-GALERKIN-Verfahren ersetzen wir den unendlichdimensionalen Raum V durch

$$V_h := \text{span} \{ \varphi_i \mid i \in I \} \subset V, \quad I := \{1, \dots, N\}, \quad \|\cdot\|_{V_h} = \|\cdot\|_V$$

und formulieren als neue Aufgabe

$$\text{finde } u_h \in V_h, \text{ soda\ss } a(u_h, v_h) = F(v_h) \quad \forall v_h \in V_h. \quad (3.3)$$

Das ist erlaubt, weil a und F wegen $V_h \subset V$ auch auf V_h definiert sind. Wir nennen die Funktion u_h die **Ritz-Galerkin-Lösung von (3.2) auf V_h** . Die Existenz und Eindeutigkeit von u_h gewährleistet der Satz von LAX-MILGRAM.

Für einen Koeffizientenvektor $\mathbf{w} := (\mathbf{w}_i)_{i=1, \dots, N} \in \mathbb{R}^N$ definieren wir die lineare und stetige Abbildung

$$J : \mathbb{R}^N \longrightarrow V_h, \quad \mathbf{w} \longmapsto J\mathbf{w} := w_h := \sum_{i=1}^N \mathbf{w}_i \varphi_i. \quad (3.4)$$

Die Abbildung J selbst ist isomorph [Hb96, Bem. 8.1.1], weshalb wir sie fortan **Koeffizientenisomorphismus** nennen. Sie kann als Zeilenvektor $(\varphi_1 \ \dots \ \varphi_N)$ dargestellt werden und besitzt die Umkehrung $J^{-1} : V_h \longrightarrow \mathbb{R}^N$. Weitere wichtige Eigenschaften tragen wir im Lemma 3.2 zusammen.

Setzen wir in (3.3) für v_h nun φ_i ein, ist $a(u_h, \varphi_i) = F(\varphi_i)$. Umgekehrt folgt für $v_h = J\mathbf{v}$

$$a(u_h, v_h) - F(v_h) = a(u_h, J\mathbf{v}) - F(J\mathbf{v}) = \sum_{i=1}^N \mathbf{v}_i (a(u_h, \varphi_i) - F(\varphi_i)) = \sum_{i=1}^N \mathbf{v}_i \cdot 0 = 0.$$

Damit lautet unser neues, zu (3.3) äquivalentes Problem

$$\text{finde } u_h \in V_h, \text{ soda\ss } a(u_h, \varphi_i) = F(\varphi_i) \quad \forall i \in I. \quad (3.5)$$

In der Konsequenz aus (3.5) suchen wir einen Koeffizientenvektor $\mathbf{u} = (\mathbf{u}_j)_{j=1, \dots, N} \in \mathbb{R}^N$ mit $J\mathbf{u} = u_h$, d.h. $a(u_h, \varphi_i) = a(J\mathbf{u}, \varphi_i) = \sum_{j=1}^N \mathbf{u}_j a(\varphi_j, \varphi_i)$. Setzen wir also

$$\left. \begin{array}{lll} l_{ij} := a(\varphi_j, \varphi_i), & L_h := (l_{ij})_{i,j=1, \dots, N} & \text{„Steifigkeitsmatrix“} \\ m_{ij} := (\varphi_i, \varphi_j)_{L^2(\Omega)}, & M := (m_{ij})_{i,j=1, \dots, N} & \text{„Massenmatrix“} \\ F_i := F(\varphi_i), & F_h := (F_i)_{i=1, \dots, N} & \text{„Lastvektor“} \end{array} \right\} \quad (3.6)$$

so sind (3.5) bzw. (3.3) äquivalent zur Lösung des linearen Gleichungssystems

$$L_h \mathbf{u} = F_h. \quad (3.7)$$

Zwischen L_h und $a(\cdot, \cdot)$ zum einen und F_h und $F(\cdot)$ zum anderen besteht der Zusammenhang

$$a(u_h, v_h) = a(J\mathbf{u}, J\mathbf{v}) = \sum_{i,j=1}^N u_j l_{ij} v_i = \langle L_h \mathbf{u}, \mathbf{v} \rangle, \quad (3.8a)$$

$$F(v_h) = F(J\mathbf{v}) = \sum_{i=1}^N v_i F(\varphi_i) = \langle F_h, \mathbf{v} \rangle. \quad (3.8b)$$

3.2 Der adjungierte Koeffizientenisomorphismus

Dem GELFAND-Dreier entspricht im Endlichdimensionalen

$$V_h = U_h = V'_h = \text{span} \{ \varphi_1, \dots, \varphi_N \},$$

jedoch trägt jeder dieser Räume eine andere Norm, nämlich die des Originals. Wir müssen beachten, daß die Dualnorm $\|\cdot\|_{V'_h}$ zwar auf V' definiert ist [Hb96, S. 148],

$$\|v'\|_{V'_h} := \sup \left\{ \frac{|(v', \tilde{v})_U|}{\|\tilde{v}\|_V} \mid \tilde{v} \in V_h \setminus \{0\} \right\},$$

aber nur auf V'_h eine Norm darstellt, weil nicht nur der $0 \in V'$ der Wert $0 \in \mathbb{R}$ zugeordnet wird:

$$\forall v' \in V'_h \subset V' : (v', \tilde{v})_U = 0 \implies \|v'\|_{V'_h} = 0.$$

Wir hatten ursprünglich $J : \mathbb{R}^N \longrightarrow V_h$ definiert; wegen der Inklusion $V_h \subset V = H_0^1(\Omega)$ können wir aber auch

$$J : \mathbb{R}^N \longrightarrow H_0^1(\Omega) \quad (3.9)$$

auslegen. Zwar ist J dann kein Isomorphismus mehr, aber diese Auffassung erlaubt die zu J adjungierte Abbildung

$$J^* : H^{-1}(\Omega) \longrightarrow \mathbb{R}^N, \quad \langle J^* u, \mathbf{w} \rangle := (u, J\mathbf{w})_{H^{-1}(\Omega) \times H_0^1(\Omega)} \quad \forall \mathbf{w} \in \mathbb{R}^N \quad (3.10)$$

nicht nur für Funktionen aus dem Ansatzraum $V_h = V'_h$, sondern für alle Funktionale aus $H^{-1}(\Omega)$ wie z.B. LJ zu erklären, vgl. Lemma 3.3.

Lemma 3.2 (Eigenschaften von J und J^*) Seien J und J^* die in (3.4) und (3.10) definierten Abbildungen.

- (a) $\ker(J^*) = \{u \in H^{-1}(\Omega) \mid J^*u = 0 \in \mathbb{R}^N\} = V_h^\perp$.
- (b) $J^* : V_h' = V_h \longrightarrow \mathbb{R}^N$ ist isomorph (aber $J^*|_{V_h'} \neq J^{-1}$, weil J^* von der Wahl des Skalarproduktes abhängt, nicht so aber J^{-1}).
- (c) Die Komposition $J^*J : \mathbb{R}^N \longrightarrow \mathbb{R}^N$ ist eine invertierbare $N \times N$ -Matrix.

Beweis: siehe [Hb96, S. 149].

□

Aussagen über die Beschränktheit von J hängen von der Triangulation ab, vgl. Satz 3.4.

3.3 Koeffizientenisomorphismus und FE-Matrizen

Lemma 3.3 (FE-Matrizen) Sei L der elliptische Differentialoperator aus (2.3) und J der Koeffizientenisomorphismus gemäß (3.4) und (3.9) mit seiner Adjungierten J^* . Dann gilt für die RITZ-GALERKIN-Diskretisierung von L , also für die Steifigkeitsmatrix,

$$L_h = J^*LJ, \quad (3.11)$$

und für die Inverse der FE-Matrix

$$(L_h)^{-1} = (J^*LJ)^{-1}. \quad (3.12)$$

Von dieser unterscheidet sich die RITZ-GALERKIN-Diskretisierung des inversen Differentialoperators

$$(L^{-1})_h = J^*L^{-1}J. \quad (3.13)$$

Die Massenmatrix erhalten wir zu

$$M = J^*J \quad (3.14)$$

und den Lastvektor zu

$$F_h = J^*f. \quad (3.15)$$

Beweis: Wir beginnen mit der Steifigkeitsmatrix.

$$a(u_h, v_h) = (LJ\mathbf{u}, J\mathbf{v})_{L^2(\Omega)} = \langle J^*LJ\mathbf{u}, \mathbf{v} \rangle \stackrel{(3.8a)}{\implies} L_h = J^*LJ$$

Die Aussage für die Inverse der FE-Matrix erhalten wir schlichtweg durch Inversion und die für den inversen Differentialoperator durch Einsetzen von L^{-1} statt L . Kommen wir zur Massenmatrix. Seien $v_h, w_h \in V_h$ beliebig. Dann existieren genau zwei Vektoren $\mathbf{v}, \mathbf{w} \in \mathbb{R}^N$, sodaß $v_h = J\mathbf{v}, w_h = J\mathbf{w}$ und

$$\begin{aligned} (v_h, w_h)_{L^2(\Omega)} &= \int_{\Omega} v_h \cdot w_h = \int_{\Omega} \sum_{k=1}^N \mathbf{v}_k \varphi_k \cdot \sum_{l=1}^N \mathbf{w}_l \varphi_l \\ &= \sum_{k,l=1}^N \mathbf{v}_k \mathbf{w}_l \int_{\Omega} \varphi_k \varphi_l \\ &= \langle M\mathbf{v}, \mathbf{w} \rangle. \end{aligned}$$

Andererseits ist für alle $v_h, w_h \in V_h$

$$(v_h, w_h)_{L^2(\Omega)} = (J\mathbf{v}, J\mathbf{w})_{L^2(\Omega)} = \langle J^* J\mathbf{v}, \mathbf{w} \rangle,$$

also folgt $M = J^* J$. Schließlich bleibt die Formel für den Lastvektor zu zeigen. Für alle $v_h \in V_h$ existiert ein Vektor $\mathbf{v} \in \mathbb{R}^N$ mit $v_h = J\mathbf{v}$, sodaß

$$\langle F_h, \mathbf{v} \rangle = \sum_{k=1}^N F(\varphi_k) \mathbf{v}_k = F(v_h) = (f, v_h)_{L^2(\Omega)} = (f, J\mathbf{v})_{L^2(\Omega)} = \langle J^* f, \mathbf{v} \rangle,$$

also folgt mit (3.8b) die Behauptung. **QED**

Für die nächste Aussage benötigen wir das sog. **skalierte Skalarprodukt**

$$\langle \mathbf{x}, \mathbf{\eta} \rangle_h := h^n \cdot \sum_{i=1}^N \mathbf{x}_i \eta_i = h^n \cdot \langle \mathbf{x}, \mathbf{\eta} \rangle, \quad \mathbf{x}, \mathbf{\eta} \in \mathbb{R}^N, \quad (3.16)$$

mit der zugehörigen **skalierten euklidischen Norm**

$$\|\mathbf{x}\|_h := \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle_h} = \sqrt{h^n} \|\mathbf{x}\|_2. \quad (3.17)$$

Die Multiplikation mit $\sqrt{h^n}$ entspricht der Integration über Ω und erlaubt damit den (einfacheren) Vergleich mit der L^2 -Norm. Diskretisieren wir nämlich Ω mit N Elementen, deren Durchmesser zu h proportional ist, so ist deren Volumen zu h^n proportional, d.h. $\text{vol}(\Omega) \sim N \cdot h^n$. Also folgt

$$\frac{1}{Nh^n} \langle F_h, F_h \rangle_h = \frac{1}{N} \langle F_h, F_h \rangle \sim \frac{1}{\text{vol}(\Omega)} \int_{\Omega} f^2(x) dx = \frac{1}{\text{vol}(\Omega)} (f, f)_{L^2(\Omega)},$$

d.h. bis auf einen konstanten Faktor gilt $\langle F_h, F_h \rangle_h = (f, f)_{L^2(\Omega)}$. Die zugehörige Matrixnorm ist die Spektralnorm.

Satz 3.4 *Für eine quasi-gleichmäßige Triangulation und lineare Ansatzfunktionen existiert eine von der Maschen-/Schrittweite h bzw. der Anzahl N der Freiheitsgrade unabhängige Konstante $c_J > 0$ mit*

$$\frac{1}{c_J} \|\mathbf{x}\|_h \leq \|J\mathbf{x}\|_{L^2(\Omega)} \leq c_J \|\mathbf{x}\|_h \quad \forall \mathbf{x} \in \mathbb{R}^N.$$

Beweis: Siehe [Hb96, Satz 8.8.1]. **QED**

Die Konstante c_J ist also eine obere Schranke für $\|J\|_{\mathcal{L}(\mathbb{R}^N, L^2(\Omega))}$. Für normale Matrizen B , d.h. $B^\top B = BB^\top$, gilt $\rho(B) = \|B\|_2$; insbesondere sind symmetrisch-positiv-definite Matrizen normal, sodaß

$$\mu_{\max} = \rho(M) = \|M\|_2 = \|J^* J\|_2 = \|J\|_{\mathcal{L}(\mathbb{R}^N, L^2(\Omega))}^2 \leq c_J^2.$$

folgt. Mit $\mu_{\min} = \|M^{-1}\|_2^{-1}$ erhalten wir die Ungleichungskette

$$\sqrt{\mu_{\min}} \|\mathbf{x}\|_h \leq \|J\mathbf{x}\|_{L^2(\Omega)} \leq \sqrt{\mu_{\max}} \|\mathbf{x}\|_h \quad \forall \mathbf{x} \in \mathbb{R}^N.$$

Die Kondition der Massenmatrix schätzen wir unabhängig von ihrer Dimension,

$$\text{cond}(M) = \|M\|_2 \cdot \|M^{-1}\|_2 = \frac{\mu_{\max}}{\mu_{\min}}.$$

3.4 Fehler- und Konvergenzanalyse

Eine Projektion ist eine lineare Abbildung Q von einem linearen Raum V in einen Teilraum V_Q desselbigen, d.h. $Qv = v$ für alle $v \in V_Q$ bzw. $Q^2 = Q$. Die Projektion heißt (bzgl. des V -Skalarproduktes) orthogonal (auf V_Q), falls $(v - Qv, w)_V = 0$ für alle $v \in V$ und $w \in V_Q$. Das bedeutet, daß Q selbstadjungiert, also $Q = Q^*$ ist.

Lemma 3.5 Sei J die Abbildung gemäß (3.4) bzw. (3.9). Dann definiert

$$Q_h := J(J^*J)^{-1}J^*$$

eine $(\cdot, \cdot)_{L^2}$ -orthogonale Projektion von $L^2(\Omega)$ auf den Ansatzraum V_h . Aufgefaßt als Abbildung von V' nach V ist Q_h stetig.

Beweis: Daß Q_h eine orthogonale Projektion und somit stetig ist, ist klar. \square

Wenden wir Q_h auf die rechte Seite f der Differentialgleichung an, erhalten wir eine Funktion f_h im Ansatzraum V_h . Andererseits gilt offensichtlich $J^* = J^*Q_h$. Das bedeutet für die rechte Seite des linearen Gleichungssystems (3.7)

$$F_h \stackrel{\text{Lemma 3.3}}{=} J^*f = J^*Q_h f = J^*f_h,$$

d.h. Q_h projiziert f auf eine Funktion $f_h \in V_h$, die für das lineare Gleichungssystem denselben Effekt hat wie f .

Die Funktion $Q_h u$ ist die Bestapproximation von u im Ansatzraum V_h . Damit beträgt der Fehler in der $L^2(\Omega)$ -Norm, den wir bei der Ersetzung von V durch V_h begehen, mindestens

$$\|u - Q_h u\|_{L^2(\Omega)}. \tag{3.18}$$

Die **RITZ-Projektion** $R_h : V \rightarrow V_h$ ist definiert durch

$$R_h u \in V_h \text{ und } a(u - R_h u, v_h) = 0 \quad \forall v_h \in V_h \quad \forall u \in V$$

und somit eine bzgl. $a(\cdot, \cdot)$ orthogonale Projektion auf den Ansatzraum V_h .

Lemma 3.6 Sei J der Koeffizientenisomorphismus, u die schwache Lösung und L_h die Steifigkeitsmatrix. Dann gelten

$$\begin{aligned} R_h u &= u_h, \\ R_h &= J(L_h)^{-1}J^*L, \end{aligned}$$

d.h. R_h bildet die schwache Lösung der Variationsaufgabe (3.2) ab auf die Lösung von (3.3), die sog. **RITZ-GALERKIN-Lösung**.

Beweis: Siehe [Hb96, S. 155].

\square

Den sog. **FE-** oder **Diskretisierungsfehler**

$$\|u - R_h u\| = \|u - u_h\| \quad (3.19)$$

können/dürfen wir sowohl in der L^2 - als auch in der H^1 -Norm messen. Wir entscheiden uns für die L^2 -Norm, weil diese für die Abschätzung des \mathcal{H} -Approximationsfehlers die geeignete Wahl darstellt. Außerdem gilt in der L^2 -Norm

$$\|u - Q_h u\|_{L^2(\Omega)} \leq \|u - R_h u\|_{L^2(\Omega)}. \quad (3.20)$$

Das CÉA-Lemma [BS02, Abs. 2.8] liefert

$$\|u - u_h\|_{H^1(\Omega)} \leq \sqrt{\frac{c_S}{c_E}} \cdot \min_{v_h \in V_h} \|u - v_h\|_{H^1(\Omega)}$$

(diese Aussage gilt auch für die zugeordnete H^1 -Seminorm). Nun wissen wir zwar, daß u_h gegen u konvergiert, aber nicht mit welcher Geschwindigkeit. Bei einer quasi-uniformen Triangulation mit linearen finiten Elementen gilt

$$\min_{v_h \in V_h} \|u - v_h\|_{H^1(\Omega)} \leq c_1 \cdot h^\alpha \cdot |u|_{H^{1+\alpha}(\Omega)}$$

für $\alpha \in [0, 1]$ und $u \in V \cap H^{1+\alpha}(\Omega)$, d.h. wir nehmen $H^{1+\alpha}$ -Regularität der Lösung u an, die z.B. von [Hb96, Satz 9.1.16] gewährleistet wird. Sie bedingt per definitionem

$$|u|_{H^{1+\alpha}(\Omega)} \leq c_2 \cdot \|f\|_{H^{\alpha-1}(\Omega)}$$

[Br97, S. 83]. In der Folge haben wir

$$\|u - u_h\|_{H^1(\Omega)} \leq c_3 \cdot h^\alpha \cdot \|f\|_{H^{\alpha-1}(\Omega)} \quad \forall u \in V \cap H^{1+\alpha}(\Omega).$$

Die linke Seite wollen wir auf nichttriviale Weise nach unten durch $\|u - u_h\|_{L^2(\Omega)}$ abschätzen. Das gelingt mit Hilfe des NITSCHKE-Tricks, der zu

$$\frac{1}{h^\alpha} \|u - u_h\|_{H^{1-\alpha}(\Omega)} \leq \|u - u_h\|_{H^1(\Omega)}$$

führt. Offensichtlich gelten

$$\|u - u_h\|_{L^2(\Omega)} \leq \|u - u_h\|_{H^{1-\alpha}(\Omega)} \quad \text{und} \quad \|f\|_{H^{\alpha-1}(\Omega)} \leq \|f\|_{L^2(\Omega)}$$

und wir erhalten insgesamt

$$\|u - u_h\|_{L^2(\Omega)} \leq c_{FEM} \cdot h^{2\alpha} \cdot \|f\|_{L^2(\Omega)}, \quad \alpha \in [0, 1]. \quad (3.21)$$

3.5 Finite-Element-Methode bei Systemen von Differentialgleichungen

3.5.1 Ansatz- und Testräume

Wir legen nach wie vor N innere Gitterpunkte zugrunde, sodaß wir wegen der q -Dimensionalität des Differentialgleichungssystems zum k -ten Gitterpunkt die Freiheitsgrade

$$(k-1)q+1, \dots, (k-1)q+q = kq, \quad k = 1, \dots, N,$$

bekommen (knotenorientierte Diskretisierung, siehe Absatz 3.5.2). Nun brauchen wir für jeden Freiheitsgrad $s \in \{1, \dots, qN\}$ eine Ansatzfunktion

$$\vec{\eta}_s : \Omega \longrightarrow \mathbb{R}^q,$$

die wir aus den bereits bekannten, mit dem k -ten Gitterpunkt assoziierten Ansatzfunktionen φ_k und den kanonischen Einheitsvektoren $e_i \in \mathbb{R}^q$, $i = 1, \dots, q$, konstruieren. Für $s = (k-1)q + i$ setzen wir

$$\vec{\eta}(x) := \varphi_k(x) \cdot e_i.$$

Den Test- und Lösungsraum $V^q := H_0^1(\Omega, \mathbb{R}^q)$ approximieren wir durch den Ansatzraum

$$V_h^q := \langle \vec{\eta}_1, \dots, \vec{\eta}_{qN} \rangle = \text{span} \{ \varphi_1 e_1, \dots, \varphi_1 e_q, \varphi_2 e_1, \dots, \varphi_2 e_q, \dots, \varphi_N e_1, \dots, \varphi_N e_q \}.$$

Dann ist zur schwachen Formulierung (2.26a) äquivalent

$$\vec{a}(\vec{u}_h, \vec{v}_h) = \vec{F}(\vec{v}_h) \quad \vec{v}_h \in V_h^q$$

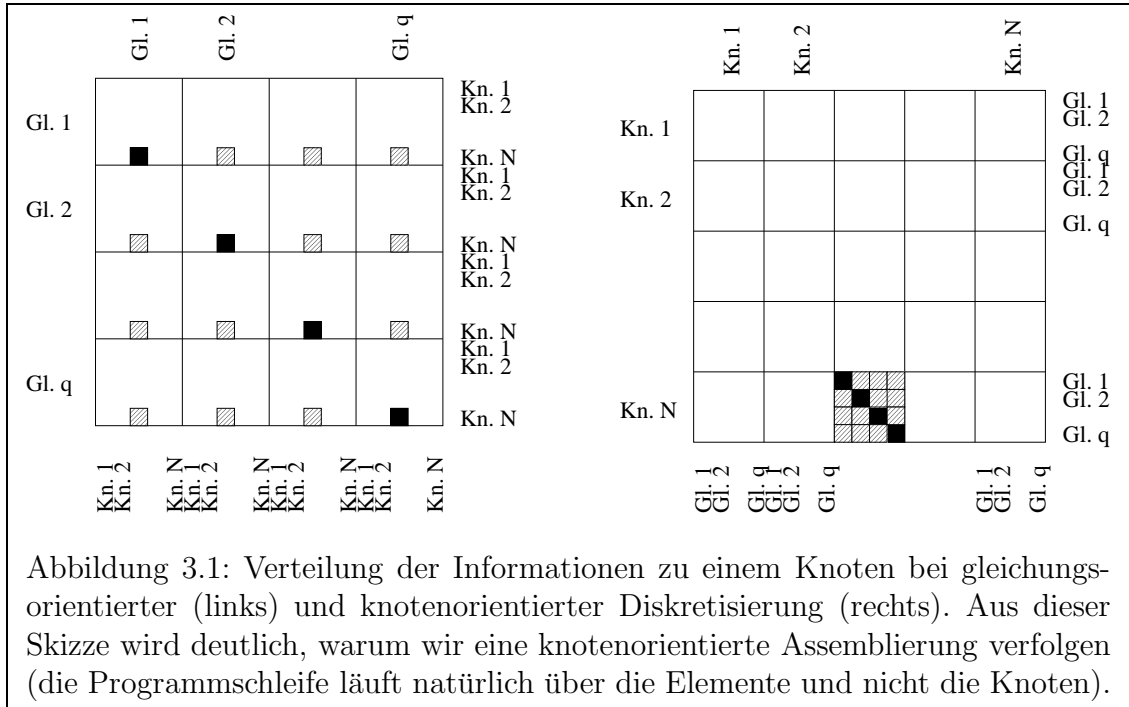
und das wiederum ist gleichbedeutend mit

$$\text{finde } \vec{u}_h \in V_h^q, \text{ sodaß } \vec{a}(\vec{u}_h, \vec{\eta}_s) = \vec{F}(\vec{\eta}_s) \quad \forall \vec{\eta}_s \in V_h^q. \quad (3.22)$$

Das sind qN -viele Gleichungen und exakt dieselben Gleichungen wie später im linearen Gleichungssystem.

3.5.2 Gleichungs- und knotenorientierte Diskretisierung

Bei der Konstruktion des Koeffizientenisomorphismus haben wir zwei Möglichkeiten: entweder diskretisieren wir die q Differentialgleichungen des Systems Gleichung für Gleichung, d.h. wir müssen jeden Freiheitsgrad q -mal anfassen, oder knotenweise, d.h. wir fassen jeden Freiheitsgrad nur einmal an und behandeln alle q Gleichungen. Den resultierenden Koeffizientenisomorphismus bezeichnen wir im ersten Fall mit \vec{J}_G (G für gleichungsweise) und im zweiten Fall mit \vec{J}_K (K für knotenweise) bzw. \vec{J} (weil diese Vorgehensweise unser Standard sein wird).



Jede Vorgehensweise erzeugt unterschiedliche Matrixstrukturen und \mathcal{H} -Bäume. Bei der knotenorientierten Vorgehensweise finden sich die Informationen zum k -ten Gitterpunkt in den Zeilen $(k-1)q+1$ bis kq , $k=1, \dots, N$, also in q aufeinanderfolgenden Zeilen, während die gleichungswise Diskretisierung die Informationen über die ganze Matrix verteilt und in die Zeilen $k+(i-1)N$, $i=1, \dots, q$, schreibt, siehe Abb. 3.1. Die Matrizen unterscheiden sich also nur durch Zeilen- und Spaltenvertauschungen und sind durch Multiplikation mit Permutationsmatrizen ineinander überführbar.

Obschon diese Wahl bzgl. der späteren Verwendung die ungünstige von beiden ist, gehen wir kurz auf \vec{J}_G ein, um die Eigenwerte der Massenmatrix leichter einzusehen. Wenn wir gleichungswise vorgehen, können wir auf den bereits aus (3.4) bekannten Koeffizientenisomorphismus J zurückgreifen. Bezeichnet e_i den i -ten Einheitsvektor, definieren wir

$$\vec{J}_G : \mathbb{R}^{qN} \longrightarrow V_h^q, \quad \vec{J}_G = \text{diag}(J, \dots, J), \quad \vec{\mathfrak{w}} \longmapsto \sum_{i=1}^q \sum_{k=1}^N \mathfrak{w}_{(i-1)N+k} \cdot \varphi_k \cdot e_i.$$

Der Vektor $\vec{\mathfrak{w}} \in \mathbb{R}^{qN}$ besteht aus den q Komponenten $\vec{\mathfrak{w}}_i \in \mathbb{R}^N$, die sich ihrerseits aus den N Komponenten $\mathfrak{w}_{i,k} = \mathfrak{w}_{(i-1)N+k} \in \mathbb{R}$ zusammensetzen.

Die Massenmatrix \vec{M}_G erhalten wir hierbei wieder durch die Hintereinanderausführung von \vec{J}_G und der Adjungierten

$$\vec{J}_G^* : (H_0^{-1}(\Omega))^q \longrightarrow \mathbb{R}^{qN}.$$

Seien $\vec{v}_h, \vec{w}_h \in V_h^q$ beliebig. Dann existieren genau zwei Vektoren $\vec{\mathbf{v}}, \vec{\mathbf{w}} \in \mathbb{R}^{qN}$, sodaß $\vec{v}_h = \vec{J}_G \vec{\mathbf{v}}, \vec{w}_h = \vec{J}_G \vec{\mathbf{w}}$ und

$$\begin{aligned}
(\vec{v}_h, \vec{w}_h)_{L^2(\Omega, \mathbb{R}^q)} &= \sum_{i=1}^q ((\vec{v}_h)_i, (\vec{w}_h)_i)_{L^2(\Omega)} \\
&= \sum_{i=1}^q \int_{\Omega} \sum_{k=1}^N \mathbf{v}^{(i-1)N+k} \varphi_k \cdot \sum_{k=1}^N \mathbf{w}^{(i-1)N+k} \varphi_k \\
&= \sum_{i=1}^q \sum_{k,l=1}^N \mathbf{v}^{(i-1)N+k} \mathbf{w}^{(i-1)N+l} \int_{\Omega} \varphi_k \varphi_l \\
&= \sum_{i=1}^q \langle M \vec{\mathbf{v}}_i, \vec{\mathbf{w}}_i \rangle \\
&= \langle \vec{M}_G \vec{\mathbf{v}}, \vec{\mathbf{w}} \rangle
\end{aligned}$$

mit der Blockdiagonalmatrix $\vec{M}_G = \text{diag}(M, \dots, M) \in \mathbb{R}^{qN \times qN}$. Andererseits ist

$$(\vec{v}_h, \vec{w}_h)_{L^2(\Omega, \mathbb{R}^q)} = \left(\vec{J}_G \vec{\mathbf{v}}, \vec{J}_G \vec{\mathbf{w}} \right)_{L^2(\Omega, \mathbb{R}^q)} = \left\langle \vec{J}_G^* \vec{J}_G \vec{\mathbf{v}}, \vec{\mathbf{w}} \right\rangle,$$

also gilt auch hier wieder

$$\vec{J}_G^* \vec{J}_G = \vec{M}_G.$$

Die durch \vec{J}_G induzierten und anhand von \vec{M}_G illustrierten Matrixstrukturen sind für die geometriebasierte Clusterungs-Strategie der \mathcal{H} -Matrizen nachteilhaft, insbesondere müßten wir alle Aussagen über die Arithmetik von \mathcal{H} -Matrizen neu untersuchen und beweisen. Deshalb verwenden wir überwiegend die knotenbasierte Vorgehensweise. Hier werden, wie bereits ausführlich dargelegt, alle Informationen einen bestimmten Knoten betreffend in q aufeinanderfolgenden Zeilen codiert, d.h. wir halten $k \in \{1, \dots, N\}$ fest und lassen $i \in \{1, \dots, q\}$ laufen:

$$J_k = \text{diag}(\varphi_k, \dots, \varphi_k)_{q \times q} : \mathbb{R}^q \longrightarrow V_h^q, \quad (\mathbf{w}^{(i-1)N+k})_{i=1, \dots, q} \longmapsto \varphi_k \sum_{i=1}^q \mathbf{w}^{(i-1)N+k} e_i.$$

Als Gesamtabbildung erhalten wir

$$\vec{J} = \left(J_1 \quad \dots \quad J_N \right)_{q \times qN} : \mathbb{R}^{qN} \longrightarrow V_h^q, \quad \vec{\mathbf{u}} \longmapsto \sum_{s=1}^{qN} \mathbf{u}_s \vec{\eta}_s = \sum_{i=1}^q \sum_{k=1}^N \mathbf{u}_{(k-1)q+i} \varphi_k e_i.$$

Vergleichen wir den Aufbau von \vec{J} mit \vec{J}_G , folgt die Existenz von Permutationsmatrizen $P \in \{0, 1\}^{qN \times qN}$ mit $P^{-1} = P^\top$ und

$$\vec{J} = \vec{J}_G P \quad \text{bzw.} \quad \vec{J}_G = \vec{J} P^\top. \quad (3.23)$$

Zwischen der zu \vec{J} adjungierten Abbildung $\vec{J}^* : (V_h^q)' = V_h^q \longrightarrow \mathbb{R}^{qN}$ und \vec{J}_G^* besteht der Zusammenhang

$$\vec{J}^* = \left(\vec{J}_G P \right)^* = P^\top \vec{J}_G^*, \quad \vec{J}_G^* = \left(\vec{J} P^\top \right)^* = P \vec{J}^*. \quad (3.24)$$

Um z.B. die Steifigkeitsmatrix als Komposition von Operatoren schreiben zu können, müssen wir den Wertebereich von \vec{J} erweitern. Interpretieren wir

$$\vec{J} : \mathbb{R}^{qN} \longrightarrow H^1(\Omega, \mathbb{R}^q),$$

so ist

$$\vec{J}^* : H^{-1}(\Omega, \mathbb{R}^q) \longrightarrow \mathbb{R}^{qN}, \quad \left\langle \vec{J}^* \vec{u}, \vec{w} \right\rangle := \left(\vec{u}, \vec{J} \vec{w} \right)_{L^2(\Omega, \mathbb{R}^q)}.$$

Lemma 3.7 (Eigenschaften von \vec{J} und \vec{J}^*) Seien μ_{\min} und μ_{\max} kleinster und größter Eigenwert von M . Dann gelten:

(a) Der kleinste und größte Eigenwert von \vec{M} ist μ_{\min} bzw. μ_{\max} .

(b) Die Abbildung \vec{J} ist beschränkt, genauer

$$\sqrt{\mu_{\min}} \|\vec{w}\|_h \leq \left\| \vec{J} \vec{w} \right\|_{L^2(\Omega, \mathbb{R}^q)} \leq \sqrt{\mu_{\max}} \|\vec{w}\|_h \quad \forall \vec{w} \in \mathbb{R}^{qN}.$$

(c) Die bei der Finiten-Element-Methode auftretenden Matrizen besitzen die Operatordarstellung

$$\begin{aligned} \vec{L}_h &= \vec{J}^* \vec{L} \vec{J} && \text{(Steifigkeitsmatrix),} \\ \left(\vec{L}^{-1} \right)_{\frac{h}{}} &= \vec{J}^* \vec{L}^{-1} \vec{J} && \text{(RITZ-GALERKIN-Diskretisierung von } \vec{L}^{-1}), \\ \vec{M} &= \vec{J}^* \vec{J} && \text{(Massenmatrix),} \\ \vec{F}_h &= \vec{J}^* \vec{f} && \text{(Lastvektor).} \end{aligned}$$

Beweis: Wir beginnen mit (c). Aus (3.24) folgt

$$\vec{M}_G = \vec{J}_G^* \vec{J}_G = P \vec{J}^* \vec{J} P^\top.$$

Multiplizieren wir von links und rechts mit $P^{-1} = P^\top$ und P , erhalten wir die gewünschte Darstellung

$$\vec{M} = P^\top \vec{M}_G P = \vec{J}^* \vec{J}.$$

Bezeichnet $\vec{L}_h = (l_{rs})_{r,s=1,\dots,qN}$ die Steifigkeitsmatrix, so ist für $r := (k-1)q + i$ und $s := (l-1)q + j$ gerade

$$l_{rs} = \vec{a}(\varphi_k e_i, \varphi_l e_j).$$

Zu allen $\vec{u}, \vec{v} \in V_h^q$ existieren Koeffizientenvektoren $\vec{u}, \vec{v} \in \mathbb{R}^{qN}$, soda

$$\begin{aligned} \bar{a}(\vec{u}_h, \vec{v}_h) &= \bar{a}(\vec{J}\vec{u}, \vec{J}\vec{v}) \\ &= \sum_{i,j=1}^q \sum_{k,l=1}^N \mathbf{u}^{(k-1)q+i} \cdot \mathbf{v}^{(l-1)q+j} \cdot l^{(k-1)q+i, (l-1)q+j} \\ &= \langle \vec{L}_h \vec{u}, \vec{v} \rangle \end{aligned}$$

gilt. Andererseits ist nach Definition von \vec{J}^*

$$\langle \vec{J}^* \vec{L} \vec{J} \vec{u}, \vec{v} \rangle := \left(\vec{L} \vec{J} \vec{u}, \vec{J} \vec{v} \right)_{L^2(\Omega, \mathbb{R}^q)} = \bar{a}(\vec{u}_h, \vec{v}_h),$$

also folgt $\vec{L}_h = \vec{J}^* \vec{L} \vec{J}$. Fr den Lastvektor $\vec{F}_h = (F_s)_{s=1, \dots, qN}$ gilt

$$F_s = \left(\vec{f}, \vec{\eta}_s \right)_{L^2(\Omega, \mathbb{R}^q)} \stackrel{s=(k-1)q+i}{=} \left(\vec{f}, \varphi_k e_i \right)_{L^2(\Omega, \mathbb{R}^q)} = (f_i, \varphi_k)_{L^2(\Omega)}.$$

Sei nun $\vec{v}_h \in V_h^q$ beliebig. Dann existiert ein Koeffizientenvektor $\vec{v} \in \mathbb{R}^{qN}$ mit $\vec{v}_h = \vec{J}\vec{v}$ und

$$\left(\vec{f}, \vec{v}_h \right)_{L^2(\Omega, \mathbb{R}^q)} = \sum_{s=1}^{qN} \mathbf{v}_s \left(\vec{f}, \vec{\eta}_s \right)_{L^2(\Omega, \mathbb{R}^q)} = \sum_{s=1}^{qN} \mathbf{v}_s F_s = \langle \vec{F}_h, \vec{v} \rangle$$

sowie

$$\left(\vec{f}, \vec{v}_h \right)_{L^2(\Omega, \mathbb{R}^q)} = \left(\vec{f}, \vec{J}\vec{v} \right)_{L^2(\Omega, \mathbb{R}^q)} = \langle \vec{J}^* \vec{f}, \vec{v} \rangle,$$

was $\vec{F}_h = \vec{J}^* \vec{f}$ zeigt.

Kommen wir zu (a). Wir haben eben $\vec{M} = P^\top \vec{M}_G P = \vec{J}^* \vec{J}$ bewiesen, die beiden Matrizen \vec{M} und \vec{M}_G sind also hnlich und besitzen dieselben Eigenwerte, insbesondere ist μ_{min} der kleinste und μ_{max} der grte Eigenwert.

Verbleibt (b) zu zeigen. Fr $\vec{w} \in \mathbb{R}^{qN}$ gilt

$$\begin{aligned} \left\| \vec{J}\vec{w} \right\|_{L^2(\Omega, \mathbb{R}^q)}^2 &= \left(\vec{J}\vec{w}, \vec{J}\vec{w} \right)_{L^2(\Omega, \mathbb{R}^q)} \\ &= \left\langle \vec{w}, \vec{J}^* \vec{J}\vec{w} \right\rangle_h \\ &= \left\langle \vec{w}, \vec{M}\vec{w} \right\rangle_h \\ &\leq \mu_{max} \left\langle \vec{w}, \vec{w} \right\rangle_h \\ &= \mu_{max} \left\| \vec{w} \right\|_h^2. \end{aligned}$$

Entsprechend erhalten wir die Abschtzung nach unten. ◻

Eigenwerte und Kondition der Massenmatrix sind also unabhngig von der Anordnung der Freiheitsgrade. Die Eigenvektoren von \vec{M} sind diejenigen von \vec{M}_G multipliziert mit P^\top . Aber was sind die Eigenwerte und -vektoren von \vec{M} und \vec{M}_G ? Seien

$0 < \mu_1 \leq \dots \leq \mu_N < \infty$ die Eigenwerte der Massenmatrix M und $x_1, \dots, x_N \in \mathbb{R}^N$ die zugehörigen Eigenvektoren, so sind die μ_i auch die Eigenwerte von \vec{M}_G mit den jeweils q Eigenvektoren

$$\begin{pmatrix} x_i^\top & 0^\top & \dots & 0^\top \end{pmatrix}^\top, \begin{pmatrix} 0^\top & x_i^\top & 0^\top & \dots & 0^\top \end{pmatrix}^\top, \dots, \begin{pmatrix} 0^\top & \dots & 0^\top & x_i^\top \end{pmatrix}^\top,$$

denn

$$\vec{M}_G \begin{pmatrix} \vdots \\ x_i \\ \vdots \end{pmatrix} = \text{diag}(0, \dots, 0, Mx_i, 0, \dots, 0) = \text{diag}(0, \dots, 0, \mu_i x_i, 0, \dots, 0) = \mu_i \begin{pmatrix} \vdots \\ x_i \\ \vdots \end{pmatrix}.$$

3.5.3 Fehler- und Konvergenzanalyse

Die in Abschnitt 3.4 definierten Projektionen erweitern wir auf Differentialgleichungssysteme durch Arrangierung in Blockdiagonalmatrizen.

Lemma 3.8 Sei Q_h die Projektion aus Lemma 3.5 und \vec{Q}_h die L^2 -orthogonale Projektion von $L^2(\Omega, \mathbb{R}^q)$ auf V_h^q . Dann besitzt \vec{Q}_h die Darstellung

$$\vec{Q}_h = \text{diag}(Q_h, \dots, Q_h) = \vec{J} \left(\vec{J}^* \vec{J} \right)^{-1} \vec{J}^*. \quad (3.25)$$

Sie bildet $H^{-1}(\Omega, \mathbb{R}^q)$ stetig ab in $H^1(\Omega, \mathbb{R}^q)$.

Beweis: Sei $\vec{u} \in V^q = H^1(\Omega, \mathbb{R}^q)$ und $\vec{v}_h \in V_h^q$. Dann ist

$$\left(\vec{u} - \vec{Q}_h \vec{u}, \vec{v}_h \right)_{L^2(\Omega, \mathbb{R}^q)} = \sum_{i=1}^q (\vec{u}_i - Q_h \vec{u}_i, \vec{v}_{h,i})_{L^2(\Omega)} = 0,$$

da $\vec{u}_i - Q_h \vec{u}_i \in V_h^\perp$ und $\vec{v}_{h,i} \in V_h$. Also ist $\vec{Q}_h = \text{diag}(Q_h, \dots, Q_h)$ die bzgl. $(\cdot, \cdot)_{L^2(\Omega, \mathbb{R}^q)}$ orthogonale Projektion auf V_h^q . Wir müssen uns noch ihrer Darstellung durch den Koeffizientenisomorphismus widmen. Nach Lemma 3.5 ist $Q_h = JM^{-1}J^*$, also

$$\begin{aligned} \vec{Q}_h &= \text{diag}(JM^{-1}J^*, \dots, JM^{-1}J^*) \\ &= \vec{J}_G \vec{M}_G^{-1} \vec{J}_G^* \\ &= \left(\vec{J} P^\top \right) \left(P \vec{M} P^\top \right)^{-1} \left(P \vec{J}^* \right) \\ &= \vec{J} P^\top P \vec{M}^{-1} P^\top P \vec{J}^* \\ &= \vec{J} \vec{M}^{-1} \vec{J}^*. \end{aligned}$$

Schließlich folgern wir $\vec{Q}_h \in \mathcal{L}(H^{-1}(\Omega, \mathbb{R}^q), H^1(\Omega, \mathbb{R}^q))$ aus der Linearität und Stetigkeit von Q_h . **QED**

Wie nicht anders zu erwarten, zeigt (3.25), daß \vec{Q}_h unabhängig von der Vorgehensweise ist, ob wir also gleichungs- oder knotenweise diskretisieren. Somit ist wie im skalaren Fall $\vec{Q}_h \vec{u}$ die Bestapproximation von \vec{u} über V_h^q , d.h. einen kleineren Fehler als

$$\left\| \vec{u} - \vec{Q}_h \vec{u} \right\|_{L^2(\Omega, \mathbb{R}^q)} \quad (3.26)$$

können wir bei der Ersetzung von V^q durch V_h^q gar nicht begehen.

Die definierende Gleichung für die **RITZ-Projektion**

$$\vec{R}_h : V^q = H_0^1(\Omega, \mathbb{R}^q) \longrightarrow V_h^q$$

ist

$$\vec{a}(\vec{u} - \vec{R}_h \vec{u}, \vec{v}_h) = 0 \quad \forall \vec{v}_h \in V_h^q \quad \forall \vec{u} \in V^q.$$

Offensichtlich projiziert \vec{R}_h bzgl. $\vec{a}(\cdot, \cdot)$ orthogonal auf den Ansatzraum V_h^q .

Lemma 3.9 Die RITZ-Projektion \vec{R}_h bildet die Lösung \vec{u} der Variationsgleichung (2.26a) ab auf die RITZ-GALERKIN-Lösung \vec{u}_h , d.h.

$$\vec{R}_h \vec{u} = \vec{u}_h.$$

Sie ist beschränkt und besitzt die Darstellung

$$\vec{R}_h = \text{diag}(R_h, \dots, R_h) = \vec{J} \left(\vec{L}_h \right)^{-1} \vec{J}^* \vec{L}. \quad (3.27)$$

Beweis: Sei \vec{u}_h die RITZ-GALERKIN-Lösung. Dann gilt für alle $\vec{v}_h \in V_h^q$

$$\begin{aligned} 0 &= \vec{F}(\vec{v}_h) - \vec{a}(\vec{u}_h, \vec{v}_h) \\ &= \left(\vec{f}, \vec{v}_h \right)_{H^{-1}(\Omega, \mathbb{R}^q) \times H^1(\Omega, \mathbb{R}^q)} - \left(\vec{L} \vec{u}_h, \vec{v}_h \right)_{H^{-1}(\Omega, \mathbb{R}^q) \times H^1(\Omega, \mathbb{R}^q)} \\ &= \left(\vec{f} - \vec{L} \vec{u}_h, \vec{v}_h \right)_{H^{-1}(\Omega, \mathbb{R}^q) \times H^1(\Omega, \mathbb{R}^q)} \\ &= \left(\vec{L}(\vec{u} - \vec{u}_h), \vec{v}_h \right)_{H^{-1}(\Omega, \mathbb{R}^q) \times H^1(\Omega, \mathbb{R}^q)} \\ &= \vec{a}(\vec{u} - \vec{u}_h, \vec{v}_h), \end{aligned}$$

also folgt nach Definition $\vec{u}_h = \vec{R}_h \vec{u}$. Um die Operatordarstellung von \vec{R}_h zu gewinnen, überlegen wir uns zunächst, daß

$$\begin{aligned} \vec{a}(\vec{u}_h, \vec{v}_h) &= \left(\vec{L} \vec{u}_h, \vec{v}_h \right)_{H^{-1}(\Omega, \mathbb{R}^q) \times H^1(\Omega, \mathbb{R}^q)} \\ &= \left(\vec{L} \vec{J} \vec{u}, \vec{J} \vec{v} \right)_{H^{-1}(\Omega, \mathbb{R}^q) \times H^1(\Omega, \mathbb{R}^q)} \\ &= \left\langle \vec{L}_h \vec{u}, \vec{v} \right\rangle_h, \\ \vec{F}(\vec{v}_h) &= \left(\vec{f}, \vec{v}_h \right)_{H^{-1}(\Omega, \mathbb{R}^q) \times H^1(\Omega, \mathbb{R}^q)} \\ &= \left(\vec{L} \vec{u}, \vec{J} \vec{v} \right)_{H^{-1}(\Omega, \mathbb{R}^q) \times H^1(\Omega, \mathbb{R}^q)} \\ &= \left\langle \vec{J}^* \vec{L} \vec{u}, \vec{v} \right\rangle_h \end{aligned}$$

für alle $\vec{v}_h \in V_h^q$ gelten. Bilden wir dann die Differenz, ist

$$\begin{aligned}
0 &= \vec{a}(\vec{u}_h, \vec{v}_h) - \vec{F}(\vec{v}_h) \\
&= \left\langle \vec{L}_h \vec{u} - \vec{J}^* \vec{L} \vec{u}, \vec{v} \right\rangle_h \\
&= \left\langle \vec{L}_h \cdot \left(\vec{u} - \left(\vec{L}_h \right)^{-1} \vec{J}^* \vec{L} \vec{u} \right), \vec{v} \right\rangle_h \\
&=_{\vec{L}_h = (\vec{L}_h)^\top} \left\langle \vec{u} - \left(\vec{L}_h \right)^{-1} \vec{J}^* \vec{L} \vec{u}, \vec{L}_h \cdot \vec{v} \right\rangle_h \\
&=_{\vec{L}_h = (\vec{J}^* \vec{L} \vec{J})} \left(\vec{J} \left(\vec{u} - \left(\vec{L}_h \right)^{-1} \vec{J}^* \vec{L} \vec{u} \right), \vec{L} \vec{J} \vec{v} \right)_{H^1(\Omega, \mathbb{R}^q) \times H^{-1}(\Omega, \mathbb{R}^q)}.
\end{aligned}$$

Wegen $\vec{J} \vec{u} = \vec{u}_h = \vec{R}_h \vec{u}_h = \vec{R}_h \vec{u}$ erhalten wir

$$0 = \left(\vec{R}_h \vec{u} - \vec{J} \left(\vec{L}_h \right)^{-1} \vec{J}^* \vec{L} \vec{u}, \vec{L} \vec{v}_h \right)_{H^1(\Omega, \mathbb{R}^q) \times H^{-1}(\Omega, \mathbb{R}^q)} \quad \forall \vec{v}_h \in V_h^q,$$

was (3.27) beweist. **QED**

Wir definieren als **FE-** oder **Diskretisierungsfehler**

$$\left\| \vec{u} - \vec{R}_h \vec{u} \right\|_{L^2(\Omega, \mathbb{R}^q)} = \left\| \vec{u} - \vec{u}_h \right\|_{L^2(\Omega, \mathbb{R}^q)}. \quad (3.28)$$

Natürlich können wir den Fehler auch in der H^1 -Norm messen, doch ist die L^2 -Norm unserer Sache dienlicher. Da bereits \vec{Q}_h die optimale Approximation liefert, ist

$$\left\| \vec{u} - \vec{Q}_h \vec{u} \right\|_{L^2(\Omega, \mathbb{R}^q)} \leq \left\| \vec{u} - \vec{u}_h \right\|_{L^2(\Omega, \mathbb{R}^q)}. \quad (3.29)$$

FE-Konvergenz

Für die Bestimmung der Konvergenzgeschwindigkeit müssen wir den FE-Fehler in der L^2 -Norm durch einen Ausdruck abschätzen, der eine möglichst hohe Potenz der Schrittweite h beinhaltet.

Unter Verwendung von V^q -Elliptizität und Stetigkeit der Bilinearform $\vec{a}(\cdot, \cdot)$ können wir die Gültigkeit des CÉA-Lemmas

$$\left\| \vec{u} - \vec{u}_h \right\|_{H^1(\Omega, \mathbb{R}^q)} \leq \sqrt{\frac{c_S}{c_E}} \cdot \min_{\vec{v}_h \in V_h^q} \left\| \vec{u} - \vec{v}_h \right\|_{H^1(\Omega, \mathbb{R}^q)}$$

für den Systemfall beweisen [Ba02, S. 282]. Alle weiteren Abschätzungen verlaufen analog zu Absatz 3.4 (vgl. auch [Ba02, S. 286ff]).

Lemma 3.10 (Konvergenzrate) *Sei \vec{u} die schwache Lösung der Variationsgleichung (2.26a) und $H^{1+\alpha}$ -regulär, d.h. $\vec{u} \in H_0^1(\Omega, \mathbb{R}^q) \cap H^{1+\alpha}(\Omega, \mathbb{R}^q)$, $\alpha \in [0, 1]$. Dann ist bei einer quasi-uniformen Triangulation mit linearen finiten Elementen die Konvergenz des Verfahrens von der Ordnung 2α , d.h.*

$$\left\| \vec{u} - \vec{u}_h \right\|_{L^2(\Omega, \mathbb{R}^q)} \leq c_{FEM} \cdot h^{2\alpha} \cdot \left\| \vec{f} \right\|_{L^2(\Omega, \mathbb{R}^q)}.$$

3.5.4 FE-Formulierung der LAMÉ-Gleichung

Aus (3.22) ergibt sich für uns die Suche eines Koeffizientenvektors $\vec{\mathbf{u}} \in \mathbb{R}^q$, sodaß

$$\vec{\mathbf{u}}_h = \vec{J}\vec{\mathbf{u}} = \sum_{r=1}^{qN} \mathbf{u}_r \vec{\eta}_r.$$

Wegen der Linearität der Bilinearform folgt

$$\vec{a}(\vec{\mathbf{u}}_h, \vec{\eta}_s) = \sum_{r=1}^{qN} \mathbf{u}_r \vec{a}(\vec{\eta}_r, \vec{\eta}_s) \stackrel{(3.22)}{=} \vec{F}(\vec{\eta}_s), \quad s = 1, \dots, qN,$$

und wir können $\vec{\mathbf{u}}$ durch Lösen des Gleichungssystems $\vec{L}_h \vec{\mathbf{u}} = \vec{F}_h$ berechnen.

Seien $i, j \in \{1, \dots, N\}$ und $k, l \in \{1, \dots, q\}$. Dann existiert zu jedem Paar $(r, s) \in \{1, \dots, qN\}^2$ genau ein Paar (i, j) und genau ein Paar (k, l) , sodaß

$$r = (i-1)q + k, \quad s = (j-1)q + l$$

gilt.

Die Steifigkeitsmatrix erhalten wir per

$$\vec{L}_h = (l_{rs})_{r,s=1,\dots,qN}, \quad l_{rs} := \vec{a}(\vec{\eta}_r, \vec{\eta}_s).$$

Mit dem KRONECKER-Delta δ_{kl} folgt

$$\begin{aligned} \vec{a}(\vec{\eta}_r, \vec{\eta}_s) &= (\lambda + \mu) \int_{\Omega} \langle \nabla, \vec{\eta}_r \rangle \langle \nabla, \vec{\eta}_s \rangle + \mu \int_{\Omega} \langle \nabla \vec{\eta}_r, \nabla \vec{\eta}_s \rangle \\ &= (\lambda + \mu) \int_{\Omega} \operatorname{div} \vec{\eta}_r \operatorname{div} \vec{\eta}_s + \mu \int_{\Omega} \langle \nabla(\varphi_i e_k), \nabla(\varphi_j e_l) \rangle \\ &= (\lambda + \mu) \int_{\Omega} \operatorname{div}(\varphi_i e_k) \operatorname{div}(\varphi_j e_l) + \mu \int_{\Omega} \langle \nabla(\varphi_i), \nabla(\varphi_j) \rangle \delta_{kl} \\ &= (\lambda + \mu) \int_{\Omega} \partial_k \varphi_i e_k \cdot \partial_l \varphi_j e_l + \mu \int_{\Omega} \langle \nabla(\varphi_i), \nabla(\varphi_j) \rangle \delta_{kl}. \end{aligned}$$

Kapitel 4

\mathcal{H} -Matrizen

In diesem Kapitel tragen wir nicht nur später benötigte Aussagen zusammen, sondern gewähren auch einen kleinen abrundenden Einblick in das Konzept der \mathcal{H} -Matrizen. Wer gewillt ist, sich tiefgehend mit dieser Thematik auseinanderzusetzen, sei auf die lehrbuchmäßige Einführung [BGH02] verwiesen, in der auch die Anwendungen nicht zu kurz kommen. Eine sehr schöne und ausführliche Darstellung der Arithmetik bietet [Gr01]; eine komprimierte Fassung dieser Arbeit ist [GH02]. Daneben wollen wir die beiden Originalarbeiten [Hb98] und [HK99] nicht vergessen.

4.1 Idee

Interpretieren wir die Steifigkeitsmatrix als Verknüpfungstabelle, so bedeutet ein Nichtnulleintrag an der Stelle (i, j) , daß die beiden Freiheitsgrade i und j bzw. die assoziierten Gitterpunkte miteinander kommunizieren. Dieser Informationsaustausch ist offenbar umso intensiver, je näher sich die Gitterpunkte sind.

Eine intelligente Informationsreduktion besteht gewiß nicht aus dem Wegwerfen von Einträgen, die gleich der Null oder ihr sehr nahe sind. Vielmehr müssen wir uns überlegen, wo sich Reduktion lohnt — bei einer dünnbesetzten Matrix wird kaum etwas zu gewinnen sein, aber die Inverse einer dünnbesetzten Matrix ist vollbesetzt — und wie wir reduzieren, damit die Kosten des Reduktionsalgorithmus das Ersparte nicht einfordern.

Eines der in Frage kommenden Verfahren ist die Methode der hierarchischen Matrizen, kurz \mathcal{H} -Matrizen. Sie entstammt den Randelementmethoden (BEM), die vollbesetzte Steifigkeitsmatrizen produzieren.

4.1.1 Eine Integralgleichung zur Motivation

Wir wollen Sinn und Zweck der \mathcal{H} -Matrizen anhand eines kurzen Beispiels, wie es in [BGH02] zu finden ist, motivieren. Gleichzeitig wird uns in Ergänzung zu Absatz

2.2.3 verdeutlicht, warum wir lieber mit schwachen Lösungen als mit Singularitäten- oder GREENSche Funktionen arbeiten.

Zu lösen ist die Integralgleichung

$$\int_0^1 \log|x-y| \cdot u(y) dy = f(x), \quad x \in [0, 1] \subset \mathbb{R}, \quad (4.1)$$

d.h. für eine rechte Seite $f : [0, 1] \rightarrow \mathbb{R}$ müssen wir $u(y)$ berechnen. Der Kern

$$k(x, y) := \log|x-y|$$

ist auf $[0, 1]^2$ definiert und entlang der Diagonalen $x = y$ singularär. Wählen wir zur numerischen Lösung einen GALERKIN-Ansatz, multiplizieren wir die Gleichung mit einer Testfunktion $v(x)$ und integrieren bzgl. x ,

$$\int_0^1 v(x) \cdot \int_0^1 \log|x-y| \cdot u(y) dy dx = \int_0^1 v(x) \cdot f(x) dx.$$

Anschließend ersetzen wir u, v durch eine Linearkombination von Ansatzfunktionen aus dem Ansatzraum $V_h := \{\varphi_1, \dots, \varphi_N\}$,

$$u(y) = \sum_{i=1}^N \mathbf{u}_i \varphi_i(y), \quad v(x) = \sum_{j=1}^N \mathbf{v}_j \varphi_j(x)$$

mit $\mathbf{u} = (\mathbf{u}_i)_{i=1, \dots, N}$, $\mathbf{v} = (\mathbf{v}_j)_{j=1, \dots, N} \in \mathbb{R}^N$. Also haben wir

$$\sum_{i,j=1}^N \mathbf{u}_i \mathbf{v}_j \int_0^1 \int_0^1 \log|x-y| \cdot \varphi_j(x) \cdot \varphi_i(y) dy dx = \sum_{j=1}^N \mathbf{v}_j \int_0^1 \varphi_j(x) \cdot f(x) dx$$

zu lösen. Setzen wir

$$G_{ij} := \int_0^1 \int_0^1 \log|x-y| \cdot \varphi_j(x) \cdot \varphi_i(y) dy dx, \quad F_j := \int_0^1 \varphi_j(x) \cdot f(x) dx,$$

so ist die linke Seite gleich

$$\sum_{i,j=1}^N \mathbf{u}_i \mathbf{v}_j G_{ij} = \langle \mathbf{u}, G\mathbf{v} \rangle = \langle G\mathbf{u}, \mathbf{v} \rangle, \quad G := (G_{ij})_{i,j=1, \dots, N}$$

und die rechte Seite gleich

$$\sum_{j=1}^N \mathbf{v}_j F_j = \langle F, \mathbf{v} \rangle, \quad F := (F_j)_{j=1, \dots, N},$$

also insgesamt

$$\langle G\mathbf{u}, \mathbf{v} \rangle = \langle F, \mathbf{v} \rangle \quad \text{bzw.} \quad G\mathbf{u} = F.$$

Problematisch ist dabei die Berechnung der Matrix G wegen der Kernsingularität.

Bemerkung 4.1 Bei genauerem Hinsehen stellen wir fest, daß $k(x, y)$ fern der Singularität in großen Teilgebieten und nahe der Singularität in kleinen Teilgebieten glatt ist. Diese Erkenntnis ist für das Verständnis der \mathcal{H} -Matrizen fundamental und findet sich in den \mathcal{H} -Strukturen wieder.

Wir nähern den Kern nun durch seine TAYLOR-Entwicklung

$$\tilde{g}(x, y) := \sum_{\nu=0}^{k-1} \frac{\partial_x^\nu g(x_0, y)}{\nu!} \cdot (x - x_0)^\nu = \sum_{\nu=0}^{k-1} g_\nu(x) \cdot h_\nu(y)$$

k -ter Ordnung, die ihrerseits, das suggeriert das Istgleichzeichen, ein **entarteter Kern** und deshalb von Vorteil ist [Hb97, S. 62]. Seien

$$x \in \tau := [a, b] \subset [0, 1] \quad \text{und} \quad y \in \sigma := [c, d] \subset [0, 1]$$

sowie $b < c$, d.h. die Intervalle τ und σ sind disjunkt, und $x_0 := 0.5(a + b)$. Dann gilt nach [BGH03, Lemma 1.3]

$$|g(x, y) - \tilde{g}(x, y)| \leq \frac{|x_0 - a|}{|c - b|} \cdot \left(1 + \frac{|c - b|}{|x_0 - a|}\right)^{-k}.$$

Da die geklammerte Summe größer als Eins ist, haben wir **exponentielle Konvergenz** auf $[a, b]$. Die Approximation selbst wird umso schlechter, je näher sich b und c kommen. Abhilfe schaffen wir, indem wir die Bedingung $b < c$ durch die stärkere Bedingung, die sog. (Standard-) Zulässigkeitsbedingung

$$\emptyset_\tau \leq \text{dist}(\sigma, \tau)$$

ersetzen. In diesem Falle gilt $|g - \tilde{g}| \leq 3^{-k}$, d.h. solange die Zulässigkeitsbedingung erfüllt ist, ist der Approximationsfehler unabhängig von den Intervallen beschränkt und exponentiell konvergent.

Wir definieren zwei Indexteilmengen $s, t \subset I := \{1, \dots, N\}$ und setzen

$$\sigma := \bigcup_{i \in s} \text{supp}(\varphi_i), \quad \tau = \bigcup_{j \in t} \text{supp}(\varphi_j).$$

Falls τ und σ zulässig sind, also die Zulässigkeitsbedingung erfüllen, ist g so glatt, daß wir es durch das leichter auszuwertende \tilde{g} ersetzen dürfen. Das ist äquivalent mit der Ersetzung von G_{ij} durch

$$\begin{aligned} \tilde{G}_{ij} &:= \int_0^1 \int_0^1 \tilde{g}(x, y) \cdot \varphi_j(x) \cdot \varphi_i(y) \, dy dx \\ &= \sum_{\nu=0}^{k-1} \int_0^1 g_\nu(x) \cdot \varphi_j(x) \, dx \cdot \int_0^1 h_\nu(y) \cdot \varphi_i(y) \, dy \end{aligned}$$

sofern (i, j) in einem zulässigen Cluster $t \times s$ liegen. In den unzulässigen, aber kleinen Blöcken müssen wir G_{ij} wie gewohnt berechnen.

Statt des Zweifachintegrals haben wir jetzt zwei Einfachintegrale, weil \tilde{g} entartet ist. Setzen wir noch

$$A := (A_{i\nu})_{i,\nu} \in \mathbb{R}^{N \times k}, \quad A_{i\nu} := \int_0^1 h_\nu(y) \cdot \varphi_i(y) dy,$$

$$B := (B_{j\nu})_{j,\nu} \in \mathbb{R}^{N \times k}, \quad B_{j\nu} := \int_0^1 g_\nu(x) \cdot \varphi_j(x) dx,$$

gilt $\tilde{G}_{ij} = (AB^\top)_{ij}$ und wir dürfen den Matrixblock $G|_{t \times s}$ durch das Produkt zweier Rang- k -Matrizen ersetzen.

Beispiel 4.2 Für die numerische Lösung unserer Integralgleichung (4.1) verwenden wir $N = 8$ stückweise konstante Ansatzfunktionen.

- (a) Zu den Indexteilmengen $t = \{1, 2\}$ und $s = \{3, 4\}$ gehören die Teilgebiete $\tau = [0, \frac{1}{4}]$ und $\sigma = [\frac{1}{4}, \frac{1}{2}]$. Wegen $\varnothing_\tau = \frac{1}{4}$ und $\text{dist}(\sigma, \tau) = 0$ ist die Zulässigkeitsbedingung unerfüllt, also sind die Blöcke unzulässig. Wir können diesen Block entweder nochmals unterteilen oder wir müssen die „exakten“ Matrixeinträge G_{ij} berechnen.
- (b) Zu den Indexteilmengen $t = \{1, 2\}$ und $s = \{5, 6\}$ gehören die Teilgebiete $\tau = [0, \frac{1}{4}]$ und $\sigma = [\frac{1}{2}, \frac{3}{4}]$. Wegen $\varnothing_\tau = \frac{1}{4}$ und $\text{dist}(\sigma, \tau) = \frac{1}{4}$ ist die Zulässigkeitsbedingung erfüllt, sodaß $t \times s$ zulässig ist und wir G_{ij} durch \tilde{G}_{ij} approximieren dürfen.

Im nachstehenden Bild sind die für unzulässig befundenen Blöcke grau gefärbt.

	1	2	3	4	5	6	7	8
1								
2								
3								
4								
5								
6								
7								
8								

4.1.2 Zielsetzung und Vorgehensweise

Wir wollen nun die in der (Steifigkeits-) Matrix enthaltene Information so reduzieren, daß der Rechen- und Speicheraufwand abnimmt und die informationsreduzierte Lösung mit der richtigen Lösung „verwandt“ ist, der neu hinzugekommene und von den \mathcal{H} -Matrizen verursachte Fehler sich also in der Größenordnung der schon vorhandenen Fehler bewegt.

Wir skizzieren grob die Transformation einer Matrix in eine \mathcal{H} -Matrix. Die genaue Bedeutung der Begriffe lernen wir in den Folgeabsätzen kennen.

Die Umstrukturierung einer Matrix in eine \mathcal{H} -Matrix erfolgt über einen \mathcal{H} -Produktbaum, dessen (Knoten und) Blätter Mengen von Indexpaaren sind. Mit einer *Zulässigkeitsbedingung* beurteilen wir die Knoten und Blätter als unzulässig oder zulässig.

Die zu zulässigen Blättern gehörenden Matrixblöcke werden als rangreduzierte Matrizen abgespeichert und heißen fortan *Rk-Matrizen*. Die unzulässigen Blätter erlauben uns nicht, in den entsprechenden Matrixblöcken eine Informationsreduktion vorzunehmen, ohne einen empfindlichen Fehler hinnehmen zu müssen. Wir müssen sie also unverändert, d.h. voll abspeichern und nennen sie deshalb auch *volle Matrizen*. Die Partitionierung einer Matrix in Rk- und volle Matrizen nennen wir \mathcal{H} -Matrix.

4.2 Hierarchische Strukturen

Wir haben eine Menge von Gitterpunkten, denen einen gewissen Anzahl von Freiheitsgraden zugeordnet ist. Die Freiheitsgrade zusammen ergeben dann die Indexmenge $I = \{1, \dots, N\}$ der Matrix.

Einer Konvertierung einer Matrix in eine \mathcal{H} -Matrix geht eine Partitionierung von I voraus, die durch einen Baum T_I dargestellt wird. Nun wird die Matrix aber durch Indexpaare $(i, j) \in I \times I$ beschrieben, weshalb wir einen Baum $T_{I \times I}$ benötigen, der eine Partitionierung von $I \times I$ angibt. Einen solchen gewinnen wir durch „Kreuzen“ von T_I mit sich selbst unter Beachtung einer Zulässigkeitsbedingung.

4.2.1 \mathcal{H} -Bäume

Wir zerteilen die Indexmenge I — und damit auch die Menge der Gitterpunkte — disjunkt in zwei Teilmengen I_1 und I_2 , entweder kardinalitätsbalanciert, d.h. $|I_1| \approx |I_2|$, oder geometrieorientiert, d.h. $\mathcal{O}_{\Omega_1} \approx \mathcal{O}_{\Omega_2}$, wobei

$$\Omega_i := \{x_k \in \Omega^{(h)} \mid k \in I_i\}$$

ein sog. **Cluster** ist. Im zweiten Schritt teilen wir I_1 und I_2 entsprechend auf usw. Diese Vorgehensweise heißt **binäre Raumzerlegung** [Gr01, Beispiel 3.10]. Die Unterteilung endet im Extremfall bei einelementigen Mengen, sie kann aber bei Vorgabe einer Mindestgröße bereits vorher abbrechen. Dargestellt wird diese Aufteilung durch den **hierarchischen Partitionsbaum**, kurz \mathcal{H} -Baum.

Definition 4.3 (\mathcal{H} -Baum) Sei $T_I = (V, E)$ ein Baum mit Knotenmenge V und Kantenmenge E . Für einen Knoten $v \in V$ definieren wir die Menge seiner Söhne zu $S(v) := \{w \in V \mid (v, w) \in E\}$. Wir nennen T_I einen \mathcal{H} -Baum (von I), falls

- (a) I die Wurzel von T_I ist,
- (b) jeder Knoten eine nichtleere Teilmenge von I ist und
- (c) jeder Knoten entweder keinen Sohn hat, also ein Blatt ist, oder die disjunkte Vereinigung seiner Söhne ist.

Notation 4.4 (Baumparameter und -strukturen) *Bezeichne*

- p_{T_I} die Höhe des Baumes T_I ,
- L_{T_I} die Menge der Stufennummern von T_I , auf denen sich Blätter befinden,
- $\mathcal{L}(T_I, l)$ die Menge der Blätter von T_I , die sich auf der Stufe l befinden,
- $\mathcal{L}(T_I)$ die Menge der Blätter von T_I ,
- $\mathcal{L}^+(T_I)$ die Menge der zulässigen Blätter von T_I .

Lemma 4.5 (Eigenschaften) *Sei T_I ein mit binärer Raumzerlegung kardinalitätsbalanciert aufgebauter \mathcal{H} -Baum der Indexmenge $I = \{1, \dots, N\}$. Dann beträgt die Baumhöhe $p_T = \mathcal{O}(\log N)$ und die Aufbaukosten für T_I belaufen sich auf $\mathcal{O}(N \log N)$.*

Beweis: Siehe [Gr01, Lemma 3.11].

QED

4.2.2 Zulässigkeitsbedingung

Definition 4.6 (Gebietsdurchmesser) *Sei $\Omega \subset \mathbb{R}^n$ ein Gebiet. Dann bezeichnen wir als Durchmesser von Ω*

$$\varnothing_\Omega := 2 \cdot \inf \{ \rho > 0 \mid \exists z \in \mathbb{R}^n : \Omega \subset B_\rho(z) \}.$$

Gelegentlich liest man für $\frac{1}{2}\varnothing_\Omega$ den Namen ČEBYŠEV-Radius. Eine gängige, aber hiervon verschiedene Definition des Durchmessers ist

$$\text{diam}(\Omega) := \sup \{ \|x - z\| \mid x, y \in \Omega \}.$$

Korollar 4.7 *In der Situation von Definition 4.6 gilt*

$$\frac{1}{2}\varnothing_\Omega \leq \text{diam}(\Omega) \leq \varnothing_\Omega.$$

Beweis: Wir setzen $r := \frac{1}{2}\varnothing_\Omega$. Seien $x_0, z_0 \in \Omega$ derart, daß $\|x_0 - z_0\|_2 = \text{diam}(\Omega)$. Wegen $\Omega \subset B_r(\zeta)$ für ein $\zeta \in \mathbb{R}^n$ muß aber auch $\|x_0 - z_0\|_2 \leq 2r = \varnothing_\Omega$ gelten. Widmen wir uns der ersten Ungleichung:

$$\begin{aligned} \frac{1}{2}\varnothing_\Omega &= \inf_{z \in \mathbb{R}^n} \{ \rho > 0 \mid \Omega \subset B_\rho(z) \} \\ &= \inf_{z \in \mathbb{R}^n} \{ \rho > 0 \mid \forall x \in \Omega : \|x - z\|_2 < \rho \} \\ &\leq \sup_{z \in \Omega} \{ \|x - z\|_2 \mid x \in \Omega \} \\ &= \text{diam}(\Omega). \end{aligned}$$

QED

Das Kreuzprodukt zweier \mathcal{H} -Bäume spiegelt die Struktur der \mathcal{H} -Matrix wider. Wie bereits angedeutet, besteht eine \mathcal{H} -Matrix aus vollen und ranggekürzten Untermatrizen. Eine Untermatrix wird ranggekürzt, wenn die sie beschreibende Menge von Indexpaaren zulässig ist.

Die Zulässigkeit überprüfen wir mit einer **Zulässigkeitsbedingung**. Sie kann mit einer TAYLOR-Entwicklung aus der Singularitätenfunktion des Differentialoperators abgeleitet werden. Die Praxis zeigt, daß die **Standardzulässigkeitsbedingung**

$$b = t \times s \subset I \times I \text{ ist } \begin{cases} \text{zulässig} & \text{für } \min \{ \varnothing_\sigma, \varnothing_\tau \} \leq 2\eta \text{dist}(\sigma, \tau) \\ \text{unzulässig} & \text{sonst} \end{cases} \quad (4.2)$$

mit den Clustern $\sigma := \{ x_i \in \Omega^{(h)} \mid i \in s \}$ und τ analog für alle Probleme eine zufriedenstellende Partitionierung erzeugt. Der Parameter $\eta > 0$ hängt vom Differentialoperator ab; die Wahl $\eta = 2$ hat sich bewährt (vgl. Bemerkung 4.11).

Durchmesser und Abstand zweier Cluster zu bestimmen ist keine simple Aufgabe. Damit wir bei der Auswertung der Zulässigkeitsbedingung unsere Zeit und Rechenleistung nicht vergeuden, verwenden wir die kleinstmöglichen Quader

$$Q(\sigma) := \left\{ x \in \mathbb{R}^n \mid \inf_{y \in \sigma} y_i \leq x_i \leq \sup_{y \in \sigma} y_i, i = 1, \dots, n \right\},$$

die alle Gitterpunkte beinhalten. Nun kann aber der Fall $Q(\sigma) \subset Q(\tau)$ eintreten, was zu $\text{dist}(Q(\sigma), Q(\tau)) = 0$ führt, obwohl vielleicht $\text{dist}(\sigma, \tau) > 0$ sein mag, d.h. ein eigentlich zulässiges Paar von Clustern wird als unzulässig eingestuft! Damit unser Verfahren an dieser Stelle nicht unnötig verfeinert und **ineffizient** wird, verwenden wir im Programm als **Standardzulässigkeitsbedingung**

$$\left. \begin{aligned} &\text{falls } \varnothing_{Q(\tau)} \leq \varnothing_{Q(\sigma)}, \\ &\text{dann } \min \{ \varnothing_{Q(\sigma)}, \varnothing_{Q(\tau)} \} \leq 2\eta \cdot \text{dist}(\sigma, Q(\tau)), \\ &\text{sonst } \min \{ \varnothing_{Q(\sigma)}, \varnothing_{Q(\tau)} \} \leq 2\eta \cdot \text{dist}(Q(\sigma), \tau), \end{aligned} \right\} \quad (4.3)$$

mit $\eta > 0$. Eine ausführliche Diskussion von Zulässigkeitsbedingungen findet sich u.a. in [Gr01, Absatz 3.2].

4.2.3 Produkt von \mathcal{H} -Bäumen

Definition 4.8 (a) Wir nennen einen Baum $T_{I \times I}$ einen \mathcal{H} -Produktbaum von $I \times I$, falls $T_{I \times I}$ ein \mathcal{H} -Baum von $I \times I$ und jeder Knoten von $T_{I \times I}$ von der Form $b = t \times s$ für $s, t \in I$ ist.

(b) Sei T_I ein \mathcal{H} -Baum von I . Dann zeichnet sich das **Kreuzprodukt** $T_I \otimes T_I$ durch

$$\forall t \times s \in T_I \otimes T_I : \begin{cases} s, t \in T_I, \\ S_{T_I \otimes T_I}(t \times s) = \{\tilde{t} \times \tilde{s} \mid \tilde{s} \in S_{T_I}(s), \tilde{t} \in S_{T_I}(t)\} \end{cases}$$

aus und ist selbst ein \mathcal{H} -Produktbaum.

Für die Konvertierung einer Matrix in eine \mathcal{H} -Matrix ist dieser Produktbaum viel zu komplex. Wir wollen stattdessen einen Baum haben, dessen Blätter entweder alle zulässig sind oder eine gewisse Größe unterschreiten [Gr01, Absatz 3.3].

Definition 4.9 (minimal zulässiger \mathcal{H} -Produktbaum) Sei T_I der \mathcal{H} -Baum von I . Wir nennen einen \mathcal{H} -Produktbaum $T = T_{I \times I}$ **minimal zulässig**, falls

- seine Knoten der i -ten Stufe nur Produkte von Knoten von T_I der i -ten Stufe sind und
- jedes Blatt $b = t \times s$ von T entweder zulässig ist oder die **Mindestblockgröße** $b_{\min} \in \mathbb{N}$ unterschreitet.

Offensichtlich gilt $T \neq T_I \otimes T_I$ im allgemeinen. Die Wahl von b_{\min} hängt von den Hardware-Eigenschaften (Cache-Größe) des Rechners ab.

Matrizen $A = (a_{ij})$ sind oftmals dünnbesetzt, d.h. sie weisen erheblich mehr Null- als Nichtnulleinträge auf. Bezeichnet n_Z die maximale Anzahl von Nichtnulleinträgen, die eine Matrixzeile haben kann, also

$$n_Z := \max_{i \in I} |\{j \in I \mid a_{ij} \neq 0\}|,$$

so ist die Anzahl der Nichtnulleinträge von A durch $n_Z \cdot |I|$ beschränkt. Etwas Vergleichbares wollen wir für \mathcal{H} -Produktbäume erreichen.

Definition 4.10 (Schwachbesetztheit von \mathcal{H} -Produktbäumen) Sei $T_{I \times I}$ der aus T_I erzeugte \mathcal{H} -Produktbaum. Seien

$$C_{sp,row} := \max_{t \in T_I} |\{s \in T_I \mid t \times s \in T_{I \times I}\}|, \quad C_{sp,col} := \max_{s \in T_I} |\{t \in T_I \mid t \times s \in T_{I \times I}\}|.$$

Wir nennen

$$C_{sp} := \max \{C_{sp,row}, C_{sp,col}\}$$

die **Schwachbesetztheitskonstante** und den Baum $T_{I \times I}$ **schwachbesetzt zur Konstanten** C_{sp} .

In vielen Abschätzungen laufen Summen über die Anzahl von Indexpaarteilmengen, die wir mit der Konstanten C_{sp} vereinfachen können, vgl. [BGH02, Absatz 6.2] und [GH02, Definition 2.1]. Diese Konstante ist sowohl für skalare wie auch für Systeme von Differentialgleichungen dieselbe. In [GH02, Absatz 4] finden sich Abschätzungen, in der vorliegenden Arbeit führen wir im Rahmen der numerischen Experimente Protokoll über die Größe von C_{sp} .

Bemerkung 4.11 *Aus der Abschätzung von C_{sp} in [GH02, Lemma 4.5] folgern wir, daß sich η in der Größenordnung von $\sqrt{n} \cdot \sqrt[n]{C_{sp}}$ bewegt (n ist die Raumdimension gewesen). Nun sind aber ausschließlich Mengen, die sich nicht berühren oder schneiden, (irgendwann) zulässig. Für $\eta \rightarrow 0$ werden die zulässigen Mengen sehr klein, was wir aber (aus Effizienzgründen, vgl. Definition 4.9) nicht wollen. Wir dürfen also annehmen, daß sich η nicht allzuweit von der Eins entfernt, und den Fall $\eta \approx 1$ diskutieren. Die bereits erwähnte Wahl $\eta = 2$ im Kontext von FE-Rechnungen ist empirischer Natur.*

4.3 Rk-Matrizen

Daß ein Blatt des \mathcal{H} -Produktbaumes T zulässig ist, bedeutet, daß die zugehörigen Cluster einen gewissen Mindestabstand zueinander einhalten. Deshalb dürfen wir die im entsprechenden Matrixblock $A|_b$, $b = t \times s$, enthaltene Information reduzieren.

Unser Approximationsverfahren besteht darin, den Block $A|_b$, der vom Maximalrang $\min\{|s|, |t|\}$ ist, auf einen Rang $k \in \mathbb{N}$ zu kürzen. Dieses k können wir für alle zulässigen Blöcke einheitlich vorgeben oder von Block zu Block unterschiedlich bestimmen (*Rangadaptivität*). Wir definieren

$$\mathcal{R}(k, n, m) := \{M \in \mathbb{R}^{m \times n} \mid \text{rnk } M \leq k\}$$

als die Menge aller reellen $m \times n$ -Matrizen mit Maximalrang k . Eine Matrix $A \in \mathcal{R}(k, n, m)$ nennen wir **Rk-Matrix**.

4.3.1 Die Kürzungsmethode

Wir können eine Matrix $A \in \mathcal{R}(k, n, m)$ darstellen als

$$A = \sum_{i=1}^k a_i b_i^\top, \quad a_i \in \mathbb{R}^m, b_i \in \mathbb{R}^n,$$

was sich günstig auf den Speicherverbrauch auswirkt. Um eine beliebige Matrix $A \in \mathbb{R}^{m \times n}$ in eine Rk-Matrix zu konvertieren, wählen wir eine *gekürzte Singulärwertzerlegung*. Berechnen wir von A die Singulärwertzerlegung, so ist

$$A = USV^\top \tag{4.4}$$

für zwei orthogonale Matrizen $U \in \mathbb{R}^{m \times m}$ und $V \in \mathbb{R}^{n \times n}$ sowie eine Diagonalmatrix

$$S = \text{diag}(s_1, s_2, \dots, s_{\min\{m,n\}}) \in \mathbb{R}^{m \times n}. \quad (4.5)$$

Die Singulärwerte s_i sind der Größe nach geordnet, d.h.

$$s_1 \geq s_2 \geq \dots \geq s_{\min\{m,n\}} \geq 0.$$

Aus (4.4) und (4.5) erhalten wir

$$A = \sum_{i=1}^{\min\{m,n\}} s_i u_i v_i^\top, \quad (4.6)$$

wobei u_i, v_i die Spaltenvektoren von U, V sind. Hier drängt sich die Idee auf, für eine Kürzung von A auf Rang k die letzten $\min\{m, n\} - k \geq 0$ Singulärwerte wegzulassen. Diese Vorgehensweise nennen wir **gekürzte Singulärwertzerlegung** (genaugenommen wird die exakte Summe (4.6) gekürzt). Die konvertierte Matrix

$$A_{Rk} = \sum_{i=1}^k s_i u_i v_i^\top$$

stellt zugleich die **Bestapproximation** von A in $\mathcal{R}(k, n, m)$ dar, denn nach [Gr01, Satz 2.10] gilt

$$\|A - A_{Rk}\|_2 = s_{k+1}.$$

4.3.2 Arithmetik

Bei der Arithmetik von Rk -Matrizen unterscheiden wir zwischen den formatierten und den exakten/unformatierten Operationen. Für die (exakte) Addition gilt

$$+ : \mathcal{R}(k_1, n, m) \times \mathcal{R}(k_2, n, m) \longrightarrow \mathcal{R}(k_1 + k_2, n, m).$$

Soll ein gegebener Rang k nicht überschritten werden, müssen wir eine Rangkürzung durchführen, d.h. wir wenden auf die exakte Summe eine gekürzte Singulärwertzerlegung an. Wir sprechen dann von der formatierten Addition

$$\oplus : \mathcal{R}(k_1, n, m) \times \mathcal{R}(k_2, n, m) \longrightarrow \mathcal{R}(k, n, m).$$

Da die (exakte) Multiplikation

$$\cdot : \mathcal{R}(k_1, l, m) \times \mathcal{R}(k_2, m, n) \longrightarrow \mathcal{R}(\min\{k_1, k_2\}, l, n)$$

auf den kleineren Rang „kürzt“ (was man brachial, aber simpel über die Singulärwertzerlegung zeigen kann), müssen wir keine formatierte Multiplikation erklären.

4.3.3 Rangadaptivität

Bei Finite-Element-Matrizen ist es geschickter, keinen für alle Rk -Matrizen geltenden Rang k vorzugeben, sondern einen von einer lokalen Approximationsgenauigkeit $\varepsilon_{lok} > 0$ abhängenden Rang

$$k(\varepsilon_{lok}) := \min \{k \in \mathbb{N} \mid \|A|_{t \times s} - R\|_2 \leq \varepsilon_{lok} \quad \forall R \in \mathcal{R}(k, |t|, |s|)\} \quad (4.7)$$

zu berechnen. Der Vorteil gerade bei der Invertierung der Steifigkeitsmatrix ist, daß die Dünnbesetztheit länger gewahrt werden kann. Rangadaptivität wirkt sich — zumindest bei Verwendung der Standardzulässigkeitsbedingung — nicht auf den \mathcal{H} -Produktbaum aus. Siehe auch Bemerkung 4.14.

4.4 \mathcal{H} -Matrizen

Die Matrixblöcke, die über den \mathcal{H} -Produktbaum als unzulässig eingestuft werden, müssen wir unverändert belassen und als volle Matrizen abspeichern. Diese vollen Matrizen ergeben zusammen mit den Rk -Matrizen eine \mathcal{H} -Matrix.

Definition 4.12 (\mathcal{H} -Matrix) Sei $I = \{1, \dots, N\}$ eine Indexmenge und $T_{I \times I}$ der aus einer Zulässigkeitsbedingung Z gebildete \mathcal{H} -Produktbaum zu $I \times I$. Sei $k \in \mathbb{N}$.

- (a) Wir nennen $A \in \mathbb{R}^{N \times N}$ eine **\mathcal{H} -Matrix vom Blockrang k** , falls jeder bzgl. Z für zulässig befundene Matrixblock $A|_b$, $b \in \mathcal{L}(T_{I \times I})$, vom Rang k ist.
- (b) Die Menge aller \mathcal{H} -Matrizen, die als **\mathcal{H} -Matrizen bzgl. $T_{I \times I}$ und Z darstellbar** sind, bezeichnen wir mit

$$\mathcal{H}(k, T_{I \times I}, Z) := \{A \in \mathbb{R}^{N \times N} \mid \text{rnk}(A|_b) \leq k \quad \forall b \in \mathcal{L}^+(T_{I \times I})\}$$

4.4.1 \mathcal{H} -Approximation von Matrizen

Sei $A \in \mathbb{R}^{n \times n}$ eine Matrix und durch $A_{\mathcal{H}} \in \mathcal{H}(k, T, Z)$ zu approximieren, d.h. zu zeigen ist, daß wir den Fehler $\|A - A_{\mathcal{H}}\|_2$ kontrollieren können. Der Nachweis der \mathcal{H} -Approximation kann nur für die FROBENIUS-Norm $\|\cdot\|_F$ für jedes Blatt separat geführt werden, vgl. [Gr01, Absatz 4.2.1]. Für andere Normen wie die Spektralnorm ist der Nachweis schwieriger, weil die Blätter dann nicht mehr voneinander unabhängig sind. Jedoch gilt für schwachbesetzte \mathcal{H} -Bäume der

Satz 4.13 (Schranken für die Spektralnorm) Sei T ein aus T_I und der Zulässigkeitsbedingung Z gebildeter \mathcal{H} -Produktbaum. Sei L_T die Menge der Stufen von T , auf denen sich Blätter befinden, und $\mathcal{L}(T, l)$ die Menge der Blätter von T , die sich

auf der Stufe l befinden. Sei $A \in \mathcal{H}(k, T, Z)$.

Dann existiert eine Schwachbesetztheitskonstante C_{sp} , sodaß

$$\|A\|_2 \leq C_{sp} \cdot \sum_{l \in L_T} \max_{b \in \mathcal{L}(T, l)} \|A|_b\|_2$$

gilt.

Beweis: Siehe [Gr01, S. 88]. \square

Die Abschätzung angewandt auf unsere Fehlermatrix $A - A_{\mathcal{H}}$ ergibt

$$\|A - A_{\mathcal{H}}\|_2 \leq C_{sp} \cdot \sum_{l \in L_T} \max_{b \in \mathcal{L}(T, l)} \|[A - A_{\mathcal{H}}]_b\|_2.$$

In den unzulässigen Blöcken bzw. Blättern dürfen wir nicht approximieren und können dort folglich keine Fehler produzieren. Damit müssen wir nur über alle zulässigen Blätter laufen, die in der Menge $\mathcal{L}^+(T)$ versammelt sind, und erhalten

$$\sum_{l \in L_T} \max_{b \in \mathcal{L}(T, l)} \|[A - A_{\mathcal{H}}]_b\|_2 = \sum_{l \in L_T} \max_{b \in \mathcal{L}^+(T)} \|[A - A_{\mathcal{H}}]_b\|_2 \leq |L_T| \cdot \max_{b \in \mathcal{L}^+(T)} \|[A - A_{\mathcal{H}}]_b\|_2.$$

Soll also $\|A - A_{\mathcal{H}}\|_2 < \varepsilon$ erfüllt werden, haben wir als hinreichende Bedingung für die globale Approximationseigenschaft die lokale Approximationseigenschaft

$$\|[A - A_{\mathcal{H}}]_b\|_2 \leq \frac{\varepsilon}{C_{sp} \cdot |L_T|} \quad \forall b \in \mathcal{L}^+(T).$$

Sie verlangt, daß die Singulärwerte der zulässigen Blätter beschränkt sind bzw. rasch abklingen. In der Tat sind im Falle der Spektralnorm die lokale und die globale Approximationseigenschaft äquivalent.

Bemerkung 4.14 (Rangadaptivität) *Rechnen wir mit Rangadaptivität, so ist $\varepsilon_{lok} := \varepsilon (C_{sp} |L_T|)^{-1}$ in (4.7) eine gute Wahl.*

4.4.2 Arithmetik

Aus der Aufgabenstellung ergibt sich, daß wir immer mit demselben \mathcal{H} -Produktbaum T arbeiten können. Ebenso ausnahmslos verwenden wir als Zulässigkeitsbedingung Z die Standardzulässigkeitsbedingung, weshalb wir die arithmetischen Operationen nur für T und Z betrachten müssen. Was unterschiedlich sein kann, ist der Blockrang k .

Wir beginnen mit der (exakten) Addition und sehen sofort ein, daß

$$+ : \mathcal{H}(k_1, T_{I \times I}, Z) \times \mathcal{H}(k_2, T_{I \times I}, Z) \longrightarrow \mathcal{H}(k_1 + k_2, T_{I \times I}, Z)$$

gilt. Weniger offensichtlich ist das Ergebnis der (exakten) Multiplikation, die einzig für unsere Theorie in Kapitel 5 von Bedeutung ist. Dort beweisen wir die Existenz einer \mathcal{H} -Inversen der Steifigkeitsmatrix und benötigen für die Fehlerabschätzung eine solche Aussage.

Satz 4.15 (Exakte Multiplikation von \mathcal{H} -Matrizen) Sei $I = \{1, \dots, N\}$ eine Indexmenge und T der bzgl. der Standardzulässigkeitsbedingung aufgebaute \mathcal{H} -Produktbaum von $I \times I$. Sei b_{\min} die minimale Blockgröße, $p_T = \mathcal{O}(\log N)$ die Höhe von T sowie C_{sp} die Schwachbesetztheitskonstante. Dann besitzt die exakte Multiplikation

$$\cdot : \mathcal{H}(k_1, T_{I \times I}, Z) \times \mathcal{H}(k_2, T_{I \times I}, Z) \longrightarrow \mathcal{H}(k, T_{I \times I}, Z)$$

den maximalen Blockrang

$$k = (p_T + 1) \cdot C_{sp} \cdot \max\{k_1, k_2, b_{\min}\}.$$

Beweis: Siehe [Gr01, Lemma 4.11] oder [GH02, Theorem 2.20]. □

In der Praxis verwenden wir oftmals keine exakten, sondern formatierte Operationen. Bei der formatierten Addition

$$\oplus : \mathcal{H}(k_1, T_{I \times I}, Z) \times \mathcal{H}(k_2, T_{I \times I}, Z) \longrightarrow \mathcal{H}(k, T_{I \times I}, Z)$$

benutzen wir für zulässige Bätter die formatierte Addition von Rk -Matrizen und für unzulässige Bätter die exakte Addition. Daß ein zulässiger mit einem unzulässigen Block verrechnet werden muß, kann in unserer Konstellation nicht auftreten.

Die formatierte Multiplikation

$$\odot : \mathcal{H}(k_1, T_{I \times I}, Z) \times \mathcal{H}(k_2, T_{I \times I}, Z) \longrightarrow \mathcal{H}(k, T_{I \times I}, Z)$$

besteht aus den zwei Schritten Berechnung des exakten Produktes und Konvertierung des exakten Produktes in eine $\mathcal{H}(k, T_{I \times I}, Z)$ -Matrix. In der Praxis wiederum vermeiden wir aus Kostengründen tunlichst die exakte Berechnung des Produktes und verwenden möglichst viele formatierte Operationen.

4.4.3 \mathcal{H} -Inversion

Für die Invertierung einer Matrix A nehmen wir Blockgestalt an, d.h.

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}.$$

Wenden wir nun die Block-GAUSS-Elimination an, erhalten wir für die Inverse die Blockdarstellung

$$A^{-1} = \begin{pmatrix} (A_{11})^{-1} (Id + A_{12} S^{-1} A_{21} (A_{11})^{-1}) & (A_{11})^{-1} A_{12} S^{-1} \\ S^{-1} A_{21} (A_{11})^{-1} & S^{-1} \end{pmatrix},$$

wobei

$$S = A_{22} - A_{21} (A_{11})^{-1} A_{12}$$

SCHUR-Komplement genannt wird. Für die Invertierung von A benötigen wir also nur die beiden Inversen von A_{11} und S und sonst nur Multiplikationen und Additionen von Matrizen. Zur Berechnung von $(A_{11})^{-1}$ und S^{-1} nehmen wir wieder Blockgestalt von A_{11} und S an usw.

Bei der Bestapproximation werden die benötigten (Zwischen-) Ergebnisse exakt berechnet und dann gekürzt. So wollen wir das auch für den Beweis der \mathcal{H} -Approximation in Kapitel 5 handhaben. In der Praxis zeigt sich jedoch, daß die Verwendung formatierter Operationen zu gleichwertigen Ergebnissen bei gleichzeitiger Kostenersparnis führt [Gr01, S. 55].

Falls A_{11} eine volle Matrix ist, invertieren wir mittels eines der klassischen Verfahren. Falls A_{11} eine \mathcal{H} -Matrix ist, rufen wir die \mathcal{H} -Invertierungsroutine auf und konstruieren auf diese Weise einen rekursiven Algorithmus.

Haben wir $(A_{11})^{-1}$ bestimmt, berechnen wir ausschließlich mit formatierten Operationen die Matrix S sowie die vier Teilblöcke von A^{-1} .

Kapitel 5

Approximationseigenschaft elliptischer Differentialoperatoren

Eine Matrix $A \in \mathbb{R}^{m \times n}$ nennen wir als \mathcal{H} -Matrix **darstellbar**, falls in der Menge der \mathcal{H} -Matrizen ein $A_{\mathcal{H}}$ existiert, das die **globale Approximationseigenschaft**

$$\|A - A_{\mathcal{H}}\| \leq \varepsilon$$

einhält. Dementsprechend nennen wir einen **Differentialoperator darstellbar**, falls die bei seiner Diskretisierung anfallenden Matrizen als \mathcal{H} -Matrizen darstellbar sind.

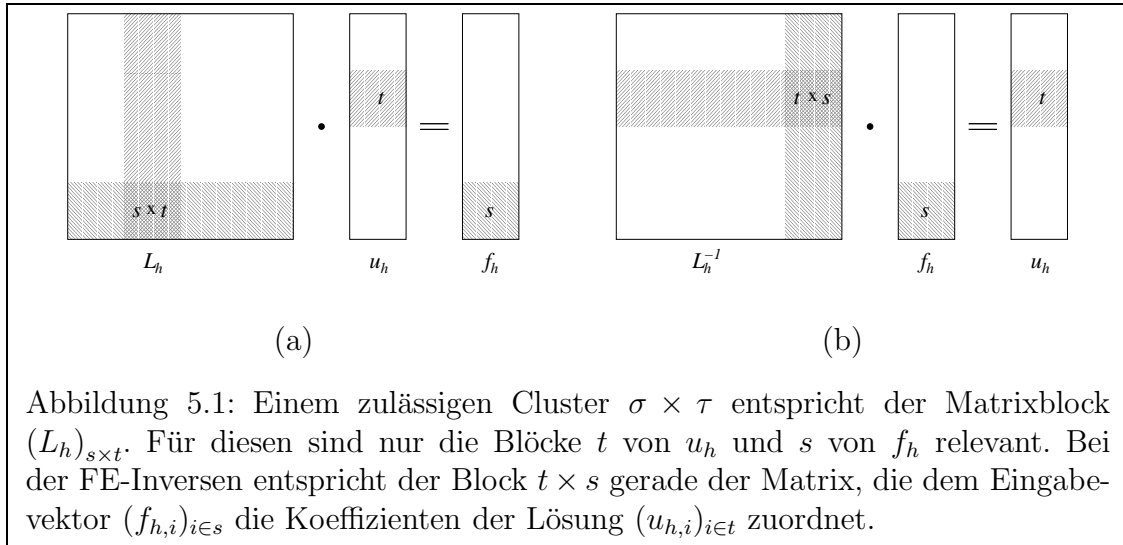
Läßt sich A durch eine \mathcal{H} -Matrix darstellen, so gilt das auch für die Inverse A^{-1} (die Invertierung braucht ja nur innerhalb der \mathcal{H} -Arithmetik zu erfolgen). Zu klären bleibt, wie gut die Approximation ist.

Schauen wir uns die Situation bei den FE-Matrizen an. Sei L_h eine RITZ-GALERKIN-Diskretisierung (FE-Matrix) des Differentialoperators L . Wegen des beschränkten Trägers der Basisfunktionen ist L_h eine dünnbesetzte Bandmatrix. Die Bänder speichern wir als volle Blockmatrizen ab, die zulässigen Blöcke umfassen ausschließlich Nulleinträge¹. Bezeichnet nun $L_{\mathcal{H}}$ die \mathcal{H} -Matrixapproximation von L_h , so gilt

$$L_h - L_{\mathcal{H}} = 0,$$

d.h. jede FE-Matrix ist per se eine \mathcal{H} -Matrix. Die Inverse $(L_{\mathcal{H}})^{-1}$ von $L_{\mathcal{H}}$ berechnen wir mittels Algorithmus 6.2, also unter Verwendung der \mathcal{H} -Arithmetik, sodaß $(L_{\mathcal{H}})^{-1}$ auf jeden Fall eine \mathcal{H} -Matrix ist. Unklar ist dabei, ob wir uns durch das Kürzen zu große Fehler einhandeln, ob also die Singulärwerte der zulässigen Blöcke nicht schnell genug abklingen. Uns bleibt daher die schwierige Aufgabe, die Approximationsgüte

¹Die Zulässigkeitsbedingung Z , die den Produktbaum T aufbaut, beurteilt zwei Teilgebiete σ und τ von Ω als zulässig, wenn sie weit genug entfernt sind. Im Falle von FEM sind σ und τ die Träger der Basis-/Testfunktionen und zulässig, wenn sie disjunkt sind. Konsequenz: entweder ist ein korrespondierender Matrixblock zulässig und damit gleich 0 oder unzulässig und voll. D.h. jede FEM-Matrix ist Element von $\mathcal{H}(b_{min}, T, Z)$, b_{min} kleinste Blattgröße.



zu untersuchen, d.h. den Fehler

$$\|(L_h)^{-1} - (L_{\mathcal{H}})^{-1}\|$$

abzuschätzen (zwar ist $L_h = L_{\mathcal{H}}$, aber während $(L_h)^{-1}$ die „klassische“ Inverse bezeichnet, ist $(L_{\mathcal{H}})^{-1}$ die mit der \mathcal{H} -Arithmetik berechnete Inverse, also $(L_h)^{-1} \neq (L_{\mathcal{H}})^{-1}$). Von zentraler Bedeutung ist die Abschätzung von

$$\|(L^{-1})_h - (L^{-1})_{\mathcal{H}}\|.$$

Wir werden uns die meiste Zeit mit ihr beschäftigen müssen.

5.1 Approximationseigenschaft für skalare Differentialoperatoren

5.1.1 Beweisidee

Für den Beweis der Approximationseigenschaft sind die lokalen, nämlich die auf den zulässigen Mengen gültigen Integrierbarkeits- und Differenzierbarkeitsaussagen ausschlaggebend, weil für einen zulässigen Matrixblock $t \times s$ nur der Teil des Eingabevektors relevant ist, der zur Indexteilmenge s gehört. Man kann auch sagen, daß auf der rechten Seite f_h des Gleichungssystems alles gleich null gesetzt wird, was mit s nichts zu tun hat (vgl. Abb. 5.1).

Deshalb treffen wir die Annahme, daß f nur innerhalb von σ , das ist das mit s assoziierte Teilgebiet von Ω , von null verschieden ist. Weiterhin habe die Abbildung, die einem solchen auf σ lebenden f die Einschränkung von u auf τ zuordnet, niedrigen Rang.

Hat man analytische Koeffizienten und ein glattes Gebiet, kann man das aus der inneren Regularität folgern. Diese liefert einem Abschätzungen der Form

$$|u|_{H^m(\tau)} \leq \frac{m!}{(\text{dist}(\sigma, \tau))^m} \|f\|_{L^2(\sigma)}.$$

Durch eine gemittelte TAYLOR-Entwicklung erhalten wir

$$\|u - Q^m u\|_{L^2(\tau)} \leq \left(\frac{\text{diam}(\tau)}{\text{dist}(\sigma, \tau)} \right)^m \|f\|_{L^2(\sigma)},$$

wobei $Q^m u$ ein Polynom m -ter Ordnung ist (erstrebenswert ist die Abschätzung $\|u - Q^m u\|_{H^1(\tau)} \leq \left(\frac{\text{diam}(\tau)}{\text{dist}(\sigma, \tau)} \right)^m \|f\|_{H^{-1}(\sigma)}$). Durch Einsetzen der Zulässigkeitsbedingung erhalten wir sofort

$$\|u - Q^m u\|_{L^2(\tau)} \leq \eta^m \|f\|_{L^2(\sigma)}$$

bzw. $\|u - Q^m u\|_{H^1(\tau)} \leq \eta^m \|f\|_{H^{-1}(\sigma)}$, was uns erfreut, weil $Q^m u$ ein Polynom ist, das wir in unserer bevorzugten Polynombasis darstellen.

Bezeichnet $L^{-1} : f \mapsto u$ den Lösungsoperator und $Q^m L^{-1}$ seine Approximation, so gilt $\|L^{-1} - Q^m L^{-1}\| \leq \eta^m$. Da das gesamte Bild von $Q^m L^{-1}$ im Raum der Polynome der Ordnung m liegt, ist $Q^m L^{-1}$ ein Operator niedrigen Ranges, der exponentiell gegen L^{-1} konvergiert, weshalb wir L^{-1} durch $Q^m L^{-1}$ ersetzen dürfen. Schließlich ist $Q^m L^{-1}$ noch zu diskretisieren.

5.1.2 Innere Regularität

Satz 5.1 (Innere Regularität elliptischer Differentialoperatoren) *Sei L wie in (2.3) ein gleichmäßig elliptischer linearer Differentialoperator zweiter Ordnung mit der symmetrisch-positiv-definiten Koeffizientenmatrix $C = (c_{ij})$, $c_{ij} \in \mathbb{R}$ konstant, deren kleinster und größter Eigenwert durch λ_{\min} und λ_{\max} für alle $x \in \Omega$ gegeben sind. Seien σ, τ zwei kompakte Teilmengen von Ω (s. Abb. 5.2a) mit $\sigma \cap \tau = \emptyset$ und positivem Abstand zum Rand, d.h.*

$$\text{dist}(\sigma, \tau, \Gamma) := \min \{ \text{dist}(\tau, \sigma), \text{dist}(\tau, \Gamma) \} > 0.$$

Sei u die schwache Lösung der Differentialgleichung

$$L[u](x) = f(x) \text{ in } \Omega, \quad u = 0 \text{ auf } \Gamma$$

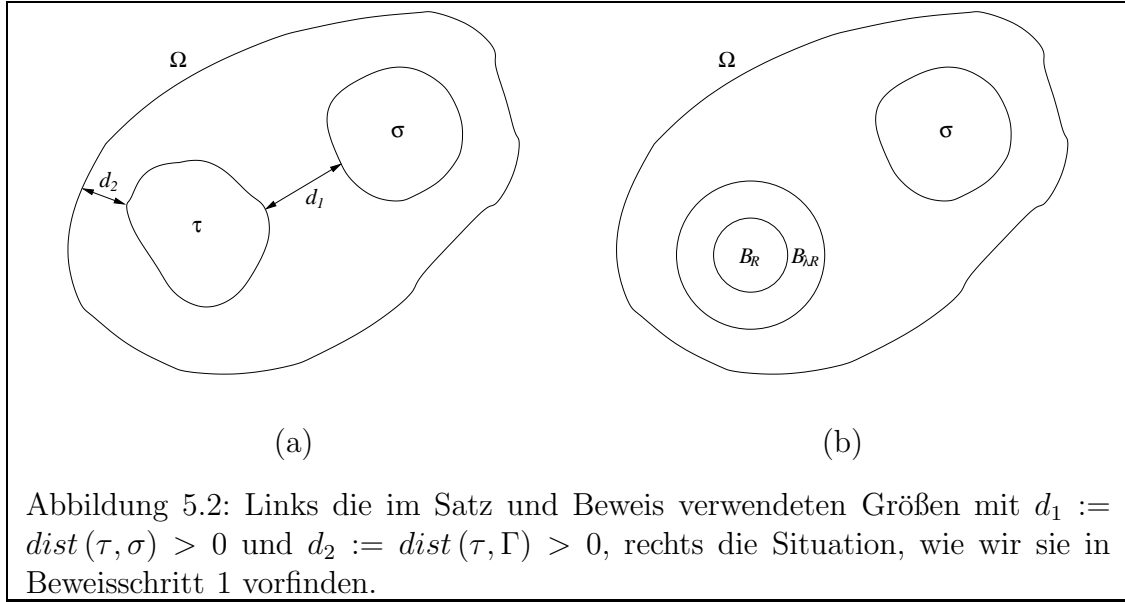
und L -harmonisch in $\Omega \setminus \sigma$. Dabei ist σ der Träger von f , d.h.

$$\text{supp}(f) \subset\subset \sigma. \tag{5.1}$$

Dann sind auch die m -ten partiellen Ableitungen, $m \in \mathbb{N}$, von u L -harmonisch in $\Omega \setminus \sigma$ und es existiert eine Konstante $c_{\text{reg}} > 0$, sodaß

$$|u|_{H^m(\tau)} \leq c_{\text{reg}} \cdot c_m \cdot \kappa_C^{m/2} \cdot \frac{m!}{\text{dist}(\sigma, \tau, \Gamma)^m} \cdot \|f\|_{L^2(\sigma)}$$

mit $c_m = (m!)^\psi$, $\psi > 1$ (vgl. Bemerkung 5.2).



Beweisschritt 1

Wir lassen in den ersten beiden Schritten τ außen vor und überlegen uns die Abschätzung für zwei Bälle

$$B_R \subset B_{\lambda R} \subset \Omega \setminus \sigma, \quad \lambda > 1 \quad (5.2)$$

$$\text{supp}(\eta) \subset\subset B_{\lambda R}, \quad \eta|_{B_R} \equiv 1 \quad (5.3)$$

und auf $B_{\lambda R} \setminus B_R$ in etwa linear abfallend, sodaß

$$\left| \nabla \eta|_{B_{\lambda R} \setminus B_R} \right| \leq \frac{1 + \delta}{\lambda R - R} \equiv \text{const}, \quad 0 < \delta \ll 1. \quad (5.4)$$

gilt. Wir schreiben L als

$$Lu = - \sum_j \partial_j \left(\sum_i c_{ij} \cdot \partial_i u \right) = - \sum_j \partial_j \cdot (C \cdot \nabla u)_j = - \text{div}(C \nabla u) = - \langle \nabla, C \nabla u \rangle.$$

Dabei bezeichnet $\langle \cdot, \cdot \rangle$ das euklidische Skalarprodukt, für das die Produktregel

$$\langle \nabla, \nabla ab \rangle = \langle \nabla, \nabla a \cdot b \rangle = \langle \nabla, \nabla a \rangle \cdot b + \langle \nabla b, \nabla a \rangle \quad (5.5)$$

und die Regel für partielle Integration

$$\int_B \langle \nabla a, \nabla b \rangle = - \int_B a \langle \nabla, \nabla b \rangle + \int_{\partial B} a \langle \vec{\nu}, \nabla b \rangle \quad (5.6)$$

gelten; da C (symmetrisch-) positiv-definit ist, gilt weiterhin

$$\lambda_{\min} \langle x, x \rangle = \langle x, \lambda_{\min} x \rangle \leq \langle x, Cx \rangle \leq \langle x, \lambda_{\max} x \rangle = \lambda_{\max} \langle x, x \rangle \quad \forall x.$$

Wegen der L -Harmonizität von u auf $\Omega \setminus \sigma$ gilt

$$\begin{aligned}
0 &\leq \lambda_{\min} \int_{B_R} |\nabla u|^2 \leq \lambda_{\min} \int_{B_{\lambda R}} |\nabla(\eta u)|^2 = \lambda_{\min} \int_{B_{\lambda R}} \langle \nabla(\eta u), \nabla(\eta u) \rangle \\
&\leq \int_{B_{\lambda R}} \langle \nabla(\eta u), C \cdot \nabla(\eta u) \rangle \\
&\stackrel{\text{part. Int.}}{=} - \int_{B_{\lambda R}} \eta u \langle \nabla, C \nabla(\eta u) \rangle + \int_{\partial B_{\lambda R}} \underbrace{\eta}_{=0} u \langle \vec{\nu}, C \nabla(\eta u) \rangle \\
&= - \int_{B_{\lambda R}} \eta [\langle \nabla, C \nabla \eta u \rangle + \langle \nabla, C \eta \nabla u \rangle] u \\
&= - \int_{B_{\lambda R}} \eta \left[\langle \nabla, C \nabla \eta \rangle u + \langle \nabla u, C \nabla \eta \rangle + \langle \nabla \eta, C \nabla u \rangle + \underbrace{\eta \langle \nabla, C \nabla u \rangle}_{=Lu=0} \right] u \\
&= - \int_{B_{\lambda R}} \eta [\langle \nabla, C \nabla \eta \rangle u + 2 \langle \nabla \eta, C \nabla u \rangle] u \\
&= - \int_{B_{\lambda R}} \eta \langle \nabla, C \nabla \eta \rangle u^2 + 2 \int_{B_{\lambda R}} \eta \langle \nabla \eta, C \nabla u \rangle u \\
&= - \int_{B_{\lambda R}} \eta \langle \nabla, C \nabla \eta \rangle u^2 - \frac{1}{2} \int_{B_{\lambda R}} \langle \nabla \eta^2, C \nabla u^2 \rangle,
\end{aligned}$$

weil in der vorletzten Zeile

$$\begin{aligned}
\eta \langle \nabla \eta, C \nabla u \rangle u &= \eta u \sum_i \partial_i \eta \cdot \sum_j c_{ij} \partial_j u = \sum_i (\eta \partial_i \eta) \left(\sum_j c_{ij} u \partial_j u \right) \\
&= \sum_i \left(\frac{1}{2} \partial_i \eta^2 \right) \left(\sum_j \frac{1}{2} c_{ij} \partial_j u^2 \right) = \frac{1}{4} \langle \nabla \eta^2, C \nabla u^2 \rangle
\end{aligned}$$

ist. Nächstes Ziel ist der Umbau des zweiten Integrals zu einem Ausdruck ähnlich oder gleich dem ersten Integral. Mit partieller Integration dieses Skalarproduktes,

$$\begin{aligned}
\int_{B_{\lambda R}} \langle \nabla \eta^2, C \nabla u^2 \rangle &= \int_{B_{\lambda R}} \langle C \nabla \eta^2, \nabla u^2 \rangle \\
&= \sum_i \int_{B_{\lambda R}} \left(\sum_j c_{ij} 2 \eta \partial_j \eta \right) \cdot \partial_i u^2 \\
&= \sum_i \int_{\partial B_{\lambda R}} \sum_j c_{ij} 2 \underbrace{\eta}_{=0} \partial_j \eta \cdot u^2 \\
&\quad - \sum_i \int_{B_{\lambda R}} \partial_i \left(\sum_j c_{ij} 2 \eta \partial_j \eta \right) \cdot u^2 \\
&= - \int_{B_{\lambda R}} 2 \langle \nabla, C \cdot \eta \nabla \eta \rangle u^2,
\end{aligned}$$

erhalten wir

$$\begin{aligned}
& - \int_{B_{\lambda R}} \eta \langle \nabla, C \nabla \eta \rangle u^2 - \frac{1}{2} \int_{B_{\lambda R}} \langle C \nabla \eta^2, \nabla u^2 \rangle \\
= & - \int_{B_{\lambda R}} \eta \langle \nabla, C \nabla \eta \rangle u^2 + \frac{1}{2} \int_{B_{\lambda R}} 2 \langle \nabla, C \eta \nabla \eta \rangle u^2 \\
= & - \int_{B_{\lambda R}} \eta \langle \nabla, C \nabla \eta \rangle u^2 + \int_{B_{\lambda R}} [\langle \nabla \eta, C \nabla \eta \rangle + \eta \langle \nabla, C \nabla \eta \rangle] u^2 \\
= & \int_{B_{\lambda R}} \langle \nabla \eta, C \nabla \eta \rangle u^2 \\
\leq & \int_{B_{\lambda R}} \langle \nabla \eta, \lambda_{max} \nabla \eta \rangle u^2 \\
= & \lambda_{max} \int_{B_{\lambda R}} \langle \nabla \eta, \nabla \eta \rangle u^2 \\
\stackrel{\text{wg. (5.3)}}{=} & \lambda_{max} \int_{B_{\lambda R} \setminus B_R} |\nabla \eta|^2 u^2 \\
\stackrel{(5.4)}{\leq} & \lambda_{max} \left(\frac{1 + \delta}{\lambda R - R} \right)^2 \int_{B_{\lambda R} \setminus B_R} u^2 \\
\leq & \lambda_{max} \left(\frac{1 + \delta}{\lambda R - R} \right)^2 \int_{B_{\lambda R}} u^2.
\end{aligned}$$

Wir dividieren mit λ_{min} , was wegen der positiven Definitheit der Koeffizientenmatrix C erlaubt ist, und erhalten als neue Abschätzung und Ergebnis des ersten Beweisschrittes

$$\int_{B_R} |\nabla u|^2 \leq \frac{\kappa_C (1 + \delta)^2}{(R(\lambda - 1))^2} \int_{B_{\lambda R}} |u|^2, \quad \kappa_C = \frac{\lambda_{max}}{\lambda_{min}}, \quad 0 < \delta \ll 1. \quad (5.7)$$

Beweisschritt 2

Wie bei Regularitätsbeweisen üblich, wollen wir nun eine Schachtelung mehrerer Bälle konstruieren, um (5.7) auf $\nabla^m u$ iterativ anwenden zu können. Als Skalierungsfaktoren für den Radius R wählen wir die Glieder einer streng monoton wachsenden Folge $(\lambda_i)_{i \in \mathbb{N}}$, um eine echte Schachtelung von Sphären zu erhalten. Die Folge muß aber auch konvergieren und ihr Grenzwert $p := \lim_{i \rightarrow \infty} \lambda_i$

$$B_{pR}(x) \subset \Omega \setminus \sigma \quad \forall x \in \tau$$

gewährleisten; andernfalls kann $B_{pR}(x) \cap \sigma \neq \{\}$ eintreten oder Teile von $B_{pR}(x)$ können aus Ω herausragen. Die geforderten Eigenschaften werden z.B. von

$$\lambda_i := \sum_{k=1}^i \frac{1}{k^\psi}, \quad i \in \mathbb{N},$$

erfüllt, sofern $\psi > 1$. Da mit $B_{pR}(x)$ später eine Überdeckung von τ konstruiert wird, muß $pR - R < \text{dist}(\sigma, \tau, \Gamma)$ gelten. Wir definieren daher

$$R := \frac{\text{dist}(\sigma, \tau, \Gamma)}{p}, \quad x \in \tau \text{ und } B_i := B_{\lambda_i R}(x), \quad (5.8)$$

was

$$B_1 \subsetneq B_2 \subsetneq \dots \subsetneq B_{m+1} \subsetneq \Omega \setminus \sigma \quad \forall m \in \mathbb{N} \quad (5.9)$$

bewirkt. Die Konstanz der Koeffizienten erlaubt die L -Harmonizität von $\nabla^m u$ aus der von u zu schließen,

$$L[\partial_k u] = - \sum_{i,j} \partial_j (c_{ij} \partial_i \partial_k u) = - \partial_k \sum_{i,j} \partial_j (c_{ij} \partial_i u) = \partial_k L[u] = 0. \quad (5.10)$$

Mit ihr, der L -Harmonizität von $\nabla^0 u, \dots, \nabla^m u$ auf $\Omega \setminus \sigma$, folgt durch Iteration von (5.7)

$$\begin{aligned} \int_{B_1} |\nabla^m u|^2 &\leq \frac{\kappa_C (1 + \delta)^2}{((\lambda_2 - \lambda_1) R)^2} \int_{B_2} |\nabla^{m-1} u|^2 \\ &\leq \frac{\kappa_C^2 (1 + \delta)^4}{((\lambda_3 - \lambda_2) \cdot (\lambda_2 - \lambda_1) R^2)^2} \int_{B_3 \setminus B_2} |\nabla^{m-2} u|^2 \\ &\leq \frac{\kappa_C^2 (1 + \delta)^4}{((\lambda_3 - \lambda_2) \cdot (\lambda_2 - \lambda_1) R^2)^2} \int_{B_3} |\nabla^{m-2} u|^2 \\ &\leq \dots \\ &\leq \frac{\kappa_C^m (1 + \delta)^{2m}}{((\lambda_{m+1} - \lambda_m) \cdot \dots \cdot (\lambda_2 - \lambda_1) R^m)^2} \int_{B_{m+1}} |u|^2 \\ &= \left(\frac{\kappa_C^{m/2} (1 + \delta)^m}{\prod_{i=1}^m (\lambda_{i+1} - \lambda_i) R^m} \right)^2 \cdot \|u\|_{L^2(B_{m+1})}^2 \end{aligned}$$

Als Ergebnis von Beweisschritt 2 erhalten wir

$$\|u\|_{H^m(B_0)} = \|\nabla^m u\|_{L^2(B_0)} \leq \frac{\kappa_C^{m/2} (1 + \delta)^m}{\prod_{i=1}^m (\lambda_{i+1} - \lambda_i) R^m} \|u\|_{L^2(B_{m+1})}. \quad (5.11)$$

Beweisschritt 3

Nächstes Zwischenziel ist die Abschätzung von u in der H^m -Seminorm auf τ gegen die L^2 -Norm von f auf σ . Sei hierzu \ddot{U} eine endliche (τ ist kompakt!) Überdeckung von τ mittels N_τ vielen Bällen, d.h.

$$\tau \subset \ddot{U} := \bigcup_{j=1}^{N_\tau} B_R(x_j), \quad x_j \in \tau. \quad (5.12)$$

Im Kontext des Satzes von LAX-MILGRAM gilt

$$\|u\|_{L^2(\Omega)} \leq c_\Omega \cdot \|f\|_{L^2(\Omega)}.$$

Wegen (5.1) ist

$$\|f\|_{L^2(\Omega)} = \|f\|_{L^2(\sigma)}$$

und zusammen mit $\|u\|_{L^2(B_{m+1})} \leq \|u\|_{L^2(\Omega)}$ erhalten wir aus (5.11) die Abschätzung

$$|u|_{H^m(B_1)} = |u|_{H^m(B_R)} \leq \frac{\kappa_C^{m/2} (1 + \delta)^m}{\prod_{i=1}^m (\lambda_{i+1} - \lambda_i) R^m} \cdot c_\Omega \|f\|_{L^2(\sigma)}.$$

Damit gilt

$$\begin{aligned} |u|_{H^m(\tau)}^2 &\stackrel{(5.12)}{\leq} |u|_{H^m(\dot{U})}^2 \\ &\leq \sum_{j=1}^{N_\tau} |u|_{H^m(B_R(x_j))}^2 \\ &\leq \sum_{j=1}^{N_\tau} \left[\frac{\kappa_C^{m/2} (1 + \delta)^m}{\prod_{i=1}^m (\lambda_{i+1} - \lambda_i) R^m} c_\Omega \|f\|_{L^2(\sigma)} \right]^2 \\ &= N_\tau \cdot \left[\frac{\kappa_C^{m/2} (1 + \delta)^m}{\prod_{i=1}^m (\lambda_{i+1} - \lambda_i) R^m} c_\Omega \|f\|_{L^2(\sigma)} \right]^2, \end{aligned}$$

also folgt mit (5.8), daß

$$|u|_{H^m(\tau)} \leq \sqrt{N_\tau} \cdot \frac{\kappa_C^{m/2} (1 + \delta)^m}{\prod_{i=1}^m (\lambda_{i+1} - \lambda_i)} \cdot \frac{p^m}{\text{dist}(\sigma, \tau, \Gamma)^m} \cdot c_\Omega \cdot \|f\|_{L^2(\sigma)}. \quad (5.13)$$

Beweisschritt 4

Wir müssen den Faktor $\prod_{i=1}^m \frac{p}{(\lambda_{i+1} - \lambda_i)}$ größenordnungsmäßig durch $m!$ abschätzen:

$$\prod_{i=1}^m \frac{p}{(\lambda_{i+1} - \lambda_i)} = \prod_{i=1}^m \left[p (i+1)^\psi \right] = p^m \cdot \left[\prod_{j=1}^{m+1} j \right]^\psi \leq c_{Fak} (m!)^{1+\psi}. \quad (5.14)$$

Beweisschritt 5

Die Verschmelzung von (5.13) mit (5.14) führt zu

$$|u|_{H^m(\tau)} \leq \kappa_C^{m/2} (1 + \delta)^m \sqrt{N_\tau} \cdot \frac{c_{Fak} (m!)^{1+\psi}}{\text{dist}(\sigma, \tau, \Gamma)^m} \cdot c_\Omega \cdot \|f\|_{L^2(\sigma)}.$$

Wegen (5.4) dürfen wir den Grenzübergang $\delta \rightarrow 0$ vollziehen, um uns des Faktors $(1 + \delta)^m$ zu entledigen. Wir setzen

$$c_{reg} := \sqrt{N_\tau} c_{Fak} c_\Omega, \quad c_m := (m!)^\psi, \quad \psi > 1, \quad (5.15)$$

und erhalten das gewünschte Resultat. \square

Bemerkung 5.2 Den Faktor $(m!)^\psi$, der artefaktischer Natur ist, versenken wir in der „Konstanten“ c_m , denn mittels einer schärferen Beweisführung sollte man $c_m \leq 1$ nachweisen können².

5.1.3 Hilfsaussagen

Lemma 5.3 Seien $x, z \in \Omega \subset \mathbb{R}^n$. Sei $m \in \mathbb{N}$ die Ableitungsordnung. Dann gilt

$$\int_{\Omega} |x - z|^{m-n} dz \leq \tilde{c}_{m,n} \cdot \varnothing_{\Omega}^m \text{ mit } \tilde{c}_{m,n} := \begin{cases} \frac{\omega_n}{2^n} & \text{für } m > n, \\ \frac{n\omega_n}{2^m m} & \text{für } m \leq n, \end{cases}$$

wobei ω_n das Volumen der n -dimensionalen Einheitskugel ist.

Beweis: Wir müssen für verschiedene Werte von m und n das Integral gesondert untersuchen. Seien $m, n \in \mathbb{N}$. Wir setzen $\rho := \frac{1}{2}\varnothing_{\Omega}$ als den halben Durchmesser des Gebietes $\Omega \subset \mathbb{R}^n$ und wählen einen Punkt $\zeta \in \mathbb{R}^n$ derart, daß $\Omega \subset B_{\rho}(\zeta)$. Wir merken uns

$$\text{vol}(B_{\rho}(\zeta)) = \rho^n \cdot \omega_n = \left(\frac{\varnothing_{\Omega}}{2}\right)^n \cdot \omega_n$$

Fall $m > n \geq 1$.

$$\int_{\Omega} |x - z|^{m-n} dz \leq \int_{\Omega} \varnothing_{\Omega}^{m-n} dz \leq \varnothing_{\Omega}^{m-n} \int_{B_{\rho}(\zeta)} dz = \frac{\omega_n}{2^n} \varnothing_{\Omega}^m$$

Fall $m = n \geq 1$.

$$\int_{\Omega} |x - z|^{m-n} dz = \int_{\Omega} dz \leq \int_{B_{\rho}(\zeta)} dz = \rho^n \omega_n \stackrel{m=n}{=} n\omega_n \frac{\varnothing_{\Omega}^m}{2^m m}$$

Fall $1 \leq m < n$. Setze $a := -(m - n) > 0$.

$$\int_{\Omega} |x - z|^{m-n} dz = \int_{\Omega} |x - z|^{-a} dz = \int_{\Omega} \frac{1}{|x - z|^a} dz.$$

Da $a = -m + n$, ist $a < n$ für $m \geq 1$ und wir können [Kö93, S. 226, Bsp. 2i]

²S. Email-Wechsel BROKATE M.-FRIESECKE G. vom 26.11.2003, 21:12.

ins Feld führen:

$$\begin{aligned}
\int_{\Omega} \frac{dz}{|x-z|^a} &= \int_{\Omega} \frac{dz}{|z-x|^a} \\
&\leq \int_{B_{\rho}(\zeta)} \frac{dz}{|z-x|^a} \\
&= \int_{B_{\rho}(0)} \frac{dz}{|z-(x+\zeta)|^a} \\
&\leq \int_{B_{\rho}(0)} \frac{dz}{|z|^a} \\
&\stackrel{[\text{K693}]}{=} \frac{n}{n-a} \omega_n \left(\frac{\varnothing_{\Omega}}{2} \right)^{n-a}.
\end{aligned}$$

Mit $n-a=m$ folgt die Behauptung. \square

Lemma 5.4 Sei $\tilde{g} \in L^p(\Omega)$ für $p \in [1, \infty]$. Sei $m \geq 1$ und

$$g(x) := \int_{\Omega} |x-z|^{m-n} |\tilde{g}(z)| dz.$$

Dann gilt

$$\|g\|_{L^p(\Omega)} \leq \tilde{c}_{m,n} \varnothing_{\Omega}^m \|\tilde{g}\|_{L^p(\Omega)}$$

bzw. in Integralschreibweise

$$\int_{\Omega} \left| \int_{\Omega} |x-z|^{m-n} |\tilde{g}(z)| dz \right|^p dx \leq (\tilde{c}_{m,n} \varnothing_{\Omega}^m)^p \int_{\Omega} |\tilde{g}(z)|^p dz$$

mit der Konstanten c_n aus dem vorigen Lemma.

Beweis: Siehe [BS02, Lemma 4.3.6]. \square

Lemma 5.5 Seien $m, n \in \mathbb{N}$ mit $m \geq n$ und $\alpha \in \mathbb{N}_0^n$. Wir vereinbaren

$$|\alpha| := \sum_{i=1}^n \alpha_i \text{ und } \alpha! := \prod_{i=1}^n \alpha_i!.$$

Dann gilt

$$s_n(m) := \sum_{|\alpha|=m} \frac{1}{\alpha!} = \frac{n^m}{m!}.$$

Beweis: Durch Induktion nach n .

$n = 2$:

$$s_2(m) = \sum_{|\alpha|=m} \frac{1}{\alpha!} = \sum_{\alpha_1+\alpha_2=m} \frac{1}{\alpha_1! \alpha_2!} = \sum_{\alpha_1=0}^m \frac{1}{\alpha_1! (m-\alpha_1)!} = \frac{2^m}{m!}$$

$n = 3$:

$$\begin{aligned}
 s_3(m) &= \sum_{\alpha_1+\alpha_2+\alpha_3=m} \frac{1}{\alpha_1!\alpha_2!\alpha_3!} \\
 &= \sum_{\alpha_1=0}^m \frac{1}{\alpha_1!} \cdot \sum_{\alpha_2+\alpha_3=m-\alpha_1} \frac{1}{\alpha_2!\alpha_3!} \\
 &\stackrel{s_2(m-\alpha_1)}{=} \sum_{\alpha_1=0}^m \frac{1}{\alpha_1!} \cdot \frac{2^{m-\alpha_1}}{(m-\alpha_1)!} \\
 &= \frac{3^m}{m!}
 \end{aligned}$$

Induktionsvoraussetzung: für $n \in \mathbb{N}$ und $m \geq n$ gelte

$$s_n(m) = \frac{n^m}{m!}.$$

Induktionsschritt $n \rightarrow n + 1$:

$$\begin{aligned}
 s_{n+1}(m) &= \sum_{\substack{\alpha_1, \dots, \alpha_{n+1}=0, \dots, m \\ \alpha_1 + \dots + \alpha_{n+1} = m}} \frac{1}{\alpha_1! \cdot \dots \cdot \alpha_{n+1}!} \\
 &= \sum_{\alpha_1=0}^m \sum_{\substack{\alpha_2, \dots, \alpha_{n+1}=0, \dots, m \\ \alpha_2 + \dots + \alpha_{n+1} = m - \alpha_1}} \frac{1}{\alpha_1! \cdot \dots \cdot \alpha_{n+1}!} \\
 &= \sum_{\alpha_1=0}^m \frac{1}{\alpha_1!} \sum_{\substack{\alpha_2, \dots, \alpha_{n+1}=0, \dots, m \\ \alpha_2 + \dots + \alpha_{n+1} = m - \alpha_1}} \frac{1}{\alpha_2! \cdot \dots \cdot \alpha_{n+1}!} \\
 &= \sum_{\alpha_1=0}^m \frac{1}{\alpha_1!} s_n(m - \alpha_1) \\
 &\stackrel{\text{IV}}{=} \sum_{\alpha_1=0}^m \frac{1}{\alpha_1!} \frac{n^{m-\alpha_1}}{(m-\alpha_1)!} \\
 &= \frac{n^m \left(\frac{n+1}{n}\right)^m}{m!} \\
 &= \frac{(n+1)^m}{m!}.
 \end{aligned}$$

□

5.1.4 Lokale Niedrigrangapproximation von L^{-1}

An dieser Stelle befassen wir uns kurz mit der schwachen TAYLOR-Entwicklung; für eine ausführliche Darstellung siehe [BS02, Abs. 4.1].

Das TAYLOR-Polynom von $u \in C^{m-1}(\mathbb{R}^n)$ im Entwicklungspunkt y der Ordnung m ist gegeben durch

$$T_y^m u(x) := \sum_{|\alpha| < m} \frac{1}{\alpha!} D^\alpha u(y) \cdot (x - y)^\alpha, \quad \alpha \in \mathbb{N}_0^n.$$

Meistens liegt u in einem SOBOLEV-Raum, weshalb $D^\alpha u$ nicht punktweise existiert. Wir passen das eben definierte TAYLOR-Polynom diesem Umstand an, indem wir eine Mittelung bzgl. y durchführen.

Definition 5.6 *Habe u schwache Ableitungen der Ordnung kleiner m in einer Menge τ , die so beschaffen sei, daß der Ball $B = B_\rho(x_0)$ kompakt in τ liegt. Wir definieren über B das (ein!) **gemittelte TAYLOR-Polynom** von u der Ordnung m durch*

$$Q^m u(x) := \int_B T_y^m u(x) \cdot \varphi(y) dy,$$

wobei $\varphi \in C_0^\infty(\mathbb{R}^n)$ eine Abschneidefunktion mit $\text{supp}(\varphi) = \overline{B}$ ist.

Korollar 5.7 (Eigenschaften) *Die gemittelte TAYLOR-Entwicklung $Q^m u$ der Ordnung m ist ein Polynom vom Grad kleiner m in der Variablen x , d.h. für $x, y \in \Omega \subset \mathbb{R}^n$ gilt*

$$\dim(\text{im}(Q^m)) \leq \frac{(m+n)!}{m!n!}.$$

Die Wohldefiniertheit von $Q^m u$ ist bereits für $u \in L^1(B)$ gewährleistet.

Beweis: Das schwache TAYLOR-Polynom $Q^m u$ ist von der Ordnung m (siehe [BS02, (4.1.9)]), d.h. die größte vorkommende Potenz in diesem Polynom ist $m-1$. Der Vektorraum der Polynome vom Grad kleiner m besitzt die Dimension $\frac{(m+n)!}{m!n!} \leq m^n$, also ist $\dim(\text{im}(Q^m)) \leq \frac{(m+n)!}{m!n!}$ für $u: \mathbb{R}^n \rightarrow \mathbb{R}$.

Durch partielle Integration wird in [BS02, S. 95] gezeigt, daß $Q^m u$ nicht nur für $u \in H^{m-1}$, sondern auch für $u \in L^1$ wohldefiniert ist und Q^m selbst eine beschränkte Abbildung von L^1 - auf $W^{k,\infty}$ -Funktionen, $k \in \mathbb{N}_0$, ist. **QED**

Die Differenz

$$u - Q^m u =: R^m u$$

ist gleich dem Fehler m -ter Ordnung, den wir in der $L^2(\tau)$ -Norm abschätzen müssen. Nach [BS02, Proposition 4.2.8] gilt für die konvexe Hülle $C_x \subset \tau$ eines Balles $B := B_\rho(x_0)$ und eines Punktes x

$$\begin{aligned} R^m u(x) &= m \sum_{|\alpha|=m} \int_{C_x} \frac{(x-z)^\alpha}{\alpha!} \cdot k(x,z) \cdot D^\alpha u(z) dz, \\ |k(x,z)| &\leq c_{TE} \cdot |x-z|^{-n}, \\ c_{TE} &:= c \cdot (1+\gamma)^n. \end{aligned} \tag{5.16}$$

Bei Funktionen f , die von zwei Variablen x und z abhängen, bedeutet die Schreibweise $\|f(x, z)\|_{V(\Omega, x)}$, daß die V -Norm bzgl. der Variablen $x \in \mathbb{R}^n$ ausgewertet wird. Mit

$$\begin{aligned} |(x - z)^\alpha| &:= \left| \prod_{i=1}^n (x_i - z_i)^{\alpha_i} \right| = \prod_{i=1}^n |(x_i - z_i)^{\alpha_i}| \\ &\leq \prod_{i=1}^n |x - z|^{\alpha_i} = |x - z|^{\sum_{i=1}^n \alpha_i} = |x - z|^m \end{aligned} \quad (5.17)$$

folgt für den Fehler der schwachen TAYLOR-Entwicklung

$$\begin{aligned} \|R^m u\|_{L^2(\tau)} &= m \left\| \sum_{|\alpha|=m} \int_{C_x} \frac{(x - z)^\alpha}{\alpha!} k(x, z) D^\alpha u(z) dz \right\|_{L^2(\tau, x)} \\ &\stackrel{\text{MINKOWSKI}}{\leq} m \sum_{|\alpha|=m} \frac{1}{\alpha!} \left\| \int_{C_x} (x - z)^\alpha k(x, z) D^\alpha u(z) dz \right\|_{L^2(\tau, x)} \\ &= m \sum_{|\alpha|=m} \frac{1}{\alpha!} \sqrt{\int_\tau \left| \int_{C_x} (x - z)^\alpha k(x, z) D^\alpha u(z) dz \right|^2 dx} \\ &\leq m \sum_{|\alpha|=m} \frac{1}{\alpha!} \sqrt{\int_\tau \left(\int_{C_x} |(x - z)^\alpha| |k(x, z)| |D^\alpha u(z)| dz \right)^2 dx} \\ &\stackrel{(5.16, 5.17)}{\leq} m \sum_{|\alpha|=m} \frac{1}{\alpha!} \sqrt{\int_\tau \left(\int_{C_x} |x - z|^{m-n} c_{TE} |D^\alpha u(z)| dz \right)^2 dx} \\ &= m c_{TE} \sum_{|\alpha|=m} \frac{1}{\alpha!} \left\| \int_{C_x} |x - z|^{m-n} |D^\alpha u(z)| dz \right\|_{L^2(\tau, x)} \\ &\stackrel{\text{Lemma 5.4}}{\leq} m c_{TE} \sum_{|\alpha|=m} \frac{1}{\alpha!} \tilde{c}_{m,n} \varnothing_\tau^m \|D^\alpha u\|_{L^2(\tau)} \\ &= c_{TE} \cdot \tilde{c}_{m,n} \cdot m \cdot \varnothing_\tau^m \left[\sum_{|\alpha|=m} \frac{1}{\alpha!} \|D^\alpha u\|_{L^2(\tau)} \right] \\ &\leq c_{TE} \cdot \tilde{c}_{m,n} \cdot m \cdot \varnothing_\tau^m \cdot \sum_{|\alpha|=m} \frac{1}{\alpha!} \cdot \max_{|\alpha|=m} \|D^\alpha u\|_{L^2(\tau)}. \end{aligned}$$

Mit Lemma 5.5 und Satz 5.1 folgen dann

$$\|R^m u\|_{L^2(\tau)} \leq c_{TE} \cdot \tilde{c}_{m,n} \cdot m \cdot \varnothing_\tau^m \cdot \left[\frac{n^m}{m!} \right] \cdot |u|_{H^m(\tau)} \quad (5.18)$$

$$\leq \kappa_C^{m/2} n^m c_{TE} c_{reg} c_m \tilde{c}_{m,n} m \left(\frac{\varnothing_\tau}{\text{dist}(\sigma, \tau, \Gamma)} \right)^m \|f\|_{L^2(\sigma)}. \quad (5.19)$$

Bemerkung 5.8 (Die liebe Müh' mit dem Rand) *Nun bekommen wir ein Problem, denn nach Satz 5.1 haben wir $\text{dist}(\sigma, \tau, \Gamma)$ anstatt*

$$d = \min \{ \text{dist}(\tau, \sigma) \}.$$

Somit dürfen wir nicht die Standardzulässigkeitsbedingung $\varnothing_\tau \leq 2\eta \cdot \text{dist}(\sigma, \tau)$ einsetzen, sondern müssen die neue Zulässigkeitsbedingung $\varnothing_\tau \leq 2\eta \text{dist}(\sigma, \tau, \Gamma)$ verwenden. Diese hat aber zur Konsequenz, daß Cluster umso kleiner werden, je näher sie am Rand liegen; das führt letztlich zur Unzulässigkeit aller Randcluster — ob schon sie nur noch aus einem einzigen Knoten bestehen — und induziert auf diese Weise eine neue Arithmetik, für die die Aufwandsabschätzungen neu anzustellen sind.

Was wir hier benötigen, ist die Erweiterung von Ω um ein Gebiet D , sodaß

$$\text{dist}(\sigma, \tau, \Gamma) = \text{dist}(\sigma, \tau).$$

Auf D müssen wir dann eine Fortsetzung von u und den Koeffizienten c_{ij} definieren, sodaß κ_C erhalten bleibt und u mitsamt seinen m -ten partiellen Ableitungen bzgl. L und des neuen auf $D \setminus \Omega$ konstruierten Differentialoperators harmonisch ist. Leider ist dies unmöglich, weshalb wir für den Augenblick $d = \text{dist}(\sigma, \tau)$ annehmen und eine Umgehung der beiden Hindernisse in Absatz 5.3 vorstellen.

Aus der Standardzulässigkeitsbedingung $\varnothing_\tau \leq 2\eta \cdot \text{dist}(\sigma, \tau)$ folgt

$$\frac{\varnothing_\tau}{\text{dist}(\sigma, \tau)} \leq 2\eta.$$

Wir verwenden diese Abschätzung in (5.19) und erreichen

$$\|R^m u\|_{L^2(\tau)} \leq c_{(5.20)} c_m c_{m,n} (c_0 \eta)^m \|f\|_{L^2(\sigma)} \quad \forall f \in L^2(\sigma) \quad (5.20)$$

mit den Konstanten

$$c_{(5.20)} := c_{TE} c_{reg}, \quad c_{m,n} := 2^m m \tilde{c}_{m,n}$$

und — in historischer Würdigung der Pionierarbeiten [Hb98], [HK99] —

$$c_0 := \sqrt{\kappa_C n^2},$$

d.h. wir haben (wegen der bereits begründeten Annahme $c_m c_{m,n} \approx 1$) für $\eta < c_0^{-1}$ den Fehler, den wir bei der Approximation von u durch $Q^m u$ begehen, auf dem Cluster $\beta := \tau \times \sigma$ abgeschätzt und erreichen exponentielle Konvergenz.

Im nächsten Schritt stellen wir den Bezug zu den Operatoren her und springen dann mittels des Koeffizientenisomorphismus J direkt in die entsprechenden Matrixblöcke. Aus der Differentialgleichung ergibt sich $u = L^{-1} f$ für alle $f \in L^2(\sigma)$. Wenden wir auf diese Gleichheit den Operator Q^m an, führt uns das zur gesuchten **Niedrigrangapproximation** $Q^m L^{-1}$ von L^{-1} .

Korollar 5.9 *Der Operator $Q^m L^{-1}$ besitzt den Maximalrang $\frac{(m+n)!}{m!n!}$ und konvergiert auf dem Cluster $\beta := \tau \times \sigma$ exponentiell gegen L^{-1} , d.h.*

$$\|L^{-1} - Q^m L^{-1}\|_{\mathcal{L}(L^2(\sigma), L^2(\tau))} \leq c_{(5.20)} c_m c_{m,n} (c_0 \eta)^m.$$

Beweis: Wegen $\dim \operatorname{im} (Q^m) \leq \frac{(m+n)!}{m!n!}$ folgt unverzüglich

$$\dim \operatorname{im} (Q^m L^{-1}) \leq \min \{ \dim \operatorname{im} (Q^m), \dim \operatorname{im} (L^{-1}) \} \leq \frac{(m+n)!}{m!n!}.$$

Für alle $f \in L^2(\sigma)$ ist

$$\|R^m u\|_{L^2(\tau)} = \|u - Q^m u\|_{L^2(\tau)} = \|L^{-1} f - Q^m L^{-1} f\|_{L^2(\tau)},$$

weshalb mit (5.20) die Behauptung folgt. □

Jetzt übertragen wir dieses Ergebnis auf den Matrixblock. Seien $s, t \subset I := \{1, \dots, N\}$ die zu σ bzw. τ gehörenden Indexteilmengen. Den zum Cluster $b := t \times s$ gehörenden Matrixblock $[(L^{-1})_h]_b$ approximieren wir durch

$$[(L^{-1})_{\mathcal{H}}]_b := [J^* Q^m L^{-1} J]_b$$

Diese Matrix $[(L^{-1})_{\mathcal{H}}]_b$ ist eine \mathcal{H} -, genauer Rk -Approximation von $[(L^{-1})_h]_b$! Der Operator $Q^m L^{-1}$ ist wegen Korollar 5.7 höchstens vom Rang $\frac{(m+n)!}{m!n!}$, also kann auch die Verkettung $J^* Q^m L^{-1} J$ maximal diesen Rang haben, d.h. für beliebige Indexteilmengen $s, t \subset I$ gilt

$$\operatorname{rk} \left([(L^{-1})_{\mathcal{H}}]_b \right) = \min \left\{ |t|, \frac{(m+n)!}{m!n!}, |s| \right\} \leq \min \{ |t|, |s| \} = \operatorname{rk} \left([(L^{-1})_h]_b \right).$$

Daher folgt für Koeffizientenvektoren $\mathfrak{x} \in \mathbb{R}^{|t|}$, $\mathfrak{y} \in \mathbb{R}^{|s|}$ und $\Omega_t := \bigcup_{i \in t} \operatorname{supp}(\varphi_i)$

$$\begin{aligned} & \left| \left\langle [(L^{-1})_h - (L^{-1})_{\mathcal{H}}]_{t \times s} \mathfrak{y}, \mathfrak{x} \right\rangle_h \right| \\ &= \left| \left\langle [J^* (L^{-1} - Q^m L^{-1}) J]_{t \times s} \mathfrak{y}, \mathfrak{x} \right\rangle_h \right| \\ &= \left| \left((L^{-1} - Q^m L^{-1}) J_s \mathfrak{y}, J_t \mathfrak{x} \right)_{L^2(\Omega_t)} \right| \\ &\leq \| (L^{-1} - Q^m L^{-1}) J_s \mathfrak{y} \|_{L^2(\Omega_s)} \cdot \| J_t \mathfrak{x} \|_{L^2(\Omega_t)} \\ &\leq \| L^{-1} - Q^m L^{-1} \|_{\mathcal{L}(L^2(\sigma), L^2(\tau))} \cdot \| J_s \mathfrak{y} \|_{L^2(\Omega_s)} \cdot \| J_t \mathfrak{x} \|_{L^2(\Omega_t)} \\ &\stackrel{\text{Kor. 5.9, Satz 3.4}}{\leq} c_{(5.20)} c_m c_{m,n} (c_0 \eta)^m \cdot c_J \| \mathfrak{y} \|_h \cdot c_J \| \mathfrak{x} \|_h \end{aligned}$$

und damit

$$\begin{aligned} \left\| [(L^{-1})_h - (L^{-1})_{\mathcal{H}}]_b \right\|_2 &:= \sup \left\{ \left| \frac{\left\langle [(L^{-1})_h - (L^{-1})_{\mathcal{H}}]_b \mathfrak{y}, \mathfrak{x} \right\rangle_h}{\| \mathfrak{y} \|_h \cdot \| \mathfrak{x} \|_h} \right| \mid \mathfrak{x} \in \mathbb{R}^{|t|}, \mathfrak{y} \in \mathbb{R}^{|s|} \right\} \\ &\leq \sup \left\{ \frac{c_{(5.20)} (c_0 \eta)^m \cdot c_J^2 \| \mathfrak{y} \|_h \cdot \| \mathfrak{x} \|_h}{\| \mathfrak{y} \|_h \cdot \| \mathfrak{x} \|_h} \mid \mathfrak{x} \in \mathbb{R}^{|t|}, \mathfrak{y} \in \mathbb{R}^{|s|} \right\} \\ &= c_{(5.20)} c_J^2 c_m c_{m,n} (c_0 \eta)^m. \end{aligned}$$

Wir fassen das Ergebnis zusammen in

Satz 5.10 (Lokale \mathcal{H} -Approximation von $(L^{-1})_h$) Seien $\sigma, \tau \subset \Omega$ zulässige Teilgebiete und $s, t \subset I$ die assoziierten Indermengen. Bezeichne $(L^{-1})_h$ die diskrete Inverse. Dann ist

$$[(L^{-1})_{\mathcal{H}}]_b := [J^* Q^m L^{-1} J]_b$$

die \mathcal{H} -Approximation von $(L^{-1})_h$ auf dem Matrixteilblock $b := t \times s$ mit

$$\left\| [(L^{-1})_h - (L^{-1})_{\mathcal{H}}]_b \right\|_2 \leq c_{(5.20)} c_J^2 c_m c_{m,n} (c_0 \eta)^m \quad \text{und} \quad \text{rk} \left([(L^{-1})_{\mathcal{H}}]_b \right) \leq \frac{(m+n)!}{m!n!}.$$

5.2 Lokale Approximationseigenschaft für Matrixoperatoren

Teil I: konstante Koeffizienten

Gegeben sei das lineare und gleichmäßig elliptische Differentialgleichungssystem zweiter Ordnung (2.18). Idee und Führung des Beweises basieren auf dem in Abschnitt 5.1 vorgestellten Muster.

5.2.1 Innere Regularität

Satz 5.11 (Innere Regularität elliptischer Matrixdifferentialoperatoren)

Sei \vec{L} wie in (2.18) ein gleichmäßig elliptischer linearer Matrixdifferentialoperator zweiter Ordnung mit den konstanten Koeffizienten $c_{ijkl} \in \mathbb{R}$, die der Symmetriebedingung (2.19) genügen. Seien $\vec{\lambda}_{\min}$ und $\vec{\lambda}_{\max}$ der kleinste und größte Eigenwert der Matrix $\vec{C} := (C_{ij})_{i,j=1,\dots,q}$ für alle $x \in \Omega$ sowie $\kappa_{\vec{C}} := \vec{\lambda}_{\max} \cdot \vec{\lambda}_{\min}^{-1}$. Seien σ, τ zwei kompakte Teilmengen von Ω (s. Abb. 5.2a) mit $\sigma \cap \tau = \{\}$ und positivem Abstand zum Rand, d.h.

$$\text{dist}(\sigma, \tau, \Gamma) := \min \{ \text{dist}(\tau, \sigma), \text{dist}(\tau, \Gamma) \} > 0.$$

Sei \vec{u} die schwache Lösung der Differentialgleichung

$$\vec{L}[\vec{u}](x) = \vec{f}(x) \quad \text{in } \Omega, \quad \vec{u} = 0 \quad \text{auf } \Gamma$$

und \vec{L} -harmonisch in $\Omega \setminus \sigma$, wobei $\text{supp}(\vec{f}) \subset\subset \sigma$.

Dann sind die m -ten partiellen Ableitungen von \vec{u} \vec{L} -harmonisch auf $\Omega \setminus \sigma$ und es existiert eine Konstante $c_{\text{reg}} > 0$, sodaß

$$|\vec{u}|_{H^m(\tau, \mathbb{R}^q)} \leq c_{\text{reg}} \cdot c_m \cdot \kappa_{\vec{C}}^{m/2} \cdot \frac{m!}{\text{dist}(\sigma, \tau, \Gamma)^m} \cdot \left\| \vec{f} \right\|_{L^2(\sigma, \mathbb{R}^q)}$$

mit $c_m = (m!)^\psi$, $\psi > 1$, und $m \in \mathbb{N}$.

Beweis

Obwohl $\nabla \vec{u}$ eine Matrix ist, interpretieren wir hier $\nabla \vec{u}$ als Spaltenvektor der Länge nq , d.h.

$$\nabla \vec{u}(x) := \left(\nabla u_1(x)^\top \quad \dots \quad \nabla u_q(x)^\top \right)^\top.$$

Dies erlaubt uns, in einfacher Weise an die vorigen Abschnitte anzuknüpfen:

$$\vec{\lambda}_{\min} \langle \nabla \vec{u}, \nabla \vec{u} \rangle = \vec{\lambda}_{\min} \sum_{i=1}^q \sum_{k=1}^n (\partial_k u_i)^2 \leq \sum_{i,j=1}^q \sum_{k,l=1}^n c_{ijkl} \partial_k u_i \partial_l u_j = \langle \vec{C} \nabla \vec{u}, \nabla \vec{u} \rangle.$$

Wir vergewissern uns noch, daß

$$\begin{aligned} \|\nabla \vec{u}\|_{L^2(\Omega, \mathbb{R}^q)}^2 &= \int_{\Omega} \langle \nabla \vec{u}, \nabla \vec{u} \rangle = \int_{\Omega} \sum_{i=1}^q \sum_{k=1}^n (\partial_k u_i)^2 \\ &= \sum_{i=1}^q \int_{\Omega} \sum_{k=1}^n (\partial_k u_i)^2 = \sum_{i=1}^q \int_{\Omega} \langle \nabla u_i, \nabla u_i \rangle = \sum_{i=1}^q |u_i|_{H_0^1(\Omega)}^2 = |\vec{u}|_{H_0^1(\Omega, \mathbb{R}^q)}^2 \end{aligned}$$

gilt. Sei $\eta \in C^\infty(\Omega)$ eine Abschneidefunktion mit den Eigenschaften (5.3) und (5.4).

Schritt 1

Mit den Vorüberlegungen folgt unverzüglich

$$\begin{aligned} & \vec{\lambda}_{\min} |\vec{u}|_{H^1(B_R, \mathbb{R}^q)}^2 \\ &= \vec{\lambda}_{\min} \int_{B_R} \langle \nabla \vec{u}, \nabla \vec{u} \rangle \\ &\leq \vec{\lambda}_{\min} \int_{B_{\lambda R}} \langle \nabla(\eta \vec{u}), \nabla(\eta \vec{u}) \rangle \\ &\stackrel{\vec{C} \text{ pos.-def.}}{\leq} \int_{B_{\lambda R}} \langle \vec{C} \nabla(\eta \vec{u}), \nabla(\eta \vec{u}) \rangle \\ &= \sum_{i,j=1}^q \int_{B_{\lambda R}} \langle C_{ij} \nabla(\eta u_j), \nabla(\eta u_i) \rangle \\ &= \sum_{i,j=1}^q \int_{B_{\lambda R}} \sum_{k,l=1}^n c_{ijkl} \cdot \partial_k(\eta u_i) \cdot \partial_l(\eta u_j) \\ &= \sum_{i,j=1}^q \int_{B_{\lambda R}} \sum_{k=1}^n \partial_k(\eta u_i) \cdot \left(\sum_{l=1}^n c_{ijkl} \cdot \partial_l(\eta u_j) \right) \\ &\stackrel{\text{part. Int.}}{=} \sum_{i,j=1}^q \sum_{k=1}^n \left[\int_{\partial B_{\lambda R}} \eta u_i \sum_{l=1}^n c_{ijkl} \partial_l(\eta u_j) - \int_{B_{\lambda R}} \eta u_i \cdot \partial_k \left(\sum_{l=1}^n c_{ijkl} \partial_l(\eta u_j) \right) \right] \\ &\stackrel{\eta|_{\partial B_{\lambda R}}=0}{=} - \sum_{i,j=1}^q \sum_{k=1}^n \int_{B_{\lambda R}} \eta u_i \cdot \partial_k \left(\sum_{l=1}^n c_{ijkl} \partial_l(\eta u_j) \right) \end{aligned}$$

$$\begin{aligned}
&= - \sum_{i,j=1}^q \int_{B_{\lambda R}} \eta u_i \cdot \sum_{k=1}^n \partial_k \left(\sum_{l=1}^n c_{ijkl} \partial_l (\eta u_j) \right) \\
&= - \sum_{i,j=1}^q \int_{B_{\lambda R}} \eta u_i \cdot \langle \nabla, C_{ij} \nabla (\eta u_j) \rangle \\
&\stackrel{(5.5)}{=} - \sum_{i,j=1}^q \int_{B_{\lambda R}} \eta u_i [\langle \nabla, C_{ij} \nabla \eta \rangle u_j + \langle \nabla u_j, C_{ij} \nabla \eta \rangle] \tag{5.21} \\
&\quad - \sum_{i,j=1}^q \int_{B_{\lambda R}} \eta u_i \left[\underbrace{\langle \nabla, C_{ij} \nabla u_j \rangle \eta}_{=L_{ij} u_j} + \underbrace{\langle \nabla \eta, C_{ij} \nabla u_j \rangle}_{=\langle C_{ij} \nabla \eta, \nabla u_j \rangle \text{ wg. (2.19)}} \right] \\
&\stackrel{\vec{u} \text{ } \vec{L}\text{-harm.}}{=} - \sum_{i,j=1}^q \int_{B_{\lambda R}} \eta u_i [\langle \nabla, C_{ij} \nabla \eta \rangle u_j + 2 \langle \nabla u_j, C_{ij} \nabla \eta \rangle] \\
&= - \sum_{i,j=1}^q \int_{B_{\lambda R}} \eta u_i \langle \nabla, C_{ij} \nabla \eta \rangle u_j - \sum_{i,j=1}^q \int_{B_{\lambda R}} 2 \eta u_i \langle \nabla u_j, C_{ij} \nabla \eta \rangle. \tag{5.22}
\end{aligned}$$

Wir stellen für das Integral der zweiten Summe separate Rechnungen an:

$$\begin{aligned}
&\int_{B_{\lambda R}} 2 \eta u_i \langle \nabla u_j, C_{ij} \nabla \eta \rangle \\
&= \int_{B_{\lambda R}} \langle u_i \nabla u_j, C_{ij} 2 \eta \nabla \eta \rangle \\
&= \int_{B_{\lambda R}} \langle \nabla u_j, u_i C_{ij} \nabla (\eta^2) \rangle \\
&= \int_{B_{\lambda R}} \sum_{k=1}^n \left[\partial_k u_j \cdot u_i \sum_{l=1}^n c_{ijkl} \partial_l (\eta^2) \right] \\
&= \sum_{k=1}^n \int_{B_{\lambda R}} \partial_k u_j \cdot u_i \sum_{l=1}^n c_{ijkl} \partial_l (\eta^2) \\
&\stackrel{\text{part. Int.}}{=} \sum_{k=1}^n \int_{\partial B_{\lambda R}} u_j \cdot u_i \sum_{l=1}^n c_{ijkl} \underbrace{\partial_l (\eta^2)}_{=2\eta \partial_l \eta=0} - \sum_{k=1}^n \int_{B_{\lambda R}} u_j \cdot \partial_k \left[u_i \sum_{l=1}^n c_{ijkl} \partial_l (\eta^2) \right] \\
&= - \sum_{k=1}^n \int_{B_{\lambda R}} u_j \cdot \left[\partial_k u_i \cdot \sum_{l=1}^n c_{ijkl} \partial_l (\eta^2) + u_i \cdot \partial_k \left(\sum_{l=1}^n c_{ijkl} \partial_l (\eta^2) \right) \right] \\
&= - \int_{B_{\lambda R}} u_j \cdot \sum_{k=1}^n \left[\partial_k u_i \cdot \sum_{l=1}^n c_{ijkl} \partial_l (\eta^2) \right] \\
&\quad - \int_{B_{\lambda R}} u_j u_i \cdot \sum_{k=1}^n \left[\partial_k \left(\sum_{l=1}^n c_{ijkl} \partial_l (\eta^2) \right) \right] \\
&= - \int_{B_{\lambda R}} u_j \langle \nabla u_i, C_{ij} \nabla (\eta^2) \rangle - \int_{B_{\lambda R}} u_i u_j \langle \nabla, C_{ij} \nabla (\eta^2) \rangle.
\end{aligned}$$

Betrachten wir die letzte und die dritte Zeile des Formelblockes, folgt

$$- \int_{B_{\lambda R}} u_i u_j \langle \nabla, C_{ij} \nabla (\eta^2) \rangle = \int_{B_{\lambda R}} \langle u_i \nabla u_j, C_{ij} \nabla (\eta^2) \rangle + \int_{B_{\lambda R}} u_j \langle \nabla u_i, C_{ij} \nabla (\eta^2) \rangle$$

und mit Summenbildung über i, j

$$\begin{aligned} & - \sum_{i,j=1}^q \int_{B_{\lambda R}} u_i u_j \langle \nabla, C_{ij} \nabla (\eta^2) \rangle \\ &= \sum_{i,j=1}^q \int_{B_{\lambda R}} \langle u_i \nabla u_j, C_{ij} \nabla (\eta^2) \rangle + \sum_{i,j=1}^q \int_{B_{\lambda R}} u_j \langle \nabla u_i, C_{ij} \nabla (\eta^2) \rangle \\ &\stackrel{(2.19)}{=} \sum_{i,j=1}^q \int_{B_{\lambda R}} \langle u_i \nabla u_j, C_{ij} \nabla (\eta^2) \rangle + \sum_{i,j=1}^q \int_{B_{\lambda R}} u_j \langle \nabla u_i, C_{ji} \nabla (\eta^2) \rangle \\ &= 2 \sum_{i,j=1}^q \int_{B_{\lambda R}} 2\eta u_i \langle \nabla u_j, C_{ij} \nabla \eta \rangle. \end{aligned}$$

Dieses Ergebnis pflügen wir in die Hauptüberlegung (5.22) ein:

$$\begin{aligned} & - \sum_{i,j=1}^q \int_{B_{\lambda R}} \eta u_i \langle \nabla, C_{ij} \nabla \eta \rangle u_j - \sum_{i,j=1}^q \int_{B_{\lambda R}} 2\eta u_i \langle \nabla u_j, C_{ij} \nabla \eta \rangle \\ &= - \sum_{i,j=1}^q \int_{B_{\lambda R}} \eta \langle \nabla, C_{ij} \nabla \eta \rangle u_i u_j + \frac{1}{2} \sum_{i,j=1}^q \int_{B_{\lambda R}} \langle \nabla, C_{ij} \nabla (\eta^2) \rangle u_i u_j \\ &= \sum_{i,j=1}^q \int_{B_{\lambda R}} \langle \nabla, \eta C_{ij} \nabla \eta \rangle u_i u_j - \eta \langle \nabla, C_{ij} \nabla \eta \rangle u_i u_j \\ &= \sum_{i,j=1}^q \int_{B_{\lambda R}} \eta \langle \nabla, C_{ij} \nabla \eta \rangle u_i u_j + \langle \nabla \eta, C_{ij} \nabla \eta \rangle u_i u_j - \eta \langle \nabla, C_{ij} \nabla \eta \rangle u_i u_j \\ &= \sum_{i,j=1}^q \int_{B_{\lambda R}} \langle \nabla \eta, C_{ij} \nabla \eta \rangle u_i u_j \\ &= \sum_{i,j=1}^q \int_{B_{\lambda R}} \sum_{k,l=1}^n \partial_k \eta \cdot c_{ijkl} \cdot \partial_l \eta \cdot u_i u_j \tag{5.23} \\ &= \int_{B_{\lambda R}} \sum_{i,j=1}^q \sum_{k,l=1}^n c_{ijkl} \cdot (\partial_k \eta) u_i \cdot (\partial_l \eta) u_j \\ &\leq \vec{\lambda}_{max} \int_{B_{\lambda R}} \sum_{i=1}^q \sum_{k=1}^n (\partial_k \eta \cdot u_i)^2 \\ &= \vec{\lambda}_{max} \int_{B_{\lambda R}} |\nabla \eta|^2 \cdot |\vec{u}|^2 \\ &\stackrel{(5.4)}{\leq} \frac{\vec{\lambda}_{max} (1 + \delta)^2}{(\lambda R - R)^2} \int_{B_{\lambda R} \setminus B_R} |\vec{u}|^2 \end{aligned}$$

mit $0 < \delta \ll 1$. Nach der die Ungleichungsrelation erhaltenden Erweiterung des Integrationsbereichs von $B_{\lambda R} \setminus B_R$ auf $B_{\lambda R}$ und der Division mit $\vec{\lambda}_{min}$ erhalten wir als Ergebnis des ersten Schrittes

$$|\vec{u}|_{H^1(B_R, \mathbb{R}^q)} \leq \frac{\sqrt{\kappa \vec{c}} (1 + \delta)}{\lambda R - R} \|\vec{u}\|_{L^2(B_{\lambda R}, \mathbb{R}^q)}. \quad (5.24)$$

Schritt 2

Wir übernehmen die Schachtelung (5.9). Die \vec{L} -Harmonizität der m -ten partiellen Ableitung folgt wie in (5.10). Wir iterieren (5.24):

$$\begin{aligned} |\vec{u}|_{H^m(B_1, \mathbb{R}^q)} &\leq \frac{1}{\lambda_2 - \lambda_1} \frac{\sqrt{\kappa \vec{c}} (1 + \delta)}{R} |\vec{u}|_{H^{m-1}(B_2, \mathbb{R}^q)} \\ &\leq \frac{1}{\lambda_3 - \lambda_2} \frac{1}{\lambda_2 - \lambda_1} \left(\frac{\sqrt{\kappa \vec{c}} (1 + \delta)}{R} \right)^2 |\vec{u}|_{H^{m-2}(B_3, \mathbb{R}^q)} \\ &\leq \dots \\ &\leq \prod_{i=1}^m \frac{1}{\lambda_{i+1} - \lambda_i} \left(\frac{\sqrt{\kappa \vec{c}} (1 + \delta)}{R} \right)^m |\vec{u}|_{H^0(B_{m+1}, \mathbb{R}^q)}. \end{aligned} \quad (5.25)$$

Schritt 3

Unser Ziel ist eine Abschätzung der Form $|\vec{u}|_{H^m(\tau, \mathbb{R}^q)} \leq \text{const} \cdot \|\vec{f}\|_{L^2(\sigma, \mathbb{R}^q)}$. Da nach Voraussetzung der Träger von \vec{f} kompakt in σ ist, folgt sofort

$$\|\vec{f}\|_{L^2(\sigma, \mathbb{R}^q)} = \|\vec{f}\|_{L^2(\Omega, \mathbb{R}^q)}. \quad (5.26)$$

Andererseits ist $\|\vec{u}\|_{L^2(B_m(x_k), \mathbb{R}^q)} \leq \|\vec{u}\|_{L^2(\Omega, \mathbb{R}^q)}$, sodaß aus (5.25)

$$|\vec{u}|_{H^m(B_0(x_k), \mathbb{R}^q)} \leq \prod_{i=1}^m \frac{1}{\lambda_{i+1} - \lambda_i} \left(\frac{\sqrt{\kappa \vec{c}} (1 + \delta)}{R} \right)^m \|\vec{u}\|_{L^2(\Omega, \mathbb{R}^q)}$$

folgt und mit der gemäß (5.12) definierten Überdeckung \vec{U} von τ

$$\begin{aligned} |\vec{u}|_{H^m(\tau, \mathbb{R}^q)} &\leq |\vec{u}|_{H^m(\vec{U}, \mathbb{R}^q)} \\ &= \sqrt{\int_{\vec{U}} |\nabla^m \vec{u}|^2} \\ &\leq \sqrt{\sum_{k=1}^{N_\tau} \int_{B_1(x_k)} |\nabla^m \vec{u}|^2} \\ &= \sqrt{\sum_{k=1}^{N_\tau} |\vec{u}|_{H^m(B_1(x_k), \mathbb{R}^q)}^2} \end{aligned}$$

$$\leq \sqrt{N_\tau} \prod_{i=1}^m \frac{1}{\lambda_{i+1} - \lambda_i} \left(\frac{\sqrt{\kappa \bar{C}} (1 + \delta)}{R} \right)^m \|\vec{u}\|_{L^2(\Omega, \mathbb{R}^q)}. \quad (5.27)$$

Jetzt müssen wir (5.27) mit (5.26) verbinden, d.h. eine Beziehung zwischen $\|\vec{u}\|_{L^2}$ und $\|\vec{f}\|_{L^2}$ herstellen. Das gelingt uns mit Hilfe des Satzes von LAX-MILGRAM,

$$\|\vec{u}\|_{L^2}^2 \leq c_\Omega^2 \cdot \|\vec{f}\|_{L^2}^2.$$

Diese Ungleichung ist das gesuchte Verbindungsglied und führt uns mit $\frac{1}{R} = \frac{p}{\text{dist}(\sigma, \tau, \Gamma)}$ zum Ziel des dritten Beweisschrittes, nämlich

$$\|\vec{u}\|_{H^m(\tau, \mathbb{R}^q)} \leq \sqrt{N_\tau} \prod_{i=1}^m \frac{p}{\lambda_{i+1} - \lambda_i} \left(\frac{\sqrt{\kappa \bar{C}} (1 + \delta)}{\text{dist}(\sigma, \tau, \Gamma)} \right)^m c_\Omega \|\vec{f}\|_{L^2(\sigma, \mathbb{R}^q)}. \quad (5.28)$$

Schritt 4

... behandelt die Abschätzung des Faktors $\prod_{i=1}^m \frac{1}{\lambda_{i+1} R - \lambda_i R}$. Auch hier gilt (5.14).

Schritt 5

... faßt (5.14) mit (5.28) zusammen,

$$\|\vec{u}\|_{H^m(\tau, \mathbb{R}^q)} \leq c_{reg} \kappa \frac{m/2}{\bar{C}} (1 + \delta)^m \frac{(m!)^{1+\psi}}{\text{dist}(\sigma, \tau, \Gamma)^m} \|\vec{f}\|_{L^2(\sigma, \mathbb{R}^q)}, \quad \psi > 1,$$

und bildet so das Ende des Beweises. Wie bei skalaren Differentialoperatoren eliminieren wir den Faktor $(1 + \delta)^m$ per Grenzwertbildung $\delta \rightarrow 0$ und setzen

$$c_{reg} := \sqrt{N_\tau} c_\Omega c(5.14), \quad c_m := (m!)^\psi.$$

□

5.2.2 Lokale Niedrigrangapproximation von \vec{L}^{-1}

Sei $Q^m u_j$ das gemittelte TAYLOR-Polynom von u_j der Ordnung m und $R^m u_j := u_j - Q^m u_j$ der Fehler m -ter Ordnung. Über diese Komponentenfunktionen definieren wir das vektorielle TAYLOR-Polynom

$$\vec{Q}^m \vec{u} := (Q^m u_j)_{j=1, \dots, q} : \Omega \rightarrow \mathbb{R}^q$$

mit dem Fehler (-vektor)

$$\vec{R}^m \vec{u} := \vec{u} - \vec{Q}^m \vec{u} : \Omega \rightarrow \mathbb{R}^q,$$

der in der $L^2(\tau, \mathbb{R}^q)$ -Norm abzuschätzen ist:

$$\begin{aligned}
\left\| \vec{R}^m \vec{u} \right\|_{L^2(\tau, \mathbb{R}^q)}^2 &:= \sum_{j=1}^q \|R^m u_j\|_{L^2(\tau)}^2 \\
&\stackrel{(5.18)}{\leq} \sum_{j=1}^q \left[c_j (1 + \gamma_j)^n \tilde{c}_{m,n} \cdot m \cdot \varnothing_\tau^m \cdot \frac{n^m}{m!} \right]^2 \cdot |u_j|_{H^m(\tau)}^2 \\
&\leq \left[c (1 + \gamma)^n \tilde{c}_{m,n} \cdot m \cdot \varnothing_\tau^m \cdot \frac{n^m}{m!} \right]^2 \cdot \sum_{j=1}^q |u_j|_{H^m(\tau)}^2 \\
&= \left[c_{TE} \tilde{c}_{m,n} \cdot m \cdot \varnothing_\tau^m \cdot \frac{n^m}{m!} \right]^2 \cdot \sum_{j=1}^q |u_j|_{H^m(\tau)}^2
\end{aligned}$$

mit $c := \max\{c_1, \dots, c_q\}$, $\gamma := \max\{\gamma_1, \dots, \gamma_q\}$ und $c_{TE} := c(1 + \gamma)^n$. Zum einen ist $\sum_{j=1}^q |u_j|_{H^m(\tau)}^2 = |\vec{u}|_{H^m(\tau, \mathbb{R}^q)}^2$ und mit dem Regularitätssatz 5.11 aus dem letzten Abschnitt abschätzbar, zum anderen ist die Standardzulässigkeitsbedingung unabhängig vom Differentialoperator. Wir nehmen wieder $\text{dist}(\sigma, \tau, \Gamma) = \text{dist}(\sigma, \tau)$ an, vgl. Bemerkung 5.8. Für den Fehlervektor, der bei der Approximation von \vec{u} durch $\vec{Q}^m \vec{u}$ auf τ entsteht, gilt mit der Standardzulässigkeitsbedingung

$$\begin{aligned}
\left\| \vec{R}^m \vec{u} \right\|_{L^2(\tau, \mathbb{R}^q)} &\leq \left[c_{TE} \tilde{c}_{m,n} m \varnothing_\tau^m \frac{n^m}{m!} \right] \cdot \left[c_{reg} c_m \kappa_{\vec{c}}^{m/2} \frac{m!}{\text{dist}(\sigma, \tau, \Gamma)^m} \left\| \vec{f} \right\|_{L^2(\sigma, \mathbb{R}^q)} \right] \\
&= c_{TE} \tilde{c}_{m,n} m n^m c_{reg} c_m \kappa_{\vec{c}}^{m/2} \left(\frac{\varnothing_\tau}{\text{dist}(\sigma, \tau)} \right)^m \left\| \vec{f} \right\|_{L^2(\sigma, \mathbb{R}^q)} \\
&\leq c_{TE} \tilde{c}_{m,n} m n^m c_{reg} c_m \kappa_{\vec{c}}^{m/2} (2\eta)^m \left\| \vec{f} \right\|_{L^2(\sigma, \mathbb{R}^q)} \\
&= (c_{TE} c_{reg}) \cdot c_m \cdot (2^m m \tilde{c}_{m,n}) \cdot \left(\sqrt{\kappa_{\vec{c}} n^2} \right)^m \cdot \eta^m \cdot \left\| \vec{f} \right\|_{L^2(\sigma, \mathbb{R}^q)}.
\end{aligned}$$

Wir setzen

$$c_{m,n} := 2^m m \tilde{c}_{m,n}, \quad \vec{c}_0 := \sqrt{\kappa_{\vec{c}} n^2} \quad \text{und} \quad c_{(5.29)} := c_{TE} c_{reg} c_n, \quad (5.29)$$

vgl. S. 62. Wegen $\vec{R}^m \vec{u} = \vec{u} - \vec{Q}^m \vec{u}$ gilt also kurzum

$$\left\| \vec{u} - \vec{Q}^m \vec{u} \right\|_{L^2(\tau, \mathbb{R}^q)} \leq c_{(5.29)} c_m c_{m,n} (\vec{c}_0 \eta)^m \left\| \vec{f} \right\|_{L^2(\sigma, \mathbb{R}^q)} \quad \forall \vec{f} \in L^2(\sigma, \mathbb{R}^q). \quad (5.30)$$

Lösen wir die Differentialgleichung $\vec{L} \vec{u} = \vec{f}$ nach \vec{u} auf und wenden \vec{Q}^m an, erhalten wir $\vec{Q}^m \vec{u} = \vec{Q}^m \vec{L}^{-1} \vec{f}$, d.h. $\vec{Q}^m \vec{L}^{-1}$ ist die Niedrigrangapproximation von \vec{L}^{-1} . Mit dieser Erkenntnis wenden wir uns nochmals dem Approximationsfehler durch das TAYLOR-Polynom zu und stellen fest, daß

$$\left\| \vec{u} - \vec{Q}^m \vec{u} \right\|_{L^2(\tau, \mathbb{R}^q)} = \left\| \vec{L}^{-1} \vec{f} - \vec{Q}^m \vec{L}^{-1} \vec{f} \right\|_{L^2(\sigma, \mathbb{R}^q)} \quad (5.31)$$

für alle $\vec{f} \in L^2(\sigma, \mathbb{R}^q)$ ist. Wir fassen (5.30) und (5.31) zusammen im

Korollar 5.12 Die Abbildung $\vec{Q}^m \vec{L}^{-1}$ besitzt den Maximalrang $q \frac{(m+n)!}{m!n!}$ und erfüllt

$$\left\| \vec{L}^{-1} - \vec{Q}^m \vec{L}^{-1} \right\|_{\mathcal{L}(L^2(\sigma, \mathbb{R}^q), L^2(\tau, \mathbb{R}^q))} \leq c_{(5.29)} c_m c_{m,n} (\vec{c}_0 \eta)^m.$$

Dieses Ergebnis entspricht dem Korollar 5.9 aus dem skalaren Fall und ist nun auf einen Matrixblock zu übertragen. Dazu benötigen wir die in Abschnitt 3.5.2 definierte Erweiterung des bisherigen Koeffizientenisomorphismus, die so ausgelegt ist, daß bei Differentialgleichungssystemen dem Cluster $\beta := \tau \times \sigma$ ein Matrixblock entspricht, d.h. die Informationen zu einem Gitterpunkt sind in q aufeinanderfolgenden Zeilen kodiert, was dem Prinzip der geometrischen Clustering entgegenkommt.

Satz 5.13 (Lokale \mathcal{H} -Approximation von $(\vec{L}^{-1})_h$) Die RITZ-GALERKIN-Approximation von \vec{L}^{-1} sei mit $(\vec{L}^{-1})_h$ und ihre \mathcal{H} -Approximation mit $(\vec{L}^{-1})_{\mathcal{H}}$ bezeichnet. Mit den zulässigen und kompakten Teilmengen $\sigma, \tau \subset \Omega$ seien die Indexmengen

$$s := \{s_1, \dots, s_a\}, \quad t := \{t_1, \dots, t_b\} \subset I = \{1, \dots, N\}$$

assoziiert sowie die Menge der Spaltennummern,

$$\vec{s} := \{(s_i - 1)q + 1, \dots, s_i q \mid i = 1, \dots, a\}, \quad |\vec{s}| = q |s| = qa,$$

und die Menge der Zeilennummern,

$$\vec{t} := \{(t_i - 1)q + 1, \dots, t_i q \mid i = 1, \dots, b\}, \quad |\vec{t}| = q |t| = qb,$$

d.h. $\vec{b} := \vec{t} \times \vec{s}$ ist die Menge aller Indices von Matrixeinträgen, die Informationen zum Cluster $\beta := \tau \times \sigma$ bergen.

Dann ist

$$\left[(\vec{L}^{-1})_{\mathcal{H}} \right]_{\vec{b}} := \left[\vec{J}^* \vec{Q}^m \vec{L}^{-1} \vec{J} \right]_{\vec{b}}$$

eine \mathcal{H} -Approximation von $(\vec{L}^{-1})_h$ auf dem Matrixteilblock \vec{b} mit dem Fehler

$$\left\| \left[(\vec{L}^{-1})_h - (\vec{L}^{-1})_{\mathcal{H}} \right]_{\vec{b}} \right\|_2 \leq c_{(5.29)} \mu_{\max} c_m c_{m,n} (\vec{c}_0 \eta)^m$$

und dem Maximalrang

$$\text{rnk} \left(\left[(\vec{L}^{-1})_{\mathcal{H}} \right]_{\vec{b}} \right) \leq q \frac{(m+n)!}{m!n!}.$$

Beweis:

(a) Wir eröffnen mit der Rangabschätzung:

$$\begin{aligned}
 \dim \left(\text{im} \left(\vec{Q}^m \right) \right) &= \dim \left(\text{im} \left(Q^m \right) \times \dots \times \text{im} \left(Q^m \right) \right) \\
 &= \sum_{i=1}^q \dim \left(\text{im} \left(Q^m \right) \right) \\
 &\stackrel{\text{Kor. 5.7}}{\leq} \sum_{i=1}^q \frac{(m+n)!}{m!n!} \\
 &= q \frac{(m+n)!}{m!n!},
 \end{aligned}$$

das bedeutet $\text{rnk} \left(\left[\left(\vec{L}^{-1} \right)_{\mathcal{H}} \right]_{\vec{b}} \right) \leq q \frac{(m+n)!}{m!n!}$.

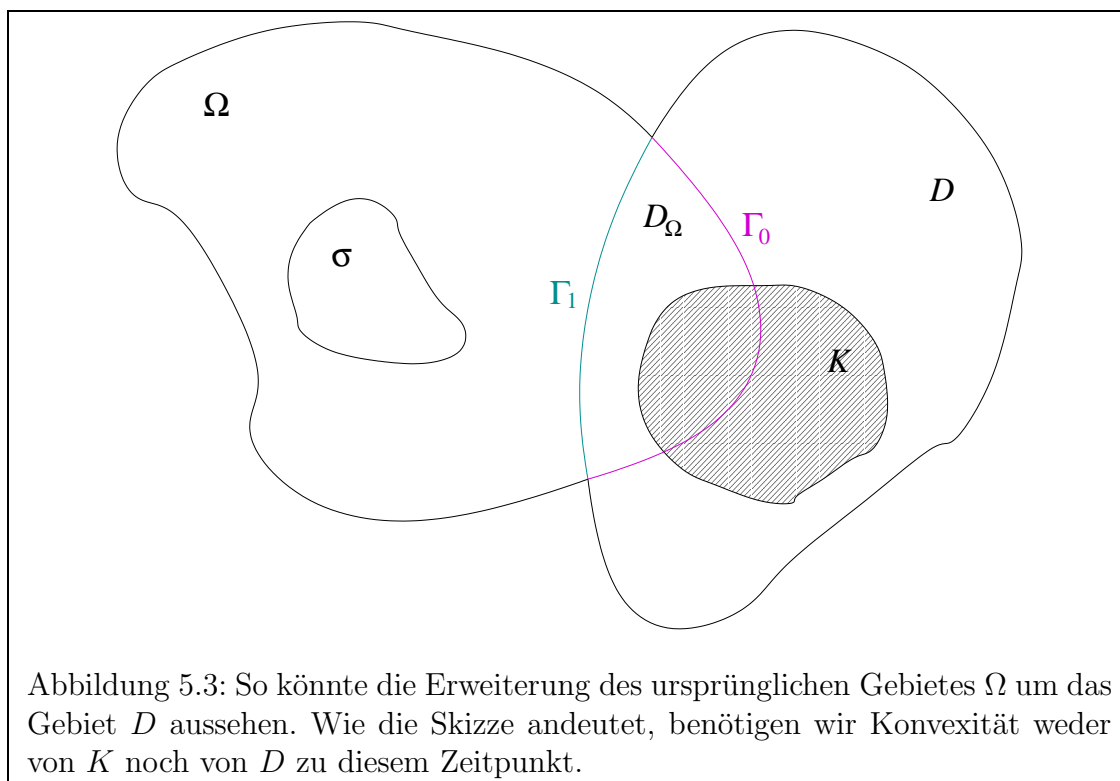
(b) Nach Lemma 3.7 ist $\left\| \vec{J} \vec{\mathfrak{w}} \right\|_{L^2(\Omega, \mathbb{R}^q)} \leq \sqrt{\mu_{max}} \left\| \vec{\mathfrak{w}} \right\|_h$ für alle $\vec{\mathfrak{w}} \in \mathbb{R}^{qN}$ mit μ_{max} als größtem Eigenwert der Massenmatrix M . Seien $\vec{\mathfrak{s}} \in \mathbb{R}^{qa}$ und $\vec{\mathfrak{t}} \in \mathbb{R}^{qb}$ zwei Koeffizientenvektoren und $\Omega_{\vec{t}} = \Omega_t := \bigcup_{i \in t} \text{supp}(\varphi_i)$. Dann ist

$$\begin{aligned}
 &\left| \left\langle \left[\left(\vec{L}^{-1} \right)_h - \left(\vec{L}^{-1} \right)_{\mathcal{H}} \right]_{\vec{b}} \vec{\mathfrak{s}}, \vec{\mathfrak{t}} \right\rangle_h \right| \\
 = &\left| \left\langle \left[\vec{J}^* \left(\vec{L}^{-1} - \vec{Q}^m \vec{L}^{-1} \right) \vec{J} \right]_{\vec{b}} \vec{\mathfrak{s}}, \vec{\mathfrak{t}} \right\rangle_h \right| \\
 = &\left| \left(\left\langle \left(\vec{L}^{-1} - \vec{Q}^m \vec{L}^{-1} \right) \vec{J}_{\vec{s}} \vec{\mathfrak{s}}, \vec{J}_{\vec{t}} \vec{\mathfrak{t}} \right\rangle \right)_{L^2(\Omega_t, \mathbb{R}^q)} \right| \\
 \leq &\left\| \vec{L}^{-1} - \vec{Q}^m \vec{L}^{-1} \right\|_{\mathcal{L}(L^2(\sigma, \mathbb{R}^q), L^2(\tau, \mathbb{R}^q))} \cdot \left\| \vec{J}_{\vec{s}} \vec{\mathfrak{s}} \right\|_{L^2(\Omega_s, \mathbb{R}^q)} \cdot \left\| \vec{J}_{\vec{t}} \vec{\mathfrak{t}} \right\|_{L^2(\Omega_t, \mathbb{R}^q)} \\
 \stackrel{\text{Kor. 5.12}}{\leq} &c_{(5.29)} c_m c_{m,n} (\vec{c}_0 \eta)^m \cdot \sqrt{\mu_{max}} \left\| \vec{\mathfrak{s}} \right\|_h \cdot \sqrt{\mu_{max}} \left\| \vec{\mathfrak{t}} \right\|_h
 \end{aligned}$$

(c) Mit dem vorigen Schritt folgt kurzerhand

$$\begin{aligned}
 &\left\| \left[\left(\vec{L}^{-1} \right)_h - \left(\vec{L}^{-1} \right)_{\mathcal{H}} \right]_{\vec{b}} \right\|_2 \\
 := &\sup \left\{ \frac{\left| \left\langle \left[\left(\vec{L}^{-1} \right)_h - \left(\vec{L}^{-1} \right)_{\mathcal{H}} \right]_{\vec{b}} \vec{\mathfrak{s}}, \vec{\mathfrak{t}} \right\rangle_h \right|}{\left\| \vec{\mathfrak{s}} \right\|_h \cdot \left\| \vec{\mathfrak{t}} \right\|_h} \mid \vec{\mathfrak{s}} \in \mathbb{R}^{qa}, \vec{\mathfrak{t}} \in \mathbb{R}^{qb} \right\} \\
 \leq &c_{(5.29)} c_m c_{m,n} (\vec{c}_0 \eta)^m \mu_{max}.
 \end{aligned}$$

QED



5.3 Lokale Approximationseigenschaft für Matrixoperatoren

Teil II: L^∞ -Koeffizienten

In diesem Abschnitt zeigen wir, daß wir auch elliptische Matrixdifferentialoperatoren zweiter Ordnung mit L^∞ -Koeffizienten approximieren können. Allerdings müssen wir für den Beweis dieser Aussage die Idee von BEBENDORF und HACKBUSCH adaptieren, daß die Differenz einer \vec{L} -harmonischen Funktion und die geeignete Projektion einer stückweise konstanten Funktion ebenfalls \vec{L} -harmonisch ist.

5.3.1 Ein weiteres Regularitätsresultat

Unser letzter Beweisansatz scheitert zunächst daran, daß wir den Gebietsrand erreichen und die Lösung der Differentialgleichung außerhalb des Gebietes nicht so fortzusetzen wissen, daß sie mitsamt ihren m -ten schwachen Ableitungen \vec{L} -harmonisch ist. Setzen wir sie durch Null fort, so ist sie am Rand im allgemeinen zwar geknickt, aber insgesamt immer noch schwach differenzierbar, was uns in diesem Fall genügt.

Wir beginnen mit der Erweiterung des Definitionsbereiches $\Omega \subset \mathbb{R}^n$ der Differen-

tialgleichung. Seien $D, K \subset \mathbb{R}^n$ zwei Gebiete mit $K \subset D$ und

$$\begin{aligned} D_\Omega &:= D \cap \Omega \neq \{\}, \\ \Gamma_0 &:= \Gamma_0(D) := D \cap \partial\Omega, \\ \Gamma_1 &:= \Gamma_1(D) := \partial D \cap \overline{\Omega}, \end{aligned}$$

vgl. Abbildung 5.3. Wir stellen fest, daß

$$\begin{aligned} \Gamma_0(D) = \{\} &\iff D \subset \Omega, \\ \Gamma_1(D) = \{\} &\iff \Omega \subset D \iff \Gamma_0(D) = \partial D. \end{aligned}$$

Im Falle $D \subset \Omega$ folgt $\Gamma_1(D) = \partial D$ und im Falle $\Omega \subset D$ ist $\Gamma_1(D) = \{\}$. Da wir später $\text{dist}(K, \Gamma_1(D)) > 0$ bzw. $\text{dist}(K, \sigma) > 0$ benötigen, muß $\Gamma_1(D) \neq \{\}$ sein, also schließen wir den Fall $\Omega \subset D$ aus.

Im Falle $D \setminus \Omega \neq \{\}$ bedarf die Lösung \vec{u} der Differentialgleichung einer Fortsetzung außerhalb von Ω . Diese sei die Nullfunktion,

$$\vec{u}^{(ext)}(x) := \begin{cases} \vec{u}(x) & \text{für } x \in \overline{\Omega}, \\ 0 & \text{für } x \in D \setminus \overline{\Omega}. \end{cases}$$

Damit ist die DIRICHLET-Nullrandbedingung auf $\Gamma_0(D)$ unversehrt und $\vec{u}^{(ext)} \in H_0^1(\Omega \cup D, \mathbb{R}^q)$. Die Differentialgleichung wird auf $D \setminus \Omega$ im allgemeinen nicht erfüllt sein; wollten wir das, müßten wir uns auch Gedanken über die Koeffizienten c_{ijkl} machen, die unter Umständen sogar von \vec{u} abhängen.

Lemma 5.14 (CACCIOPPOLI-Ungleichung für vektorwertige Funktionen)

Seien Ω, K, D und $\Gamma_1(D)$ wie gerade eben definiert sowie $\text{dist}(K, \partial D) > 0$. Sei \vec{L} der in (2.18) erklärte Matrixdifferentialoperator. Sei $\vec{u}^{(ext)}$ im schwachen Sinne \vec{L} -harmonisch auf D_Ω , d.h.

$$\int_{D_\Omega} \langle \vec{C} \nabla \vec{u}^{(ext)}, \nabla \vec{\varphi} \rangle = 0 \quad \forall \vec{\varphi} \in H_0^1(D_\Omega, \mathbb{R}^q).$$

Dann gilt

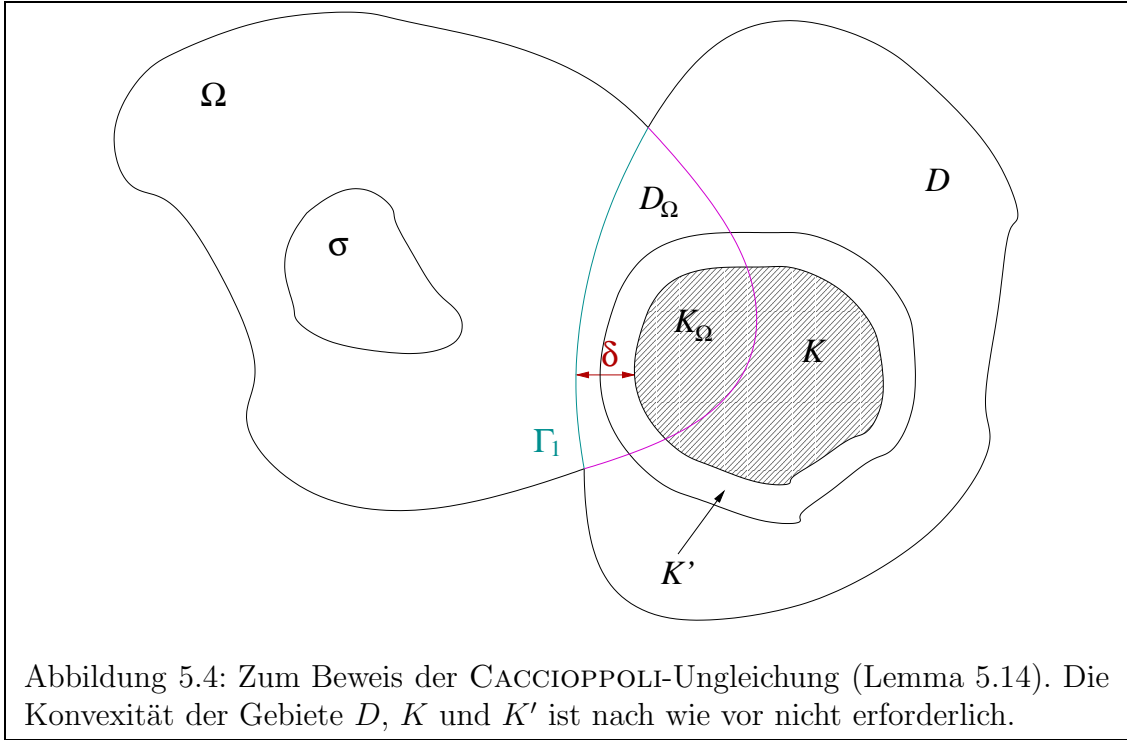
$$\|\vec{u}^{(ext)}\|_{H^1(K, \mathbb{R}^q)} \leq \frac{2}{\text{dist}(K, \partial D)} \sqrt{\kappa_{\vec{C}}} \cdot \|\vec{u}^{(ext)}\|_{L^2(D, \mathbb{R}^q)}$$

für alle $\vec{u}^{(ext)} \in H^1(D_\Omega, \mathbb{R}^q) \cap \left\{ \vec{v} \in L^2(D, \mathbb{R}^q) \mid \|\vec{v}\|_{L^2(D \setminus \Omega, \mathbb{R}^q)} = 0 \right\}$.

Bemerkung 5.15 Daß wir um einen Faktor 2 besser abschätzen als [BH03], liegt in einer anderen Beweistechnik begründet.

Beweis: Sei $\eta : D \rightarrow [0, 1]$ eine stetig differenzierbare Abschneidefunktion mit

$$\eta|_K \equiv 1 \text{ und } \eta = 0 \text{ nahe } \Gamma_1(D).$$



Setzen wir $\delta := \text{dist}(K, \Gamma_1(D))$, so ist nach Voraussetzung $\delta > 0$ (falls der Abstand von K zu ∂D kleiner als δ ist, so stellt das kein Ärgernis dar, weil wir D beliebig vergrößern dürfen, solange nur $\Omega \not\subset D$ gewahrt ist). Wir setzen $K_\Omega := \Omega \cap K$ und $D_\Omega := \Omega \cap D$, vgl. Abbildung 5.4. Angenommen, η falle fast überall (f.ü.) auf $D_\Omega \setminus K_\Omega$ linear ab, so wäre $|\nabla \eta|_{D_\Omega \setminus K_\Omega} \approx \frac{1-\eta}{\delta}$. Da η aber etwas schneller abfallen soll, um $\eta = 0$ nahe $\Gamma_1(D)$ bzw. ∂D zu garantieren, verlangen wir

$$|\nabla \eta|_{D_\Omega \setminus K_\Omega} \leq \frac{2}{\delta} \equiv \text{const.}$$

Wir setzen nun

$$K' := \text{supp}(\eta)$$

(vgl. Abbildung 5.4). Da Träger per definitionem abgeschlossen sind und D bzw. Ω sinnigerweise beschränkt sind, ist K' kompakt, also

$$\delta := \text{dist}(K, \partial D) > \frac{\delta}{2} \geq \text{dist}(K', \partial D) > 0.$$

Nach Voraussetzung ist $\vec{u}^{(ext)} \in H_0^1(\Omega \cup D, \mathbb{R}^q)$, also insbesondere $\vec{u}^{(ext)} \in H^1(K', \mathbb{R}^q)$. Wir definieren die Testfunktion

$$\vec{\varphi} : D \longrightarrow \mathbb{R}^q, \quad x \longmapsto \eta^2(x) \cdot \vec{u}^{(ext)}(x).$$

Wegen $\text{supp}(\eta) = K'$ und $\text{supp}(\vec{u}^{(ext)}) \subset \Omega$ ist

$$\text{supp}(\vec{\varphi}) = \text{supp}(\eta) \cap \text{supp}(\vec{u}^{(ext)}) \subset D_\Omega.$$

Mit $\eta \in C_0^1(D)$ und $\vec{u}^{(ext)} \in H^1(K', \mathbb{R}^q) = H^1(\text{supp}(\eta), \mathbb{R}^q)$ folgt

$$\vec{\varphi} \in H_0^1(D_\Omega, \mathbb{R}^q).$$

Mit den Erfahrungen aus Beweisschritt 1 des Satzes 5.11 ist

$$\begin{aligned}
0 &\leq \vec{\lambda}_{min} \cdot \|\nabla \vec{u}^{(ext)}\|_{L^2(K_\Omega, \mathbb{R}^q)}^2 \\
&\leq \vec{\lambda}_{min} \cdot \|\nabla(\eta \vec{u}^{(ext)})\|_{L^2(D_\Omega, \mathbb{R}^q)}^2 \\
&= \vec{\lambda}_{min} \cdot \int_{D_\Omega} \langle \nabla(\eta \vec{u}^{(ext)}), \nabla(\eta \vec{u}^{(ext)}) \rangle \\
&\leq \int_{D_\Omega} \langle \vec{C} \nabla(\eta \vec{u}^{(ext)}), \nabla(\eta \vec{u}^{(ext)}) \rangle \\
&= \int_{D_\Omega} \sum_{i,j=1}^q \sum_{k,l=1}^n c_{ijkl} \cdot \partial_k(\eta u_i^{(ext)}) \cdot \partial_l(\eta u_j^{(ext)}) \\
&= \sum_{i,j} \sum_k \int_{D_\Omega} \partial_k(\eta u_i^{(ext)}) \cdot \sum_l c_{ijkl} \cdot \partial_l(\eta u_j^{(ext)}) \\
&\stackrel{\text{part. Int.}}{=} \sum_{i,j} \sum_k \left[\int_{\partial D_\Omega} \eta u_i^{(ext)} \cdot \sum_l c_{ijkl} \partial_l(\eta u_j^{(ext)}) \right] \\
&\quad - \sum_{i,j} \sum_k \left[\int_{D_\Omega} \eta u_i^{(ext)} \partial_k \left(\sum_l c_{ijkl} \partial_l(\eta u_j^{(ext)}) \right) \right] \\
&\stackrel{\eta|_{\partial D_\Omega}=0}{=} - \sum_{i,j} \int_{D_\Omega} \eta u_i^{(ext)} \cdot \sum_k \partial_k \left(\sum_l c_{ijkl} \partial_l(\eta u_j^{(ext)}) \right) \\
&\stackrel{(5.21)}{=} \sum_{i,j} \int_{D_\Omega} \sum_{k,l} \partial_k \eta \cdot c_{ijkl} \cdot \partial_l \eta \cdot u_i^{(ext)} \cdot u_j^{(ext)} \\
&\leq \vec{\lambda}_{max} \int_{D_\Omega} \sum_i \sum_k \left(\partial_k \eta \cdot u_i^{(ext)} \right)^2 \\
&= \vec{\lambda}_{max} \int_{K_\Omega} \underbrace{|\nabla \eta|^2}_{=0} \cdot |\vec{u}^{(ext)}|^2 + \vec{\lambda}_{max} \int_{D_\Omega \setminus K_\Omega} \underbrace{|\nabla \eta|^2}_{\leq 4\delta^{-2}} \cdot |\vec{u}^{(ext)}|^2 \\
&\leq \frac{4\vec{\lambda}_{max}}{\delta^2} \cdot \int_{D_\Omega \setminus K_\Omega} |\vec{u}^{(ext)}|^2 \\
&\leq \frac{4\vec{\lambda}_{max}}{\delta^2} \cdot \|\vec{u}^{(ext)}\|_{L^2(D_\Omega, \mathbb{R}^q)}^2.
\end{aligned}$$

Nach Division mit $\vec{\lambda}_{min}$ folgt

$$|\vec{u}^{(ext)}|_{H^1(K_\Omega, \mathbb{R}^q)} \leq \frac{2}{\delta} \sqrt{\kappa \vec{C}} \cdot \|\vec{u}^{(ext)}\|_{L^2(D_\Omega, \mathbb{R}^q)}$$

bzw.

$$|\vec{u}^{(ext)}|_{H^1(K, \mathbb{R}^q)} \leq \frac{2}{\delta} \sqrt{\kappa \vec{C}} \cdot \|\vec{u}^{(ext)}\|_{L^2(D, \mathbb{R}^q)},$$

weil $\vec{u}|_{D \setminus \Omega} = 0$ und $\nabla \vec{u}|_{D \setminus \Omega} = 0$.

QED

5.3.2 Funktionalanalytische Betrachtungen

Endliche Approximation von $\vec{u}^{(ext)}$ durch L^2 -Funktionen

Mit Hilfe einer POINCARÉ-Ungleichung beweisen wir zunächst die Aussage, daß der Fehler, der bei der Approximation von \vec{u} durch L^2 -Funktionen entsteht, durch die ersten Ableitungen von \vec{u} beschränkt ist.

Lemma 5.16 *Sei $D \subset \mathbb{R}^n$, $n \in \mathbb{N}$, ein konvexes³ Gebiet. Sei $X = X(D, \mathbb{R}^q)$ ein abgeschlossener⁴ Teilraum von $L^2(D, \mathbb{R}^q)$.*

Dann existiert für alle $k \in \mathbb{N}$ ein abgeschlossener Teilraum

$$V := V(D, \mathbb{R}^q, k) \subset X(D, \mathbb{R}^q)$$

mit

$$\dim(V) \leq qk, \quad (5.32)$$

$$\inf_{\vec{v} \in V} \|\vec{u} - \vec{v}\|_{L^2(D, \mathbb{R}^q)} \leq \frac{2\sqrt{n}}{\pi} \cdot \frac{\varnothing_D}{\sqrt{k}} \cdot \|\nabla \vec{u}\|_{L^2(D, \mathbb{R}^q)} \quad (5.33)$$

für alle $\vec{u} \in X \cap H^1(D, \mathbb{R}^q)$. Auf diese Weise wird eine orthogonale Projektion

$$\vec{S} = \vec{S}_{X(D, \mathbb{R}^q)} : L^2(D, \mathbb{R}^q) \longrightarrow X(D, \mathbb{R}^q)$$

definiert und es gilt

$$\inf_{\vec{v} \in V} \|\vec{u} - \vec{v}\| = \|\vec{u} - \vec{S}\vec{u}\|.$$

Beweis:

- (a) Wir packen D in einen Würfel⁵ Q , dessen Kantenlänge gleich dem Durchmesser von D ist, kurz \varnothing_D ; als **Durchmesser eines Gebietes** bezeichnen wir den Durchmesser des kleinsten umgebenden Balles. Für $k = l^n$, $l \in \mathbb{N}$ (anderer Fall später), unterteilen wir Q in k Teilwürfel Q_i der Kantenlänge $\varnothing_D \cdot l^{-1}$, vgl. Abbildung 5.5. Setzen wir $D_i := D \cap Q_i$, folgt, daß

- $D = \bigcup_{i=1}^k D_i$,
- jedes D_i konvex ist, weil D und Q_i konvex sind,

³Die Konvexität wird für die POINCARÉ-Ungleichung benötigt.

⁴Wir wollen auf X projizieren und deshalb muß X abgeschlossen sein.

⁵In der Tat müssen wir anstelle der Würfel und Teilwürfel die aus der Maßtheorie bekannten **Figuren** verwenden, damit wir im weiteren Beweisverlauf keine Überlagerung von Randflächen haben.

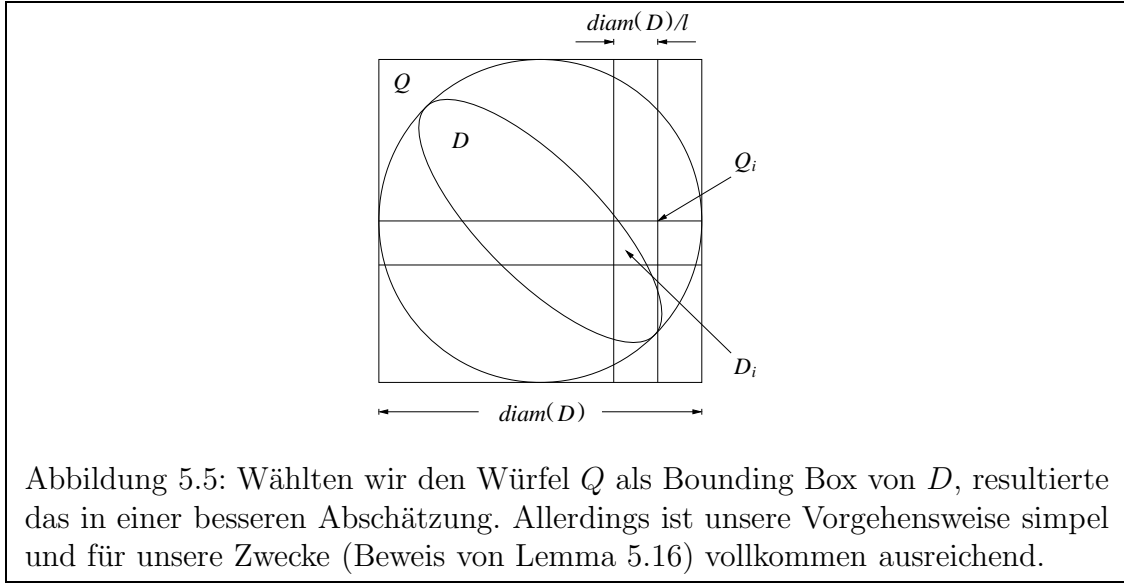


Abbildung 5.5: Wählen wir den Würfel Q als Bounding Box von D , resultierte das in einer besseren Abschätzung. Allerdings ist unsere Vorgehensweise simpel und für unsere Zwecke (Beweis von Lemma 5.16) vollkommen ausreichend.

- sowie

$$\varnothing_{D_i} \leq \varnothing_{Q_i} = \sqrt{\dim(Q_i) \cdot (\text{Kantenlänge})^2} = \sqrt{n \cdot \left(\frac{\varnothing_D}{l}\right)^2}. \quad (5.34)$$

Wir setzen

$$W(k) := \left\{ \vec{v} \in L^2(D, \mathbb{R}^q) \mid v_1, \dots, v_q \text{ sind konstant auf } D_1, \dots, D_k \right\},$$

d.h. die Menge der auf D stückweise konstanten Funktionen. Somit ist $W(k)$ isomorph zum \mathbb{R}^{qk} , also abgeschlossen, und $\dim(W(k)) = qk$. Für die Abschätzung von \vec{u} gegen $\nabla \vec{u}$ definieren wir

$$\begin{aligned} \vec{u} : D &\longrightarrow \mathbb{R}^q, & x &\longmapsto \sum_{i=1}^k \vec{u}_i \cdot \text{Ind}_{D_i}(x), \\ \vec{u}_i &:= \begin{pmatrix} \bar{u}_{1,i} \\ \vdots \\ \bar{u}_{q,i} \end{pmatrix} = \text{const} \in \mathbb{R}^q, & i &= 1, \dots, k, \\ \bar{u}_{j,i} &:= \frac{1}{\text{vol}(D_i)} \cdot \int_{D_i} u_j(x) dx = \text{const} \in \mathbb{R}, & j &= 1, \dots, q. \end{aligned}$$

Damit ist $\vec{u} \in W(k)$ und $\|\vec{u} - \vec{u}\|$ zugleich der Fehler, den wir bei der Approximation von \vec{u} durch eine stückweise konstante Funktion \vec{u} kassieren. Mit der auf vektorwertige Funktionen verallgemeinerten POINCARÉ-Ungleichung nach

PAYNE und WEINBERGER, siehe Formel (A.8), folgt sofort

$$\begin{aligned}
\|\vec{u} - \bar{\vec{u}}\|_{L^2(D, \mathbb{R}^q)}^2 &= \sum_{i=1}^k \|\vec{u} - \bar{\vec{u}}\|_{L^2(D_i, \mathbb{R}^q)}^2 \\
&= \sum_{i=1}^k \sum_{j=1}^q \|u_j - \bar{u}_j\|_{L^2(D_i)}^2 \\
&\leq \sum_{i=1}^k \sum_{j=1}^q \left(\frac{\varnothing_{D_i}}{\pi} \right)^2 \|\nabla(u_j)\|_{L^2(D_i)}^2 \\
&\stackrel{(5.34)}{\leq} \sum_{i=1}^k \sum_{j=1}^q \left(\frac{\sqrt{n} \cdot \varnothing_D}{\pi \cdot l} \right)^2 \|\nabla(u_j)\|_{L^2(D_i)}^2 \\
&= \left(\frac{\sqrt{n} \cdot \varnothing_D}{\pi \cdot l} \right)^2 \cdot \sum_{i=1}^k \|\nabla \vec{u}\|_{L^2(D_i)}^2,
\end{aligned}$$

also ist wegen $\bigcup_{i=1}^k D_i \subsetneq \bar{D} = \bigcup_{i=1}^k \bar{D}_i$

$$\|\vec{u} - \bar{\vec{u}}\|_{L^2(D, \mathbb{R}^q)} \leq \frac{\sqrt{n} \cdot \varnothing_D}{\pi \cdot l} \cdot \|\nabla \vec{u}\|_{L^2(D, \mathbb{R}^q)}$$

und erst recht

$$\inf_{\vec{w} \in W(k)} \|\vec{u} - \vec{w}\| \leq \frac{\sqrt{n} \cdot \varnothing_D}{\pi \cdot l} \cdot \|\nabla \vec{u}\|_{L^2(D, \mathbb{R}^q)}.$$

- (b) Jetzt müssen wir den Beweis aus dem ersten Schritt für k 's anpassen, die nicht die n -te Potenz einer natürlichen Zahl l sind. Sei l diejenige natürliche Zahl, die wir durch Abrunden von $\sqrt[n]{k}$ auf die nächstgrößere natürliche Zahl erhalten, d.h.

$$l < \sqrt[n]{k} < l + 1.$$

Zum einen ist k die Anzahl der Teilwürfel Q_i und deshalb direkt verbunden mit der Approximationsgenauigkeit von u über $W(k)$, zum anderen wird k später den Blockrang bei der \mathcal{H} -Approximation verkörpern. Weil wir den Approximationsfehler aber nicht verkleinern, sondern nach oben abschätzen wollen, teilen wir Q nicht in $k^+ := (l+1)^n$, sondern nur in $k^- := l^n$ Unterwürfel auf. Entsprechend zum ersten Beweisschritt haben wir

$$\varnothing_{D_i} \leq \frac{\sqrt{n} \cdot \varnothing_D}{l},$$

$W(k^-)$ anstelle von $W(k)$ sowie

$$\dim(W(k^-)) = qk^- \leq qk.$$

Unsere Funktion \vec{u} würde also, wenn's denn $W(k)$ gäbe, über $W(k)$ genauer approximiert als über $W(k^-)$, d.h.

$$\begin{aligned} \inf_{\vec{w} \in W(k)} \|\vec{u} - \vec{w}\| &< \inf_{\vec{w} \in W(k^-)} \|\vec{u} - \vec{w}\| \\ &\stackrel{\text{Schritt (a)}}{\leq} \frac{\sqrt{n} \cdot \varnothing_D}{\pi \cdot l} \cdot \|\nabla \vec{u}\|_{L^2(D, \mathbb{R}^q)} \\ &< 2 \cdot \frac{\sqrt{n} \cdot \varnothing_D}{\pi \cdot \sqrt[n]{k}} \cdot \|\nabla \vec{u}\|_{L^2(D, \mathbb{R}^q)}, \end{aligned}$$

weil

$$\frac{1}{l} \leq \frac{2}{l+1} < \frac{2}{\sqrt[n]{k}}.$$

- (c) Bisher erfolgte die Approximation von \vec{u} durch Funktionen $W(k)$, d.h. die Approximierenden verfügen nicht über die Eigenschaften, die die Elemente in $X(D, \mathbb{R}^q)$ auszeichnen und die wir wünschen. Deshalb müssen wir $W(k)$ auf $X(D, \mathbb{R}^q)$ projizieren.

Sei \vec{S} die orthogonale Projektion von $L^2(D, \mathbb{R}^n)$ auf den abgeschlossenen Teilraum $X(D, \mathbb{R}^n)$, d.h.

$$\vec{S}^2 = \vec{S}, \quad \vec{S}^* = \vec{S}, \quad (v - \vec{S}v, \vec{S}v)_{L^2(D, \mathbb{R}^n)} = 0 \quad \forall v \in L^2(D, \mathbb{R}^q).$$

Da $W(k) \subset L^2(D, \mathbb{R}^q)$ und abgeschlossen ist, dürfen wir

$$V := \vec{S}(W(k))$$

setzen. Insbesondere ist $V \subset X(D, \mathbb{R}^q)$ abgeschlossen und $\dim(V) \leq qk$. Mit $\vec{u} \in X(D, \mathbb{R}^q)$ gilt $\vec{u} = \vec{S}\vec{u}$ und

$$\left\| \vec{u} - \vec{S}\vec{u} \right\|_{L^2(D, \mathbb{R}^q)} = \left\| \vec{S}(\vec{u} - \vec{u}) \right\|_{L^2(D, \mathbb{R}^q)} \leq \left\| \vec{u} - \vec{u} \right\|_{L^2(D, \mathbb{R}^q)}.$$

Zugleich ist $\vec{S}\vec{u} \in V \subset X$, also gilt auch

$$\left\| \vec{u} - \vec{S}\vec{u} \right\|_{L^2(D, \mathbb{R}^q)} \geq \inf_{\vec{v} \in V} \|\vec{u} - \vec{v}\|$$

und insgesamt folgt

$$\left\| \vec{u} - \vec{S}\vec{u} \right\|_{L^2(D, \mathbb{R}^q)} = \text{dist}_{L^2(D, \mathbb{R}^q)}(\vec{u}, V) < \frac{\sqrt{n}}{\pi} \cdot \frac{\varnothing_D}{l} \cdot \|\nabla \vec{u}\|_{L^2(D, \mathbb{R}^q)}.$$

QED

Räume \vec{L} -harmonischer Funktionen

In Lemma 5.16 war $X(D, \mathbb{R}^q)$ ein beliebiger Teilraum von $L^2(D, \mathbb{R}^q)$. Für die Approximation von $\vec{u}^{(ext)}$ benötigen wir jedoch einen ganz bestimmten Teilraum, nämlich einen Raum, dessen Elemente auch \vec{L} -harmonisch auf D_Ω sind. Auf dem Wege dorthin definieren wir

$$\tilde{X}(D, \mathbb{R}^q) := \left\{ \vec{v} \in L^2(D, \mathbb{R}^q) \left| \begin{array}{l} \bullet \|\vec{v}\|_{L^2(D \setminus \Omega, \mathbb{R}^q)} = 0, \\ \bullet \forall K \subset D, \text{dist}(K, \Gamma_1(D)) > 0 : \\ \vec{v} \in H^1(K, \mathbb{R}^q). \end{array} \right. \right\}.$$

Unsere Funktion $\vec{u}^{(ext)}$ ist nach Konstruktion \vec{L} -harmonisch in D_Ω — sie liegt sogar in $H_0^1(\Omega \cup D, \mathbb{R}^q)$ —, und erfüllt die Variationsgleichung

$$\int_{D_\Omega} \langle \vec{C} \nabla \vec{u}^{(ext)}, \nabla \vec{\varphi} \rangle = 0 \quad \forall \vec{\varphi} \in C_0^\infty \subset H_0^1(D_\Omega, \mathbb{R}^q). \quad (5.35)$$

Wir wollen diese berücksichtigt haben und das führt uns zu

$$\begin{aligned} X(D, \mathbb{R}^q) &:= \left\{ \vec{v} \in \tilde{X}(D, \mathbb{R}^q) \mid \vec{v} \text{ erfüllt (5.35)} \right\} \\ &= \left\{ \vec{v} \in L^2(D, \mathbb{R}^q) \left| \begin{array}{l} \bullet \|\vec{v}\|_{L^2(D \setminus \Omega, \mathbb{R}^q)} = 0, \\ \bullet \forall K \subset D, \text{dist}(K, \Gamma_1(D)) > 0 : \\ \vec{v} \in H^1(K, \mathbb{R}^q), \\ \bullet \vec{v} \text{ erfüllt (5.35)}. \end{array} \right. \right\}. \end{aligned} \quad (5.36)$$

Wir formulieren nun ein Lemma, das uns die Existenz gewisser im Abschnitt 5.3.3 benötigter Projektionen garantiert.

Lemma 5.17 *Die Menge $X(D, \mathbb{R}^q)$ ist ein vollständiger Teilraum des $L^2(D, \mathbb{R}^q)$.*

Beweis: Die CAUCHY-Folge $(\vec{v}^{(i)})_{i \in \mathbb{N}} \subset X(D, \mathbb{R}^q)$ konvergiere gegen die Grenzfunktion $\vec{v} \in L^2(D, \mathbb{R}^q)$. Wir müssen zeigen, daß das $\vec{v} \in X(D, \mathbb{R}^q)$ bedingt.

(a) *Zu zeigen:* $\|\vec{v}\|_{L^2(D \setminus \Omega, \mathbb{R}^q)} = 0$.

Nach Voraussetzung ist $(\vec{v}^{(i)})_{i \in \mathbb{N}}$ eine CAUCHY-Folge, d.h. für alle $\varepsilon > 0$ existiert ein $i_0 \in \mathbb{N}$, sodaß

$$\|\vec{v}^{(i)} - \vec{v}\|_{L^2(D, \mathbb{R}^q)} < \varepsilon \quad \forall i > i_0.$$

Andererseits gilt

$$\begin{aligned} \|\vec{v}^{(i)} - \vec{v}\|_{L^2(D, \mathbb{R}^q)} &\geq \|\vec{v}^{(i)} - \vec{v}\|_{L^2(D \setminus \Omega, \mathbb{R}^q)} \\ &\geq \left| \|\vec{v}^{(i)}\|_{L^2(D \setminus \Omega, \mathbb{R}^q)} - \|\vec{v}\|_{L^2(D \setminus \Omega, \mathbb{R}^q)} \right| \\ &= \|\vec{v}\|_{L^2(D \setminus \Omega, \mathbb{R}^q)}, \end{aligned}$$

weil $\|\vec{v}^{(i)}\|_{L^2(D \setminus \Omega, \mathbb{R}^q)} = 0$. Das zeigt $\vec{v} = 0$ im L^2 -Sinne.

- (b) *Zu zeigen: Für alle $K \subset D$ mit $\text{dist}(K, \Gamma_1(D)) > 0$ ist $\vec{v} \in H^1(K, \mathbb{R}^q)$.*
 Nach Konstruktion von $X(D, \mathbb{R}^q)$ liegt jedes Folgenglied $\vec{v}^{(i)}$ in $H^1(K, \mathbb{R}^q)$. Nun ist zum einen dieser Raum bekanntlich ein HILBERT-Raum, also insbesondere vollständig, zum anderen ist $(\vec{v}^{(i)})_{i \in \mathbb{N}}$ auch in $H^1(K, \mathbb{R}^q)$ eine CAUCHY-Folge, denn für alle $\varepsilon > 0$ existiert ein $i_0 \in \mathbb{N}$, sodaß

$$\|\vec{v}^{(k)} - \vec{v}^{(l)}\|_{L^2(D, \mathbb{R}^q)} < \varepsilon \quad \forall k, l > i_0.$$

Wegen der CACCIOPPOLI-Ungleichung (Lemma 5.14) existiert eine Konstante $c_1 > 0$ derart, daß

$$\|\vec{v}^{(k)} - \vec{v}^{(l)}\|_{H^1(K, \mathbb{R}^q)} \leq c_1 \cdot \|\vec{v}^{(k)} - \vec{v}^{(l)}\|_{L^2(D, \mathbb{R}^q)}$$

bzw.

$$\|\vec{v}^{(k)} - \vec{v}^{(l)}\|_{H^1(K, \mathbb{R}^q)} \leq (1 + c_1) \cdot \|\vec{v}^{(k)} - \vec{v}^{(l)}\|_{L^2(D, \mathbb{R}^q)}.$$

Folglich existiert für alle $\tilde{\varepsilon} = (1 + c_1)\varepsilon > 0$ ein $i_0 \in \mathbb{N}$, sodaß

$$\|\vec{v}^{(k)} - \vec{v}^{(l)}\|_{H^1(K, \mathbb{R}^q)} < \tilde{\varepsilon} \quad \forall k, l > i_0.$$

Damit haben wir $\vec{v} \in H^1(K, \mathbb{R}^q)$ gezeigt.

- (c) *Zu zeigen: Für alle $\vec{\varphi} \in C_0^\infty(D_\Omega, \mathbb{R}^q)$ ist $\int_{D_\Omega} \langle \vec{C} \nabla \vec{v}, \nabla \vec{\varphi} \rangle = 0$.*

Zu Beginn des Beweises hatten wir $\vec{v} \in H^1(K, \mathbb{R}^q)$ für alle $K \subset D$ mit $\text{dist}(K, \Gamma_1(D)) > 0$ festgestellt. Das impliziert die schwache Konvergenz

$$\int_{K_\Omega} \langle \vec{C} \nabla \vec{v}^{(i)}, \nabla \vec{\varphi} \rangle \xrightarrow{i \rightarrow \infty} \int_{K_\Omega} \langle \vec{C} \nabla \vec{v}, \nabla \vec{\varphi} \rangle \quad \forall \vec{\varphi} \in C_0^\infty(D_\Omega, \mathbb{R}^q)$$

(auf $K \setminus \Omega$ ist \vec{C} nicht definiert und $\vec{v}^{(i)}$ sowie \vec{v} gleich null). Nach Voraussetzung erfüllt jedes $\vec{v}^{(i)}$ die Variationsgleichung (5.35) und mit $K = \text{supp}(\vec{\varphi})$ folgt

$$0 = \int_{D_\Omega} \langle \vec{C} \nabla \vec{v}^{(i)}, \nabla \vec{\varphi} \rangle = \int_{K_\Omega} \langle \vec{C} \nabla \vec{v}^{(i)}, \nabla \vec{\varphi} \rangle$$

und weiter

$$\int_{D_\Omega} \langle \vec{C} \nabla \vec{v}, \nabla \vec{\varphi} \rangle = 0,$$

was den Beweis schließt. **QED**

5.3.3 \vec{L} -harmonische Approximation von $\vec{u}^{(ext)}$

Wir kommen nun zur Approximation von \vec{u} durch die auf die Menge der \vec{L} -harmonischen Funktionen projizierten stückweise konstanten Funktionen. Somit ist auch die Differenz von \vec{u} und diesen speziellen Funktionen \vec{L} -harmonisch (obschon zu anderen Randbedingungen). Diese Eigenschaft nutzen wir für eine rekursive Beweisführung aus (siehe auch Abb. 5.8 für die Beweisidee).

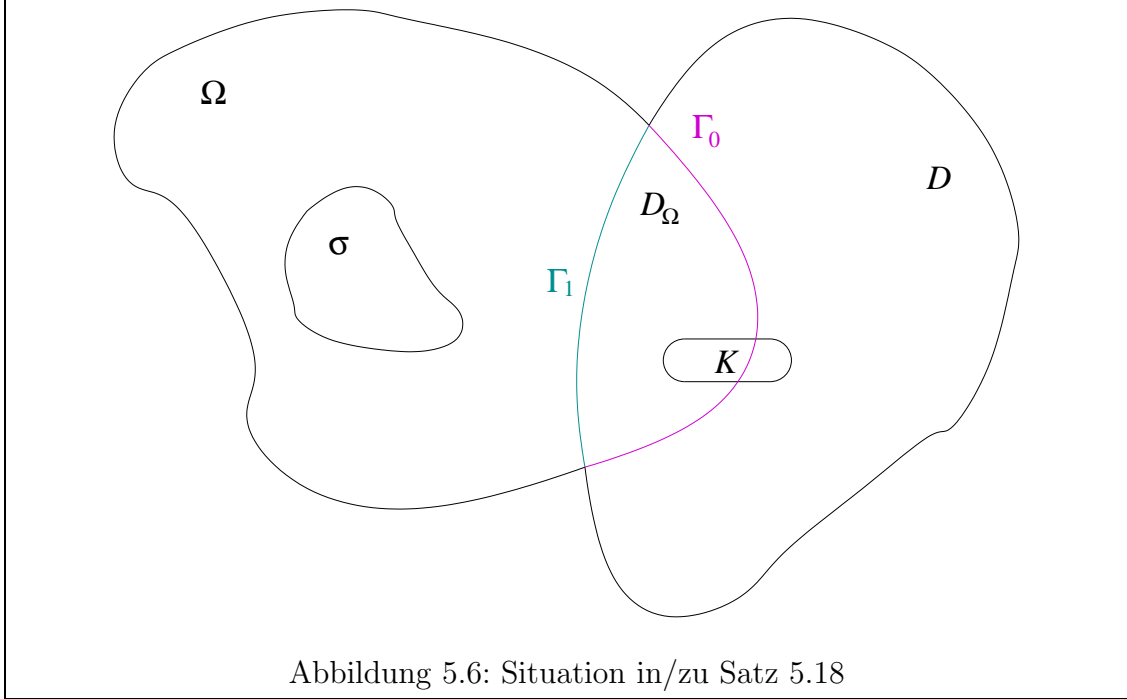


Abbildung 5.6: Situation in/zu Satz 5.18

Satz 5.18 Sei $\Omega \subset \mathbb{R}^n$ ein Gebiet und \vec{L} der Differentialoperator aus Absatz 2.4. Seien $K, D \subset \mathbb{R}^n$ zwei Gebiete mit $K \subset D$ konvex (vgl. Abbildung 5.6) und

$$\text{dist}(K, \partial D) \geq \frac{\varnothing_K}{2\eta} > 0. \quad (5.37)$$

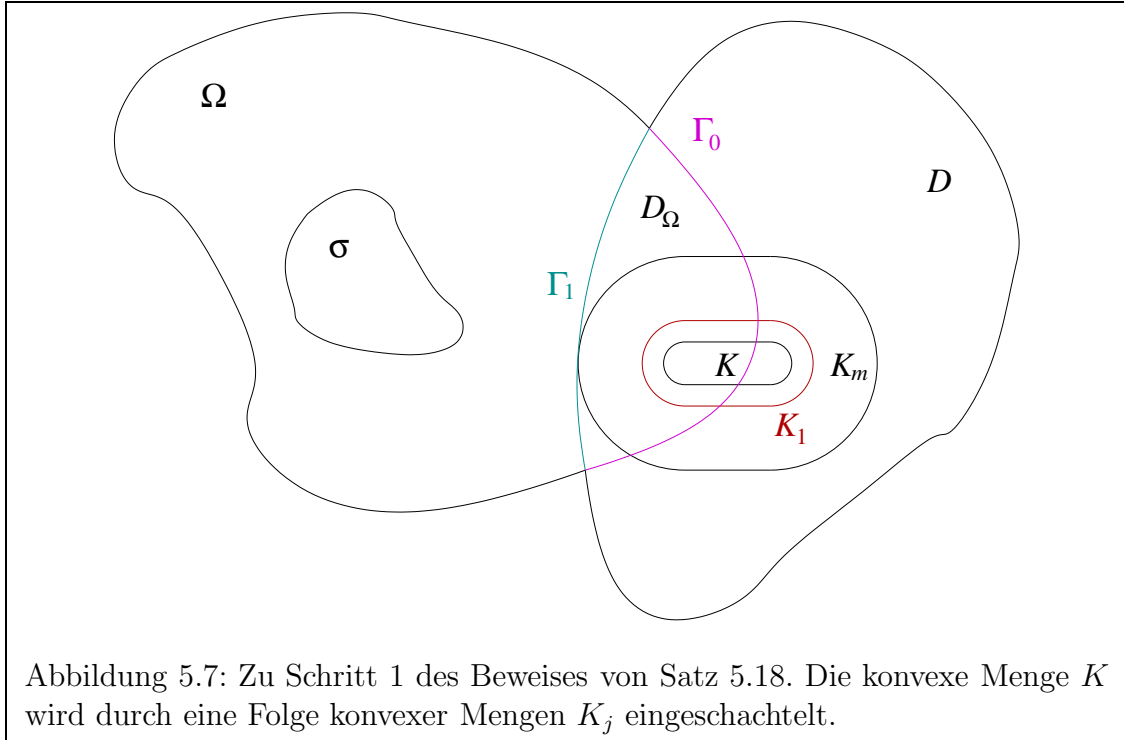
Dann existiert ein Teilraum $V \subset X(D, \mathbb{R}^q)$ und eine orthogonale Projektion $\vec{S}^m : L^2 \rightarrow X$ mit

$$\begin{aligned} \dim(V) &\leq m q k, \\ \left\| \vec{u}^{(ext)} - \vec{S}^m \vec{u}^{(ext)} \right\|_{L^2(K, \mathbb{R}^q)} &\leq \left(\frac{8}{\pi^n \sqrt{\vartheta_1}} \right)^m \cdot c_{Pot} \cdot (\vec{c}_0 \eta)^m \left\| \vec{u}^{(ext)} \right\|_{L^2(D, \mathbb{R}^q)}, \\ \vartheta_1 &:= k m^{-n} \\ \vec{c}_0 &:= \sqrt{\kappa \vec{c}^n}, \\ \vec{S}^m \vec{u}^{(ext)} &\in V. \end{aligned}$$

Beweisschritt 1: Konstruktion der konvexen Schachtelung

Wir schachteln das Teilgebiet K wie folgt ein:

$$\begin{aligned} K &=: K_0 \subsetneq \dots \subsetneq K_{m-1} \subsetneq K_m \subset D, \\ K_j &:= \{x \in \mathbb{R}^n \mid \text{dist}(x, K) < r_j\}, \quad j = 1, \dots, m, \\ r_0 &:= 0, \\ r_j &:= \text{dist}(K, \partial D) \cdot \frac{j}{m}, \quad j = 1, \dots, m. \end{aligned}$$



Nach Konstruktion von K_m ist also $\text{dist}(K_m, \partial D) = 0$; die Menge K_m dient nur zur technischen Vereinfachung der Beweisführung. Die Szenerie mit allen für diesen Beweis erforderlichen Mengen ist in Abbildung 5.7 dargestellt. Damit gelten, daß die Mengen K_1, \dots, K_m konvex sind und

$$\begin{aligned} \text{dist}(K_{j-1}, \partial K_j) &= r_j - r_{j-1} = \frac{\text{dist}(K, \partial D)}{m}, \\ \varnothing_{K_j} &\leq \varnothing_{K_0} + 2\text{dist}(K_0, \partial K_j) \\ &\stackrel{K_0=K}{=} \varnothing_K + 2(r_j - r_0) \\ &\stackrel{r_0=0}{=} \varnothing_K + \frac{2j}{m}\text{dist}(K, \partial D). \end{aligned}$$

Wir definieren die Funktionenräume

$$X_j := X(K_j, \mathbb{R}^q), \quad j = 0, \dots, m,$$

vgl. (5.36). Die Räume $V_j := V(K_j, \mathbb{R}^q, k) \subset X_j$ seien wie in Lemma 5.16 die L^2 -Projektion der stückweise konstanten Funktionen auf den Raum der \vec{L} -harmonischen Funktionen, also $\dim(V_j) \leq qk$.

Beweisschritt 2a: Approximation von $\vec{u}^{(ext)}$ auf K_{m-1}

Wir wollen $\vec{u}^{(ext)}$ über K durch Treppenfunktionen, die auf den Raum der \vec{L} -harmonischen Funktionen projiziert wurden, möglichst genau approximieren. Um auf

K_{m-1} abschätzen zu können, benötigen wir noch die nur für diesen Zweck eingeführte Obermenge K_m . Weiterhin seien X_{m-1} und X_m wie gerade eben definiert und

$$\vec{u}^{(ext,m)} := \vec{u}^{(ext)}.$$

Dann existieren nach Lemma 5.16 für alle $k \in \mathbb{N}$ ein Teilraum V_{m-1} von \vec{L} -harmonischen Funktionen mit

$$\begin{aligned} V_{m-1} &:= V(K_{m-1}, \mathbb{R}^q, k) \subset X_{m-1}, \\ \dim(V_{m-1}) &\leq qk \end{aligned}$$

und eine orthogonale Projektion

$$\vec{S}_{m-1} : L^2(D, \mathbb{R}^q) \longrightarrow X_{m-1}, \quad \vec{S}_{m-1} \vec{u}^{(ext,m)} := \vec{v}^{(ext,m-1)} \in V_{m-1}$$

und es gilt

$$\left\| \vec{u}^{(ext,m)} - \vec{S}_{m-1} \vec{u}^{(ext,m)} \right\|_{L^2(K_{m-1}, \mathbb{R}^q)} \leq \frac{2\sqrt{n} \varnothing_{K_{m-1}}}{\pi \sqrt[k]{k}} \left\| \nabla \vec{u}^{(ext)} \right\|_{L^2(K_{m-1}, \mathbb{R}^q)}. \quad (5.38)$$

Die linke Seite ist der Fehler, den wir bei der Approximation von $\vec{u}^{(ext,m)}$ durch die projizierten stückweise konstanten Funktionen begehen. Wir setzen daher

$$\text{Fehlerfunktion} = \vec{u}^{(ext,m-1)} := \vec{u}^{(ext,m)} - \vec{v}^{(ext,m-1)}.$$

Wir prüfen die Voraussetzungen für die CACCIOPPOLI-Ungleichung: K_{m-1} übernimmt die Rolle von K und K_m die von D ; außerdem sind $\text{dist}(K_{m-1}, \partial K_m) > 0$ und $K_{m-1} \subset K_m$; also folgt

$$\left\| \nabla \vec{u}^{(ext,m)} \right\|_{L^2(K_{m-1}, \mathbb{R}^q)} \leq \frac{2\sqrt{k\bar{C}}}{\text{dist}(K_{m-1}, \partial K_m)} \cdot \left\| \vec{u}^{(ext,m)} \right\|_{L^2(K_m, \mathbb{R}^q)}.$$

Mit $\text{dist}(K_{m-1}, \partial K_m) = \frac{1}{m} \text{dist}(K, \partial D)$ und $\left\| \vec{u}^{(ext,m)} \right\|_{L^2(K_m, \mathbb{R}^q)} \leq \left\| \vec{u}^{(ext,m)} \right\|_{L^2(D, \mathbb{R}^q)}$ erreichen wir

$$\left\| \nabla \vec{u}^{(ext,m)} \right\|_{L^2(K_{m-1}, \mathbb{R}^q)} \leq \frac{2m\sqrt{k\bar{C}}}{\text{dist}(K, \partial D)} \cdot \left\| \vec{u}^{(ext,m)} \right\|_{L^2(D, \mathbb{R}^q)}.$$

Zusammen mit (5.38) erzielen wir die gewünschte Abschätzung

$$\left\| \vec{u}^{(ext,m)} - \vec{S}_{m-1} \vec{u}^{(ext,m)} \right\|_{L^2(K_{m-1}, \mathbb{R}^q)} \leq \frac{4m\sqrt{k\bar{C}n}}{\pi \sqrt[k]{k}} \cdot \frac{\varnothing_{K_{m-1}}}{\text{dist}(K, \partial D)} \left\| \vec{u}^{(ext)} \right\|_{L^2(D, \mathbb{R}^q)}.$$

Beweisschritt 2b: Approximation von $\vec{u}^{(ext,m-1)}$ auf K_{m-2}

Nach Lemma 5.16 existiert für alle $k \in \mathbb{N}$ wieder ein Teilraum V_{m-2} mit $\dim(V_{m-2}) \leq qk$ und eine orthogonale Projektion

$$\vec{S}_{m-2} : L^2(D, \mathbb{R}^q) \longrightarrow X_{m-2}, \quad \vec{S}_{m-2} \vec{u}^{(ext,m-1)} := \vec{v}^{(ext,m-2)} \in V_{m-2},$$

sodaß gilt

$$\left\| \vec{u}^{(ext,m-1)} - \vec{S}_{m-2} \vec{u}^{(ext,m-1)} \right\|_{L^2(K_{m-2}, \mathbb{R}^q)} \leq \frac{2\sqrt{n}}{\pi} \cdot \frac{\varnothing_{K_{m-2}}}{\sqrt{k}} \cdot \left\| \nabla \vec{u}^{(ext,m-1)} \right\|_{L^2(K_{m-2}, \mathbb{R}^q)}.$$

Die Voraussetzungen für die CACCIOPOLI-Ungleichung sind auch hier erfüllt, d.h.

$$\left\| \nabla \vec{u}^{(ext,m-1)} \right\|_{L^2(K_{m-2}, \mathbb{R}^q)} \leq \frac{2\sqrt{\kappa_{\vec{C}}}}{\text{dist}(K_{m-2}, \partial K_{m-1})} \cdot \left\| \vec{u}^{(ext,m-1)} \right\|_{L^2(K_{m-1}, \mathbb{R}^q)}.$$

Kombinieren wir die beiden Abschätzungen mit

$$\text{dist}(K_{m-2}, \partial K_{m-1}) = \frac{1}{m} \text{dist}(K, \partial D),$$

erhalten wir

$$\begin{aligned} & \left\| \vec{u}^{(ext,m-1)} - \vec{S}_{m-2} \vec{u}^{(ext,m-1)} \right\|_{L^2(K_{m-2}, \mathbb{R}^q)} \\ & \leq \frac{4m\sqrt{\kappa_{\vec{C}}n}}{\pi \cdot \sqrt{k}} \cdot \frac{\varnothing_{K_{m-2}}}{\text{dist}(K, \partial D)} \cdot \left\| \vec{u}^{(ext,m-1)} \right\|_{L^2(K_{m-1}, \mathbb{R}^q)}. \end{aligned}$$

Da aber

$$\begin{aligned} & \left\| \vec{u}^{(ext,m-1)} - \vec{S}_{m-2} \vec{u}^{(ext,m-1)} \right\|_{L^2(K_{m-2}, \mathbb{R}^q)} \\ & = \text{dist}_{L^2(K_{m-2}, \mathbb{R}^q)} \left(\vec{u}^{(ext,m-1)} + \vec{v}^{(ext,m-1)}, \left\langle V_{m-2}, V_{m-1} \Big|_{K_{m-2}} \right\rangle \right), \end{aligned}$$

erhalten wir mit

$$\begin{aligned} \vec{u}^{(ext,m-1)} &= \vec{u}^{(ext,m-2)} + \underbrace{\vec{v}^{(ext,m-2)}}_{\in V_{m-2}}, \\ \vec{u}^{(ext,m)} &= \vec{u}^{(ext,m-2)} + \vec{v}^{(ext,m-2)} + \vec{v}^{(ext,m-1)} \end{aligned}$$

nach zwei Rekursionen die Abschätzung

$$\begin{aligned} \left\| \vec{u}^{(ext,m-2)} \right\|_{L^2(K_{m-2}, \mathbb{R}^q)} &= \text{dist}_{L^2(K_{m-2}, \mathbb{R}^q)} \left(\vec{u}^{(ext)}, \left\langle V_{m-2}, V_{m-1} \Big|_{K_{m-2}} \right\rangle \right) \\ &\leq \left(\frac{4m\sqrt{\kappa_{\vec{C}}n}}{\pi \cdot \sqrt{k}} \right)^2 \cdot \frac{\varnothing_{K_{m-2}} \cdot \varnothing_{K_{m-1}}}{(\text{dist}(K, \partial D))^2} \cdot \left\| \vec{u}^{(ext)} \right\|_{L^2(D, \mathbb{R}^q)}. \end{aligned}$$

Beweisschritt 2c: Approximation auf K_0

Die Abschätzungsprozedur aus den Schritten 2a und 2b können wir solange wiederholen, bis wir bei K_0 angelangt sind, also noch $(m-3)$ -Mal. Damit haben wir am Ende der Iteration die Fortsetzung $\vec{u}^{(ext)}$ von \vec{u} aufgesplittet, nämlich

$$\vec{u}^{(ext)} = \vec{u}^{(ext,m)} = \vec{u}^{(ext,0)} + \sum_{j=0}^{m-1} \vec{v}^{(ext,j)}$$

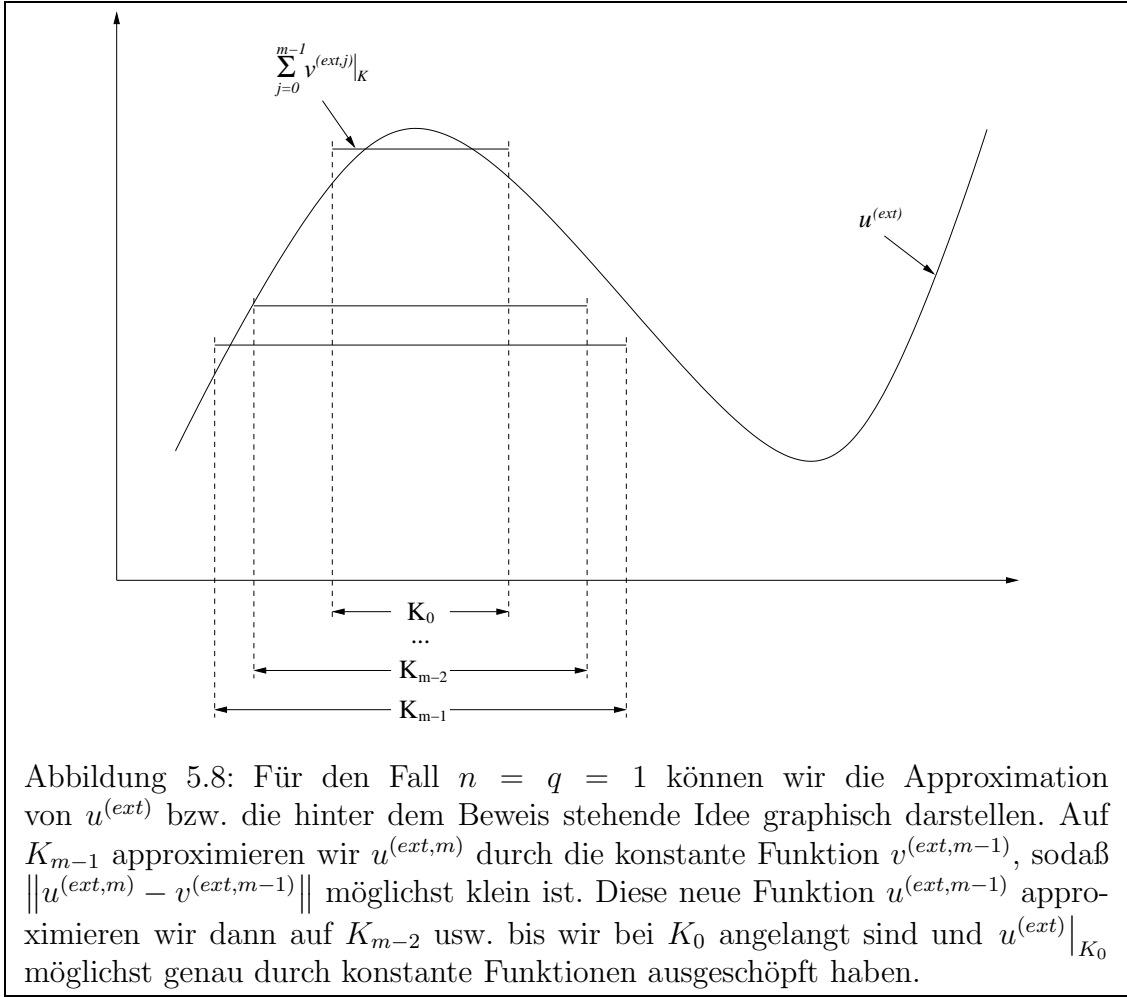


Abbildung 5.8: Für den Fall $n = q = 1$ können wir die Approximation von $u^{(ext)}$ bzw. die hinter dem Beweis stehende Idee graphisch darstellen. Auf K_{m-1} approximieren wir $u^{(ext,m)}$ durch die konstante Funktion $v^{(ext,m-1)}$, sodaß $\|u^{(ext,m)} - v^{(ext,m-1)}\|$ möglichst klein ist. Diese neue Funktion $u^{(ext,m-1)}$ approximieren wir dann auf K_{m-2} usw. bis wir bei K_0 angelangt sind und $u^{(ext)}|_{K_0}$ möglichst genau durch konstante Funktionen ausgeschöpft haben.

(vgl. Abb. 5.8). Die Funktionen $\vec{v}^{(ext,j)}$ sind Elemente des Raumes

$$V := \langle V_0, V_1|_K, \dots, V_{m-1}|_K \rangle,$$

d.h. wir können die Genauigkeit der Approximation von $\vec{u}^{(ext)}$ durch Funktionen aus V abschätzen durch

$$\begin{aligned} \|\vec{u}^{(ext,0)}\|_{L^2(K, \mathbb{R}^q)} &= \text{dist}_{L^2(K, \mathbb{R}^q)}(\vec{u}^{(ext)}, V) \\ &= \vec{u}^{(ext,1)} - \vec{S}_0 \vec{u}^{(ext,1)} \\ &\leq \left(\frac{4m\sqrt{\kappa \bar{c} \bar{n}}}{\pi \cdot \sqrt{k}} \right)^m \cdot \frac{\prod_{j=0}^{m-1} \varnothing_{K_j}}{(\text{dist}(K, \partial D))^m} \cdot \|\vec{u}^{(ext)}\|_{L^2(D, \mathbb{R}^q)}. \end{aligned}$$

Lemma 5.16 erlaubt beliebige $k \in \mathbb{N}$, was wir ausnutzen, um uns des Faktors m^m zu entledigen, indem wir

$$k := \vartheta_1 \cdot m^n, \quad \vartheta_1 > 0,$$

wählen (ϑ_1 bzw. Index 1 deshalb, um etwaige Verwechslungen mit der Elliptizitätskonstanten ϑ zu vermeiden).

Beweisschritt 3: weitere Abschätzungen

Für $m \in \mathbb{N}$ und $j = 0, \dots, m-1$ gilt mit den Überlegungen aus Schritt 1

$$\begin{aligned} \frac{\varnothing_{K_j}}{\text{dist}(K, \partial D)} &\leq \frac{\varnothing_K + 2\frac{j}{m}\text{dist}(K, \partial D)}{\text{dist}(K, \partial D)} \\ &= \varnothing_K \cdot \frac{1}{\text{dist}(K, \partial D)} + 2\frac{j}{m} \\ &\stackrel{(5.37)}{\leq} \varnothing_K \cdot \frac{2\eta}{\varnothing_K} + 2\frac{j}{m} \\ &= 2\left(\eta + \frac{j}{m}\right). \end{aligned}$$

Damit ist

$$\prod_{j=0}^{m-1} \frac{\varnothing_{K_j}}{\text{dist}(K, \partial D)} \leq \prod_{j=0}^{m-1} \left[2\left(\eta + \frac{j}{m}\right) \right] = \left(\frac{2}{m}\right)^m \cdot \prod_{j=0}^{m-1} (\eta m + j) \leq c_{Pot} 2^m \eta^m$$

für eine positive Konstante c_{Pot} ⁶ und wir haben für den Approximationsfehler die Obergrenze

$$\begin{aligned} \text{dist}_{L^2(K, \mathbb{R}^q)}(\vec{u}^{(ext)}, V) &\leq c_{Pot} \left(\frac{4\sqrt{\kappa \vec{c}} n}{\pi \sqrt[n]{\vartheta_1}}\right)^m 2^m \eta^m \|\vec{u}^{(ext)}\|_{L^2(D, \mathbb{R}^q)} \\ &\leq c_{Pot} \left(\frac{8}{\pi \sqrt[n]{\vartheta_1}}\right)^m (\vec{c}_0 \eta)^m \|\vec{u}^{(ext)}\|_{L^2(D, \mathbb{R}^q)} \end{aligned}$$

mit der Konstanten

$$\vec{c}_0 := \sqrt{\kappa \vec{c} n}.$$

Definieren wir \vec{S}^m als die Verkettung von $(Id - \vec{S}_i)$, $i = 0, \dots, m-1$, so ist der Gesamtfehler

$$\|\vec{u}^{(ext,0)}\|_{L^2(K, \mathbb{R}^q)} = \|\vec{u}^{(ext)} - \vec{S}^m \vec{u}^{(ext)}\|_{L^2(K, \mathbb{R}^q)} \leq c_{Pot} \cdot \left(\frac{8\vec{c}_0 \eta}{\pi \sqrt[n]{\vartheta_1}}\right)^m \cdot \|\vec{u}^{(ext)}\|_{L^2(D, \mathbb{R}^q)}.$$

Beweisschritt 4: Dimension des Approximationsraumes

Der Raum, in dem $\vec{u}^{(ext)}$ approximiert wird, ist

$$V := \langle V_0, V_1|_K, \dots, V_{m-1}|_K \rangle.$$

Da $\dim(V_j) \leq qk$, folgt $\dim(V) \leq mqk$. **QED**

⁶Diese Konstante hängt von η ab, was aber nicht dramatisch ist, weil wir uns, wie in Bemerkung 4.11 begründet, für $\eta \approx 1$ interessieren.

5.3.4 Lokale Niedrigrang- und \mathcal{H} -Approximation von \vec{L}^{-1}

Satz 5.18 hat in unserer Beweistechnik noch kein Pendant, wir müssen also noch etwas Arbeit investieren, um vergleichbare Aussagen in der Hand zu halten.

Korollar 5.19 (Lokale Niedrigrangapproximation von \vec{L}^{-1}) *Seien $\sigma, \tau \subset \Omega$ zwei zulässige Mengen. Dann konvergiert der Operator $\vec{S}^m \vec{L}^{-1}$ exponentiell gegen \vec{L}^{-1} und ist vom Maximalrang $m q k$, genauer*

$$\begin{aligned} \left\| \vec{L}^{-1} - \vec{S}^m \vec{L}^{-1} \right\|_{\mathcal{L}(L^2(\sigma, \mathbb{R}^q), L^2(\tau, \mathbb{R}^q))} &\leq c_{(5.39)} (\vec{c}_0 \eta)^m, \\ \text{rnk} \left(\vec{S}^m \vec{L}^{-1} \right) &\leq m q k \end{aligned}$$

mit der Konstanten

$$c_{(5.39)} := c_{Pot} \cdot \left(\frac{8}{\pi \sqrt[n]{\vartheta_1}} \right)^m \cdot c_\Omega. \quad (5.39)$$

Beweis:

- (a) Seien $\sigma, \tau \subset \Omega$ zwei zulässige Mengen. Wir packen τ in eine konvexe Obermenge $K \subset D \setminus \sigma$ und schließen

$$\begin{aligned} \left\| \vec{u} - \vec{S}^m \vec{u} \right\|_{L^2(\tau, \mathbb{R}^q)} &\stackrel{\tau \subset K}{\leq} \left\| \vec{u}^{(ext)} - \vec{S}^m \vec{u}^{(ext)} \right\|_{L^2(K, \mathbb{R}^q)} \\ &\stackrel{\text{Satz 5.18}}{\leq} c_{Pot} \left(\frac{8}{\pi \sqrt[n]{\vartheta_1}} \right)^m \cdot (\vec{c}_0 \eta)^m \cdot \left\| \vec{u}^{(ext)} \right\|_{L^2(D, \mathbb{R}^q)}. \end{aligned}$$

- (b) Mit dem Satz von LAX-MILGRAM gilt

$$\begin{aligned} \left\| \vec{u}^{(ext)} \right\|_{L^2(D, \mathbb{R}^q)} &= \left\| \vec{u} \right\|_{L^2(D_\Omega, \mathbb{R}^q)} \\ &\leq \left\| \vec{u} \right\|_{L^2(\Omega, \mathbb{R}^q)} \\ &\leq c_\Omega \left\| \vec{f} \right\|_{L^2(\Omega, \mathbb{R}^q)} \\ &=_{\text{supp}(\vec{f}) \subset \sigma} c_\Omega \left\| \vec{f} \right\|_{L^2(\sigma, \mathbb{R}^q)}. \end{aligned}$$

- (c) Fassen wir die ersten zwei Beweisschritte zusammen, haben wir

$$\begin{aligned} \left\| \vec{u} - \vec{S}^m \vec{u} \right\|_{L^2(\tau, \mathbb{R}^q)} &\leq c_{Pot} \left(\frac{8}{\pi \sqrt[n]{\vartheta_1}} \right)^m \cdot (\vec{c}_0 \eta)^m \cdot c_\Omega \left\| \vec{f} \right\|_{L^2(\sigma, \mathbb{R}^q)} \\ &= c_{(5.39)} (\vec{c}_0 \eta)^m \left\| \vec{f} \right\|_{L^2(\sigma, \mathbb{R}^q)}. \end{aligned}$$

Andererseits ist

$$\begin{aligned} \left\| \vec{u} - \vec{S}^m \vec{u} \right\|_{L^2(\tau, \mathbb{R}^q)} &= \left\| \vec{L}^{-1} \vec{f} - \vec{S}^m \vec{L}^{-1} \vec{f} \right\|_{L^2(\sigma, \mathbb{R}^q)} \\ &\leq \left\| \vec{L}^{-1} - \vec{S}^m \vec{L}^{-1} \right\|_{\mathcal{L}(L^2(\sigma, \mathbb{R}^q), H^1(\tau, \mathbb{R}^q))} \cdot \left\| \vec{f} \right\|_{L^2(\sigma, \mathbb{R}^q)}. \end{aligned}$$

Da beide Abschätzungen für alle $\vec{f} \in L^2(\sigma, \mathbb{R}^q)$ gelten, folgt

$$\left\| \vec{L}^{-1} - \vec{S}^m \vec{L}^{-1} \right\|_{\mathcal{L}(L^2(\sigma, \mathbb{R}^q), L^2(\tau, \mathbb{R}^q))} \leq c_{(5.39)} \cdot (\vec{c}_0 \eta)^m.$$

- (d) Nach Satz 5.18 liegt $\vec{S}^m \vec{u}^{(ext)}$ in einem Raum der Dimension höchstens $m q k$, also ist $\text{rnk}(\vec{S}^m) \leq m q k$ und somit

$$\text{rnk}(\vec{S}^m \vec{L}^{-1}) \leq \min \left\{ \text{rnk}(\vec{S}^m), \text{rnk}(\vec{L}^{-1}) \right\} \leq m q k.$$

□

Satz 5.20 (Lokale \mathcal{H} -Approximation von $(\vec{L}^{-1})_h$) Seien $\sigma, \tau \subset \Omega$ zwei zulässige Mengen und s, t die assoziierten Indexteilmengen sowie \vec{s} die Menge der Spaltennummern und \vec{t} die Menge der Zeilennummern, vgl. Satz 5.13. Dann umfaßt $\vec{b} := \vec{t} \times \vec{s}$ die Menge aller Indices von Matrixeinträgen mit Informationen zum Cluster $\beta := \tau \times \sigma$. Wir definieren den Matrixteilstück

$$\left[(\vec{L}^{-1})_h \right]_{\vec{b}} := \left[\vec{J}^* \vec{L}^{-1} \vec{J} \right]_{\vec{b}}$$

der RITZ-GALERKIN-Diskretisierung von \vec{L}^{-1} . Sei μ_{max} der größte Eigenwert der Massenmatrix.

Dann lautet eine lokale \mathcal{H} -Approximation von $\left[(\vec{L}^{-1})_h \right]_{\vec{b}}$

$$\left[(\vec{L}^{-1})_{\mathcal{H}} \right]_{\vec{b}} := \left[\vec{J}^* \vec{S}^m \vec{L}^{-1} \vec{J} \right]_{\vec{b}}$$

und erfüllt

$$\left\| \left[(\vec{L}^{-1})_h - (\vec{L}^{-1})_{\mathcal{H}} \right]_{\vec{b}} \right\|_2 \leq c_{(5.39)} (\vec{c}_0 \eta)^m \mu_{max}, \quad (5.40)$$

$$\text{rnk} \left(\left[(\vec{L}^{-1})_{\mathcal{H}} \right]_{\vec{b}} \right) \leq m q k. \quad (5.41)$$

Beweis: Nach Korollar 5.19 gilt

$$\dim(\vec{S}^m \vec{L}^{-1}) \leq m q k.$$

Da $\vec{J} : \mathbb{R}^{qN} \rightarrow H_0^1(\Omega, \mathbb{R}^q)$ und $\vec{J}^* : H^{-1}(\Omega, \mathbb{R}^q) \rightarrow \mathbb{R}^{qN}$, folgt (5.41). Die Abschätzung (5.40) beweisen wir analog zu Satz 5.13. □

5.4 Globale Approximationseigenschaft

Wie eingangs erläutert, müssen wir den Fehler $\|(L_h)^{-1} - C_{\mathcal{H}}\|_2$ abschätzen, um die \mathcal{H} -Approximierbarkeit der Inversen der FE-Matrix zu beweisen. Dazu führen wir eine Hilfsmatrix C_h ein, sodaß nach Dreiecksungleichung statt obiger Differenz eben $\|(L_h)^{-1} - C_h\|_2$ und $\|C_h - C_{\mathcal{H}}\|_2$ abzuschätzen sind. Im einzelnen müssen wir uns noch mit den Punkten

- Abschätzung von $\|(L^{-1})_h - (L^{-1})_{\mathcal{H}}\|_2$,
- Abschätzung von $\|M^{-1} - (M^{-1})_{\mathcal{H}}\|_2$,
- Definition der Hilfsmatrix $C_h := M^{-1} (L^{-1})_h M^{-1}$,
- Abschätzung von $\|(L_h)^{-1} - C_h\|_2$,
- Definition der \mathcal{H} -Approximation $C_{\mathcal{H}}$ von C_h und
- Bilanzierung des Gesamtfehlers $\|(L_h)^{-1} - C_{\mathcal{H}}\|_2$

auseinandersetzen.

5.4.1 Steifigkeitsmatrix des inversen Differentialoperators

In den beiden vorigen Absätzen haben wir gezeigt, daß wir die RITZ-GALERKIN-Diskretisierung des inversen (Matrix-) Differentialoperators lokal, d.h. auf Clustern $\beta := \tau \times \sigma$, $\sigma, \tau \subset \Omega$ zulässig, durch Rk -Matrizen darstellen können. Dieses in den Sätzen 5.10, 5.13 und 5.20 formulierte Ergebnis verwenden wir, um die globale Approximierbarkeit des inversen (Matrix-) Differentialoperators zu zeigen und einen Kandidaten $C_{\mathcal{H}}$ für die Approximation der inversen Steifigkeitsmatrizen $(L_h)^{-1}$ bzw. $(\vec{L}_h)^{-1}$ zu konstruieren.

Bemerkung 5.21 Sei c_{lokAF} die Konstante, die bei der Abschätzung der TAYLOR-Entwicklung anfällt, d.h.

$$c_{\text{lokAF}} := \begin{cases} c_{(5.20)} c_m c_{m,n} & \text{für skalare DG,} \\ c_{(5.29)} c_m c_{m,n} \text{ bzw. } c_{(5.39)} & \text{für DG-Systeme,} \end{cases} \quad (5.42)$$

Die eigentlich von m abhängenden Faktoren c_m und $c_{m,n}$ schlagen wir jetzt der Konstante c_{lokAF} zu, denn zum einen ist $c_m c_{m,n}$ in der Größenordnung von bzw. kleiner eins, zum anderen ist der in 5.3 vorgestellte Zugang allgemeiner (als die in 5.1 und 5.2 unternommenen Versuche) und kommt ohne diese Faktoren aus, d.h. durch diese Maßnahme wird nichts vertuscht oder verfälscht, schlimmstenfalls geschönt, weil die Notation einfacher wird.

Bezeichne \vec{L}^{-1} sowohl den inversen Matrix- als auch den inversen skalaren Differentialoperator L^{-1} . Lassen wir schließlich die Unterscheidung zwischen c_0 und \vec{c}_0

fallen und schreiben nur noch \vec{c}_0 , so schließen die Sätze 5.13 und 5.20 den Satz 5.10 ein und mit (5.42) gilt

$$\left\| \left[\left(\vec{L}^{-1} \right)_h - \left(\vec{L}^{-1} \right)_{\mathcal{H} \downarrow \vec{b}} \right] \right\|_2 \leq c_{lokAF} \mu_{max} (\vec{c}_0 \eta)^m.$$

Wir brauchen fortan nur noch \vec{L}^{-1} zu betrachten. Den Maximalrang der Blöcke $\left[\left(\vec{L}^{-1} \right)_{\mathcal{H} \downarrow \vec{b}} \right]$ bezeichnen wir mit $k_{\mathcal{H}}^{\vec{L}^{-1}}$; je nach Vorgehensweise haben wir zwei unterschiedliche Angaben, nämlich

$$k_{\mathcal{H}}^{\vec{L}^{-1}} \stackrel{\text{Satz 5.13}}{=} q \frac{(m+n)!}{m!n!} \quad \text{oder} \quad k_{\mathcal{H}}^{\vec{L}^{-1}} \stackrel{\text{Satz 5.20}}{=} m q k. \quad (5.43)$$

Korollar 5.22 (Globale \mathcal{H} -Approximation von \vec{L}^{-1}) Seien die Bezeichnungen wie in den Sätzen 4.13 und 5.13 bzw. 5.20. Dann kann die RITZ-GALERKIN-Approximation von \vec{L}^{-1} durch die \mathcal{H} -Matrix

$$\left(\vec{L}^{-1} \right)_{\mathcal{H}} \in \mathcal{H} \left(k_{\mathcal{H}}^{\vec{L}^{-1}}, T, Z \right) \quad (5.44)$$

dargestellt und der Fehler durch die Konstante

$$err_{L^{-1}, \mathcal{H}} := C_{sp} (1 + p_T) c_{lokAF} \mu_{max} (\vec{c}_0 \eta)^m \quad (5.45)$$

abgeschätzt werden, d.h.

$$\left\| \left(\vec{L}^{-1} \right)_h - \left(\vec{L}^{-1} \right)_{\mathcal{H} \downarrow \vec{b}} \right\|_2 \leq err_{L^{-1}, \mathcal{H}}. \quad (5.46)$$

Weiterhin gilt

$$\left\| \left(\vec{L}^{-1} \right)_{\mathcal{H}} \right\|_2 \leq err_{L^{-1}, \mathcal{H}} + \frac{\mu_{max}}{c_E}. \quad (5.47)$$

Beweis: Wir merken an, daß die Struktur des \mathcal{H} -Produktbaumes für skalare Differentialgleichungen wie für Systeme dieselbe ist, sofern diese nicht gleichungs-, sondern knotenweise diskretisiert werden.

Der maximale Rang der zulässigen Blöcke ist durch $k_{\mathcal{H}}^{\vec{L}^{-1}}$ nach oben begrenzt, also haben wir (5.44) gezeigt.

Mit dem Schrankensatz 4.13 folgt für $A := \left(\vec{L}^{-1} \right)_h - \left(\vec{L}^{-1} \right)_{\mathcal{H}}$

$$\begin{aligned} \left\| \left(\vec{L}^{-1} \right)_h - \left(\vec{L}^{-1} \right)_{\mathcal{H} \downarrow \vec{b}} \right\|_2 &\leq C_{sp} \cdot \sum_{l=0}^{p_T} \max_{b \in \mathcal{L}(T, l)} \left\| \left[\left(\vec{L}^{-1} \right)_h - \left(\vec{L}^{-1} \right)_{\mathcal{H} \downarrow b} \right] \right\|_2 \\ &\leq C_{sp} \cdot \sum_{l=0}^{p_T} c_{lokAF} \mu_{max} (\vec{c}_0 \eta)^m \\ &= C_{sp} (1 + p_T) c_{lokAF} \mu_{max} (\vec{c}_0 \eta)^m \\ &= err_{L^{-1}, \mathcal{H}}. \end{aligned}$$

Nach Lemma 3.7 ist $(\vec{L}^{-1})_h = \vec{J}^* \vec{L}^{-1} \vec{J}$, also ist

$$\left\| (\vec{L}^{-1})_h \right\|_2 \leq \left\| \vec{L}^{-1} \right\|_{\mathcal{L}(H^{-1}(\Omega, \mathbb{R}^q), H_0^1(\Omega, \mathbb{R}^q))} \cdot \left\| \vec{J} \right\|_{\mathcal{L}(\mathbb{R}^{qN}, V_h^q)}^2 \stackrel{\text{Satz 2.15}}{\leq} \frac{\mu_{\max}}{c_E}. \quad (5.48)$$

Mit der Dreiecksungleichung

$$\left\| (\vec{L}^{-1})_{\mathcal{H}} \right\|_2 \leq \left\| (\vec{L}^{-1})_{\mathcal{H}} - (\vec{L}^{-1})_k \right\|_2 + \left\| (\vec{L}^{-1})_h \right\|_2$$

folgt dann die verbleibende Behauptung. **QED**

5.4.2 Massen- und Steifigkeitsmatrix

Als Bindeglied zwischen der inversen Steifigkeitsmatrix und der RITZ-GALERKIN-Diskretisierung des inversen Differentialoperators tritt die Massenmatrix auf, deren Inverse durch eine \mathcal{H} -Matrix approximiert werden kann und muß.

Lemma 5.23 (\mathcal{H} -Approximation der inversen Massenmatrix) *Es gelte Satz 3.4 mit $h := \max_{i \in I} \text{supp}(\varphi_i)$ und $\text{vol}(\text{supp}(\varphi_i)) \geq ch^n$, $c > 0$. Sei T ein bzgl. der Standardzulässigkeitsbedingung aufgebauter \mathcal{H} -Produktbaum von $I \times I$ der Höhe p_T und $\mathcal{L}^+(T)$ die Menge seiner zulässigen Blätter.*

Dann existiert für alle $\varepsilon_M > 0$

(a) *ein $k(\varepsilon_M) \in \mathbb{N}$ mit*

$$k(\varepsilon_M) = \mathcal{O} \left(\left(\log \left(\frac{p_T}{\varepsilon_M} \right) \right)^n \right), \quad (5.49)$$

(b) *eine Matrix*

$$(M^{-1})_{\mathcal{H}} := \begin{cases} M^{-1}|_b & \text{für } |b| \leq k(\varepsilon_M)^2 \\ 0 & \text{sonst} \end{cases} \quad (5.50a)$$

mit den Eigenschaften

$$\left\| M^{-1} - (M^{-1})_{\mathcal{H}} \right\|_2 \leq \varepsilon_M \left\| M^{-1} \right\|_2 = \frac{\varepsilon_M}{\mu_{\min}} =: \varepsilon_{M^{-1}} =: \text{err}_{M^{-1}, \mathcal{H}}, \quad (5.50b)$$

$$(M^{-1})_{\mathcal{H}} \in \mathcal{H}(k(\varepsilon_M), T, Z), \quad (5.50c)$$

(c) *eine Matrix*

$$(\vec{M}^{-1})_{\mathcal{H}} := P^\top (\vec{M}_G^{-1})_{\mathcal{H}} P, \quad (5.51a)$$

wobei P die Permutationsmatrix aus (3.23) ist. Die Matrix $(\vec{M}^{-1})_{\mathcal{H}}$ besitzt die Eigenschaften

$$\|\vec{M}^{-1} - (\vec{M}^{-1})_{\mathcal{H}}\|_2 \leq \frac{\varepsilon_M}{\mu_{\min}}, \quad (5.51b)$$

$$(\vec{M}^{-1})_{\mathcal{H}} \in \mathcal{H}(qk(\varepsilon_M), T, Z), \quad (5.51c)$$

$$\|(\vec{M}^{-1})_{\mathcal{H}}\|_2 \leq \frac{1 + \varepsilon_M}{\mu_{\min}}. \quad (5.51d)$$

Beweis: Für (a) und (b) siehe [BH03, Satz 4.3]. Bleibt uns, (c) zu zeigen, indem wir die Ergebnisse aus dem skalaren Fall mit Hilfe des Absatzes 3.5.2 auf \vec{M}^{-1} übertragen. Von dort wissen wir, daß

$$\vec{M} = P^\top \vec{M}_G P, \quad \vec{M}_G := \text{diag}(M, \dots, M),$$

gilt, also ist

$$\vec{M}^{-1} = P^\top \vec{M}_G^{-1} P, \quad \vec{M}_G^{-1} = \text{diag}(M^{-1}, \dots, M^{-1}).$$

Wir setzen daher

$$(\vec{M}_G^{-1})_{\mathcal{H}} := \text{diag}((M^{-1})_{\mathcal{H}}, \dots, (M^{-1})_{\mathcal{H}}).$$

Aufgrund dieser Konstruktion folgt mit dem Satz über die Eigenwerte von Blockdiagonalmatrizen

$$\begin{aligned} \|\vec{M}_G^{-1} - (\vec{M}_G^{-1})_{\mathcal{H}}\|_2 &= \|\text{diag}(M^{-1} - (M^{-1})_{\mathcal{H}}, \dots, M^{-1} - (M^{-1})_{\mathcal{H}})\|_2 \\ &= \|M^{-1} - (M^{-1})_{\mathcal{H}}\|_2 \\ &\stackrel{(5.50b)}{\leq} \frac{\varepsilon_M}{\mu_{\min}}. \end{aligned} \quad (5.52)$$

Die Blocknebendiagonalen von $(\vec{M}_G^{-1})_{\mathcal{H}}$ bestehen nur aus Nullen, also Blockmatrizen mit Rang 0. Dadurch wird der Maximalrang der zulässigen Blöcke von $(\vec{M}_G^{-1})_{\mathcal{H}}$ durch den Maximalrang der zulässigen Blöcke von $(M^{-1})_{\mathcal{H}}$ festgelegt, nämlich $k(\varepsilon_M)$. Allerdings gehört unser \mathcal{H} -Produktbaum T nicht zu $(M^{-1})_{\mathcal{H}}$.

Der zu einem zulässigen Blatt von T gehörende Teilblock von \vec{M}^{-1} ist q -mal so groß wie der entsprechende Teilblock von M^{-1} und besitzt daher einen q -mal so großen Rang, d.h. $q \cdot k(\varepsilon_M)$, womit (5.51c) gezeigt ist. Der in (5.51b) beschriebene \mathcal{H} -Näherungsfehler folgt aus

$$\begin{aligned} \|\vec{M}^{-1} - (\vec{M}^{-1})_{\mathcal{H}}\|_2 &= \left\| \left(P^\top \vec{M}_G P \right)^{-1} - P^\top (\vec{M}_G^{-1})_{\mathcal{H}} P \right\|_2 \\ &= \left\| P^\top \left(\vec{M}_G^{-1} - (\vec{M}_G^{-1})_{\mathcal{H}} \right) P \right\|_2 \\ &\leq \|P^\top\|_2 \left\| \vec{M}_G^{-1} - (\vec{M}_G^{-1})_{\mathcal{H}} \right\|_2 \|P\|_2 \\ &\leq \frac{\varepsilon_M}{\mu_{\min}} \end{aligned}$$

wegen $\|P^\top\|_2 = \|P\|_2 = 1$ und (5.52). Schlußendlich bleibt die Abschätzung des maximalen Eigenwertes von $\left(\vec{M}^{-1}\right)_\mathcal{H}$:

$$\begin{aligned} \left\| \left(\vec{M}^{-1}\right)_\mathcal{H} \right\|_2 &\stackrel{\text{Dreiecksungl.}}{\leq} \left\| \left(\vec{M}^{-1}\right)_\mathcal{H} - \vec{M}^{-1} \right\|_2 + \left\| \vec{M}^{-1} \right\|_2 \\ &\stackrel{(5.51b)}{\leq} \frac{\varepsilon_M}{\mu_{\min}} + \frac{1}{\mu_{\min}} \\ &= \frac{1 + \varepsilon_M}{\mu_{\min}}. \end{aligned}$$

QED

Lemma 5.24 (Approximation der inversen FE- dr. eine Hilfsmatrix) Sei $\varepsilon_h = c_{FEM} \cdot h^{2\alpha}$ die FE-Konvergenzgeschwindigkeit aus Lemma 3.10. Dann kann die inverse Steifigkeitsmatrix $\left(\vec{L}_h\right)^{-1}$ durch die Hilfsmatrix

$$\vec{C}_h := \vec{M}^{-1} \left(\vec{L}^{-1}\right)_h \vec{M}^{-1} \quad (5.53a)$$

angenähert werden mit dem Fehler

$$\left\| \left(\vec{L}_h\right)^{-1} - \vec{C}_h \right\|_2 \leq 2c_{FEM} h^{2\alpha} \frac{\mu_{\max}}{\mu_{\min}^2}. \quad (5.53b)$$

Beweis: Wegen der Darstellbarkeit diskreter Operatoren durch die Verkettung kontinuierlicher Operatoren mit dem Koeffizientenmorphismus, vgl. Lemmata 3.3 und 3.7, bietet sich (5.53a) als kanonische Wahl zur Approximation von $\left(\vec{L}_h\right)^{-1}$ an. Wir schätzen den verursachten Fehler: wegen

$$\left(\vec{L}_h\right)^{-1} - \vec{C}_h = \vec{M}^{-1} \cdot \left(\vec{M} \left(\vec{L}_h\right)^{-1} \vec{M} - \left(\vec{L}^{-1}\right)_h \right) \cdot \vec{M}^{-1}$$

und $\left\| \vec{M}^{-1} \right\|_2 = \mu_{\min}^{-1}$ gilt

$$\left\| \left(\vec{L}_h\right)^{-1} - \vec{C}_h \right\|_2 \leq \frac{1}{\mu_{\min}^2} \left\| \vec{M} \left(\vec{L}_h\right)^{-1} \vec{M} - \left(\vec{L}^{-1}\right)_h \right\|_2. \quad (5.54)$$

Nach Definition der Operatornorm ist

$$\left\| \vec{M} \left(\vec{L}_h\right)^{-1} \vec{M} - \left(\vec{L}^{-1}\right)_h \right\|_2 := \sup_{\vec{v}, \vec{w} \in \mathbb{R}^{qN}} \frac{\left\langle \left[\vec{M} \left(\vec{L}_h\right)^{-1} \vec{M} - \left(\vec{L}^{-1}\right)_h \right] \vec{v}, \vec{w} \right\rangle_h}{\|\vec{v}\|_h \cdot \|\vec{w}\|_h},$$

Wir behandeln aus Gründen der Übersichtlichkeit zuerst den Zähler. Seien hierzu $\vec{\mathbf{v}}, \vec{\mathbf{w}} \in \mathbb{R}^{qN}$ und $\vec{v}_h := \vec{J}\vec{\mathbf{v}}, \vec{w}_h := \vec{J}\vec{\mathbf{w}}$. Dann gelten

$$\begin{aligned}\vec{\mathbf{v}} &= Id \cdot \vec{\mathbf{v}} \\ &= \vec{M}^{-1} \cdot \vec{M} \cdot \vec{\mathbf{v}} \\ &= \vec{M}^{-1} \cdot \vec{J}^* \vec{J} \cdot \vec{\mathbf{v}} \\ &= \vec{M}^{-1} \cdot \vec{J}^* \vec{v}_h, \\ \vec{\mathbf{w}} &= \vec{M}^{-1} \cdot \vec{J}^* \vec{w}_h.\end{aligned}$$

und wegen $\vec{M}^{-1} = (\vec{M}^{-1})^\top$ sowie $\langle \vec{J}^* \vec{v}_h, \vec{J}^* \vec{w}_h \rangle_h = (\vec{J} \vec{J}^* \vec{v}_h, \vec{w}_h)_{L^2}$ folgt

$$\begin{aligned}& \left\langle \left[\vec{M} (\vec{L}_h)^{-1} \vec{M} - (\vec{L}^{-1})_h \right] \vec{\mathbf{v}}, \vec{\mathbf{w}} \right\rangle_h \\ &= \left\langle \left[\vec{M} (\vec{L}_h)^{-1} \vec{M} - (\vec{L}^{-1})_h \right] \vec{M}^{-1} \vec{J}^* \vec{v}_h, \vec{M}^{-1} \vec{J}^* \vec{w}_h \right\rangle_h \\ &= \left\langle \vec{M}^{-1} \left[\vec{M} (\vec{L}_h)^{-1} \vec{M} - (\vec{L}^{-1})_h \right] \vec{M}^{-1} \vec{J}^* \vec{v}_h, \vec{J}^* \vec{w}_h \right\rangle_h \\ &= \left(\vec{J} \left[(\vec{L}_h)^{-1} - \vec{M}^{-1} \cdot (\vec{L}^{-1})_h \cdot \vec{M}^{-1} \right] \vec{J}^* \vec{v}_h, \vec{w}_h \right)_{L^2(\Omega, \mathbb{R}^q)} \\ &= \left(\left[\vec{J} (\vec{L}_h)^{-1} \vec{J}^* - \vec{J} \vec{M}^{-1} \cdot (\vec{L}^{-1})_h \cdot \vec{M}^{-1} \vec{J}^* \right] \vec{v}_h, \vec{w}_h \right)_{L^2(\Omega, \mathbb{R}^q)} \\ &\stackrel{(3.27)}{=} \left(\left[\vec{R}_h \vec{L}^{-1} - \vec{J} \vec{M}^{-1} \cdot (\vec{J}^* \vec{L}^{-1} \vec{J}) \cdot \vec{M}^{-1} \vec{J}^* \right] \vec{v}_h, \vec{w}_h \right)_{L^2(\Omega, \mathbb{R}^q)} \\ &\stackrel{(3.25)}{=} \left(\left[\vec{R}_h \vec{L}^{-1} - \vec{Q}_h \vec{L}^{-1} \vec{Q}_h \right] \vec{v}_h, \vec{w}_h \right)_{L^2(\Omega, \mathbb{R}^q)} \\ &\stackrel{\vec{Q}_h \vec{v}_h = \vec{v}_h}{=} \left(\left[\vec{R}_h - Id + Id - \vec{Q}_h \right] \vec{L}^{-1} \vec{v}_h, \vec{w}_h \right)_{L^2(\Omega, \mathbb{R}^q)} \\ &\leq \left[\left\| (Id - \vec{R}_h) \vec{L}^{-1} \vec{v}_h \right\|_{L^2(\Omega, \mathbb{R}^q)} + \left\| (Id - \vec{Q}_h) \vec{L}^{-1} \vec{v}_h \right\|_{L^2(\Omega, \mathbb{R}^q)} \right] \|\vec{w}_h\|_{L^2(\Omega, \mathbb{R}^q)} \\ &\stackrel{(3.29)}{\leq} 2 \cdot \left\| (Id - \vec{R}_h) \vec{L}^{-1} \vec{v}_h \right\|_{L^2(\Omega, \mathbb{R}^q)} \cdot \|\vec{w}_h\|_{L^2(\Omega, \mathbb{R}^q)} \\ &\stackrel{\text{Lemma 3.10}}{\leq} 2 \cdot c_{FEM} h^{2\alpha} \cdot \|\vec{v}_h\|_{L^2(\Omega, \mathbb{R}^q)} \cdot \|\vec{w}_h\|_{L^2(\Omega, \mathbb{R}^q)} \\ &= 2c_{FEM} h^{2\alpha} \cdot \left\| \vec{J}\vec{\mathbf{v}} \right\|_{L^2(\Omega, \mathbb{R}^q)} \cdot \left\| \vec{J}\vec{\mathbf{w}} \right\|_{L^2(\Omega, \mathbb{R}^q)} \\ &\leq 2c_{FEM} h^{2\alpha} \cdot \mu_{max} \cdot \|\vec{\mathbf{v}}\|_h \cdot \|\vec{\mathbf{w}}\|_h.\end{aligned}$$

Damit erhalten wir

$$\left\| \vec{M} (\vec{L}_h)^{-1} \vec{M} - (\vec{L}^{-1})_h \right\|_2 \leq \sup_{\vec{\mathbf{v}}, \vec{\mathbf{w}}} \frac{2c_{FEM} h^{2\alpha} \mu_{max} \|\vec{\mathbf{v}}\|_h \|\vec{\mathbf{w}}\|_h}{\|\vec{\mathbf{v}}\|_h \cdot \|\vec{\mathbf{w}}\|_h} = 2c_{FEM} h^{2\alpha} \mu_{max}$$

und mit (5.54) folgt dann die Behauptung. \square

5.4.3 \mathcal{H} -Approximation der Hilfsmatrix

Wir hatten die Hilfsmatrix \vec{C}_h erklärt zu $\vec{M}^{-1} \left(\vec{L}^{-1} \right)_h \vec{M}^{-1}$, sodaß für ihre \mathcal{H} -Approximation die Ersetzung aller vorkommender Matrizen durch die jeweilige \mathcal{H} -Approximierende naheliegt.

Lemma 5.25 (\mathcal{H} -Approximation $\vec{C}_{\mathcal{H}}$ der Hilfsmatrix) *Das exakte Produkt*

$$\vec{C}_{\mathcal{H}} := \left(\vec{M}^{-1} \right)_{\mathcal{H}} \cdot \left(\vec{L}^{-1} \right)_{\mathcal{H}} \cdot \left(\vec{M}^{-1} \right)_{\mathcal{H}} \quad (5.55a)$$

ist eine \mathcal{H} -Matrix mit Blockrang

$$k = (1 + p_T)^2 \cdot C_{sp}^2 \cdot \max \left\{ qk(\varepsilon_M), k_{\vec{L}^{-1}}, b_{min} \right\}, \quad (5.55b)$$

also $\vec{C}_{\mathcal{H}} \in \mathcal{H}(k, T, Z)$, und approximiert die Hilfsmatrix \vec{C}_h mit dem Fehler

$$\left\| \vec{C}_h - \vec{C}_{\mathcal{H}} \right\|_2 \leq err_{C_h, \mathcal{H}}, \quad (5.55c)$$

wobei

$$err_{C_h, \mathcal{H}} = \frac{\varepsilon_M}{c_E} \cdot \frac{\mu_{max}}{\mu_{min}^2} + \frac{1 + \varepsilon_M}{\mu_{min}^2} \cdot err_{L^{-1}, \mathcal{H}} + \frac{\varepsilon_M + \varepsilon_M^2}{\mu_{min}^2} \left(err_{L^{-1}, \mathcal{H}} + \frac{\mu_{max}}{c_E} \right). \quad (5.55d)$$

Beweis: Wir rechnen nach, daß $\vec{C}_{\mathcal{H}}$ die gesuchte Matrix ist. Mittels phantasievollem Ergänzen, sodaß Differenzen von normalen und \mathcal{H} -Matrizen entstehen, ist

$$\begin{aligned} \vec{C}_h - \vec{C}_{\mathcal{H}} &= \vec{M}^{-1} \left(\vec{L}^{-1} \right)_h \vec{M}^{-1} - \left(\vec{M}^{-1} \right)_{\mathcal{H}} \left(\vec{L}^{-1} \right)_{\mathcal{H}} \left(\vec{M}^{-1} \right)_{\mathcal{H}} \\ &= \vec{M}^{-1} \left(\vec{L}^{-1} \right)_h \vec{M}^{-1} \\ &\quad - \left(\vec{M}^{-1} \right)_{\mathcal{H}} \left(\vec{L}^{-1} \right)_h \vec{M}^{-1} + \left(\vec{M}^{-1} \right)_{\mathcal{H}} \left(\vec{L}^{-1} \right)_h \vec{M}^{-1} \\ &\quad - \left(\vec{M}^{-1} \right)_{\mathcal{H}} \left(\vec{L}^{-1} \right)_{\mathcal{H}} \vec{M}^{-1} + \left(\vec{M}^{-1} \right)_{\mathcal{H}} \left(\vec{L}^{-1} \right)_{\mathcal{H}} \vec{M}^{-1} \\ &\quad - \left(\vec{M}^{-1} \right)_{\mathcal{H}} \left(\vec{L}^{-1} \right)_{\mathcal{H}} \left(\vec{M}^{-1} \right)_{\mathcal{H}} \\ &= S_1 + S_2 + S_3 \end{aligned}$$

mit den Summanden

$$\begin{aligned} S_1 &:= \left(\vec{M}^{-1} - \left(\vec{M}^{-1} \right)_{\mathcal{H}} \right) \left(\vec{L}^{-1} \right)_h \vec{M}^{-1} \\ S_2 &:= \left(\vec{M}^{-1} \right)_{\mathcal{H}} \left(\left(\vec{L}^{-1} \right)_h - \left(\vec{L}^{-1} \right)_{\mathcal{H}} \right) \vec{M}^{-1} \\ S_3 &:= \left(\vec{M}^{-1} \right)_{\mathcal{H}} \left(\vec{L}^{-1} \right)_{\mathcal{H}} \left(\vec{M}^{-1} - \left(\vec{M}^{-1} \right)_{\mathcal{H}} \right) \end{aligned}$$

Das Korollar 5.22 und das Lemma 5.23c verhelfen uns zu den Abschätzungen

$$\begin{aligned}
\|S_1\|_2 &\leq \left\| \vec{M}^{-1} - (\vec{M}^{-1})_{\mathcal{H}} \right\|_2 \left\| (\vec{L}^{-1})_h \right\|_2 \left\| \vec{M}^{-1} \right\|_2 \\
&\stackrel{(5.51b)}{\leq} \frac{\varepsilon_M}{\mu_{min}} \left\| (\vec{L}^{-1})_h \right\|_2 \frac{1}{\mu_{min}} \\
&\stackrel{(5.48)}{\leq} \frac{\varepsilon_M}{c_E} \cdot \frac{\mu_{max}}{\mu_{min}^2}, \\
\|S_2\|_2 &\leq \left\| (\vec{M}^{-1})_{\mathcal{H}} \right\|_2 \left\| (\vec{L}^{-1})_h - (\vec{L}^{-1})_{\mathcal{H}} \right\|_2 \left\| \vec{M}^{-1} \right\|_2 \\
&\stackrel{(5.51d)}{\leq} \frac{1 + \varepsilon_M}{\mu_{min}} \left\| (\vec{L}^{-1})_h - (\vec{L}^{-1})_{\mathcal{H}} \right\|_2 \frac{1}{\mu_{min}} \\
&\stackrel{\text{Kor. 5.22}}{\leq} \frac{1 + \varepsilon_M}{\mu_{min}^2} \text{err}_{L^{-1}, \mathcal{H}}, \\
\|S_3\|_2 &\leq \left\| (\vec{M}^{-1})_{\mathcal{H}} \right\|_2 \left\| (\vec{L}^{-1})_{\mathcal{H}} \right\|_2 \left\| \vec{M}^{-1} - (\vec{M}^{-1})_{\mathcal{H}} \right\|_2 \\
&\stackrel{(5.51d)}{\leq} \frac{1 + \varepsilon_M}{\mu_{min}} \left\| (\vec{L}^{-1})_{\mathcal{H}} \right\|_2 \left\| \vec{M}^{-1} - (\vec{M}^{-1})_{\mathcal{H}} \right\|_2 \\
&\stackrel{\text{Kor. 5.22}}{\leq} \frac{1 + \varepsilon_M}{\mu_{min}} \left(\text{err}_{L^{-1}, \mathcal{H}} + \frac{\mu_{max}}{c_E} \right) \left\| \vec{M}^{-1} - (\vec{M}^{-1})_{\mathcal{H}} \right\|_2 \\
&\stackrel{(5.51b)}{\leq} \frac{1 + \varepsilon_M}{\mu_{min}} \left(\text{err}_{L^{-1}, \mathcal{H}} + \frac{\mu_{max}}{c_E} \right) \frac{\varepsilon_M}{\mu_{min}} \\
&= \frac{\varepsilon_M + \varepsilon_M^2}{\mu_{min}^2} \left(\text{err}_{L^{-1}, \mathcal{H}} + \frac{\mu_{max}}{c_E} \right)
\end{aligned}$$

und somit

$$\begin{aligned}
\left\| \vec{C}_h - \vec{C}_{\mathcal{H}} \right\|_2 &\leq \|S_1\|_2 + \|S_2\|_2 + \|S_3\|_2 \\
&\leq \frac{\varepsilon_M}{c_E} \cdot \frac{\mu_{max}}{\mu_{min}^2} + \frac{1 + \varepsilon_M}{\mu_{min}^2} \cdot \text{err}_{L^{-1}, \mathcal{H}} + \frac{\varepsilon_M + \varepsilon_M^2}{\mu_{min}^2} \left(\text{err}_{L^{-1}, \mathcal{H}} + \frac{\mu_{max}}{c_E} \right) \\
&=: \text{err}_{C_h, \mathcal{H}}.
\end{aligned}$$

Wir müssen noch den Blockrang von $\vec{C}_{\mathcal{H}}$ bestimmen. Dafür ist es sinnvoll, $\vec{C}_{\mathcal{H}}$ als exaktes Produkt von \mathcal{H} -Matrizen zu berechnen, denn durch formatierte Multiplikation können wir ja jeden beliebigen Rang erzwingen. Nach Lemma 5.23c und Korollar 5.22 gelten

$$\left(\vec{M}^{-1} \right)_{\mathcal{H}} \in \mathcal{H}(qk(\varepsilon_M), T, Z), \quad \left(\vec{L}^{-1} \right)_{\mathcal{H}} \in \mathcal{H}(k_{\vec{L}^{-1}}, T, Z),$$

also folgt mit Satz 4.15 zuerst

$$\left(\vec{M}^{-1} \right)_{\mathcal{H}} \cdot \left(\vec{L}^{-1} \right)_{\mathcal{H}} \in \mathcal{H}(k_1, T, Z)$$

für

$$k_1 = (1 + p_T) \cdot C_{sp} \cdot \max \left\{ qk(\varepsilon_M), k_{\vec{L}^{-1}}, b_{min} \right\}$$

und dann

$$\vec{C}_{\mathcal{H}} \in \mathcal{H}(k, T, Z),$$

wobei

$$\begin{aligned} k &= (1 + p_T) \cdot C_{sp} \max \{ qk(\varepsilon_M), k_1, b_{min} \} \\ &= (1 + p_T) \cdot C_{sp} \max \left\{ \begin{array}{l} qk(\varepsilon_M), \\ (1 + p_T) \cdot C_{sp} \max \{ qk(\varepsilon_M), k_{\vec{L}_{\mathcal{H}}}^{-1}, b_{min} \}, \\ b_{min} \end{array} \right\} \\ &\stackrel{(1+p_T)C_{sp} \geq 1}{=} (1 + p_T) \cdot C_{sp} \cdot \max \left\{ (1 + p_T) \cdot C_{sp} \max \{ qk(\varepsilon_M), k_{\vec{L}_{\mathcal{H}}}^{-1}, b_{min} \} \right\} \\ &= (1 + p_T)^2 \cdot C_{sp}^2 \cdot \max \left\{ qk(\varepsilon_M), k_{\vec{L}_{\mathcal{H}}}^{-1}, b_{min} \right\} \end{aligned}$$

der maximale Blockrang ist. \square

5.4.4 Gesamtfehler und Rangbestimmung

Wir erreichen unser Ziel, nämlich die Existenz einer \mathcal{H} -Inversen für die inverse Steifigkeitsmatrix von elliptischen Differentialgleichungssystemen zu gewährleisten sowie den Approximationsfehler zu kontrollieren, und erhalten den wohlverdienten Lohn in Form von

Satz 5.26 (\mathcal{H} -Approximation der inversen Steifigkeitsmatrix) *Sei T ein bzgl. der Standardzulässigkeitsbedingung gebildeter \mathcal{H} -Produktbaum von $I \times I$. Sei k die Rangverteilung aus (5.55b).*

Dann kann die Inverse der Steifigkeitsmatrix durch die \mathcal{H} -Matrix

$$\vec{C}_{\mathcal{H}} \in \mathcal{H}(k, T, Z)$$

approximiert werden mit dem Fehler

$$\left\| \left(\vec{L}_h \right)^{-1} - \vec{C}_{\mathcal{H}} \right\|_2 \leq \varepsilon := 2c_{FEM} h^{2\alpha} \frac{\mu_{max}}{\mu_{min}^2} + err_{C_h, \mathcal{H}}.$$

Beweis: Der Beweis ist vollkommen unspektakulär. Den Blockrang von $\vec{C}_{\mathcal{H}}$ haben wir bereits im letzten Lemma bewiesen. Für die Abschätzung des Approximationsfehlers ergänzen wir die Hilfsmatrix \vec{C}_h . Mit der Dreiecksungleichung folgt dann

$$\left\| \left(\vec{L}_h \right)^{-1} - \vec{C}_{\mathcal{H}} \right\|_2 \leq \left\| \left(\vec{L}_h \right)^{-1} - \vec{C}_h \right\|_2 + \left\| \vec{C}_h - \vec{C}_{\mathcal{H}} \right\|_2.$$

Die beiden Summanden wurden bereits in den Lemmata 5.24 und 5.25 abgeschätzt.

\square

Kapitel 6

Numerische Tests

Zum Abschluß der Arbeit wollen wir die theoretisch erlangten Aussagen durch numerische Experimente untermauern. Hierzu geben wir den Blockrang qk vor, berechnen die \mathcal{H} -Inverse und beobachten den Invertierungsfehler $\|A \cdot A^{-\mathcal{H}} - Id\|_2$.

Wir stellen fest, daß die Lösung von linearen Gleichungssystemen wie sie bei der Finiten-Element-Methode entstehen, durch die \mathcal{H} -Invertierung nicht der Weisheit letzter Schluß sein kann, und beenden die Arbeit mit einigen Praktikabilitätsversuchen, in denen wir als Löser ein mit einem rangadaptiven \mathcal{H} -CHOLESKY-Faktor vorkonditioniertes CG-Verfahren verwenden. Zum einen sind Abschätzungen hierfür verwandt mit denjenigen für die \mathcal{H} -Inverse, zum anderen ermöglicht dieses Lösungsverfahren Kostenersparnisse in erheblichem Umfang, sodaß wir in den Bereich der praktisch relevanten Verfahren vorstoßen. Zuerst setzen wir uns mit den Algorithmen und den Testaufgaben auseinander.

6.1 Algorithmen

Zur Lösung der LAMÉ-Gleichung mit linearen Finiten Elementen mußte Software im Umfang von ca. 3000 Zeilen entwickelt werden. Soweit möglich, wurde auf die von STEFFEN BÖRM und LARS GRASEDYCK entwickelte HLib (www.hmatrix.org) zurückgegriffen, um die \mathcal{H} -Arithmetik nicht implementieren zu müssen. Jedoch sind einige Modifikationen vor allem im Bereich der Verwaltung von Gitterpunkten bzw. Freiheitsgraden und der Routinen zur Cluster-Erzeugung erforderlich gewesen, damit auch Systeme von Differentialgleichungen behandelt werden können.

Wir müssen zuerst ein Gitter bereitstellen und die \mathcal{H} -Strukturen anlegen.

Algorithmus 6.1 (Grundverfahren)

- (a) *Gitter einlesen/erzeugen*
- (b) *Cluster-Baum für PDE-Systeme erzeugen*
`buildcluster_grid2d_system`

- (c) *Assemblierung der Steifigkeitsmatrix im Sparse-Matrix-Format*
`lame_system_grid2d`
und der rechten Seite
- `rhs_lame_system_grid2d` *bei konstanten Koeffizienten*
 - `rhs_lame_system_jump_coeff_grid2d` *bei springenden Koeffizienten*
- (d) *Aufbau der \mathcal{H} -Matrix-Struktur*
`build_fem_supermatrix_from_cluster`
- (e) *Konvertierung der Steifigkeitsmatrix vom Sparse- ins \mathcal{H} -Matrix-Format*
- `convertsparse3sym_supermatrix` *für symmetrienutzende Löser*
 - `convertsparse3_supermatrix` *sonst*
- (f) *Lösen des linearen Gleichungssystems gemäß der Algorithmen 6.2 oder 6.3*

Die Lösung des linearen Gleichungssystems kann auf verschiedenen Wegen berechnet werden.

Algorithmus 6.2 (Lösen durch \mathcal{H} -Invertierung)

→ `solveEqsByHInversion`

- (a) *Berechne \mathcal{H} -Inverse der Steifigkeitsmatrix*
- (b) *Berechne Lösung des linearen Gleichungssystems durch Matrix-Vektor-Multiplikation von \mathcal{H} -Inverser mit dem Lastvektor.*

Wir werden im Laufe dieses Kapitels sehen, daß die Approximation der \mathcal{H} -Inversen mit einer zum Diskretisierungsfehler passenden Genauigkeit sehr viel Speicher und Rechenzeit erfordert. Deshalb interessieren wir uns für Modifikationen, die zu sparsameren Verfahren führen. Hier bieten sich verschiedene Möglichkeiten an. Zunächst stellen wir auf rangadaptive Rechnung um. Für diese kann man zwar nichts beweisen, aber letztlich ist uns eine gute Approximation lieber als die genaue Kontrolle bzw. Kenntnis der Rangverteilung. Dann rücken aber auch neue Lösungsverfahren wie die CHOLESKY-Faktorisierung oder das CG-Verfahren, die sich für symmetrisch-positiv-definite Matrizen geradezu aufzwingen, in unser Blickfeld.

Algorithmus 6.3 (Lösen durch $\mathbf{P}_{\mathcal{H}\text{-Inverse}}$ CG-Verfahren)

→ `solveEqsByCGMwithInvPCsparse`

- (a) *Berechne (Niedrigrang-) \mathcal{H} -Inverse der Steifigkeitsmatrix*
- (b) *Starte das CG-Verfahren mit dem eben berechneten Vorkonditionierer*

Algorithmus 6.4 (Lösen durch $\mathbf{P}_{\mathcal{H}\text{-CHOLESKY}}$ CG-Verfahren)

→ `solveEqsByCGMwithCholPCsparse`

Verfahren	Operationen	Speicher	Bemerkungen
MVM	n^2	n^2	
Gauß/LU	n^3	n^2	
FFT	$n \log n$		
Cholesky	$n^3/2$	$n^2/2$	
LU f. Bandmatrix	nm^2		Bandbreite $2m + 1$
Cholesky f. Bandm.	$nm^2/2$		Bandbreite $2m + 1$
CG-Verfahren	n		
Mehrgitter	n		
\mathcal{H} -Matrix		$C_{sp}kn \log n$	[GH02, Lemma 2.4]
\mathcal{H} -MVM	$C_{sp}kn \log n$		[GH02, Lemma 2.5]
\mathcal{H} -MMM	$C_{sp}^3 k^2 n (\log n)^2$		[GH02, Th. 2.20]
\mathcal{H} -Invertierung	$C_{sp}^3 k^2 n (\log n)^2$	3 \mathcal{H} -Matrizen	[GH02, Th. 2.29]
\mathcal{H} -Cholesky	$C_{sp}^3 k^2 n (\log n)^2$	$C_{sp}kn \log n/2$	[Be05], [GKL05]

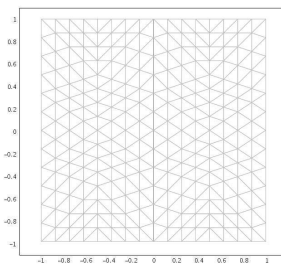
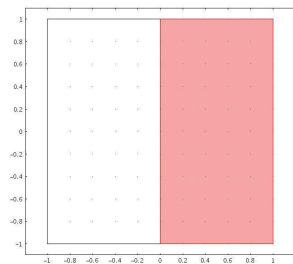
Tabelle 6.1: Gegenüberstellung des Rechen- und Speicheraufwandes verschiedener Verfahren zur Lösung eines linearen Gleichungssystems $Ax = b$ mit $A \in \mathbb{R}^{n \times n}$ und $b, x \in \mathbb{R}^n$. Alle Angaben sind im $\mathcal{O}(\cdot)$ -Sinne zu verstehen. Ob rangfest oder -adaptiv gerechnet wird, wirkt sich in den (Fehler- und) Kostenabschätzungen nicht aus, weil wir bei rangadaptiver Rechnung nicht einmal Mutmaßungen über die Rangverteilung anstellen können, also vom schlechtesten Fall ausgehen müssen und der wird bereits durch die Überlegungen bzgl. der rangfesten Arithmetik abgedeckt.

- (a) Berechne (Niedrigrang-) \mathcal{H} -CHOLESKY-Faktor der Steifigkeitsmatrix
- (b) Starte das CG-Verfahren mit dem eben berechneten Vorkonditionierer

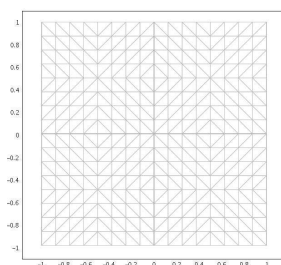
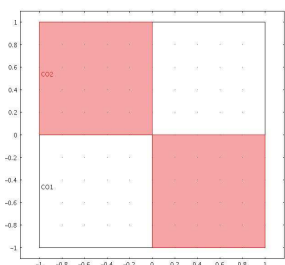
Abschätzungen für die \mathcal{H} -CHOLESKY-Zerlegung oder die Vorkonditionierung des CG-Verfahrens mit einem \mathcal{H} -CHOLESKY-Faktor basieren auf den Abschätzungen für die \mathcal{H} -Inverse, s. [Be05], [GKL05], [Hb06]. Tabelle 6.1 gibt Auskunft über die Kosten der verschiedenen Verfahren.

6.2 Die Testprobleme

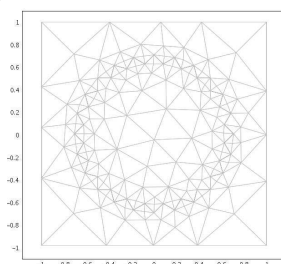
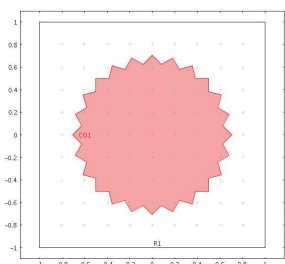
Eine Übersicht der für die vorliegende Arbeit gewählten Gebiete (im weiteren auch Geometrien genannt) und Gitter, die mit Femlab bzw. Comsol erzeugt wurden, zeigt Abbildung 6.1.



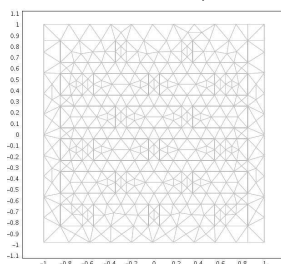
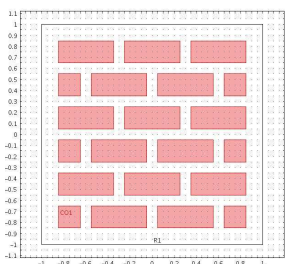
Geometrie Typ 2



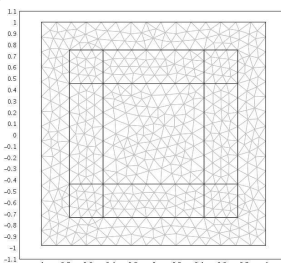
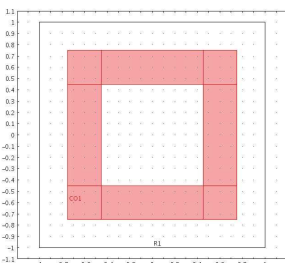
Geometrie Typ 3



Geometrie Typ 6 (15-Grad-Verschneidung)



Geometrie Typ 7



Geometrie Typ 8

Abbildung 6.1: Übersicht der gerechneten Gebiete und Gitter.

Gerechnet wird ausschließlich die zweidimensionale LAMÉ-Gleichung (siehe Abschnitt 2.4.2 bzw. Beispiel 2.16) mit DIRICHLET-Nullrandwerten. Für unsere Testrechnungen begnügen wir uns mit der polynomialen Lösung

$$\vec{u}(x, y) = \begin{pmatrix} (x^2 - 1)^2 \cdot (y^2 - 1)^2 \\ (x^2 - 1)^2 \cdot (y^2 - 1)^2 \end{pmatrix}.$$

Weil die Lösung immer glatt ist, wirkt sich das Springen der Koeffizienten λ und μ auf die rechte Seite aus. Wir setzen

$$\lambda(x) := 1, \quad \mu(x) := \begin{cases} 1 & \text{für } x \in \text{weiße Teilgebiete,} \\ 10^3 & \text{für } x \in \text{rote Teilgebiete,} \end{cases} \quad (6.1a)$$

gelegentlich **Parametertyp 3** genannt, und

$$\lambda(x) := 1, \quad \mu(x) := \begin{cases} 10^{-3} & \text{für } x \in \text{weiße Teilgebiete,} \\ 10^3 & \text{für } x \in \text{rote Teilgebiete} \end{cases} \quad (6.1b)$$

(**Parametertyp 4**).

6.3 Ergebnisevaluierung

In diesem Abschnitt fahren wir hauptsächlich Tests zur Überprüfung unserer theoretisch gewonnenen Resultate, d.h. wir geben den Blockrang k bzw. qk vor (also rangfeste Arithmetik), berechnen die inverse \mathcal{H} -Matrix und beobachten den/die auftretenden Fehler.

In den folgenden vier Unterabschnitten stellen wir die Ergebnisse für die Geometrien 2 und 3 sowie die Parametertypen 3 und 4 vor. Die Struktur ist jeweils dieselbe. Als erstes widmen wir uns dem **Invertierungsfehler**

$$\|A \cdot A^{-\mathcal{H}} - Id\|_2,$$

wobei A die zu invertierende Matrix und $A^{-\mathcal{H}}$ ihre berechnete \mathcal{H} -Inverse bezeichnet. Unseren theoretischen Ergebnissen zufolge muß dieser Fehler mit zunehmendem Blockrang k bzw. $qk = 2, 4, 8, 16, \dots, 64, 72, \dots, 128$ kleiner werden. Praktisch interessant sind der Zeitbedarf und der Gesamtspeicher. Beim **Zeitbedarf** beschränken wir uns auf eine graphische Darstellung des Aufwandes für die Invertierung und die Matrix-Vektor-Multiplikation, weil alle anderen Schritte im Vergleich hierzu kaum Zeit beanspruchen. Als **Gesamtspeicher** geben wir den tatsächlich benötigten Arbeitsspeicher an: der Invertierungsalgorithmus verlangt Speicherplatz für drei Matrizen und den wollen wir nicht unterschlagen.

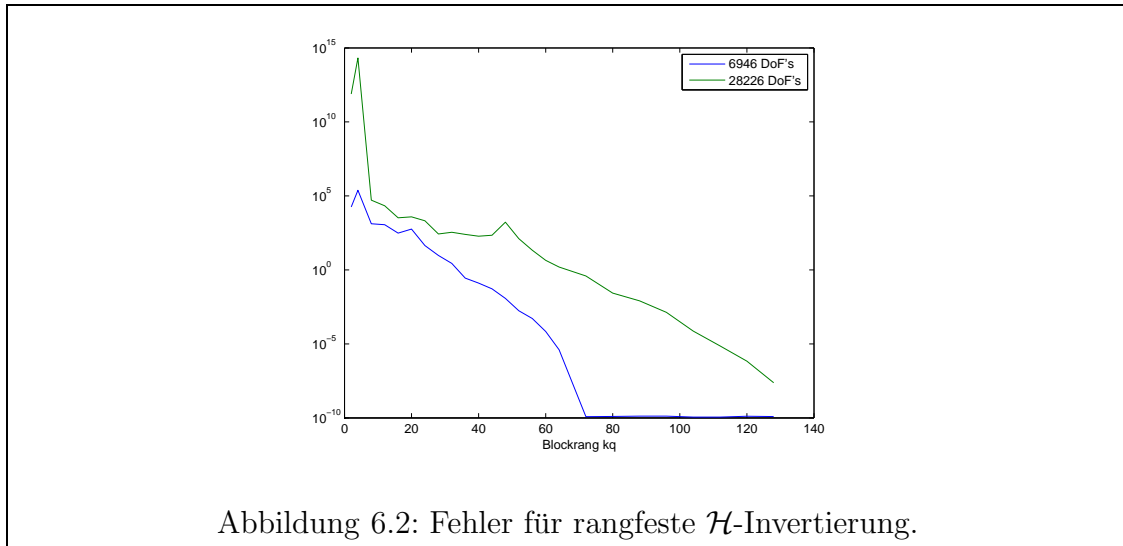
Die Meßdaten stellen wir hier des Leseflusses willen nur graphisch dar; die Meßwerte selbst sind im Anhang B.2 tabellarisiert.

Den Abschluß dieses Abschnittes bilden ein vorläufiges Resümee und einige kleinere Experimente mit rangadaptiver Berechnung der \mathcal{H} -Inversen.

6.3.1 Geometrie 2 und Parametertyp 3

Invertierungsfehler

Vgl. Tabelle B.1.



Die Invertierungsfehlerkurve ist für niedrige Blockränge oszillierend, was bedeutet, daß das Verfahren noch nicht justiert ist. Sehen wir uns $\|u - u_h\|_{L^2}$ an, stellen wir fest, daß dieser Fehler erst dann nicht mehr kleiner wird, wenn der Invertierungsfehler sich in der Größenordnung von eins bewegt, d.h. für 6946 Unbekannte benötigen wir $kq \geq 32$ (also kq fünf Tausendstel der Unbekanntenzahl oder kleiner), für 28226 Unbekannte $kq \geq 64$ (also kq ca. 2.5 Tausendstel der Unbekanntenzahl oder kleiner). Bei einer Halbierung der Schrittweite sollte also der Blockrang verdoppelt werden.

Zeitbedarf

Vgl. Tabelle B.2.

Unabhängig von der Größe des Blockranges wird der Clusterbaum für 7938 Freiheitsgrade in rund einer Hundertstel Sekunde aufgebaut, für 32258 Freiheitsgrade sind drei bis fünf Hundertstel Sekunden erforderlich.

Der Aufbau der \mathcal{H} -Matrix-Struktur hingegen ist abhängig vom Blockrang. Bei 7938 Freiheitsgraden nimmt er kontinuierlich von einer Zehntel Sekunde bis fünf Zehntel Sekunden zu. Bei 32258 Freiheitsgraden wächst der Zeitbedarf von vier Zehntel auf drei Sekunden an.

Die Konvertierung der dünnen FE-Matrix in eine \mathcal{H} -Matrix ist wiederum unabhängig vom Blockrang und beträgt zwei bzw. zehn Zehntel Sekunden.

Den Hauptteil nimmt die Berechnung der \mathcal{H} -Inversen in Anspruch. Die Matrix-Vektor-Multiplikation zur Lösung des linearen Gleichungssystems benötigt ca. ein Promill dieser Zeit.

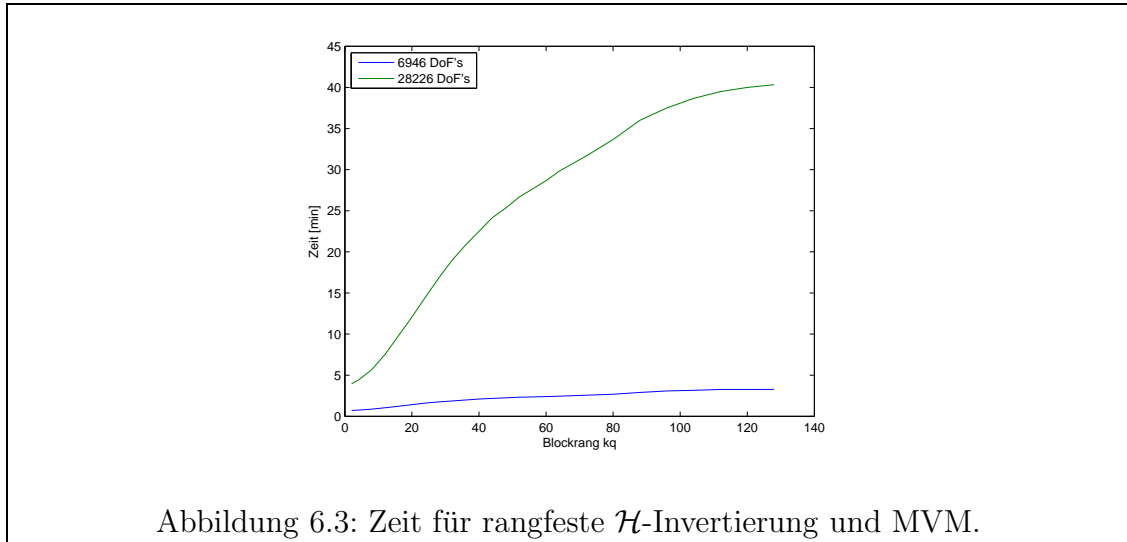


Abbildung 6.3: Zeit für rangfeste \mathcal{H} -Invertierung und MVM.

Speicherverbrauch

Vgl. Tabelle B.3.

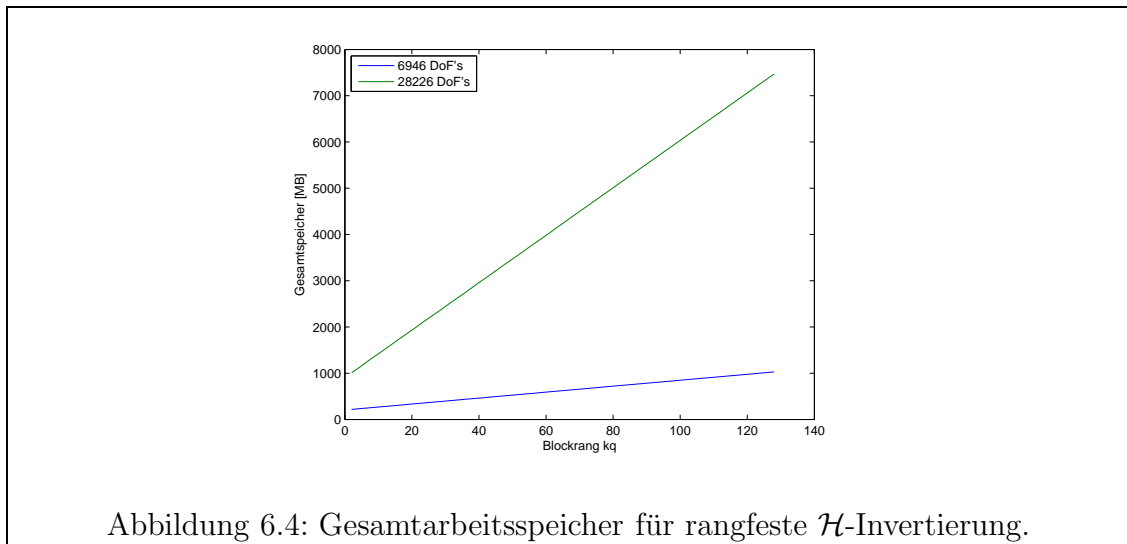
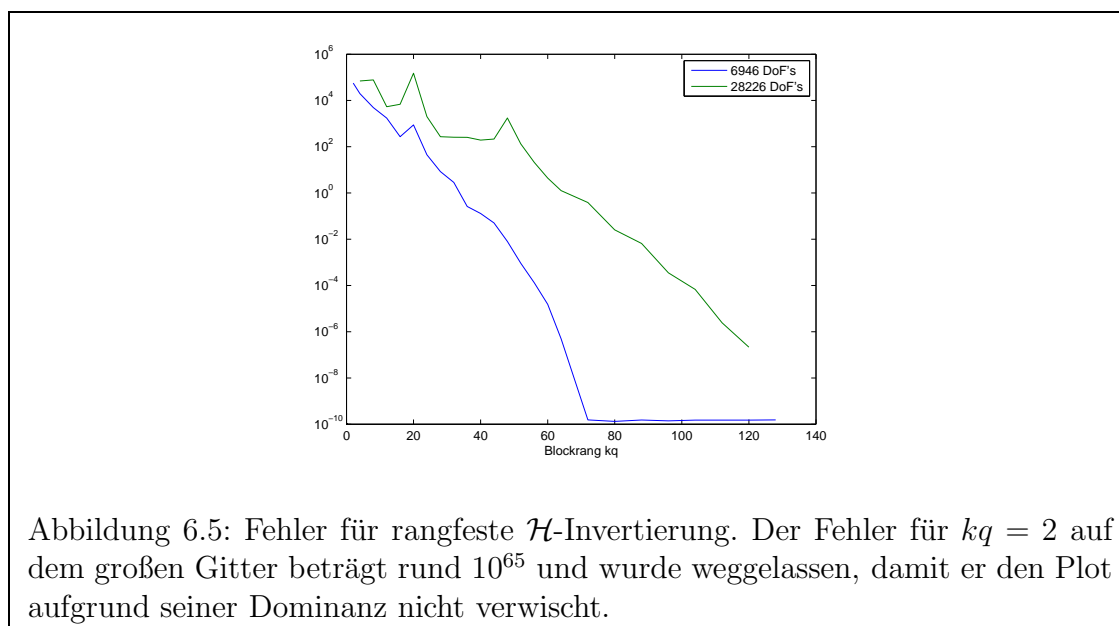


Abbildung 6.4: Gesamtarbeitsspeicher für rangfeste \mathcal{H} -Invertierung.

6.3.2 Geometrie 2 und Parametertyp 4

Invertierungsfehler

Vgl. Tabelle B.4.



Auch hier beobachten wir, daß der Invertierungsfehler mit wachsendem Blockrang fällt. Daß μ nun um den Faktor 10^6 springt, wirkt sich auf den Blockrang kaum aus: für $kq \geq 36$ wird der L^2 -Fehler nicht mehr kleiner, d.h. für ca. fünf Tausendstel der Unbekanntenzahl erreichen wir eine gute \mathcal{H} -Approximation der Inversen; ab diesem Blockrang fällt auch der Invertierungsfehler streng monoton. Entsprechend gelten die Aussagen für 28226 Unbekannte.

Zeitbedarf

Vgl. Tabelle B.5.

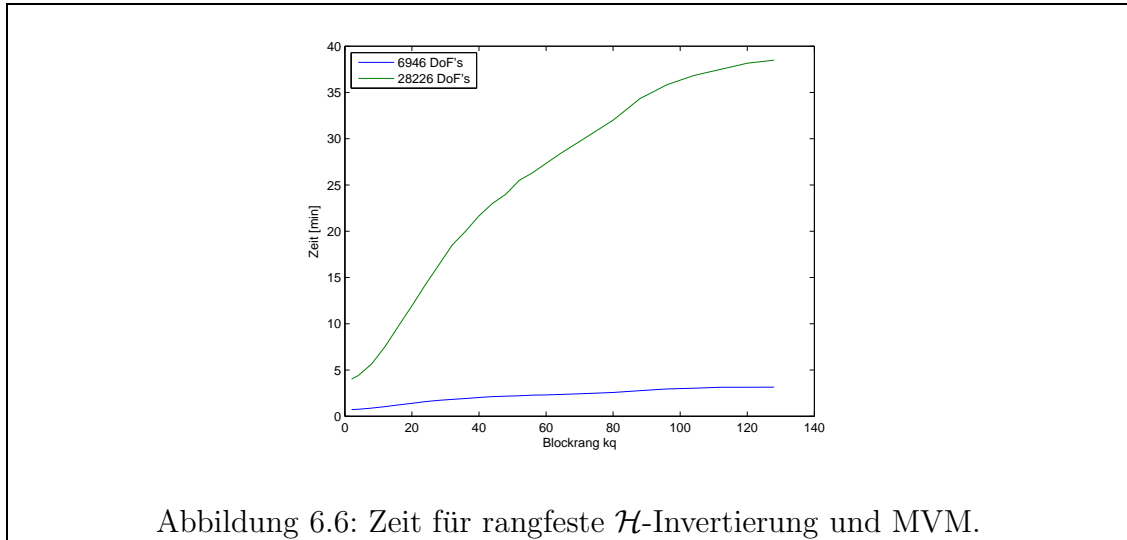


Abbildung 6.6: Zeit für rangfeste \mathcal{H} -Invertierung und MVM.

Speicherverbrauch

Vgl. Tabelle B.6.

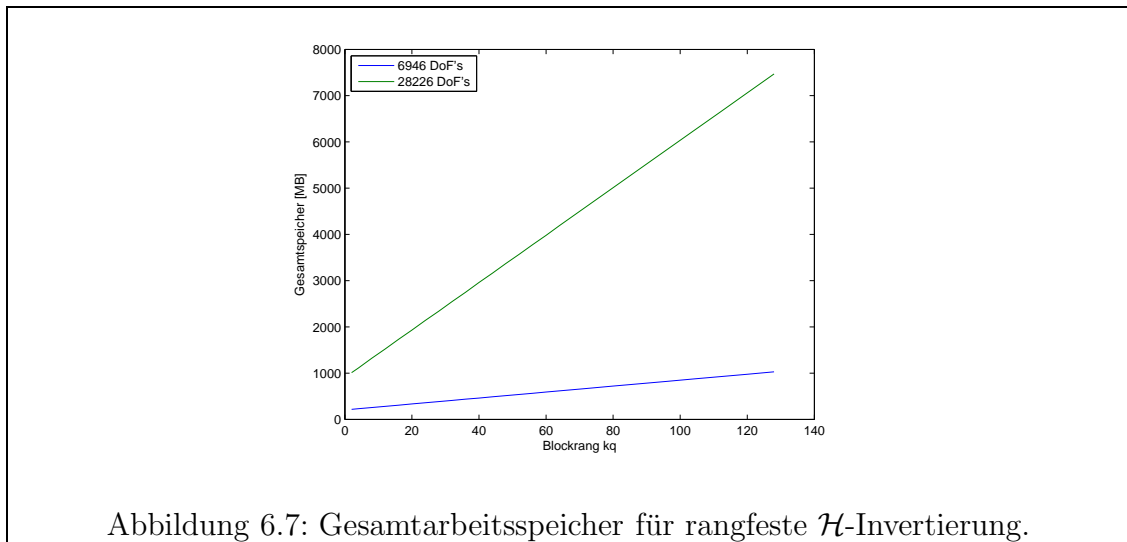
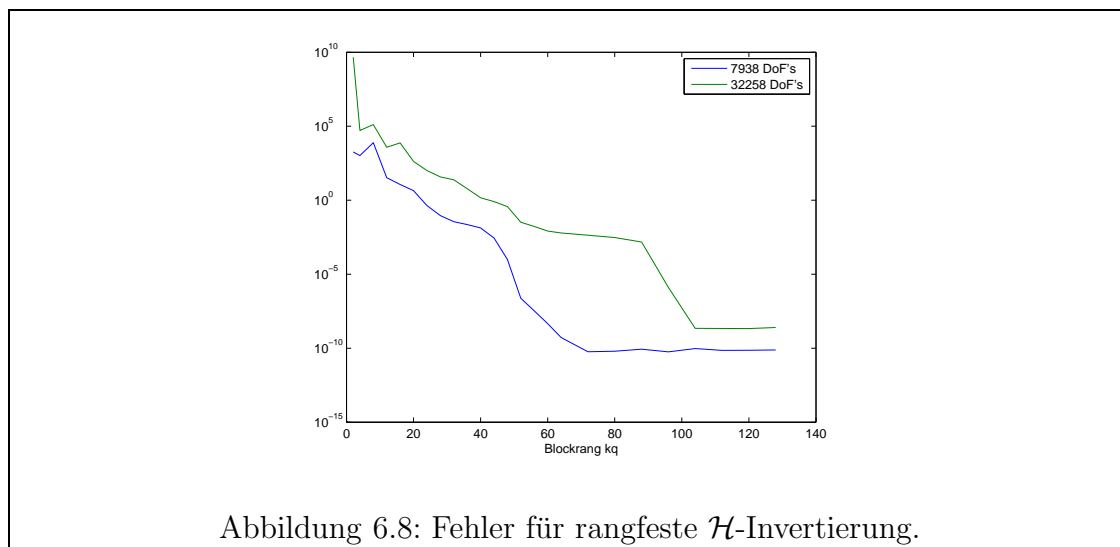


Abbildung 6.7: Gesamtarbeitsspeicher für rangfeste \mathcal{H} -Invertierung.

6.3.3 Geometrie 3 und Parametertyp 3

Invertierungsfehler

Vgl. Tabelle B.7.



Ab Blockrang $kq = 24$ (7938 Unbekannte) bzw. 48 (32258 Unbekannte) fällt der Invertierungsfehler strikt. Für diese Werte nimmt der L^2 -Fehler auch den kleinsten gemessenen Wert an. Wiederum gelten: kq ist ca. drei Tausendstel der Unbekanntenzahl und bei Halbierung der Schrittweite muß kq verdoppelt werden.

Zeitbedarf

Vgl. Tabelle B.8.

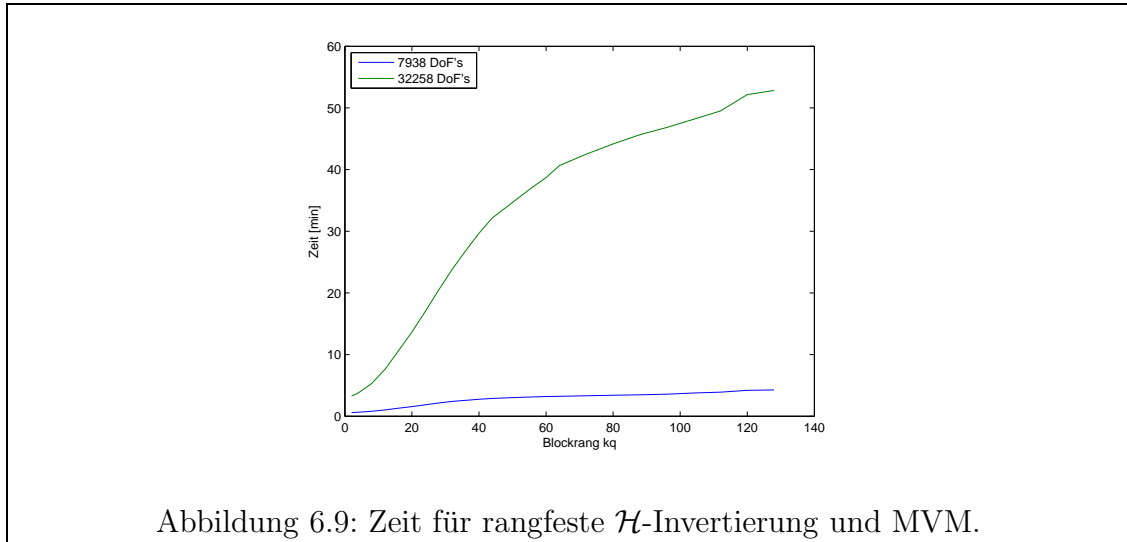


Abbildung 6.9: Zeit für rangfeste \mathcal{H} -Invertierung und MVM.

Speicherverbrauch

Vgl. Tabelle B.9.

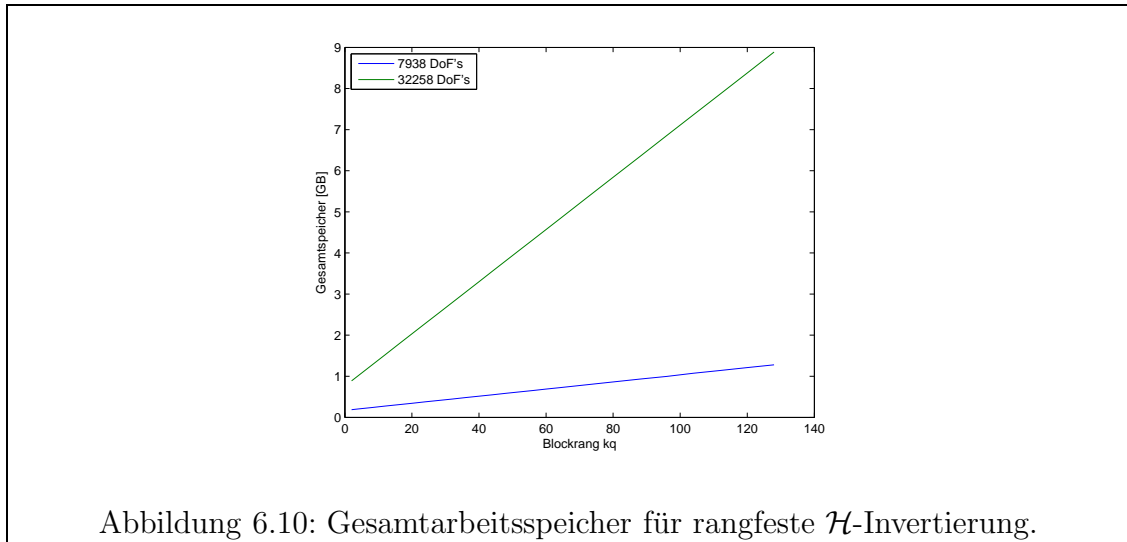
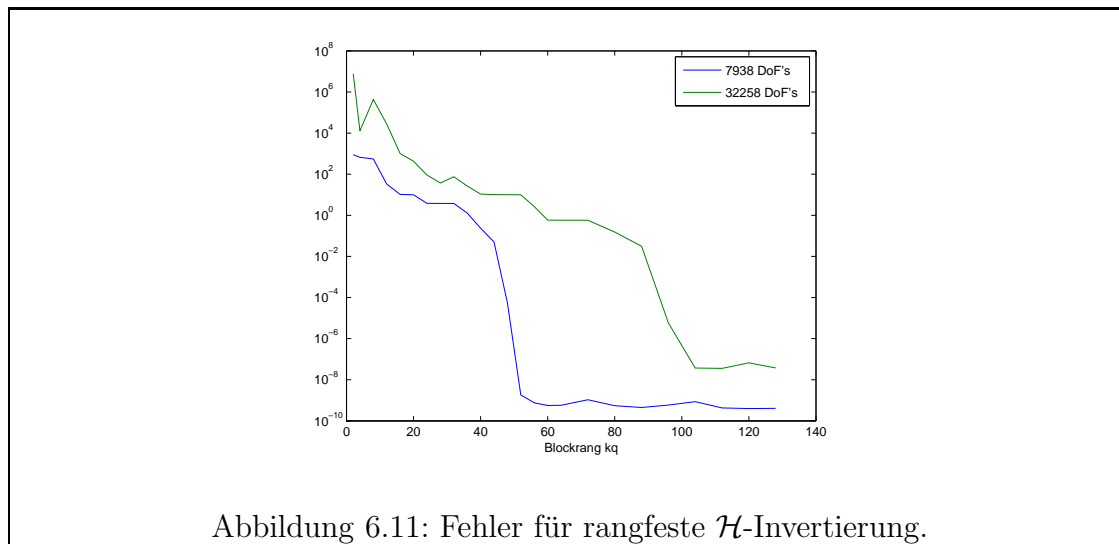


Abbildung 6.10: Gesamtarbeitsspeicher für rangfeste \mathcal{H} -Invertierung.

6.3.4 Geometrie 3 und Parametertyp 4

Invertierungsfehler

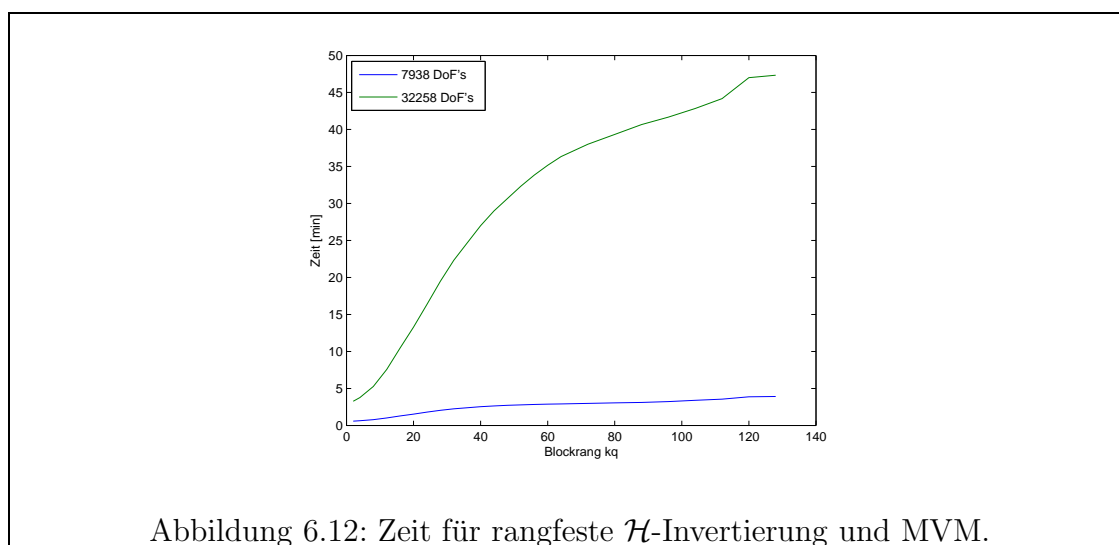
Vgl. Tabelle B.10.



Die Aussagen von 6.3.3 können übernommen werden, das Springen des Koeffizienten μ um den Faktor 10^6 statt 10^3 wirkt sich nicht aus. Ebenso wenig scheint sich die Änderung der Geometrie auszuwirken.

Zeitbedarf

Vgl. Tabelle B.11.



Speicherverbrauch

Vgl. Tabelle B.12.

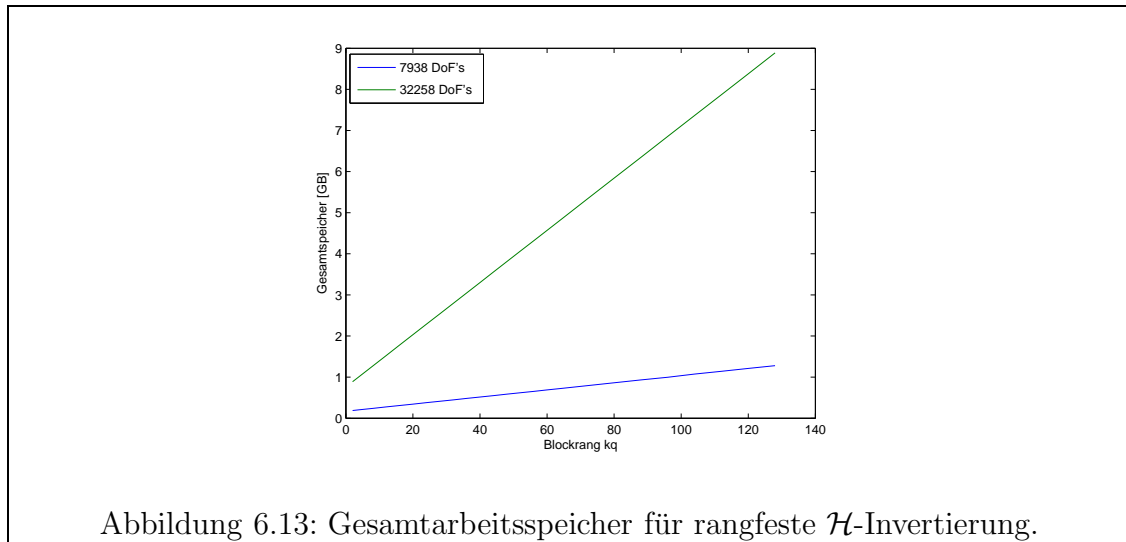


Abbildung 6.13: Gesamtarbeitsspeicher für rangfeste \mathcal{H} -Invertierung.

6.3.5 Rangadaptive \mathcal{H} -Invertierung

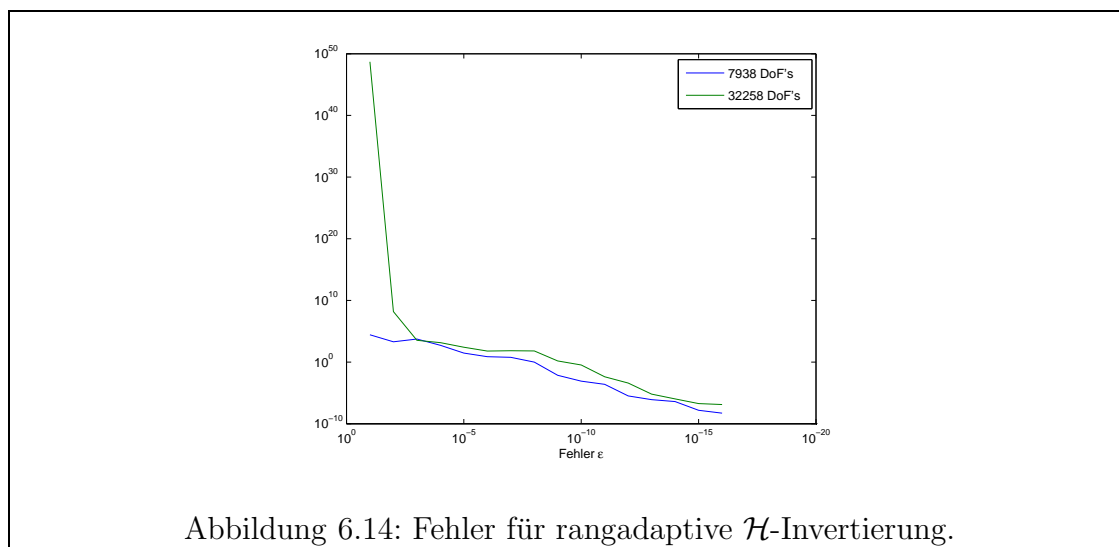
Bei rangfester Arithmetik haben wir keinen Anhaltspunkt für die Grösse des Approximationsfehlers. Da uns dieser aber wichtiger ist als die genaue Steuerung des Ranges, geben wir im folgenden den (blockweisen) Invertierungsfehler

$$\varepsilon = 10^{-1}, 10^{-2}, \dots, 10^{-16}$$

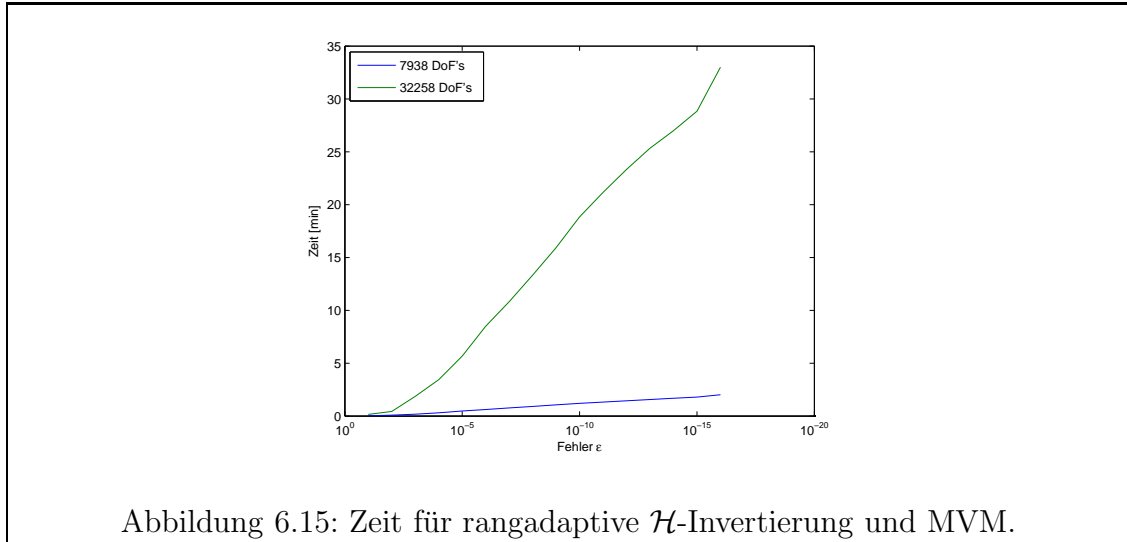
vor. Weil sich die \mathcal{H} -Matrizen bei der Behandlung von stark springenden Koeffizienten bewährt haben, testen wir gleich nur den Parametertyp 4, s. (6.1b), auf/über der Geometrie 3.

Die Ergebnisse, aus denen die Grafiken generiert wurden, stehen in den Tabellen B.13 bis B.16.

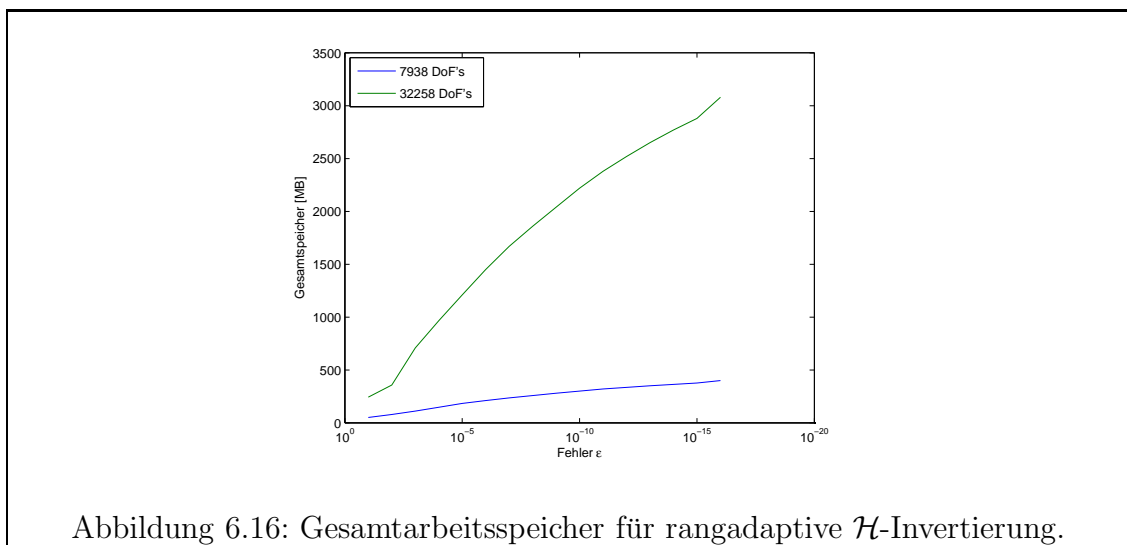
Invertierungsfehler



Zeitbedarf



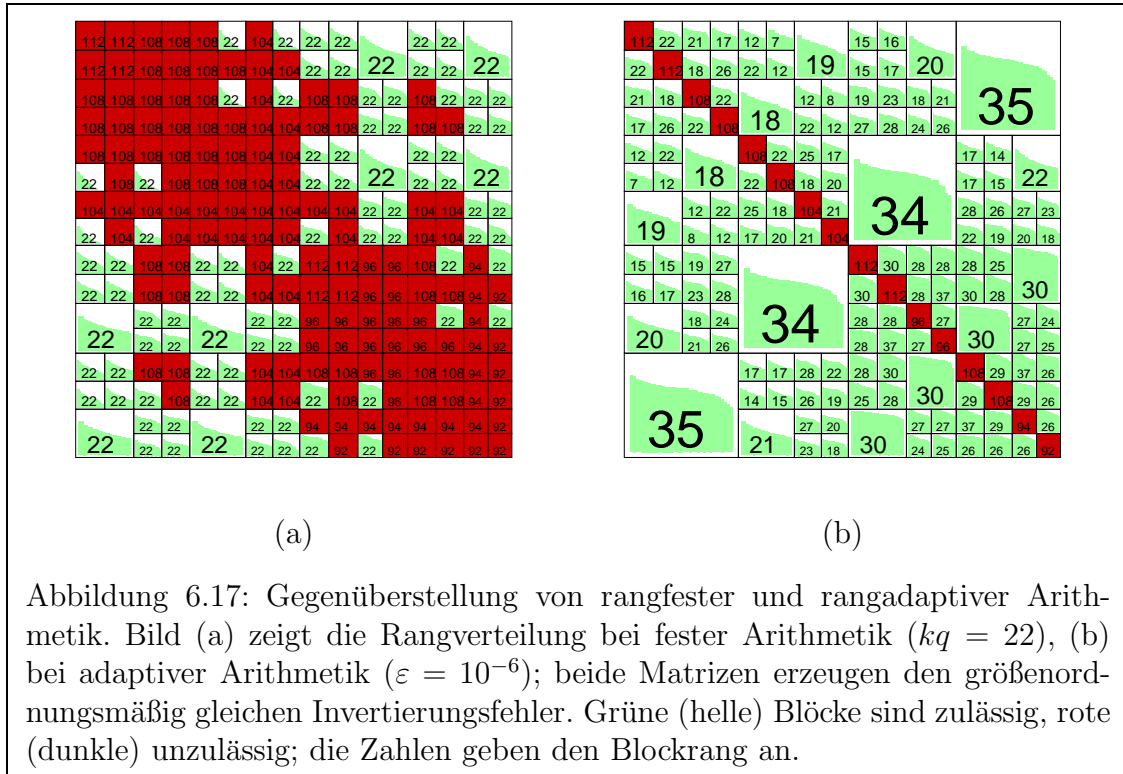
Speicherverbrauch



Vergleich von rangfester mit -adaptiver Arithmetik

Um ein bzgl. des Invertierungsfehlers gleichwertiges Ergebnis zu produzieren, benötigt die rangadaptiv berechnete \mathcal{H} -Inverse

- ca. 2/9 des Speichers
- und ca. 1/3 der Rechenzeit.



Wir resümieren, daß rangadaptive Arithmetik der rangfesten überlegen ist.

Bei rangadaptiver Arithmetik werden erheblich mehr Blöcke für zulässig befunden als bei rangfester Arithmetik (s. Abbildung 6.17).

6.4 Praktikabilitätsversuche

Wie die Versuche mit rangadaptiver Arithmetik gezeigt haben, kann die Speicherbedarfsituation entschärft werden, zufriedenstellend ist sie jedoch immer noch nicht. Im Kontext von FE-Matrizen ist das CG-Verfahren ein weit verbreiteter Löser, für den eine ganze Palette von Vorkonditionierern existiert, u.a. auch ein mit rangadaptiver Arithmetik berechneter CHOLESKY-Faktor. Dieses Verfahren nehmen wir nun unter die Lupe.

Wir rechnen wieder die LAMÉ-Gleichung mit dem Parametersatz (6.1b) und den Geometrien 3 (zu Vergleichszwecken), 6, 7 und 8. In den Grafiken (und den im Anhang zu findenden Tabellen) sind nur Messungen für $\varepsilon \leq 10^{-4}$ dargestellt: für größere Werte ist die Approximation so ungenau, daß das CG-Verfahren sie nicht mehr als Vorkonditionierer verwenden kann und abbricht.

6.4.1 $P_{\mathcal{H}\text{-Cholesky}}\text{CG}$: adaptiver Rang, Geometrie 3

S. Tabellen B.17 und B.18

Invertierungsfehler

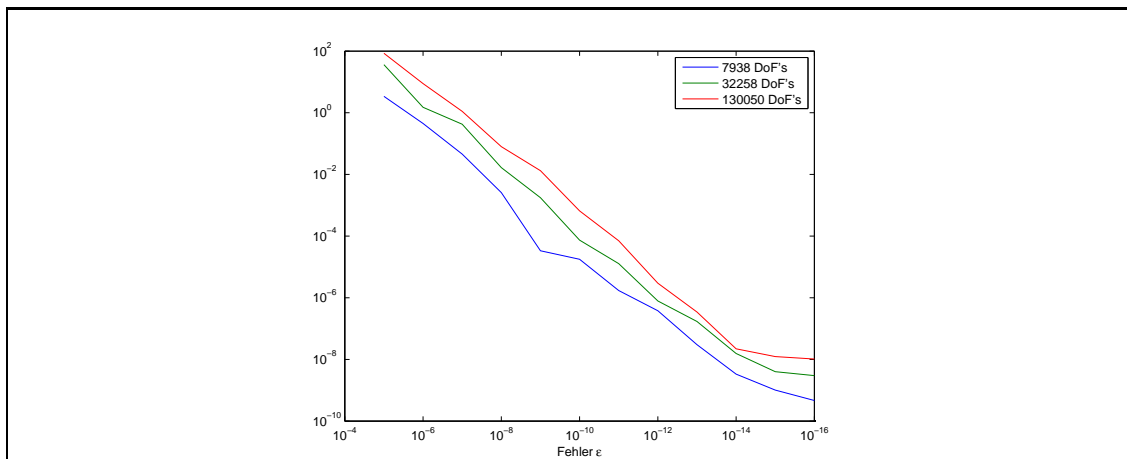


Abbildung 6.18: Fehler für rangadaptive \mathcal{H} -Cholesky-Faktorisierung und PCG-Verfahren.

Zeitbedarf

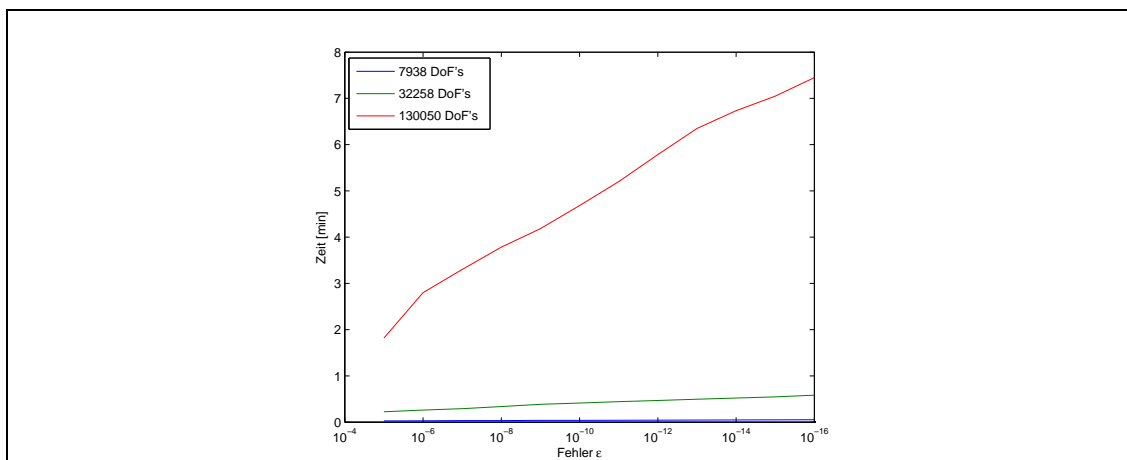
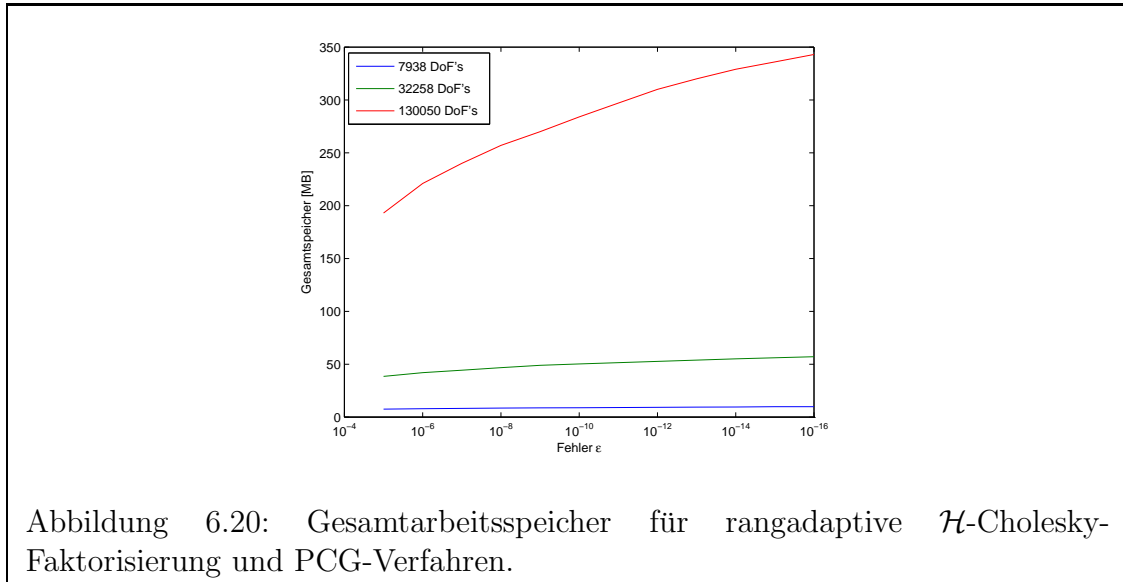


Abbildung 6.19: Zeit für rangadaptive \mathcal{H} -Cholesky-Faktorisierung und PCG-Verfahren.

Speicherverbrauch



6.4.2 $P_{\mathcal{H}\text{-Cholesky}}\text{CG}$: adaptiver Rang, Geometrie 6

S. Tabellen B.19 und B.20

Invertierungsfehler

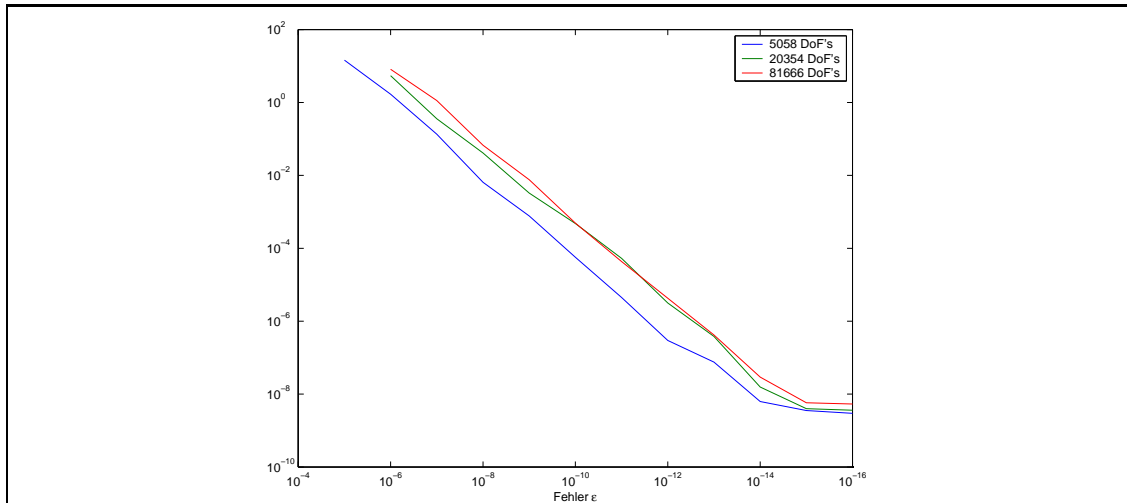


Abbildung 6.21: Fehler für rangadaptive \mathcal{H} -Cholesky-Faktorisierung und PCG-Verfahren.

Zeitbedarf

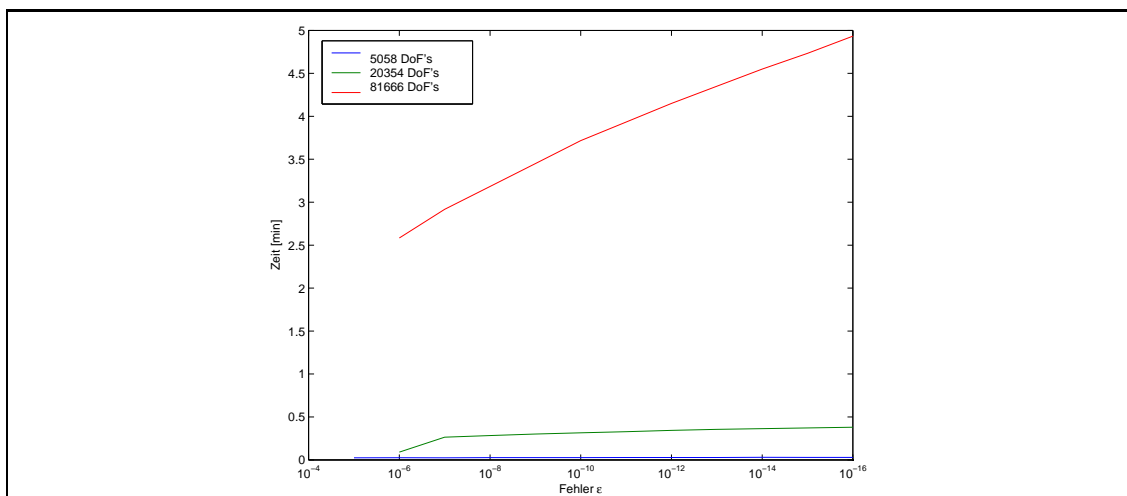
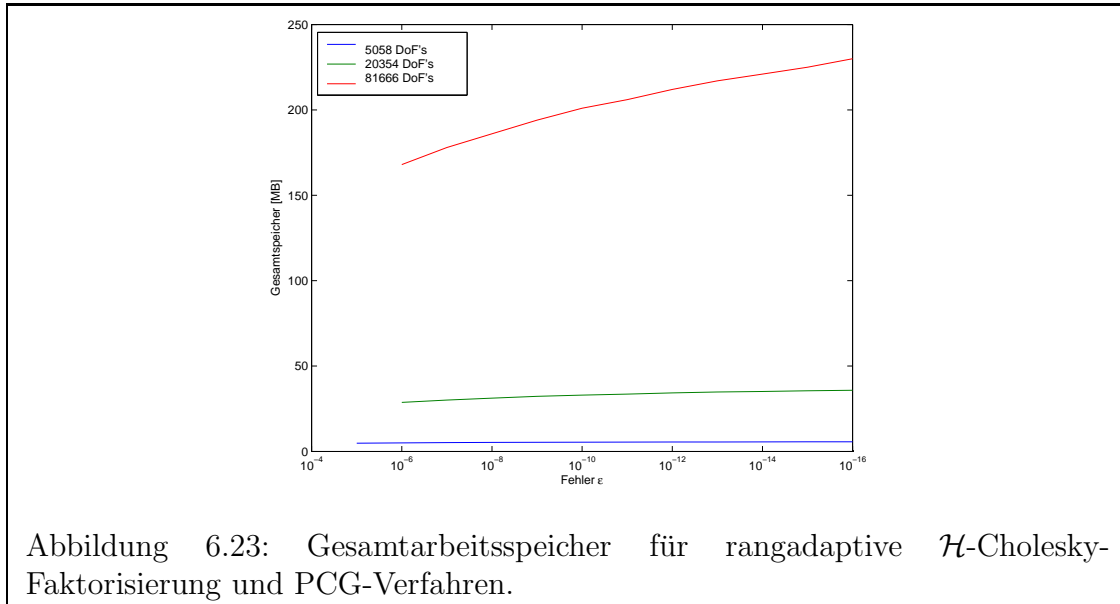


Abbildung 6.22: Zeit für rangadaptive \mathcal{H} -Cholesky-Faktorisierung und PCG-Verfahren.

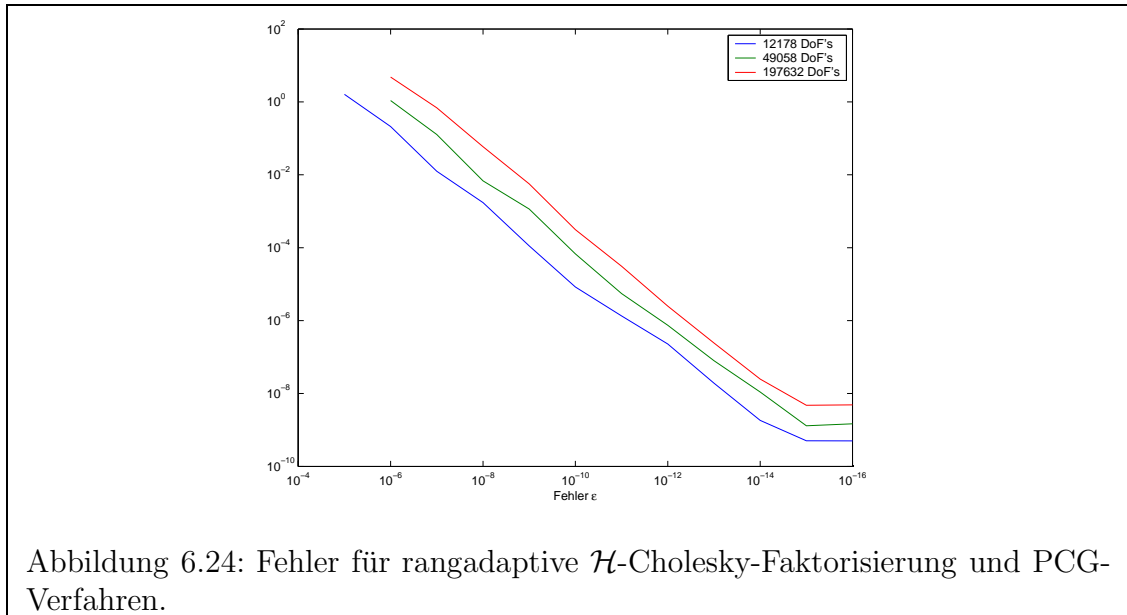
Speicherverbrauch



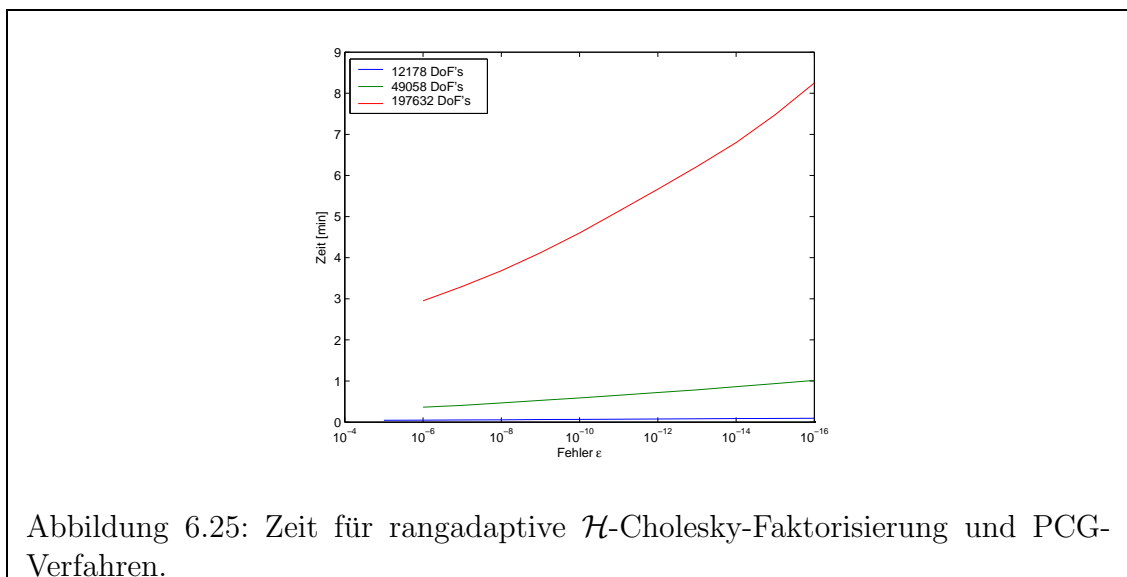
6.4.3 $P_{\mathcal{H}\text{-Cholesky}}\text{CG}$: adaptiver Rang, Geometrie 7

S. Tabellen B.21 und B.22

Invertierungsfehler



Zeitbedarf



Speicherverbrauch

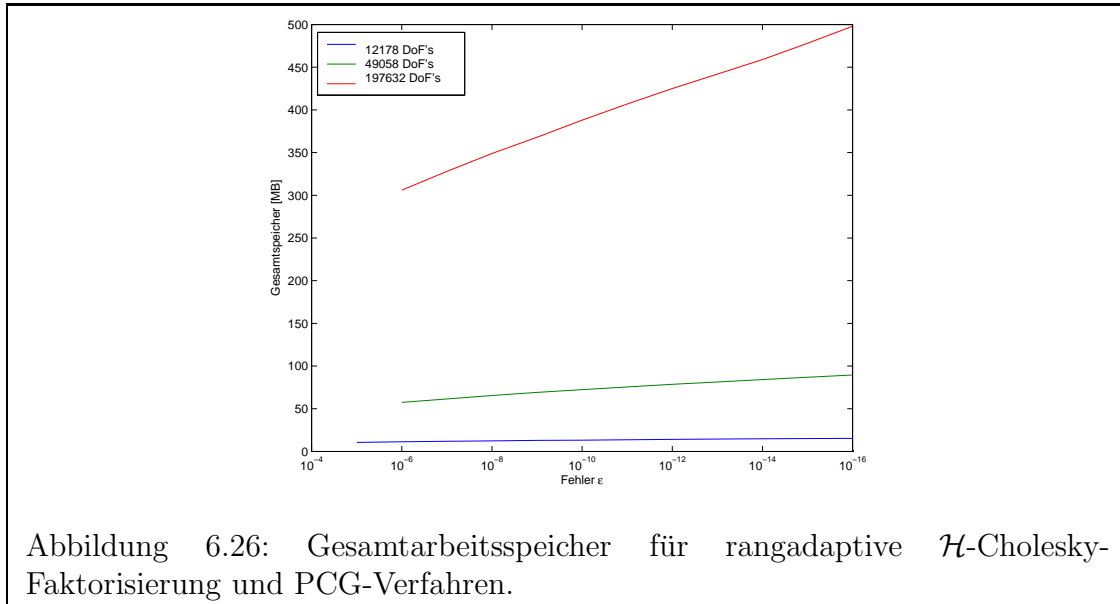
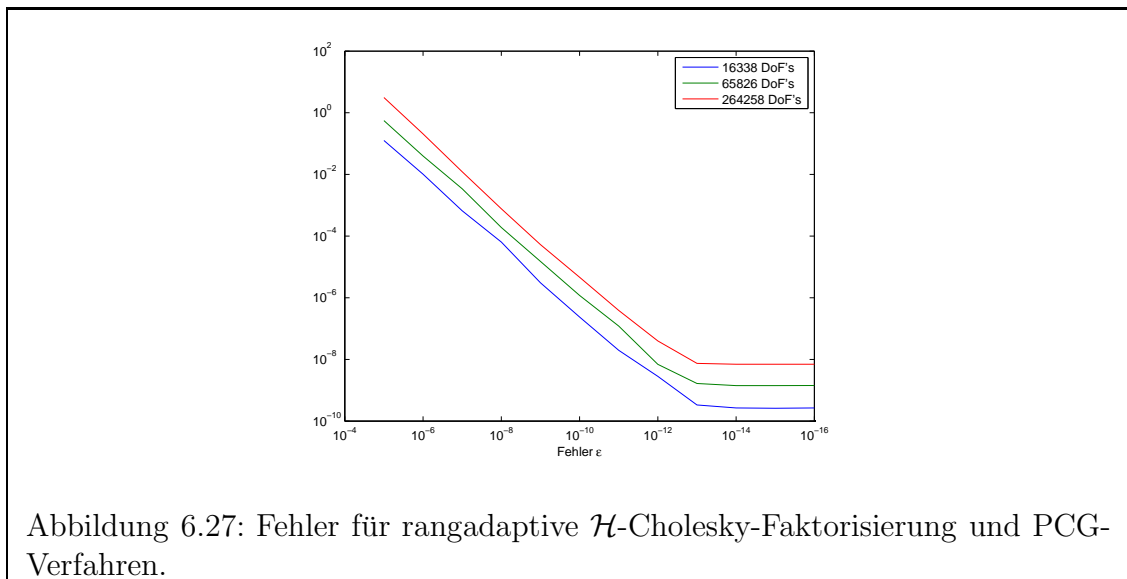


Abbildung 6.26: Gesamtarbeitsspeicher für rangadaptive \mathcal{H} -Cholesky-Faktorisierung und PCG-Verfahren.

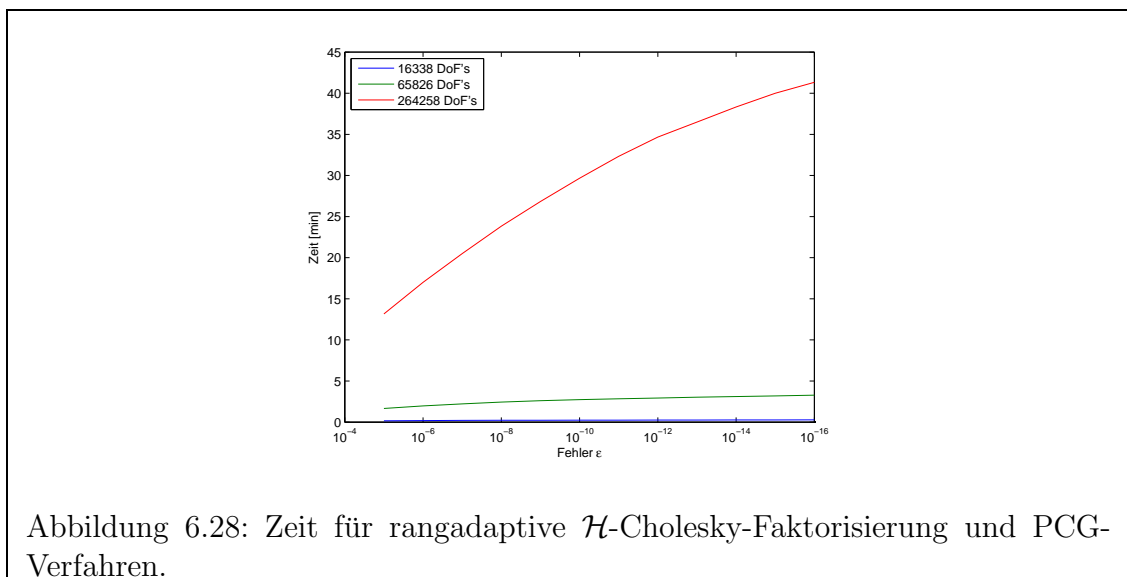
6.4.4 $P_{\mathcal{H}\text{-Cholesky}}\text{CG}$: adaptiver Rang, Geometrie 8

S. Tabellen B.23 und B.24

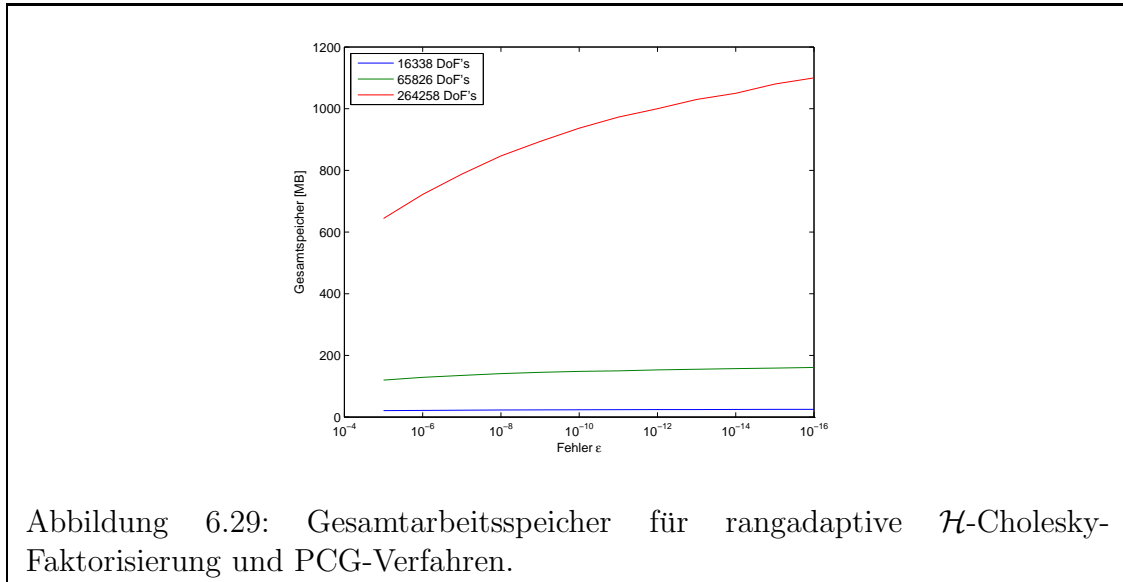
Invertierungsfehler



Zeitbedarf



Speicherverbrauch



6.5 Resümee

Zur Ergebnisevaluierung

Das Anliegen der Ergebnisevaluierung war, die in Kapitel 5 gewonnenen Theorieausagen durch numerische Tests zu stützen. Dazu mußten wir die \mathcal{H} -Approximation der Inversen einer FE-Matrix zu vorgegebenem Blockrang k bzw. kq berechnen und den Invertierungsfehler protokollieren. Die Auswertung hat das vorhergesagte und gewünschte Resultat geliefert: je höher wir den Blockrang schrauben, je genauer wir also rechnen, desto kleiner wird unser Fehler.

Interessant ist die Beobachtung, daß der Blockrang kq in der Größenordnung von einem Tausendstel der Anzahl der Freiheitsgrade sein sollte, um gute Ergebnisse zu erzielen. Das Beispiel der Geometrie 3 in Verbindung mit dem Parametertyp 4, der quasi bedingt, daß die beiden Quadrate $] -1, 0]^2$ und $] 0, 1]^2$ im wesentlichen nur über den Ursprung des Koordinatensystems miteinander verbunden sind, legt nahe, daß die \mathcal{H} -Matrizen gegenüber der Geometrie recht unempfindlich sind.

Die \mathcal{H} -Inverse als Gleichungssystemlöser hat leider zwei Haken. Für ein System mit 30000 Unbekannten müssen wir rund 50 Minuten veranschlagen. Hinzukommen kommen noch Vorarbeiten wie Aufbau des Cluster-Baumes etc., die in der Summe aber nur ein Hunderstel der Invertierungszeit benötigen.

Wegen des hohen Speicherverbrauchs konnten wir leider keine Gitter der nächstfeineren Stufe, also mit rund 120000 Unbekannten, vernünftig rechnen: ab Blockrang $kq = 60$ werden i.a. mehr als 16 GB Arbeitsspeicher benötigt, sodaß das System mit Auslagerungen in Swap-Dateien beginnt. Zeitmessungen sind hier nicht mehr sinnvoll.

Zu den Praktikabilitätsversuchen

Damit wir die Arbeit nicht mit dem Fazit *\mathcal{H} -Matrizen kommen mit stark springenden Koeffizienten und wildesten Geometrien zurecht, brauchen aber ewig viel Zeit* schließen müssen, demonstrieren wir ihre Leistungsfähigkeit anhand eines PCG-Verfahrens.

Zunächst stellen wir auf rangadaptive Arithmetik um und sparen dabei jeweils ein Drittel Rechenzeit und Arbeitsspeicher. Angesichts der ursprünglich hohen Kosten der \mathcal{H} -Invertierung scheint aber eine rangadaptiv berechnete \mathcal{H} -Inverse als Vorkonditionierer für das CG-Verfahren immer noch wenig attraktiv zu sein. Statt ihrer berechnen wir also gleich den \mathcal{H} -CHOLESKY-Faktor, der uns dann auch die ersehnten schönen Ergebnisse liefert.

Wählen wir Geometrie 3 mit Parameter 4 als Ausgangsbasis für einen Vergleich, benötigen wir mit dem PCG-Verfahren gegenüber der rangfesten \mathcal{H} -Inversen nur ca. ein Fünfzehntel des Arbeitsspeichers und ein Siebtel der Rechenzeit, um einen gleichwertigen Approximationsfehler zu erzeugen. Sind wir hingegen nur an einer guten Approximation der Lösung (und nicht der Matrix) interessiert, kommen wir

mit jeweils einem Zwanzigstel der Kosten aus.

Gesamtbewertung

Die Methode der \mathcal{H} -Matrizen eignet sich also hervorragend, um \mathcal{H} -CHOLESKY-Faktoren als Vorkonditionierer für das CG-Verfahren zur Lösung von linearen Gleichungssystemen, wie sie bei der Finiten-Element-Methode entstehen, zu berechnen.

Anhang A

Mathematische Ergänzungen und Nachträge

A.1 Matrixnormen

In diesem Abschnitt sammeln wir gezielt das für diese Arbeit relevante Wissen um Matrizen, Eigenwerte und Matrixnormen.

Wir verwenden zwei Vektornormen, nämlich die EUKLIDISCHE Norm

$$\|\cdot\|_2 : \mathbb{R}^n \longrightarrow \mathbb{R}, \quad x \longmapsto \sqrt{\sum_{i=1}^n x_i^2}$$

und die Maximums- oder Unendlichnorm

$$\|\cdot\|_\infty : \mathbb{R}^n \longrightarrow \mathbb{R}, \quad x \longmapsto \max \{|x_1|, \dots, |x_n|\}.$$

Jeder Vektornorm kann eine Matrixnorm per

$$\|\cdot\| : \mathbb{R}^{m \times n} \longrightarrow \mathbb{R}, \quad A \longmapsto \sup \left\{ \frac{\|Ax\|}{\|x\|} \mid x \in \mathbb{R}^n \setminus \{0\} \right\}$$

zugeordnet werden. Die Verallgemeinerung der EUKLIDISCHEN Norm auf Matrizen ist die FROBENIUS-Norm

$$\|\cdot\|_F : \mathbb{R}^{m \times n} \longrightarrow \mathbb{R}, \quad A \longmapsto \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2}.$$

Von der EUKLIDISCHEN Norm abgeleitet und sehr wichtig ist die Spektralnorm, die von der Notation her leicht mit der FROBENIUS-Norm verwechselt wird. Sie ist zunächst definiert als

$$\|\cdot\|_2 : \mathbb{R}^{m \times n} \longrightarrow \mathbb{R}, \quad A \longmapsto \sup \left\{ \frac{\|Ax\|_2}{\|x\|_2} \mid x \in \mathbb{R}^n \setminus \{0\} \right\}.$$

Bezeichnet $\langle \cdot, \cdot \rangle$ das EUKLIDISCHE Skalarprodukt, so folgt aus der Definition der Spektralnorm

$$\|A\|_2 \stackrel{[\text{Hb93, Lemma 2.9.3}]}{=} \max \left\{ \frac{\langle Ax, y \rangle}{\langle x, y \rangle} \mid x, y \in \mathbb{R}^n \setminus \{0\} \right\}.$$

Ist A eine normale Matrix, d.h. $AA^\top = A^\top A$, so gilt

$$\|A\|_2 \stackrel{[\text{Hb93, Satz 2.9.5}]}{=} \rho(A) := \max \{ |\lambda| : \lambda \text{ ist Eigenwert von } A \}.$$

Ist λ ein Eigenwert der Matrix A , so ist $\frac{1}{\lambda}$ ein Eigenwert der Matrix A^{-1} . Ist die Matrix A positiv definit, gilt sogar

$$\begin{aligned} \|A\|_2 &= \text{größter Eigenwert von } A = \lambda_{\max}, \\ \frac{1}{\|A^{-1}\|_2} &= \text{kleinster Eigenwert von } A = \lambda_{\min}. \end{aligned}$$

Die Kondition einer Matrix A ist normabhängig definiert als

$$\text{cond}(A) := \|A\| \cdot \|A^{-1}\|.$$

Um Unabhängigkeit von der Matrixnorm zu erlangen, führt man die Konditionszahl ein. Sie ist definiert als

$$\begin{aligned} \kappa(A) &:= \frac{\text{größter Eigenwert von } A}{\text{größter Eigenwert von } A^{-1}} = \rho(A) \cdot \rho(A^{-1}) \\ &\stackrel{A \text{ positiv definit}}{=} \frac{\text{größter Eigenwert von } A}{\text{kleinster Eigenwert von } A} = \frac{\lambda_{\max}}{\lambda_{\min}}, \end{aligned}$$

[Hb93, S. 54].

Bemerkung A.1 Die Massenmatrizen M und \vec{M}

- sind symmetrisch und positiv definit,
- besitzen als kleinsten und größten Eigenwert μ_{\min} und μ_{\max}

Deshalb besitzen die Inversen als kleinsten und größten Eigenwert μ_{\max}^{-1} und μ_{\min}^{-1} , d.h.

$$\begin{aligned} \|M\|_2 &= \|\vec{M}\|_2 = \mu_{\max}, \\ \|M^{-1}\|_2 &= \|\vec{M}^{-1}\|_2 = \frac{1}{\mu_{\min}}. \end{aligned}$$

A.2 Äquivalenz von Normen

Unter diesem Punkt tragen wir verschiedene Ungleichungen zusammen, die in der Literatur gewöhnlich im Rahmen der Äquivalenz von Normen behandelt werden oder zwischen Elliptizität und Koerzivität leben. Die Voraussetzungen und die Quantifizierung der Konstanten variieren in der Literatur sehr stark, sodaß wir an dieser Stelle versuchen wollen, möglichst viel aus der bekannten und/oder wichtigen Literatur zusammenzutragen.

Die Ungleichung von FRIEDRICHS

FRIEDRICHS-Ungleichung für DIRICHLET-Nullrandbedingung

Vgl. [Br97, S. 29] oder [Ve76, S. 59, Formel (2.66)]. Für ein beschränktes Gebiet $\Omega \subset \mathbb{R}^n$ existiert eine nur vom Gebiet abhängige Konstante $\tilde{c}_\Omega > 0$ derart, daß

$$\|u\|_{L^2(\Omega)} \leq \sqrt{\tilde{c}_\Omega} \|\partial_i u\|_{L^2(\Omega)} \leq \sqrt{\tilde{c}_\Omega} |u|_{H^1(\Omega)} \quad \forall u \in H_0^1(\Omega) \quad (\text{A.1})$$

(die zweite Ungleichung ist trivial). Nach [GT01, S. 164] ist $\tilde{c}_\Omega = (\text{vol}(\Omega) \cdot \omega_n^{-1})^{2/n}$ mit ω_n als Volumen der n -dimensionalen Einheitskugel (siehe Lemma 5.3).

FRIEDRICHS-Ungleichung für inhomogene DIRICHLET-Randbedingung

Auch **POINCARÉ-Ungleichung für $H^1(\Omega)$** oder **verallgemeinerte POINCARÉ-Ungleichung** genannt. In [BS02, Lemma 4.3.14] oder [GT01, Formel (7.45)] finden wir für ein bzgl. eines Balls B sternförmiges Gebiet (also insbesondere für konvexe Gebiete) Ω , daß

$$\|u - \bar{u}\|_{L^2(\Omega)} \leq \|u - \bar{u}\|_{H^1(\Omega)} \leq c_\Omega |u|_{H^1(\Omega)} \quad \forall u \in H^1(\Omega) \quad (\text{A.2})$$

mit der Konstanten $\bar{u} := \frac{1}{\text{vol}(\Omega)} \int_\Omega u(x) dx$ (die erste Ungleichung ist trivial). Im Falle konvexer Gebiete ist

$$c_\Omega \stackrel{[\text{GT01, Formel (7.45)}]}{=} \left(\frac{\omega_n}{\text{vol}(S)} \right)^{1-\frac{1}{n}} \cdot \varnothing_\Omega^n$$

mit $S \subset \Omega$ beliebig, aber meßbar. Beachte:

$$\|u - \bar{u}\|_{L^2(\Omega)}^2 = \|u\|_{L^2(\Omega)}^2 - \text{vol}(\Omega) \cdot \bar{u}^2.$$

FRIEDRICHS-Ungleichung für vektorwertige Funktionen

Sei $\Omega \subset \mathbb{R}^n$ beschränkt. Dann gilt

$$\|\vec{u}\|_{L^2(\Omega, \mathbb{R}^q)} \leq \sqrt{\tilde{c}_\Omega} |\vec{u}|_{H^1(\Omega, \mathbb{R}^q)} \quad \forall \vec{u} \in H_0^1(\Omega, \mathbb{R}^q) \quad (\text{A.3})$$

mit derselben gebietsunabhängigen Konstanten $\tilde{c}_\Omega > 0$ wie in (A.1).

Beweis: Aus $\vec{u} \in H_0^1(\Omega, \mathbb{R}^q)$ folgt $u_1, \dots, u_q \in H_0^1(\Omega)$, also dürfen wir (A.1) anwenden:

$$\|\vec{u}\|_{L^2(\Omega, \mathbb{R}^q)}^2 = \sum_{i=1}^q \|u_i\|_{L^2(\Omega)}^2 \leq \sum_{i=1}^q \tilde{c}_\Omega |u_i|_{H^1(\Omega)}^2 = \tilde{c}_\Omega |\vec{u}|_{H^1(\Omega, \mathbb{R}^q)}^2.$$

QED

Im Falle von inhomogenen DIRICHLET-Randbedingungen liefert die gewünschte Aussage die POINCARÉ-Ungleichung für vektorwertige Funktionen, s. S. 131.

Die Ungleichung von POINCARÉ

1. Ungleichung von POINCARÉ

Sei $\Omega \subset \mathbb{R}^n$ ein beschränktes oder — nach [Wlo82, Satz 7.6] — zumindest in eine Richtung beschränktes Gebiet (das als Vereinigung von bzgl. eines Balls sternförmiger Gebiete dargestellt werden kann). Dann existiert eine nur von m und dem Durchmesser von Ω abhängende Konstante $c_\Omega > 0$, sodaß

$$\|u\|_{H^m(\Omega)} \leq c_\Omega |u|_{H^m(\Omega)} \quad \forall u \in H_0^m(\Omega), \quad m \in \mathbb{N} \quad (\text{A.4})$$

(vgl. [BS02, Proposition 5.3.5], [DL88, S. 125/126], [GT01, Formel (7.44)], [Ve76, S. 63, Formel (2.75)], [Wlo82, Satz 7.6]).

2. Ungleichung von POINCARÉ

Vgl. [Wlo82, Satz 7.7]. Sei Ω beschränkt und erfülle die gleichmäßige Kegelbedingung. Dann existiert eine Konstante $c > 0$, sodaß

$$\|u\|_{H^m(\Omega)}^2 \leq c^2 \left(|u|_{H^m(\Omega)}^2 + \sum_{|\alpha| < m} \left\| \int_\Omega d^\alpha u \right\|^2 \right) \quad \forall u \in H^m(\Omega), \quad m \in \mathbb{N}. \quad (\text{A.5})$$

Diese Ungleichung ist die Verallgemeinerung von (A.2). Die Quantifizierung der sog. **POINCARÉschen Konstanten** c ist, wie in [DL88, S. 127 – 130, insbes. Remark 5] ausführlich dargelegt wird, nicht ohne weiteres möglich.

Unter der Annahme von Konvexität gelang jedoch PAYNE und WEINBERGER im Jahre 1960 die Bestimmung einer optimalen POINCARÉ-Konstanten [PW60]. Ein korrekter findet sich in [Be03]. Sei also $\Omega \subset \mathbb{R}^n$ ein konvexes Gebiet vom Durchmesser \varnothing_Ω . Dann gilt

$$\|u\|_{L^2(\Omega)} \leq \frac{\varnothing_\Omega}{\pi} \cdot \|\nabla u\|_{L^2(\Omega)} \quad \forall u \in H^1(\Omega) \quad \text{mit} \quad \int_\Omega u = 0. \quad (\text{A.6})$$

Wenn der Fall $\int_{\Omega} u \neq 0$ vorliegt, setzen wir

$$\begin{aligned} v &:= \frac{u}{\text{vol}(\Omega)} \longrightarrow \int_{\Omega} v = \bar{u} = 0, \\ w &:= v - \bar{v} \longrightarrow \int_{\Omega} w = \bar{u} - \frac{\bar{u} \cdot \text{vol}(\Omega)}{\text{vol}(\Omega)} = 0, \end{aligned}$$

d.h. für die Funktion $w : \Omega \rightarrow \mathbb{R}$ sind die Voraussetzungen erfüllt und die zuletzt genannte Ungleichung gilt! Die Resubstitution von w ergibt

$$\begin{aligned} \|v - \bar{v}\|_{L^2(\Omega)} &\leq \frac{\varnothing_{\Omega}}{\pi} \cdot \|\nabla(v - \bar{v})\|_{L^2(\Omega)} \\ &= \frac{\varnothing_{\Omega}}{\pi} \cdot \|\nabla v - 0\|_{L^2(\Omega)} \\ &= \frac{\varnothing_{\Omega}}{\pi} \cdot |v|_{H^1(\Omega)}, \end{aligned} \tag{A.7}$$

i.e. die von uns gewünschte Form.

3. Ungleichung von POINCARÉ für Differentialgleichungssysteme

Die Voraussetzungen bei obigen Ungleichungen beziehen sich ausschließlich auf das Gebiet Ω , weshalb wir die Ungleichungen für vektorwertige Funktionen ohne großen Aufwand formulieren können. Aus $\vec{u} \in H^1(\Omega, \mathbb{R}^q)$ folgt $u_i \in H^1(\Omega)$ für alle $i \in \{1, \dots, q\}$, sodaß wir (A.7) komponentenweise anwenden dürfen, d.h. mit

$$\bar{u}_i := \frac{1}{\text{vol}(\Omega)} \int_{\Omega} u_i(x) dx, \quad \bar{\vec{u}} := (\bar{u}_i)_{i=1, \dots, q}$$

gilt

$$\begin{aligned} \|\vec{u} - \bar{\vec{u}}\|_{L^2(\Omega, \mathbb{R}^q)}^2 &= \sum_{i=1}^q \|u_i - \bar{u}_i\|_{L^2(\Omega)}^2 \\ &\stackrel{(A.7)}{\leq} \sum_{i=1}^q \frac{\varnothing_{\Omega}^2}{\pi^2} |u_i|_{H^1(\Omega)}^2 \\ &= \frac{\varnothing_{\Omega}^2}{\pi^2} |\vec{u}|_{H^1(\Omega, \mathbb{R}^q)}^2. \end{aligned}$$

Zusammenfassend haben wir für ein konvexes Gebiet $\Omega \subset \mathbb{R}^n$ mit dem Durchmesser \varnothing_{Ω}

$$\|\vec{u} - \bar{\vec{u}}\|_{L^2(\Omega, \mathbb{R}^q)} \leq \frac{\varnothing_{\Omega}}{\pi} |\vec{u}|_{H^1(\Omega, \mathbb{R}^q)} \quad \forall \vec{u} \in H^1(\Omega, \mathbb{R}^q). \tag{A.8}$$

Anhang B

Details zu den numerischen Tests

B.1 Rechnerarchitektur und Programmatur

B.1.1 Angaben zur Hardware und Compiler

Alle Rechnungen wurden auf einer Fujitsu-Siemens V830 durchgeführt. Diese Maschine verfügt in unserer Ausstattung über einen 2.2-GHz Prozessor (AMD Opteron 275 (dual core)) und 16GB RAM.

Sowohl die HLib als auch die selbst geschriebenen Programme wurden mit dem GCC 4.1.0 und den folgenden Optionen übersetzt:

```
CFLAGS_GCC = -O3 -funroll-loops -I ${INCLUDE} -march=nocona \\  
             -ftree-vectorize -mfpmath=sse -msse3 -m64  
FC_GCC = g77  
LIBS_GCC = -lgfortran -lblas -llapack -lg2c -lm HLib/*.o  
LDFLAGS_GCC = -static
```

B.1.2 Verwendete Programme und Bibliotheken

Wir verwenden die HLib in der Version 1.3, Patch 9, vgl. www.hamtrix.org, um die \mathcal{H} -Arithmetik und die Gleichungssystemlöser nicht selbst implementieren zu müssen. Selbst programmiert wurden

- Routinen zum Aufbau des Cluster-Baumes für q -dimensionale Differentialgleichungen,
- FE-Assemblierungsroutinen für die zweidimensionale LAMÉ-Gleichung,
- der Programmrahmen

sowie Routinen für diverse Hilfsaufgaben wie Berechnung des L^2 -Fehlers o.ä.

B.1.3 Gebiets- und Gittergenerierung

Die Gebiete und Gitter wurden mit Femlab bzw. Comsol 3.2 erzeugt. Die Gitter wurden dann über ein Matlab-Skript in ein HLib-verträgliches Format konvertiert und in eine Datei geschrieben.

B.2 Numerische Resultate

Die Ergebnisse unserer Rechnungen sind in den (OpenOffice-) Tabellen

- Progs/_ergebnisse_evaluierung.sxc
 \mathcal{H} -Invertierung mit rangfester Arithmetik
- Progs/_ergebnisse_evaluierung_adaptiv.sxc
 \mathcal{H} -Invertierung mit rangadaptiver Arithmetik
- Progs/_ergebnisse_chol_cg_adaptiv.sxc
CG-Verfahren mit rangadaptiv berechnetem \mathcal{H} -CHOLESKY-Faktor

erfaßt. Die für die Arbeit für interessant befundenen und ausgewerteten Daten sind im folgenden einzeln wiedergegeben.

B.2.1 \mathcal{H} -Inverse: fester Rang, Geometrie 2, Parametertyp 3

kq	6946 DoF's		28226 DoF's	
	$\ A \cdot A^{-\mathcal{H}} - Id\ _2$	$\ u - u_h\ _{L^2}$	$\ A \cdot A^{-\mathcal{H}} - Id\ _2$	$\ u - u_h\ _{L^2}$
2	1.78E+04	5.54E+00	7.67E+11	8.40E+07
4	2.41E+05	1.72E+01	2.08E+14	7.97E+09
8	1.32E+03	1.22E+00	5.16E+04	8.20E+00
12	1.14E+03	1.53E+00	2.16E+04	1.08E+01
16	3.05E+02	9.78E-01	3.33E+03	2.40E+00
20	5.70E+02	3.01E+00	3.85E+03	1.90E+00
24	4.44E+01	5.58E-01	2.08E+03	1.48E+00
28	9.55E+00	1.74E-02	2.67E+02	3.96E-01
32	2.76E+00	1.54E-02	3.49E+02	3.82E-01
36	2.80E-01	1.52E-02	2.49E+02	1.98E-01
40	1.28E-01	1.53E-02	1.88E+02	2.37E-01
44	5.29E-02	1.53E-02	2.18E+02	4.36E-01
48	1.18E-02	1.53E-02	1.67E+03	6.91E+00
52	1.73E-03	1.53E-02	1.30E+02	9.13E-01
56	5.22E-04	1.53E-02	2.14E+01	1.43E-01
60	6.91E-05	1.53E-02	4.46E+00	7.17E-03
64	4.02E-06	1.53E-02	1.57E+00	7.41E-03
72	1.21E-10	1.53E-02	3.88E-01	7.41E-03
80	1.26E-10	1.53E-02	2.71E-02	7.41E-03
88	1.31E-10	1.53E-02	8.10E-03	7.41E-03
96	1.32E-10	1.53E-02	1.35E-03	7.41E-03
104	1.13E-10	1.53E-02	7.34E-05	7.41E-03
112	1.13E-10	1.53E-02	7.26E-06	7.41E-03
120	1.30E-10	1.53E-02	6.84E-07	7.41E-03
128	1.23E-10	1.53E-02	2.38E-08	7.41E-03

Tabelle B.1: Fehler.

kq	6946 DoF's				28226 DoF's			
	Aufbau Clusterbaum	Aufbau \mathcal{H} -Struktur	Konvertierung in \mathcal{H} -Struktur	Invertierung + MVM	Aufbau Clusterbaum	Aufbau \mathcal{H} -Struktur	Konvertierung in \mathcal{H} -Struktur	Invertierung + MVM
2	1.00E-02	1.50E-01	2.30E-01	4.27E+01	5.00E-02	4.10E-01	1.03E+00	2.38E+02
4	0.00E+00	1.10E-01	2.30E-01	4.54E+01	5.00E-02	4.60E-01	1.03E+00	2.64E+02
8	0.00E+00	1.60E-01	2.30E-01	5.23E+01	3.00E-02	5.30E-01	1.03E+00	3.40E+02
12	0.00E+00	2.00E-01	2.30E-01	6.23E+01	4.00E-02	6.30E-01	1.03E+00	4.52E+02
16	0.00E+00	2.20E-01	2.50E-01	7.37E+01	5.00E-02	7.10E-01	1.03E+00	5.88E+02
20	1.00E-02	2.10E-01	2.30E-01	8.47E+01	3.00E-02	8.00E-01	1.03E+00	7.24E+02
24	0.00E+00	1.60E-01	2.30E-01	9.66E+01	5.00E-02	8.70E-01	1.03E+00	8.68E+02
28	0.00E+00	1.70E-01	2.30E-01	1.05E+02	5.00E-02	9.60E-01	1.03E+00	1.01E+03
32	1.00E-02	2.80E-01	2.70E-01	1.12E+02	3.00E-02	1.06E+00	1.02E+00	1.14E+03
36	1.00E-02	2.30E-01	2.30E-01	1.19E+02	5.00E-02	1.12E+00	1.03E+00	1.25E+03
40	0.00E+00	3.30E-01	2.30E-01	1.26E+02	4.00E-02	1.23E+00	1.02E+00	1.35E+03
44	0.00E+00	2.20E-01	2.60E-01	1.31E+02	5.00E-02	1.31E+00	1.03E+00	1.45E+03
48	0.00E+00	3.00E-01	2.30E-01	1.35E+02	4.00E-02	1.39E+00	1.02E+00	1.52E+03
52	0.00E+00	3.50E-01	2.30E-01	1.39E+02	3.00E-02	1.48E+00	1.02E+00	1.60E+03
56	0.00E+00	2.60E-01	2.30E-01	1.42E+02	5.00E-02	1.54E+00	1.03E+00	1.66E+03
60	1.00E-02	2.70E-01	2.30E-01	1.44E+02	4.00E-02	1.63E+00	1.03E+00	1.72E+03
64	1.00E-02	2.90E-01	2.30E-01	1.47E+02	4.00E-02	1.74E+00	1.02E+00	1.79E+03
72	1.00E-02	3.40E-01	2.30E-01	1.54E+02	3.00E-02	1.87E+00	1.02E+00	1.90E+03
80	0.00E+00	3.30E-01	2.50E-01	1.61E+02	2.00E-02	2.08E+00	1.02E+00	2.02E+03
88	0.00E+00	4.30E-01	2.20E-01	1.74E+02	5.00E-02	2.25E+00	1.02E+00	2.16E+03
96	0.00E+00	4.70E-01	2.30E-01	1.85E+02	4.00E-02	2.39E+00	1.02E+00	2.25E+03
104	1.00E-02	4.90E-01	2.30E-01	1.90E+02	2.00E-02	2.50E+00	1.03E+00	2.32E+03
112	0.00E+00	3.90E-01	2.30E-01	1.96E+02	2.00E-02	2.77E+00	1.03E+00	2.37E+03
120	0.00E+00	5.30E-01	2.30E-01	1.96E+02	5.00E-02	2.91E+00	1.02E+00	2.40E+03
128	0.00E+00	4.40E-01	2.30E-01	1.97E+02	4.00E-02	3.08E+00	1.03E+00	2.42E+03

Tabelle B.2: Zeitbedarf in [sec].

kq	6946 DoF's			28226 DoF's		
	kB/DoF	gesamt [MB]	C_{sp}	kB/DoF	gesamt [MB]	C_{sp}
2	1.07E+01	2.18E+02	30	1.22E+01	1.01E+03	34
4	1.13E+01	2.31E+02	30	1.34E+01	1.11E+03	34
8	1.26E+01	2.56E+02	30	1.59E+01	1.32E+03	34
12	1.39E+01	2.82E+02	30	1.84E+01	1.52E+03	34
16	1.51E+01	3.08E+02	30	2.09E+01	1.73E+03	34
20	1.64E+01	3.34E+02	30	2.34E+01	1.93E+03	34
24	1.77E+01	3.59E+02	30	2.58E+01	2.14E+03	34
28	1.89E+01	3.85E+02	30	2.83E+01	2.34E+03	34
32	2.02E+01	4.11E+02	30	3.08E+01	2.55E+03	34
36	2.15E+01	4.37E+02	30	3.33E+01	2.75E+03	34
40	2.27E+01	4.62E+02	30	3.58E+01	2.96E+03	34
44	2.40E+01	4.88E+02	30	3.82E+01	3.16E+03	34
48	2.53E+01	5.14E+02	30	4.07E+01	3.37E+03	34
52	2.65E+01	5.40E+02	30	4.32E+01	3.57E+03	34
56	2.78E+01	5.65E+02	30	4.57E+01	3.78E+03	34
60	2.90E+01	5.91E+02	30	4.82E+01	3.98E+03	34
64	3.03E+01	6.17E+02	30	5.06E+01	4.19E+03	34
72	3.28E+01	6.68E+02	30	5.56E+01	4.60E+03	34
80	3.54E+01	7.20E+02	30	6.06E+01	5.01E+03	34
88	3.79E+01	7.71E+02	30	6.55E+01	5.42E+03	34
96	4.04E+01	8.23E+02	30	7.05E+01	5.83E+03	34
104	4.30E+01	8.74E+02	30	7.54E+01	6.24E+03	34
112	4.55E+01	9.26E+02	30	8.04E+01	6.65E+03	34
120	4.80E+01	9.77E+02	30	8.54E+01	7.06E+03	34
128	5.06E+01	1.03E+03	30	9.03E+01	7.47E+03	34

Tabelle B.3: Speicherverbrauch und Dünnbesetztheitskonstante C_{sp} .

B.2.2 \mathcal{H} -Inverse: fester Rang, Geometrie 2, Parametertyp 4

kq	6946 DoF's		28226 DoF's	
	$\ A \cdot A^{-\mathcal{H}} - Id\ _2$	$\ u - u_h\ _{L^2}$	$\ A \cdot A^{-\mathcal{H}} - Id\ _2$	$\ u - u_h\ _{L^2}$
2	5.66E+04	5.24E+00	3.35E+65	2.36E+62
4	1.97E+04	2.46E+01	6.99E+04	1.20E+01
8	4.95E+03	5.75E+00	7.72E+04	3.34E+01
12	1.74E+03	1.10E+00	5.38E+03	2.48E+00
16	2.72E+02	4.63E-01	6.84E+03	4.90E+00
20	8.77E+02	4.42E+00	1.51E+05	1.63E+00
24	4.48E+01	5.67E-01	2.00E+03	6.13E-01
28	8.40E+00	4.09E-02	2.70E+02	4.76E-01
32	2.87E+00	1.69E-02	2.58E+02	2.09E-01
36	2.61E-01	1.54E-02	2.55E+02	2.58E-01
40	1.30E-01	1.55E-02	1.93E+02	2.01E-01
44	5.07E-02	1.53E-02	2.16E+02	4.48E-01
48	7.99E-03	1.53E-02	1.75E+03	7.27E+00
52	9.16E-04	1.53E-02	1.32E+02	9.18E-01
56	1.30E-04	1.53E-02	2.09E+01	1.31E-01
60	1.53E-05	1.53E-02	4.40E+00	1.14E-02
64	4.94E-07	1.53E-02	1.26E+00	7.40E-03
72	1.54E-10	1.53E-02	3.88E-01	7.47E-03
80	1.33E-10	1.53E-02	2.55E-02	7.41E-03
88	1.52E-10	1.53E-02	6.52E-03	7.41E-03
96	1.39E-10	1.53E-02	3.54E-04	7.41E-03
104	1.50E-10	1.53E-02	6.72E-05	7.41E-03
112	1.50E-10	1.53E-02	2.44E-06	7.41E-03
120	1.51E-10	1.53E-02	2.15E-07	7.41E-03
128	1.54E-10	1.53E-02	8.61E-09	7.41E-03

Tabelle B.4: Fehler.

kq	6946 DoF's				28226 DoF's			
	Aufbau Clusterbaum	Aufbau \mathcal{H} -Struktur	Konvertierung in \mathcal{H} -Struktur	Invertierung + MVM	Aufbau Clusterbaum	Aufbau \mathcal{H} -Struktur	Konvertierung in \mathcal{H} -Struktur	Invertierung + MVM
2	1.00E-02	1.50E-01	2.30E-01	4.29E+01	4.00E-02	4.20E-01	1.02E+00	2.41E+02
4	1.00E-02	1.60E-01	2.30E-01	4.55E+01	2.00E-02	4.60E-01	1.02E+00	2.64E+02
8	1.00E-02	1.80E-01	2.30E-01	5.25E+01	3.00E-02	5.40E-01	1.02E+00	3.40E+02
12	1.00E-02	2.00E-01	2.30E-01	6.24E+01	5.00E-02	6.20E-01	1.03E+00	4.52E+02
16	1.00E-02	1.40E-01	2.80E-01	7.36E+01	2.00E-02	7.10E-01	1.03E+00	5.85E+02
20	0.00E+00	2.00E-01	2.30E-01	8.40E+01	3.00E-02	8.00E-01	1.02E+00	7.16E+02
24	0.00E+00	1.60E-01	2.20E-01	9.49E+01	2.00E-02	8.70E-01	1.03E+00	8.52E+02
28	1.00E-02	1.70E-01	2.20E-01	1.03E+02	5.00E-02	9.50E-01	1.03E+00	9.81E+02
32	0.00E+00	2.90E-01	2.70E-01	1.09E+02	3.00E-02	1.07E+00	1.03E+00	1.11E+03
36	0.00E+00	3.20E-01	2.50E-01	1.15E+02	3.00E-02	1.12E+00	1.03E+00	1.20E+03
40	0.00E+00	3.30E-01	2.40E-01	1.22E+02	5.00E-02	1.22E+00	1.03E+00	1.30E+03
44	1.00E-02	2.50E-01	2.30E-01	1.27E+02	3.00E-02	1.32E+00	1.02E+00	1.38E+03
48	0.00E+00	2.70E-01	2.20E-01	1.30E+02	4.00E-02	1.38E+00	1.02E+00	1.44E+03
52	1.00E-02	3.10E-01	2.30E-01	1.33E+02	5.00E-02	1.47E+00	1.03E+00	1.53E+03
56	1.00E-02	3.00E-01	2.30E-01	1.37E+02	2.00E-02	1.56E+00	1.02E+00	1.58E+03
60	0.00E+00	3.80E-01	2.30E-01	1.38E+02	2.00E-02	1.65E+00	1.03E+00	1.64E+03
64	0.00E+00	4.00E-01	2.20E-01	1.41E+02	4.00E-02	1.71E+00	1.03E+00	1.70E+03
72	0.00E+00	3.30E-01	2.30E-01	1.47E+02	3.00E-02	1.92E+00	1.02E+00	1.81E+03
80	0.00E+00	3.30E-01	2.30E-01	1.54E+02	2.00E-02	2.06E+00	1.03E+00	1.92E+03
88	0.00E+00	4.00E-01	2.30E-01	1.66E+02	4.00E-02	2.25E+00	1.02E+00	2.06E+03
96	1.00E-02	4.70E-01	2.30E-01	1.77E+02	2.00E-02	2.40E+00	1.03E+00	2.15E+03
104	1.00E-02	3.90E-01	2.30E-01	1.82E+02	5.00E-02	2.57E+00	1.03E+00	2.21E+03
112	1.00E-02	4.70E-01	2.20E-01	1.88E+02	4.00E-02	2.73E+00	1.03E+00	2.25E+03
120	0.00E+00	4.20E-01	2.30E-01	1.88E+02	5.00E-02	2.91E+00	1.03E+00	2.29E+03
128	1.00E-02	4.70E-01	2.30E-01	1.89E+02	3.00E-02	3.09E+00	1.02E+00	2.31E+03

Tabelle B.5: Zeitbedarf in [sec].

kq	6946 DoF's			28226 DoF's		
	kB/DoF	gesamt [MB]	C_{sp}	kB/DoF	gesamt [MB]	C_{sp}
2	1.07E+01	2.18E+02	30	1.22E+01	1.01E+03	34
4	1.13E+01	2.31E+02	30	1.34E+01	1.11E+03	34
8	1.26E+01	2.56E+02	30	1.59E+01	1.32E+03	34
12	1.39E+01	2.82E+02	30	1.84E+01	1.52E+03	34
16	1.51E+01	3.08E+02	30	2.09E+01	1.73E+03	34
20	1.64E+01	3.34E+02	30	2.34E+01	1.93E+03	34
24	1.77E+01	3.59E+02	30	2.58E+01	2.14E+03	34
28	1.89E+01	3.85E+02	30	2.83E+01	2.34E+03	34
32	2.02E+01	4.11E+02	30	3.08E+01	2.55E+03	34
36	2.15E+01	4.37E+02	30	3.33E+01	2.75E+03	34
40	2.27E+01	4.62E+02	30	3.58E+01	2.96E+03	34
44	2.40E+01	4.88E+02	30	3.82E+01	3.16E+03	34
48	2.53E+01	5.14E+02	30	4.07E+01	3.37E+03	34
52	2.65E+01	5.40E+02	30	4.32E+01	3.57E+03	34
56	2.78E+01	5.65E+02	30	4.57E+01	3.78E+03	34
60	2.90E+01	5.91E+02	30	4.82E+01	3.98E+03	34
64	3.03E+01	6.17E+02	30	5.06E+01	4.19E+03	34
72	3.28E+01	6.68E+02	30	5.56E+01	4.60E+03	34
80	3.54E+01	7.20E+02	30	6.06E+01	5.01E+03	34
88	3.79E+01	7.71E+02	30	6.55E+01	5.42E+03	34
96	4.04E+01	8.23E+02	30	7.05E+01	5.83E+03	34
104	4.30E+01	8.74E+02	30	7.54E+01	6.24E+03	34
112	4.55E+01	9.26E+02	30	8.04E+01	6.65E+03	34
120	4.80E+01	9.77E+02	30	8.54E+01	7.06E+03	34
128	5.06E+01	1.03E+03	30	9.03E+01	7.47E+03	34

Tabelle B.6: Speicherverbrauch und Dünnbesetztheitskonstante C_{sp} .

B.2.3 \mathcal{H} -Inverse: fester Rang, Geometrie 3, Parametertyp 3

kq	7938 DoF's		32258 DoF's	
	$\ A \cdot A^{-\mathcal{H}} - Id\ _2$	$\ u - u_h\ _{L^2}$	$\ A \cdot A^{-\mathcal{H}} - Id\ _2$	$\ u - u_h\ _{L^2}$
2	1.83E+03	1.17E+00	4.61E+09	4.49E+05
4	1.05E+03	1.10E+00	5.16E+04	1.11E+01
8	7.70E+03	4.96E+00	1.28E+05	7.35E+00
12	3.33E+01	7.42E-02	3.78E+03	3.82E+00
16	1.15E+01	1.03E-01	7.51E+03	1.21E+00
20	4.46E+00	4.61E-02	4.18E+02	2.06E-01
24	4.35E-01	1.69E-02	1.02E+02	2.63E-01
28	9.04E-02	1.68E-02	3.74E+01	1.91E-01
32	3.53E-02	1.69E-02	2.42E+01	1.29E-01
36	2.26E-02	1.69E-02	6.02E+00	1.74E-02
40	1.33E-02	1.69E-02	1.46E+00	1.15E-02
44	2.82E-03	1.69E-02	8.12E-01	1.29E-02
48	9.99E-05	1.69E-02	3.70E-01	1.04E-02
52	2.38E-07	1.69E-02	3.28E-02	8.97E-03
56	3.32E-08	1.69E-02	1.67E-02	8.99E-03
60	4.52E-09	1.69E-02	8.22E-03	8.99E-03
64	5.30E-10	1.69E-02	6.10E-03	8.99E-03
72	5.74E-11	1.69E-02	4.30E-03	8.99E-03
80	6.30E-11	1.69E-02	2.99E-03	8.98E-03
88	8.59E-11	1.69E-02	1.50E-03	8.98E-03
96	5.63E-11	1.69E-02	1.28E-06	8.98E-03
104	9.52E-11	1.69E-02	2.18E-09	8.98E-03
112	7.16E-11	1.69E-02	2.17E-09	8.98E-03
120	7.24E-11	1.69E-02	2.11E-09	8.98E-03
128	7.67E-11	1.69E-02	2.50E-09	8.98E-03

Tabelle B.7: Fehler.

kq	7938 DoF's				32258 DoF's			
	Aufbau Clusterbaum	Aufbau \mathcal{H} -Struktur	Konvertierung in \mathcal{H} -Struktur	Invertierung + MVM	Aufbau Clusterbaum	Aufbau \mathcal{H} -Struktur	Konvertierung in \mathcal{H} -Struktur	Invertierung + MVM
2	0.00E+00	1.40E-01	2.00E-01	3.56E+01	3.00E-02	4.30E-01	9.10E-01	1.96E+02
4	1.00E-02	1.10E-01	2.40E-01	3.85E+01	4.00E-02	4.90E-01	9.10E-01	2.27E+02
8	0.00E+00	1.90E-01	2.70E-01	4.76E+01	5.00E-02	6.20E-01	9.10E-01	3.18E+02
12	1.00E-02	1.60E-01	2.00E-01	6.12E+01	4.00E-02	7.30E-01	9.10E-01	4.58E+02
16	1.00E-02	2.30E-01	1.90E-01	7.82E+01	5.00E-02	8.70E-01	9.10E-01	6.37E+02
20	0.00E+00	2.50E-01	2.60E-01	9.37E+01	5.00E-02	9.50E-01	9.10E-01	8.19E+02
24	1.00E-02	2.30E-01	1.90E-01	1.11E+02	6.00E-02	1.10E+00	9.10E-01	1.02E+03
28	1.00E-02	2.40E-01	2.00E-01	1.29E+02	5.00E-02	1.23E+00	9.20E-01	1.23E+03
32	0.00E+00	2.40E-01	2.40E-01	1.44E+02	5.00E-02	1.34E+00	9.10E-01	1.43E+03
36	1.00E-02	3.10E-01	2.00E-01	1.55E+02	5.00E-02	1.46E+00	9.10E-01	1.61E+03
40	1.00E-02	3.30E-01	2.00E-01	1.65E+02	5.00E-02	1.58E+00	9.10E-01	1.78E+03
44	1.00E-02	3.50E-01	2.00E-01	1.73E+02	3.00E-02	1.77E+00	9.10E-01	1.93E+03
48	1.00E-02	3.40E-01	2.00E-01	1.79E+02	5.00E-02	1.77E+00	9.10E-01	2.03E+03
52	0.00E+00	3.40E-01	2.00E-01	1.84E+02	4.00E-02	1.73E+00	9.10E-01	2.13E+03
56	0.00E+00	4.30E-01	2.00E-01	1.88E+02	3.00E-02	1.83E+00	9.10E-01	2.23E+03
60	0.00E+00	4.30E-01	2.00E-01	1.92E+02	5.00E-02	1.94E+00	9.00E-01	2.32E+03
64	0.00E+00	4.30E-01	2.00E-01	1.94E+02	6.00E-02	2.11E+00	9.10E-01	2.44E+03
72	1.00E-02	4.90E-01	2.00E-01	1.99E+02	5.00E-02	2.38E+00	9.00E-01	2.55E+03
80	1.00E-02	4.90E-01	2.00E-01	2.04E+02	6.00E-02	2.68E+00	9.10E-01	2.65E+03
88	1.00E-02	4.80E-01	2.00E-01	2.09E+02	5.00E-02	2.88E+00	9.10E-01	2.74E+03
96	1.00E-02	5.30E-01	2.00E-01	2.15E+02	5.00E-02	3.03E+00	9.10E-01	2.81E+03
104	1.00E-02	5.30E-01	2.00E-01	2.26E+02	5.00E-02	3.28E+00	9.10E-01	2.89E+03
112	0.00E+00	5.80E-01	2.00E-01	2.34E+02	3.00E-02	3.57E+00	9.10E-01	2.97E+03
120	1.00E-02	7.00E-01	2.00E-01	2.52E+02	4.00E-02	3.72E+00	9.10E-01	3.13E+03
128	0.00E+00	6.20E-01	2.00E-01	2.56E+02	5.00E-02	3.93E+00	9.10E-01	3.17E+03

Tabelle B.8: Zeitbedarf in [sec].

kq	7938 DoF's			32258 DoF's		
	kB/DoF	gesamt [MB]	C_{sp}	kB/DoF	gesamt [MB]	C_{sp}
2	8.25E+00	192	30	9.63E+00	910	30
4	9.02E+00	210	30	1.10E+01	1040	30
8	1.05E+01	245	30	1.38E+01	1300	30
12	1.21E+01	281	30	1.65E+01	1560	30
16	1.36E+01	316	30	1.93E+01	1820	30
20	1.51E+01	351	30	2.20E+01	2080	30
24	1.66E+01	387	30	2.48E+01	2340	30
28	1.82E+01	422	30	2.75E+01	2600	30
32	1.97E+01	458	30	3.03E+01	2860	30
36	2.12E+01	493	30	3.30E+01	3120	30
40	2.27E+01	529	30	3.58E+01	3380	30
44	2.43E+01	564	30	3.85E+01	3640	30
48	2.58E+01	600	30	4.13E+01	3900	30
52	2.73E+01	635	30	4.40E+01	4160	30
56	2.88E+01	670	30	4.68E+01	4420	30
60	3.04E+01	706	30	4.95E+01	4680	30
64	3.19E+01	741	30	5.23E+01	4940	30
72	3.49E+01	812	30	5.78E+01	5460	30
80	3.80E+01	883	30	6.33E+01	5980	30
88	4.10E+01	954	30	6.88E+01	6500	30
96	4.41E+01	1020	30	7.43E+01	7020	30
104	4.71E+01	1100	30	7.98E+01	7540	30
112	5.02E+01	1170	30	8.53E+01	8060	30
120	5.32E+01	1240	30	9.08E+01	8580	30
128	5.63E+01	1310	30	9.63E+01	9100	30

Tabelle B.9: Speicherverbrauch und Dünnbesetztheitskonstante C_{sp} .

B.2.4 \mathcal{H} -Inverse: fester Rang, Geometrie 3, Parametertyp 4

kq	7938 DoF's		32258 DoF's	
	$\ A \cdot A^{-\mathcal{H}} - Id\ _2$	$\ u - u_h\ _{L^2}$	$\ A \cdot A^{-\mathcal{H}} - Id\ _2$	$\ u - u_h\ _{L^2}$
2	8.87E+02	1.31E+00	7.73E+06	5.90E+03
4	6.63E+02	1.16E+00	1.30E+04	3.78E+00
8	5.53E+02	2.11E+00	4.36E+05	6.49E+01
12	3.32E+01	4.98E-01	2.73E+04	3.04E+01
16	1.04E+01	1.02E-01	1.01E+03	2.05E+00
20	9.97E+00	4.32E-02	4.29E+02	6.41E-01
24	3.82E+00	2.37E-02	9.09E+01	1.40E-01
28	3.80E+00	1.67E-02	3.74E+01	2.48E-01
32	3.78E+00	1.72E-02	7.52E+01	1.16E+00
36	1.28E+00	1.70E-02	2.66E+01	3.95E-02
40	2.37E-01	1.69E-02	1.08E+01	1.45E-02
44	5.10E-02	1.69E-02	1.02E+01	1.15E-02
48	5.55E-05	1.69E-02	1.00E+01	9.47E-03
52	1.81E-09	1.69E-02	9.96E+00	9.10E-03
56	7.56E-10	1.69E-02	2.64E+00	9.10E-03
60	5.62E-10	1.69E-02	5.83E-01	9.01E-03
64	5.71E-10	1.69E-02	5.79E-01	9.01E-03
72	1.07E-09	1.69E-02	5.77E-01	9.01E-03
80	5.49E-10	1.69E-02	1.54E-01	9.00E-03
88	4.48E-10	1.69E-02	3.12E-02	9.00E-03
96	5.85E-10	1.69E-02	5.81E-06	9.00E-03
104	8.55E-10	1.69E-02	3.72E-08	9.00E-03
112	4.23E-10	1.69E-02	3.58E-08	9.00E-03
120	3.98E-10	1.69E-02	6.65E-08	9.00E-03
128	4.05E-10	1.69E-02	3.75E-08	9.00E-03

Tabelle B.10: Fehler.

kq	7938 DoF's				32258 DoF's			
	Aufbau Clusterbaum	Aufbau \mathcal{H} -Struktur	Konvertierung in \mathcal{H} -Struktur	Invertierung + MVM	Aufbau Clusterbaum	Aufbau \mathcal{H} -Struktur	Konvertierung in \mathcal{H} -Struktur	Invertierung + MVM
2	1.00E-02	1.40E-01	2.90E-01	3.53E+01	4.00E-02	4.10E-01	9.10E-01	1.96E+02
4	0.00E+00	1.40E-01	3.20E-01	3.83E+01	4.00E-02	4.80E-01	9.10E-01	2.26E+02
8	1.00E-02	1.20E-01	2.40E-01	4.72E+01	3.00E-02	5.80E-01	9.10E-01	3.16E+02
12	0.00E+00	2.00E-01	2.90E-01	6.07E+01	5.00E-02	6.70E-01	9.10E-01	4.54E+02
16	1.00E-02	1.60E-01	1.90E-01	7.73E+01	5.00E-02	8.10E-01	9.10E-01	6.28E+02
20	1.00E-02	2.30E-01	2.00E-01	9.20E+01	3.00E-02	9.30E-01	9.10E-01	7.97E+02
24	1.00E-02	2.80E-01	2.50E-01	1.08E+02	3.00E-02	1.04E+00	9.10E-01	9.81E+02
28	0.00E+00	2.60E-01	2.00E-01	1.23E+02	3.00E-02	1.14E+00	9.10E-01	1.17E+03
32	0.00E+00	3.00E-01	1.90E-01	1.35E+02	4.00E-02	1.30E+00	9.20E-01	1.34E+03
36	0.00E+00	2.80E-01	2.00E-01	1.44E+02	5.00E-02	1.34E+00	9.10E-01	1.48E+03
40	0.00E+00	2.70E-01	2.20E-01	1.52E+02	5.00E-02	1.48E+00	9.10E-01	1.62E+03
44	1.00E-02	3.60E-01	2.00E-01	1.58E+02	4.00E-02	1.63E+00	9.20E-01	1.74E+03
48	1.00E-02	3.50E-01	2.00E-01	1.63E+02	4.00E-02	1.75E+00	9.10E-01	1.84E+03
52	1.00E-02	3.10E-01	2.00E-01	1.67E+02	5.00E-02	1.81E+00	9.10E-01	1.94E+03
56	0.00E+00	3.10E-01	2.00E-01	1.70E+02	5.00E-02	1.97E+00	9.10E-01	2.03E+03
60	0.00E+00	4.10E-01	2.00E-01	1.73E+02	4.00E-02	2.17E+00	9.00E-01	2.11E+03
64	0.00E+00	4.30E-01	2.00E-01	1.75E+02	3.00E-02	2.24E+00	9.10E-01	2.18E+03
72	0.00E+00	5.00E-01	2.00E-01	1.79E+02	3.00E-02	2.45E+00	9.10E-01	2.28E+03
80	1.00E-02	5.00E-01	2.00E-01	1.83E+02	5.00E-02	2.69E+00	9.10E-01	2.36E+03
88	1.00E-02	5.00E-01	1.90E-01	1.87E+02	4.00E-02	2.83E+00	9.10E-01	2.44E+03
96	1.00E-02	4.80E-01	2.00E-01	1.94E+02	5.00E-02	3.16E+00	9.10E-01	2.50E+03
104	1.00E-02	6.30E-01	2.00E-01	2.04E+02	5.00E-02	3.39E+00	9.10E-01	2.57E+03
112	0.00E+00	5.70E-01	1.90E-01	2.13E+02	5.00E-02	3.48E+00	9.10E-01	2.65E+03
120	1.00E-02	5.60E-01	2.00E-01	2.32E+02	6.00E-02	3.68E+00	9.10E-01	2.82E+03
128	1.00E-02	6.40E-01	2.00E-01	2.35E+02	5.00E-02	3.93E+00	9.00E-01	2.84E+03

Tabelle B.11: Zeitbedarf in [sec].

kq	7938 DoF's			32258 DoF's		
	kB/DoF	gesamt [MB]	C_{sp}	kB/DoF	gesamt [MB]	C_{sp}
2	8.25E+00	1.92E+02	30	9.63E+00	9.10E+02	30
4	9.02E+00	2.10E+02	30	1.10E+01	1.04E+03	30
8	1.05E+01	2.45E+02	30	1.38E+01	1.30E+03	30
12	1.21E+01	2.81E+02	30	1.65E+01	1.56E+03	30
16	1.36E+01	3.16E+02	30	1.93E+01	1.82E+03	30
20	1.51E+01	3.51E+02	30	2.20E+01	2.08E+03	30
24	1.66E+01	3.87E+02	30	2.48E+01	2.34E+03	30
28	1.82E+01	4.22E+02	30	2.75E+01	2.60E+03	30
32	1.97E+01	4.58E+02	30	3.03E+01	2.86E+03	30
36	2.12E+01	4.93E+02	30	3.30E+01	3.12E+03	30
40	2.27E+01	5.29E+02	30	3.58E+01	3.38E+03	30
44	2.43E+01	5.64E+02	30	3.85E+01	3.64E+03	30
48	2.58E+01	6.00E+02	30	4.13E+01	3.90E+03	30
52	2.73E+01	6.35E+02	30	4.40E+01	4.16E+03	30
56	2.88E+01	6.70E+02	30	4.68E+01	4.42E+03	30
60	3.04E+01	7.06E+02	30	4.95E+01	4.68E+03	30
64	3.19E+01	7.41E+02	30	5.23E+01	4.94E+03	30
72	3.49E+01	8.12E+02	30	5.78E+01	5.46E+03	30
80	3.80E+01	8.83E+02	30	6.33E+01	5.98E+03	30
88	4.10E+01	9.54E+02	30	6.88E+01	6.50E+03	30
96	4.41E+01	1.02E+03	30	7.43E+01	7.02E+03	30
104	4.71E+01	1.10E+03	30	7.98E+01	7.54E+03	30
112	5.02E+01	1.17E+03	30	8.53E+01	8.06E+03	30
120	5.32E+01	1.24E+03	30	9.08E+01	8.58E+03	30
128	5.63E+01	1.31E+03	30	9.63E+01	9.10E+03	30

Tabelle B.12: Speicherverbrauch und Dünnesetztheitskonstante C_{sp} .

B.2.5 \mathcal{H} -Inverse: adaptiver Rang, Geometrie 3, Parameter- typ 4

Dof's	1.00E-01	1.00E-02	1.00E-03	1.00E-04
	Aufbau Clusterbaum [s]			
7938	1.00E-02	1.00E-02	0.00E+00	0.00E+00
32258	5.00E-02	5.00E-02	5.00E-02	5.00E-02
	Aufbau \mathcal{H} -Matrix-Struktur [s]			
7938	1.00E-02	2.00E-02	1.00E-02	1.00E-02
32258	4.00E-02	4.00E-02	3.00E-02	4.00E-02
	Konvertierung der Sparse- in \mathcal{H} -Matrix [s]			
7938	2.40E-01	2.70E-01	2.00E-01	2.00E-01
32258	7.50E-01	7.60E-01	7.70E-01	7.70E-01
	\mathcal{H} -Invertierung + \mathcal{H} -MVM [s]			
7938	1.64E+00	5.17E+00	1.03E+01	1.84E+01
32258	9.22E+00	2.62E+01	1.12E+02	2.07E+02
	Speicher \mathcal{H} -Inverse [kB/DoF]			
7938	2.20E+00	3.39E+00	4.77E+00	6.32E+00
32258	2.57E+00	3.79E+00	7.50E+00	1.02E+01
	Gesamtsspeicher [MB]			
7938	5.11E+01	7.88E+01	1.11E+02	1.47E+02
32258	2.43E+02	3.58E+02	7.09E+02	9.66E+02
	$\ A \cdot A^{-\mathcal{H}} - Id\ _2$			
7938	2.64E+04	1.95E+03	5.41E+03	5.18E+02
32258	4.98E+48	1.55E+08	3.39E+03	1.42E+03
	$\ u - u_h\ _{L^2}$			
7938	2.62E+01	7.32E+00	3.70E+00	2.06E+00
32258	6.31E+44	1.57E+04	3.66E+00	2.59E+00
	C_{sp}			
7938	14	14	14	14
32258	14	14	14	14

Tabelle B.13: Meßdaten für rangadaptive \mathcal{H} -Invertierung (Teil 1/4).

Dof's	1.00E-05	1.00E-06	1.00E-07	1.00E-08
	Aufbau Clusterbaum [s]			
7938	1.00E-02	1.00E-02	1.00E-02	0.00E+00
32258	5.00E-02	5.00E-02	5.00E-02	5.00E-02
	Aufbau \mathcal{H} -Matrix-Struktur [s]			
7938	1.00E-02	1.00E-02	2.00E-02	2.00E-02
32258	4.00E-02	3.00E-02	4.00E-02	4.00E-02
	Konvertierung der Sparse- in \mathcal{H} -Matrix [s]			
7938	2.90E-01	3.40E-01	2.60E-01	3.00E-01
32258	7.70E-01	7.80E-01	7.70E-01	7.80E-01
	\mathcal{H} -Invertierung + \mathcal{H} -MVM [s]			
7938	2.87E+01	3.72E+01	4.59E+01	5.47E+01
32258	3.40E+02	5.10E+02	6.49E+02	8.00E+02
	Speicher \mathcal{H} -Inverse [kB/DoF]			
7938	7.90E+00	9.09E+00	1.01E+01	1.11E+01
32258	1.28E+01	1.53E+01	1.77E+01	1.97E+01
	Gesamtpeicher [MB]			
7938	1.84E+02	2.11E+02	2.36E+02	2.58E+02
32258	1.21E+03	1.45E+03	1.67E+03	1.86E+03
	$\ A \cdot A^{-\mathcal{H}} - Id\ _2$			
7938	2.83E+01	7.44E+00	5.78E+00	9.76E-01
32258	2.54E+02	6.07E+01	7.02E+01	6.51E+01
	$\ u - u_h\ _{L^2}$			
7938	6.54E-01	4.10E-01	2.22E-01	2.42E-02
32258	9.68E-01	5.88E-01	1.00E+00	9.59E-01
	C_{sp}			
7938	14	14	14	14
32258	14	14	14	14

Tabelle B.14: Meßdaten für rangadaptive \mathcal{H} -Invertierung (Teil 2/4).

Dof's	1.00E-09	1.00E-10	1.00E-11	1.00E-12
	Aufbau Clusterbaum [s]			
7938	0.00E+00	1.00E-02	0.00E+00	1.00E-02
32258	5.00E-02	5.00E-02	5.00E-02	5.00E-02
	Aufbau \mathcal{H} -Matrix-Struktur [s]			
7938	2.00E-02	2.00E-02	1.00E-02	1.00E-02
32258	4.00E-02	4.00E-02	3.00E-02	4.00E-02
	Konvertierung der Sparse- in \mathcal{H} -Matrix [s]			
7938	3.50E-01	2.20E-01	1.90E-01	2.30E-01
32258	7.80E-01	7.80E-01	7.80E-01	7.80E-01
	\mathcal{H} -Invertierung + \mathcal{H} -MVM [s]			
7938	6.37E+01	7.20E+01	7.96E+01	8.69E+01
32258	9.56E+02	1.13E+03	1.27E+03	1.40E+03
	Speicher \mathcal{H} -Inverse [kB/DoF]			
7938	1.21E+01	1.29E+01	1.37E+01	1.45E+01
32258	2.16E+01	2.35E+01	2.51E+01	2.67E+01
	Gesamtpeicher [MB]			
7938	2.80E+02	3.00E+02	3.20E+02	3.36E+02
32258	2.04E+03	2.22E+03	2.38E+03	2.52E+03
	$\ A \cdot A^{-\mathcal{H}} - Id\ _2$			
7938	7.12E-03	8.11E-04	2.49E-04	3.18E-06
32258	1.54E+00	3.40E-01	4.04E-03	3.93E-04
	$\ u - u_h\ _{L^2}$			
7938	1.69E-02	1.69E-02	1.69E-02	1.69E-02
32258	1.61E-02	9.75E-03	9.01E-03	9.00E-03
	C_{sp}			
7938	14	14	14	14
32258	14	14	14	14

Tabelle B.15: Meßdaten für rangadaptive \mathcal{H} -Invertierung (Teil 3/4).

Dof's	1.00E-13	1.00E-14	1.00E-15	1.00E-16
	Aufbau Clusterbaum [s]			
7938	1.00E-02	0.00E+00	1.00E-02	1.00E-02
32258	4.00E-02	5.00E-02	5.00E-02	5.00E-02
	Aufbau \mathcal{H} -Matrix-Struktur [s]			
7938	1.00E-02	1.00E-02	1.00E-02	1.00E-02
32258	3.00E-02	4.00E-02	4.00E-02	3.00E-02
	Konvertierung der Sparse- in \mathcal{H} -Matrix [s]			
7938	3.60E-01	3.00E-01	3.20E-01	2.90E-01
32258	7.70E-01	7.80E-01	7.70E-01	7.80E-01
	\mathcal{H} -Invertierung + \mathcal{H} -MVM [s]			
7938	9.41E+01	1.01E+02	1.08E+02	1.21E+02
32258	1.52E+03	1.62E+03	1.73E+03	1.98E+03
	Speicher \mathcal{H} -Inverse [kB/DoF]			
7938	1.51E+01	1.57E+01	1.62E+01	1.72E+01
32258	2.80E+01	2.93E+01	3.05E+01	3.26E+01
	Gesamtpeicher [MB]			
7938	3.51E+02	3.64E+02	3.77E+02	4.00E+02
32258	2.65E+03	2.77E+03	2.88E+03	3.08E+03
	$\ A \cdot A^{-\mathcal{H}} - Id\ _2$			
7938	8.14E-07	3.98E-07	1.52E-08	5.24E-09
32258	6.31E-06	1.06E-06	1.83E-07	1.32E-07
	$\ u - u_h\ _{L^2}$			
7938	1.69E-02	1.69E-02	1.69E-02	1.69E-02
32258	9.00E-03	9.00E-03	9.00E-03	9.00E-03
	C_{sp}			
7938	14	14	14	14
32258	14	14	14	14

Tabelle B.16: Meßdaten für rangadaptive \mathcal{H} -Invertierung (Teil 4/4).

B.2.6 $P_{\mathcal{H}\text{-Cholesky}}\text{CG}$: adaptiver Rang, Geometrie 3

DoF's	1.00E-05	1.00E-06	1.00E-07	1.00E-08	1.00E-09	1.00E-10
	Aufbau Clusterbaum [s]					
7938	1.00E-02	0.00E+00	1.00E-02	1.00E-02	1.00E-02	0.00E+00
32258	5.00E-02	3.00E-02	5.00E-02	5.00E-02	5.00E-02	5.00E-02
130050	2.00E-01	2.00E-01	2.10E-01	2.00E-01	1.90E-01	2.80E-01
	Aufbau \mathcal{H} -Matrix-Struktur [s]					
7938	2.00E-02	1.00E-02	1.00E-02	1.00E-02	1.00E-02	1.00E-02
32258	3.00E-02	3.00E-02	3.00E-02	3.00E-02	3.00E-02	3.00E-02
130050	1.30E-01	1.30E-01	1.30E-01	1.20E-01	1.30E-01	1.30E-01
	Konvertierung der Sparse- in \mathcal{H} -Matrix [s]					
7938	2.10E-01	2.10E-01	2.10E-01	1.80E-01	2.00E-01	2.10E-01
32258	4.30E-01	4.50E-01	4.30E-01	4.30E-01	4.30E-01	4.30E-01
130050	1.85E+00	1.85E+00	1.86E+00	1.86E+00	1.85E+00	1.85E+00
	\mathcal{H} -Cholesky-Faktor + CGM [s]					
7938	1.63E+00	1.74E+00	1.94E+00	2.02E+00	2.22E+00	2.28E+00
32258	1.34E+01	1.56E+01	1.75E+01	2.02E+01	2.31E+01	2.48E+01
130050	1.09E+02	1.68E+02	1.98E+02	2.27E+02	2.51E+02	2.81E+02
	Iterationsschritte					
7938	14	9	6	4	3	3
32258	16	10	8	5	4	3
130050	14	13	9	5	4	3
	Speicher \mathcal{H} -Cholesky Faktor [kB/DoF]					
7938	1.94E+00	2.04E+00	2.12E+00	2.19E+00	2.24E+00	2.28E+00
32258	2.45E+00	2.67E+00	2.81E+00	2.97E+00	3.10E+00	3.19E+00
130050	3.04E+00	3.48E+00	3.78E+00	4.05E+00	4.25E+00	4.47E+00
	Gesamtsspeicher [MB]					
7938	7.51E+00	7.90E+00	8.21E+00	8.48E+00	8.68E+00	8.85E+00
32258	3.85E+01	4.20E+01	4.43E+01	4.67E+01	4.89E+01	5.03E+01
130050	1.93E+02	2.21E+02	2.40E+02	2.57E+02	2.70E+02	2.84E+02
	Invertierungsfehler					
7938	3.42E+00	4.48E-01	4.56E-02	2.57E-03	3.33E-05	1.77E-05
32258	3.62E+01	1.50E+00	4.25E-01	1.66E-02	1.76E-03	7.42E-05
130050	8.48E+01	8.83E+00	1.11E+00	7.95E-02	1.32E-02	6.56E-04
	L^2 -Fehler					
7938	1.69E-02	1.69E-02	1.69E-02	1.69E-02	1.69E-02	1.69E-02
32258	9.00E-03	9.00E-03	9.00E-03	9.00E-03	9.00E-03	9.00E-03
130050	3.31E-03	3.31E-03	3.31E-03	3.31E-03	3.31E-03	3.31E-03
	C_{sp}					
7938	14	14	14	14	14	14
32258	14	14	14	14	14	14
130050	16	16	16	16	16	16

Tabelle B.17: $P_{\mathcal{H}\text{-Cholesky}}\text{CG}$: adaptiver Rang, Geometrie 3, Parametertyp 4

DoF's	1.00E-11	1.00E-12	1.00E-13	1.00E-14	1.00E-15	1.00E-16
	Aufbau Clusterbaum [s]					
7938	1.00E-02	1.00E-02	1.00E-02	1.00E-02	1.00E-02	1.00E-02
32258	3.00E-02	5.00E-02	6.00E-02	4.00E-02	5.00E-02	5.00E-02
130050	2.10E-01	2.00E-01	1.90E-01	2.00E-01	2.00E-01	2.00E-01
	Aufbau H-Matrix-Struktur [s]					
7938	1.00E-02	1.00E-02	0.00E+00	1.00E-02	2.00E-02	1.00E-02
32258	3.00E-02	3.00E-02	3.00E-02	3.00E-02	3.00E-02	4.00E-02
130050	1.30E-01	1.30E-01	1.30E-01	1.30E-01	1.30E-01	1.30E-01
	Konvertierung der Sparse- in H-Matrix [s]					
7938	1.20E-01	1.70E-01	1.20E-01	2.10E-01	1.70E-01	2.20E-01
32258	4.30E-01	4.30E-01	4.30E-01	4.30E-01	4.40E-01	4.30E-01
130050	1.85E+00	1.85E+00	1.85E+00	1.86E+00	1.86E+00	1.85E+00
	H-Cholesky-Faktor + CGM [s]					
7938	2.44E+00	2.51E+00	2.67E+00	2.88E+00	2.95E+00	3.09E+00
32258	2.65E+01	2.81E+01	2.97E+01	3.13E+01	3.27E+01	3.50E+01
130050	3.12E+02	3.47E+02	3.81E+02	4.04E+02	4.23E+02	4.47E+02
	Iterationsschritte					
7938	3	2	2	2	2	2
32258	3	3	2	2	2	2
130050	3	2	2	2	2	2
	Speicher H-Cholesky Faktor [kB/DoF]					
7938	2.33E+00	2.37E+00	2.42E+00	2.46E+00	2.49E+00	2.53E+00
32258	3.27E+00	3.35E+00	3.42E+00	3.50E+00	3.56E+00	3.63E+00
130050	4.68E+00	4.88E+00	5.05E+00	5.18E+00	5.30E+00	5.41E+00
	Gesamtpeicher [MB]					
7938	9.02E+00	9.19E+00	9.37E+00	9.54E+00	9.80E+00	9.80E+00
32258	5.15E+01	5.27E+01	5.39E+01	5.51E+01	5.61E+01	5.71E+01
130050	2.97E+02	3.10E+02	3.20E+02	3.29E+02	3.36E+02	3.43E+02
	Invertierungsfehler					
7938	1.72E-06	3.79E-07	3.00E-08	3.32E-09	1.01E-09	4.66E-10
32258	1.28E-05	7.89E-07	1.70E-07	1.56E-08	3.99E-09	2.98E-09
130050	7.04E-05	2.97E-06	3.45E-07	2.21E-08	1.24E-08	1.03E-08
	L^2 -Fehler					
7938	1.69E-02	1.69E-02	1.69E-02	1.69E-02	1.69E-02	1.69E-02
32258	9.00E-03	9.00E-03	9.00E-03	9.00E-03	9.00E-03	9.00E-03
130050	3.31E-03	3.31E-03	3.31E-03	3.31E-03	3.31E-03	3.31E-03
	C_{sp}					
7938	14	14	14	14	14	14
32258	14	14	14	14	14	14
130050	16	16	16	16	16	16

Tabelle B.18: $P_{\mathcal{H}\text{-Cholesky}}$ CG: adaptiver Rang, Geometrie 3, Parametertyp 4

B.2.7 $P_{\mathcal{H}\text{-Cholesky}}\text{CG}$: adaptiver Rang, Geometrie 6

DoF's	1.00E-05	1.00E-06	1.00E-07	1.00E-08	1.00E-09	1.00E-10
	Aufbau Clusterbaum [s]					
5058	1.00E-02	0.00E+00	0.00E+00	1.00E-02	1.00E-02	0.00E+00
20354		3.00E-02	2.00E-02	2.00E-02	3.00E-02	1.00E-02
81666		1.80E-01	1.10E-01	1.90E-01	1.20E-01	1.20E-01
	Aufbau \mathcal{H} -Matrix-Struktur [s]					
5058	1.00E-02	0.00E+00	0.00E+00	1.00E-02	0.00E+00	0.00E+00
20354		2.00E-02	3.00E-02	2.00E-02	1.00E-02	2.00E-02
81666		5.00E-02	5.00E-02	6.00E-02	6.00E-02	5.00E-02
	Konvertierung der Sparse- in \mathcal{H} -Matrix [s]					
5058	1.70E-01	1.60E-01	1.60E-01	2.00E-01	2.00E-01	2.10E-01
20354		3.60E-01	3.90E-01	4.30E-01	3.60E-01	3.80E-01
81666		1.33E+00	1.33E+00	1.32E+00	1.32E+00	1.33E+00
	\mathcal{H} -Cholesky-Faktor + CGM [s]					
5058	1.47E+00	1.56E+00	1.48E+00	1.59E+00	1.61E+00	1.63E+00
20354		5.42E+00	1.59E+01	1.70E+01	1.81E+01	1.89E+01
81666		1.55E+02	1.75E+02	1.91E+02	2.07E+02	2.23E+02
	Iterationsschritte					
5058	22	12	6	5	4	3
20354		14	7	5	4	3
81666		14	9	6	4	4
	Speicher \mathcal{H} -Cholesky Faktor [kB/DoF]					
5058	1.96E+00	2.03E+00	2.10E+00	2.14E+00	2.17E+00	2.19E+00
20354		2.89E+00	3.03E+00	3.14E+00	3.25E+00	3.32E+00
81666		4.21E+00	4.47E+00	4.67E+00	4.86E+00	5.03E+00
	Gesamtsspeicher [MB]					
5058	4.84E+00	5.00E+00	5.18E+00	5.29E+00	5.36E+00	5.42E+00
20354	0.00E+00	2.87E+01	3.01E+01	3.12E+01	3.23E+01	3.30E+01
81666	0.00E+00	1.68E+02	1.78E+02	1.86E+02	1.94E+02	2.01E+02
	Invertierungsfehler					
5058	1.45E+01	1.68E+00	1.35E-01	6.48E-03	7.73E-04	5.69E-05
20354		5.42E+00	3.59E-01	4.12E-02	3.27E-03	4.78E-04
81666		8.17E+00	1.13E+00	6.71E-02	7.61E-03	4.90E-04
	L^2 -Fehler					
5058	1.22E-02	1.22E-02	1.22E-02	1.22E-02	1.22E-02	1.22E-02
20354		6.82E-03	6.82E-03	6.82E-03	6.82E-03	6.82E-03
81666		3.22E-03	3.22E-03	3.22E-03	3.22E-03	3.22E-03
	C_{sp}					
5058	14	14	14	14	14	14
20354		18	18	18	18	18
81666		18	18	18	18	18

Tabelle B.19: $P_{\mathcal{H}\text{-Cholesky}}\text{CG}$: adaptiver Rang, Geometrie 6, Parametertyp 4

DoF's	1.00E-11	1.00E-12	1.00E-13	1.00E-14	1.00E-15	1.00E-16
	Aufbau Clusterbaum [s]					
5058	0.00E+00	1.00E-02	1.00E-02	1.00E-02	1.00E-02	1.00E-02
20354	3.00E-02	3.00E-02	3.00E-02	3.00E-02	3.00E-02	2.00E-02
81666	1.30E-01	1.90E-01	1.10E-01	1.10E-01	1.30E-01	1.40E-01
	Aufbau \mathcal{H} -Matrix-Struktur [s]					
5058	0.00E+00	1.00E-02	1.00E-02	0.00E+00	1.00E-02	1.00E-02
20354	2.00E-02	1.00E-02	1.00E-02	2.00E-02	2.00E-02	2.00E-02
81666	6.00E-02	6.00E-02	5.00E-02	5.00E-02	6.00E-02	6.00E-02
	Konvertierung der Sparse- in \mathcal{H} -Matrix [s]					
5058	2.00E-01	1.70E-01	1.30E-01	1.80E-01	1.60E-01	1.80E-01
20354	3.60E-01	3.60E-01	3.60E-01	3.60E-01	3.50E-01	3.80E-01
81666	1.33E+00	1.32E+00	1.33E+00	1.33E+00	1.32E+00	1.33E+00
	\mathcal{H} -Cholesky-Faktor + CGM [s]					
5058	1.66E+00	1.64E+00	1.67E+00	1.80E+00	1.71E+00	1.73E+00
20354	1.97E+01	2.06E+01	2.13E+01	2.18E+01	2.23E+01	2.28E+01
81666	2.36E+02	2.49E+02	2.61E+02	2.73E+02	2.84E+02	2.96E+02
	Iterationsschritte					
5058	3	2	2	2	2	2
20354	3	3	2	2	2	2
81666	3	3	2	2	2	2
	Speicher \mathcal{H} -Cholesky Faktor [kB/DoF]					
5058	2.22E+00	2.23E+00	2.25E+00	2.26E+00	2.28E+00	2.28E+00
20354	3.38E+00	3.45E+00	3.50E+00	3.53E+00	3.57E+00	3.61E+00
81666	5.18E+00	5.31E+00	5.44E+00	5.55E+00	5.65E+00	5.76E+00
	Gesamtsspeicher [MB]					
5058	5.48E+00	5.51E+00	5.56E+00	5.59E+00	5.62E+00	5.64E+00
20354	3.36E+01	3.43E+01	3.48E+01	3.51E+01	3.55E+01	3.58E+01
81666	2.06E+02	2.12E+02	2.17E+02	2.21E+02	2.25E+02	2.30E+02
	Invertierungsfehler					
5058	4.47E-06	2.97E-07	7.54E-08	6.24E-09	3.53E-09	2.97E-09
20354	5.29E-05	3.16E-06	3.81E-07	1.56E-08	3.99E-09	3.62E-09
81666	4.32E-05	4.27E-06	4.19E-07	2.92E-08	5.77E-09	5.35E-09
	L^2 -Fehler					
5058	1.22E-02	1.22E-02	1.22E-02	1.22E-02	1.22E-02	1.22E-02
20354	6.82E-03	6.82E-03	6.82E-03	6.82E-03	6.82E-03	6.82E-03
81666	3.22E-03	3.22E-03	3.22E-03	3.22E-03	3.22E-03	3.22E-03
	C_{sp}					
5058	14	14	14	14	14	14
20354	18	18	18	18	18	18
81666	18	18	18	18	18	18

Tabelle B.20: $P_{\mathcal{H}\text{-Cholesky}}\text{CG}$: adaptiver Rang, Geometrie 6, Parametertyp 4

B.2.8 $P_{\mathcal{H}\text{-Cholesky}}\text{CG}$: adaptiver Rang, Geometrie 7

DoF's	1.00E-05	1.00E-06	1.00E-07	1.00E-08	1.00E-09	1.00E-10
	Aufbau Clusterbaum [s]					
12178	1.00E-02	1.00E-02	1.00E-02	1.00E-02	1.00E-02	1.00E-02
49058		1.00E-01	6.00E-02	8.00E-02	6.00E-02	9.00E-02
197632		3.30E-01	3.30E-01	3.30E-01	3.30E-01	3.30E-01
	Aufbau \mathcal{H} -Matrix-Struktur [s]					
12178	2.00E-02	1.00E-02	2.00E-02	2.00E-02	2.00E-02	2.00E-02
49058		3.00E-02	4.00E-02	4.00E-02	4.00E-02	3.00E-02
197632		1.30E-01	1.30E-01	1.30E-01	2.20E-01	1.30E-01
	Konvertierung der Sparse- in \mathcal{H} -Matrix [s]					
12178	2.10E-01	3.20E-01	2.90E-01	2.90E-01	2.00E-01	2.10E-01
49058		7.30E-01	7.20E-01	7.20E-01	7.20E-01	7.30E-01
197632		3.00E+00	3.01E+00	3.00E+00	6.50E+00	3.01E+00
	\mathcal{H} -Cholesky-Faktor + CGM [s]					
12178	2.76E+00	2.82E+00	3.11E+00	3.34E+00	3.73E+00	3.93E+00
49058		2.18E+01	2.45E+01	2.80E+01	3.17E+01	3.53E+01
197632		1.77E+02	1.98E+02	2.21E+02	2.47E+02	2.76E+02
	Iterationsschritte					
12178	19	9	6	4	4	3
49058		13	8	5	4	3
197632		18	11	6	4	3
	Speicher \mathcal{H} -Cholesky Faktor [kB/DoF]					
12178	1.79E+00	1.90E+00	2.01E+00	2.08E+00	2.16E+00	2.23E+00
49058		2.40E+00	2.57E+00	2.74E+00	2.89E+00	3.02E+00
197632		3.17E+00	3.40E+00	3.61E+00	3.82E+00	4.02E+00
	Gesamtspeicher [MB]					
12178	1.06E+01	1.13E+01	1.19E+01	1.24E+01	1.29E+01	1.32E+01
49058	0.00E+00	5.74E+01	6.15E+01	6.55E+01	6.91E+01	7.24E+01
197632	0.00E+00	3.06E+02	3.28E+02	3.49E+02	3.68E+02	3.88E+02
	Invertierungsfehler					
12178	1.61E+00	2.11E-01	1.25E-02	1.72E-03	1.12E-04	8.34E-06
49058		1.09E+00	1.27E-01	6.81E-03	1.15E-03	6.77E-05
197632		4.80E+00	6.88E-01	5.89E-02	5.64E-03	3.13E-04
	L^2 -Fehler					
12178	1.78E-03	1.78E-03	1.78E-03	1.78E-03	1.78E-03	1.78E-03
49058		1.15E-03	1.15E-03	1.15E-03	1.15E-03	1.15E-03
197632		6.03E-04	6.03E-04	6.03E-04	6.03E-04	6.03E-04
	C_{sp}					
12178	16	16	16	16	16	16
49058		18	18	18	18	18
197632		18	18	18	18	18

Tabelle B.21: $P_{\mathcal{H}\text{-Cholesky}}\text{CG}$: adaptiver Rang, Geometrie 7, Parametertyp 4; $\varepsilon = 10^{-5}$ ist für die beiden großen Gitter nicht rechenbar.

DoF's	1.00E-11	1.00E-12	1.00E-13	1.00E-14	1.00E-15	1.00E-16
	Aufbau Clusterbaum [s]					
12178	0.00E+00	0.00E+00	0.00E+00	1.00E-02	1.00E-02	1.00E-02
49058	1.00E-01	9.00E-02	9.00E-02	9.00E-02	9.00E-02	9.00E-02
197632	3.30E-01	3.30E-01	3.30E-01	3.30E-01	3.40E-01	3.30E-01
	Aufbau \mathcal{H} -Matrix-Struktur [s]					
12178	2.00E-02	1.00E-02	1.00E-02	1.00E-02	1.00E-02	1.00E-02
49058	4.00E-02	3.00E-02	3.00E-02	3.00E-02	3.00E-02	4.00E-02
197632	1.30E-01	1.30E-01	1.30E-01	1.30E-01	2.20E-01	1.30E-01
	Konvertierung der Sparse- in \mathcal{H} -Matrix [s]					
12178	3.20E-01	2.10E-01	2.00E-01	1.80E-01	3.30E-01	2.60E-01
49058	7.20E-01	7.20E-01	7.30E-01	7.30E-01	7.30E-01	7.30E-01
197632	3.00E+00	3.01E+00	3.01E+00	3.00E+00	6.50E+00	3.01E+00
	\mathcal{H} -Cholesky-Faktor + CGM [s]					
12178	4.30E+00	4.65E+00	4.87E+00	5.17E+00	5.39E+00	5.64E+00
49058	3.92E+01	4.33E+01	4.71E+01	5.17E+01	5.62E+01	6.09E+01
197632	3.08E+02	3.40E+02	3.73E+02	4.08E+02	4.49E+02	4.95E+02
	Iterationsschritte					
12178	3	3	2	2	2	2
49058	3	3	2	2	2	2
197632	3	3	2	2	2	2
	Speicher \mathcal{H} -Cholesky Faktor [kB/DoF]					
12178	2.30E+00	2.37E+00	2.43E+00	2.49E+00	2.53E+00	2.57E+00
49058	3.15E+00	3.28E+00	3.39E+00	3.52E+00	3.63E+00	3.74E+00
197632	4.22E+00	4.40E+00	4.58E+00	4.76E+00	4.95E+00	5.16E+00
	Gesamtpeicher [MB]					
12178	1.37E+01	1.41E+01	1.45E+01	1.48E+01	1.51E+01	1.53E+01
49058	7.55E+01	7.86E+01	8.13E+01	8.42E+01	8.69E+01	8.95E+01
197632	4.07E+02	4.25E+02	4.42E+02	4.59E+02	4.78E+02	4.98E+02
	Invertierungsfehler					
12178	1.34E-06	2.28E-07	1.93E-08	1.83E-09	5.04E-10	5.00E-10
49058	5.51E-06	7.45E-07	7.96E-08	1.10E-08	1.30E-09	1.47E-09
197632	3.09E-05	2.50E-06	2.44E-07	2.48E-08	4.74E-09	4.87E-09
	L^2 -Fehler					
12178	1.78E-03	1.78E-03	1.78E-03	1.78E-03	1.78E-03	1.78E-03
49058	1.15E-03	1.15E-03	1.15E-03	1.15E-03	1.15E-03	1.15E-03
197632	6.03E-04	6.03E-04	6.03E-04	6.03E-04	6.03E-04	6.03E-04
	C_{sp}					
12178	16	16	16	16	16	16
49058	18	18	18	18	18	18
197632	18	18	18	18	18	18

Tabelle B.22: $P_{\mathcal{H}\text{-Cholesky}}\text{CG}$: adaptiver Rang, Geometrie 7, Parametertyp 4

B.2.9 $P_{\mathcal{H}\text{-Cholesky}}\text{CG}$: adaptiver Rang, Geometrie 8

DoF's	1.00E-05	1.00E-06	1.00E-07	1.00E-08	1.00E-09	1.00E-10
	Aufbau Clusterbaum [s]					
16338	2.00E-02	1.00E-02	2.00E-02	1.00E-02	2.00E-02	1.00E-02
65826	9.00E-02	1.40E-01	9.00E-02	8.00E-02	9.00E-02	1.20E-01
264258	5.20E-01	5.10E-01	5.00E-01	5.10E-01	5.10E-01	5.20E-01
	Aufbau \mathcal{H} -Matrix-Struktur [s]					
16338	1.00E-02	1.00E-02	1.00E-02	2.00E-02	2.00E-02	2.00E-02
65826	4.00E-02	4.00E-02	4.00E-02	4.00E-02	4.00E-02	5.00E-02
264258	1.80E-01	1.90E-01	1.90E-01	1.80E-01	1.90E-01	1.80E-01
	Konvertierung der Sparse- in \mathcal{H} -Matrix [s]					
16338	3.10E-01	2.20E-01	2.20E-01	3.20E-01	2.80E-01	2.20E-01
65826	9.70E-01	9.70E-01	9.80E-01	9.80E-01	9.70E-01	9.70E-01
264258	3.95E+00	3.95E+00	3.95E+00	3.95E+00	3.94E+00	3.96E+00
	\mathcal{H} -Cholesky-Faktor + CGM [s]					
16338	9.87E+00	1.16E+01	1.27E+01	1.34E+01	1.38E+01	1.42E+01
65826	1.00E+02	1.18E+02	1.33E+02	1.46E+02	1.56E+02	1.64E+02
264258	7.90E+02	1.02E+03	1.23E+03	1.43E+03	1.61E+03	1.78E+03
	Iterationsschritte					
16338	7	5	4	3	3	2
65826	8	6	4	3	3	3
264258	11	6	5	4	3	3
	Speicher \mathcal{H} -Cholesky Faktor [kB/DoF]					
16338	2.60E+00	2.72E+00	2.80E+00	2.87E+00	2.92E+00	2.96E+00
65826	3.74E+00	4.00E+00	4.21E+00	4.38E+00	4.50E+00	4.59E+00
264258	4.99E+00	5.59E+00	6.11E+00	6.56E+00	6.93E+00	7.26E+00
	Gesamtsspeicher [MB]					
16338	2.08E+01	2.17E+01	2.23E+01	2.29E+01	2.33E+01	2.37E+01
65826	1.20E+02	1.29E+02	1.35E+02	1.41E+02	1.45E+02	1.48E+02
264258	6.44E+02	7.22E+02	7.88E+02	8.47E+02	8.94E+02	9.37E+02
	Invertierungsfehler					
16338	1.26E-01	1.02E-02	6.64E-04	6.42E-05	3.06E-06	2.37E-07
65826	5.59E-01	3.97E-02	3.42E-03	1.90E-04	1.51E-05	1.19E-06
264258	3.11E+00	2.06E-01	1.22E-02	7.68E-04	5.29E-05	4.62E-06
	L^2 -Fehler					
16338	7.51E-03	7.51E-03	7.51E-03	7.51E-03	7.51E-03	7.51E-03
65826	3.05E-03	3.05E-03	3.05E-03	3.05E-03	3.05E-03	3.05E-03
264258	1.09E-03	1.09E-03	1.09E-03	1.09E-03	1.09E-03	1.09E-03
	C_{sp}					
16338	18	18	18	18	18	18
65826	18	18	18	18	18	18
264258	18	18	18	18	18	18

Tabelle B.23: $P_{\mathcal{H}\text{-Cholesky}}\text{CG}$: adaptiver Rang, Geometrie 8, Parametertyp 4.

DoF's	1.00E-11	1.00E-12	1.00E-13	1.00E-14	1.00E-15	1.00E-16
Aufbau Clusterbaum [s]						
16338	2.00E-02	2.00E-02	1.00E-02	2.00E-02	2.00E-02	2.00E-02
65826	8.00E-02	1.30E-01	1.20E-01	1.40E-01	1.30E-01	1.30E-01
264258	5.10E-01	5.00E-01	5.30E-01	5.40E-01	5.10E-01	5.00E-01
Aufbau \mathcal{H} -Matrix-Struktur [s]						
16338	2.00E-02	1.00E-02	1.00E-02	1.00E-02	1.00E-02	2.00E-02
65826	4.00E-02	5.00E-02	5.00E-02	5.00E-02	5.00E-02	4.00E-02
264258	1.90E-01	1.80E-01	3.10E-01	3.20E-01	1.80E-01	1.90E-01
Konvertierung der Sparse- in \mathcal{H} -Matrix [s]						
16338	2.70E-01	2.20E-01	2.20E-01	2.20E-01	2.20E-01	2.30E-01
65826	9.80E-01	9.70E-01	9.60E-01	9.70E-01	9.60E-01	9.70E-01
264258	3.95E+00	3.94E+00	8.52E+00	8.53E+00	3.95E+00	3.96E+00
\mathcal{H} -Cholesky-Faktor + CGM [s]						
16338	1.46E+01	1.49E+01	1.53E+01	1.57E+01	1.56E+01	1.63E+01
65826	1.70E+02	1.76E+02	1.82E+02	1.87E+02	1.91E+02	1.97E+02
264258	1.94E+03	2.08E+03	2.19E+03	2.30E+03	2.40E+03	2.48E+03
Iterationsschritte						
16338	2	2	2	2	2	2
65826	2	2	2	2	2	2
264258	2	2	2	2	2	2
Speicher \mathcal{H} -Cholesky Faktor [kB/DoF]						
16338	3.00E+00	3.03E+00	3.06E+00	3.09E+00	3.12E+00	3.15E+00
65826	4.67E+00	4.75E+00	4.82E+00	4.88E+00	4.94E+00	5.01E+00
264258	7.54E+00	7.78E+00	7.99E+00	8.17E+00	8.35E+00	8.51E+00
Gesamtsspeicher [MB]						
16338	2.40E+01	2.42E+01	2.44E+01	2.47E+01	2.49E+01	2.51E+01
65826	1.50E+02	1.53E+02	1.55E+02	1.57E+02	1.59E+02	1.61E+02
264258	9.73E+02	1.00E+03	1.03E+03	1.05E+03	1.08E+03	1.10E+03
Invertierungsfehler						
16338	1.97E-08	2.83E-09	3.33E-10	2.67E-10	2.60E-10	2.66E-10
65826	1.21E-07	6.91E-09	1.66E-09	1.41E-09	1.41E-09	1.42E-09
264258	3.88E-07	3.96E-08	7.42E-09	6.98E-09	7.00E-09	7.00E-09
L^2 -Fehler						
16338	7.51E-03	7.51E-03	7.51E-03	7.51E-03	7.51E-03	7.51E-03
65826	3.05E-03	3.05E-03	3.05E-03	3.05E-03	3.05E-03	3.05E-03
264258	1.09E-03	1.09E-03	1.09E-03	1.09E-03	1.09E-03	1.09E-03
C_{sp}						
16338	18	18	18	18	18	18
65826	18	18	18	18	18	18
264258	18	18	18	18	18	18

Tabelle B.24: $P_{\mathcal{H}\text{-Cholesky}}\text{CG}$: adaptiver Rang, Geometrie 8, Parametertyp 4

Literaturverzeichnis

- [AA94] Altenbach J., Altenbach H.
Einführung in die Kontinuumsmechanik
Stuttgart, 1994
Zitiert auf S. -.
- [Ba02] Bathe K.-J.
Finite-Elemente-Methoden
Heidelberg, 2002
Zitiert auf S. 33.
- [Be03] Bebendorf M.
A Note on the Poincaré Inequality for Convex Domains (Technical Report)
Leipzig, 2003
Zitiert auf S. 130.
- [Be04] Bebendorf M.
Efficient Inversion of the Galerkin Matrix of General Second Order Elliptic Operators with Nonsmooth Coefficients (Technical Report)
Leipzig, 2004
Zitiert auf S. 1.
- [Be05] Bebendorf M.
Why approximate LU decompositions of finite element discretizations of elliptic operators can be computed with almost linear complexity (Technical Report)
Leipzig, 2005
Zitiert auf S. 103.
- [BGH02] Börm S., Grasedyck L., Hackbusch W.
Introduction to Hierarchical Matrices with Applications (Technical Report)
Leipzig, 2002
Zitiert auf S. 35, 43.
- [BGH03] Börm S., Grasedyck L., Hackbusch W.
Hierarchical Matrices (Lecture Note)
Leipzig, 2003
Zitiert auf S. 37.

- [BH03] Bebendorf M., Hackbusch W.
Existence of \mathcal{H} -Matrix Approximants to the Inverse FE-Matrix of Elliptic Operators with L^∞ -Coefficients
Numerische Mathematik 95 (1/2003), S. 1 – 28
Zitiert auf S. 1, 2, 4, 74, 94.
- [Br97] Braess D.
Finite Elemente
Berlin, 1997
Zitiert auf S. 19, 25, 129.
- [BS02] Brenner S. C., Scott L. R.
The Mathematical Theory of Finite Elements
New York, 2002
Zitiert auf S. 25, 58, 60, 129, 130.
- [CZ92] Chen G., Zhou J.
Boundary Element Methods
London, 1992
Zitiert auf S. 7.
- [DL88] Dautray R., Lions J.-L.
Mathematical Analysis and Numerical Methods for Science and Technology,
Vol. 2
Heidelberg, 1988
Zitiert auf S. 130.
- [DM95] Dolzmann G., Müller S.
Estimates for Green's Matrices of Elliptic Systems by L^p Theory
Manuscripta Mathematica 88 (1995), S. 261 – 273
Zitiert auf S. 15.
- [Fi72] Fichera G.
Existence Theorems in Elasticity
in/aus *Flügge S., Handbuch der Physik, Band VIa/2*
Berlin, 1972
Zitiert auf S. -.
- [GH02] Grasedyck L. , Hackbusch W.
Construction and Arithmetics of \mathcal{H} -Matrices (Technical Report)
Leipzig, 2002
Zitiert auf S. 35, 43, 47, 103.
- [Gi83] Giaquinta M.
Multiple Integrals in the Calculus of Variations and Nonlinear Elliptic Systems
Princeton, 1983
Zitiert auf S. 14.

- [Gi93] Giaquinta M.
Introduction to Regularity Theory for Nonlinear Elliptic Systems
Basel, 1993
Zitiert auf S. 14.
- [GKL05] Grasedyck L., Kriemann R., Le Borne S.
Parallel Black Box Domain Decomposition Based \mathcal{H} -LU Preconditioning (Technical Report)
Leipzig, 2005
Zitiert auf S. 103.
- [GL96] Golub G. H., van Loan C. F.
Matrix Computations
Baltimore, 1996
Zitiert auf S. -.
- [Gr01] Grasedyck L.
Theorie und Anwendung Hierarchischer Matrizen (Dissertation)
Kiel, 2001
Zitiert auf S. 8, 10, 35, 39, 40, 41, 42, 44, 45, 46, 47, 48.
- [GT01] Gilbarg D., Trudinger N. S.
Elliptic Partial Differential Equations of Second Order
Berlin, 2001
Zitiert auf S. 129, 130.
- [GW82] Grüter M., Widman K.-O.
The Green Function for Uniformly Elliptic Equations
Manuscripta Mathematica 37 (1982), S. 303 – 342
Zitiert auf S. 10.
- [Ha87] Hartmann F.
Methode der Randelemente
Berlin, 1987
Zitiert auf S. 7.
- [Hb93] Hackbusch W.
Iterative Lösung großer schwachbesetzter Gleichungssysteme
Stuttgart, 1993
Zitiert auf S. 128.
- [Hb96] Hackbusch W.
Theorie und Numerik elliptischer Differentialgleichungen
Stuttgart, 1996
Zitiert auf S. 9, 11, 12, 13, 14, 19, 20, 21, 22, 23, 24, 25.
- [Hb97] Hackbusch W.
Integralgleichungen — Theorie und Numerik

- Stuttgart, 1997
Zitiert auf S. 8, 15, 37.
- [Hb98] Hackbusch W.
A Sparse Matrix Arithmetic Based on \mathcal{H} -Matrices. Part I: Introduction to \mathcal{H} -Matrices (Technical Report)
Kiel, 1998
Zitiert auf S. 1, 35, 62.
- [Hb06] Hackbusch W.
Hierarchische Matrizen — Algorithmen und Analysis
www.mis.mpg.de/scicomp/Fulltext/hmvorlesung.ps
Leipzig, 2006
Zitiert auf S. 103.
- [HK99] Hackbusch W., Khoromskij B. N.
A Sparse \mathcal{H} -Matrix Arithmetic, Part II: Application to Multi-Dimensional Problems (Technical Report)
Leipzig, 1999
Zitiert auf S. 1, 35, 62.
- [Kh01] Khoromskij B. N.
Data-sparse Approximate Inverse in Elliptic Problems: Green's Function Approach (Technical Report)
Leipzig, 2001
Zitiert auf S. -.
- [Kö93] Königsberger K.
Analysis 2
Berlin 1993
Zitiert auf S. 57, 58.
- [Li02] Lintner M.
Lösung der 2D Wellengleichung mittels hierarchischer Matrizen (Dissertation)
München, 2002
Zitiert auf S. 1.
- [McL00] McLean W.
Strongly Elliptic Systems and Boundary Integral Equations
Cambridge, 2000
Zitiert auf S. 7, 11, 14.
- [Or80] Ortner N.
Regularisierte Faltung von Distributionen
Teil 1: Zur Berechnung von Fundamentallösungen
Teil 2: Eine Tabelle von Fundamentallösungen
Zeitschrift für angewandte Mathematik und Physik, Vol 31 (1980), S. 133 – 154

und S. 155 – 173

Zitiert auf S. 7.

- [PW60] Payne L. E., Weinberger H. F.
An Optimal Poincaré Inequality for convex Domains
Archive for Rational Mechanics and Analysis 5 (1960), S. 286 – 292
Zitiert auf S. 130.
- [RR93] Renardy M., Rogers R. C.
An Introduction to Partial Differential Equations
New York, 1993
Zitiert auf S. 10.
- [Sh92] Shimakura N.
Partial Differential Operators of Elliptic Type
Providence, 1992
Zitiert auf S. 10.
- [SS04] Sauter St., Schwab Chr.
Randelementmethoden
Wiesbaden, 2004
Zitiert auf S. 7.
- [Ve76] Velte W.
Direkte Methoden der Variationsrechnung
Stuttgart, 1976
Zitiert auf S. 129, 130.
- [Wlo82] Wloka J.
Partielle Differentialgleichungen
Stuttgart, 1982
Zitiert auf S. 130.
- [WRL95] Wloka J., Rowley B., Lawruk B.
Boundary Value Problems for Elliptic Systems
Cambridge, 1995
Zitiert auf S. 14.
- [Ze97] Zeidler E.
Applied Functional Analysis
New York, 1997
Zitiert auf S. 7.