Nichtkonforme finite Elemente und Doedel-Kollokation für elliptische Differentialgleichungen

- Diplomarbeit -

Eingereicht am Fachbereich
Mathematik und Informatik
der Philipps-Universität Marburg
von

Bastian Goldlücke

Betreuer: Prof. Dr. Klaus Böhmer

Zusammenfassung

Unter Zuhilfenahme von Techniken zu finiten Elementen mit *'Variational Crimes'* aus der Arbeit [B] von Klaus Böhmer wird eine Konvergenztheorie für eine Klasse nichtkonformer Diskretisierungen elliptischer Randwertprobleme erarbeitet. Diese zeichnet sich dadurch aus, daß die approximierenden Funktionen zwar im Inneren der finiten Elemente glatt sind, auf Übergängen zwischen zwei Elementen jedoch nur in endlich vielen Punkten stetig mit stetiger Normalenableitung sein müssen. Zu dieser Klasse zählt insbesondere ein effizientes Kollokationsverfahren, welches von Eusebius Doedel 1997 in [D] vorgestellt worden ist, für das die Konvergenz bisher aber noch nicht sichergestellt war. Einige numerische Beispielrechnungen mit einem eigens entwickelten Programmpaket illustrieren die theoretischen Resultate.

Erklärung

Ich versichere, die Arbeit selbständig verfaßt und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt zu haben.

Marburg, den 14.11.2001,

Inhaltsverzeichnis

Kapitel 0

Einführung

0.0 Problem und Diskretisierung

In dieser Arbeit geht es im wesentlichen um ein konkretes Lösungsverfahren für partielle Differentialgleichungen, welches in [D] von E.Doedel beschrieben worden ist. Es fügt sich ein in den Rahmen der Theorie der nichtkonformen finiten Elemente, wobei die zu lösenden Gleichungssysteme durch Kollokation entstehen. Im folgenden wird es sich stets darum drehen, die Gleichung

$$Au = F$$

für u zu lösen, wobei A ein partieller Differentialoperator, u die gesuchte, genügend oft differenzierbare Funktion auf einem Gebiet $\Omega \subset \mathbb{R}^D$ und $F \in L^\infty(\Omega)$ ist. Um eine eindeutige Lösung zu gewährleisten, werden später noch geeignete Randbedingungen mit ins Spiel kommen, und die Klasse der betrachteten Operatoren und Gebiete eingeschränkt.

Wollen wir diese Gleichung 'rechnertauglich' machen, so muß sie in eine Gleichung für endlichdimensionale Funktionenräume überführt werden, da Computer betrüblicherweise mit unendlichen Dingen herzlich wenig anfangen können - es mangelt zum Beispiel an unendlich großen Speichern. Die Technik der finiten Elemente basiert nun darauf, dies zu erreichen, indem das Gebiet Ω geeignet in endlich viele Teilgebiete - die finiten Elemente - zerlegt wird, und dann ein endlichdimensionaler Funktionenraum konstruiert wird, indem jedem Element ein gewisser Raum lokaler Funktionen zugeordnet wird. Dieser hat natürlich ebenfalls endlichdimensional zu sein und wird vernünftigerweise so gewählt, daß man streßfrei damit arbeiten kann, weswegen sich zumeist Polynome als Basisfunktionen anbieten.

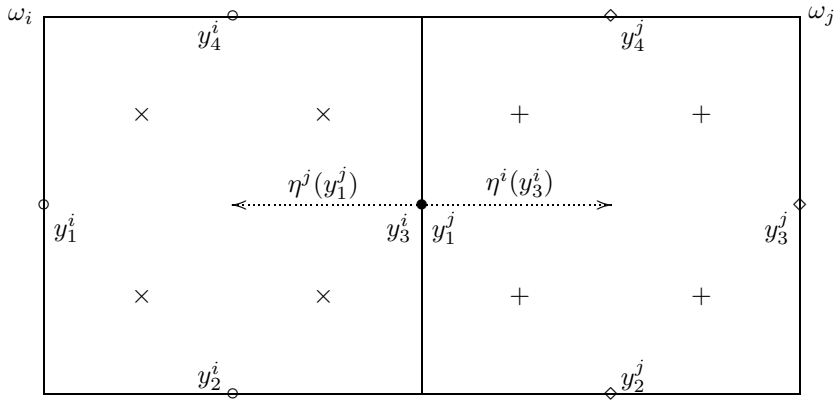
Das Gleichungssystem für diesen Funktionenraum ist dann dadurch gegeben, daß die Differentialgleichung in bestimmten endlich vielen Punkten, den sogenannten Kollokationsstellen, erfüllt ist - die dahintersteckende Idee ist, daß der Differentialoperator für die einfachen Basisfunktionen leicht auszuwerten ist. Wie diese Betrachtungen konkret in exakte Definitionen übersetzt werden, mit denen man arbeiten kann, offenbaren die restlichen Abschnitte dieses einführenden Kapitels.

0.1 Geometrische Situation und Notation

Zunächst soll die geometrische Situation in maximal möglicher Allgemeinheit beschrieben werden, notwendig werdende Restriktionen werden dann an geeigneter Stelle eingeführt. Wir fixieren im folgenden natürliche Zahlen D, F, N und M und bezeichnen mit

Ω	Ein Gebiet im $\mathbb{R}^D$, zerlegt in F paarweise disjunkte, offene Mengen $\omega_1, \ldots, \omega_F$ in dem Sinne, daß $\bigcup_{i=1}^{F} \overline{\omega_i} = \overline{\Omega}$.
X^i	Eine Menge $X^i = \{x_1^i, \ldots, x_M^i\} \subset \partial\omega_i$ von Randpunkten zu jedem ω_i, von denen jeder in einem glatten Teilstück des Randes liegt. Sie müssen weiterhin die Bedingung erfüllen, daß Punkte, die sowohl in X^i als auch auf einem $\partial\omega_j$ liegen, auch zu X^j gehören.
Y^i	Eine weitere Menge $Y^i = \{y_1^i, \ldots, y_M^i\} \subset \partial\omega_i$ von Randpunkten zu jedem ω_i, die analoge Bedingungen wie die X^i erfüllen.
Z^i	Eine Menge $Z^i = \{z_1^i, \ldots, z_N^i\} \subset \omega_i$ von inneren Punkten zu jedem ω_i.

Beispiel für zwei benachbarte finite Elemente und die Mengen Y und Z. Es ist $N = 4$ und $M = 4$. Das Bild für die Mengen X und Z sieht entsprechend aus, mit eventuell anderer Lage der Punkte aus X.



- Punkte in $Y^i \cap Y^j$
- $\circ$ Punkte in $Y^i \setminus Y^j$
- $\diamond$ Punkte in $Y^j \setminus Y^i$
- $\times$ Punkte in Z^i
- $+$ Punkte in Z^j

Zu jedem Element assoziieren wir nun noch einen Unterraum der Dimension $K := M + N$ der (reellen) Polynome in D Variablen, welche als glatte Funktionen auf $\overline{\omega}_i$ interpretiert werden. Die Basispolynome seien $\phi_1^i, \ldots, \phi_K^i$:

$$\mathcal{P}^i := \left\langle \phi_1^i, \ldots, \phi_K^i \right\rangle_{\mathbb{R}} \subset \mathcal{C}^\infty(\overline{\omega}_i) \cap \mathbb{P}[x_1, \ldots, x_D].$$

Ein Tupel $(\omega_i, X^i, Y^i, Z^i, \mathcal{P}^i)$ heißt *finites Element* mit *Verbindungspunkten* X^i and Y^i, *Kollokationspunkten* Z^i und *lokalen Funktionen* $\mathcal{P}^i$. Wenn aus dem Kontext heraus klar ist, daß ein bestimmtes Element fixiert ist, werden wir den Index '^{i}' zur Vereinfachung der Notation gewöhnlich fortlassen.

Eine Summe von lokalen Funktionen, die trivial auf $\overline{\Omega}$ fortgesetzt wurden, heißt *global*. Man beachte, daß globale Funktionen glatt sind, wenn man sie auf ein einzelnes ω_i einschränkt, im allgemeinen jedoch unstetig auf den Rändern der Elemente. Aus diesem Grund werden so konstruierte finite Elemente auch als '*nichtkonform*' bezeichnet.

Einem Vektor $c^i \in \mathbb{R}^K$ kann man nun auf kanonische Weise eine lokale Funktion in $\mathcal{P}^i$ zuordnen vermöge einer Linearkombination von Basisfunktionen, die im folgenden mit $P^i\left(c^i\right)$ bezeichnet wird:

$$P^i\left(c^i\right) := \sum_{k=1}^{K} c_k^i \phi_k^i$$

Nicht weniger kanonisch kann man dann zu einer Matrix $c = (c^i)_{1 \le i \le F} \subset \mathbb{R}^{F \times K}$ eine globale Funktion $V(c)$ konstruieren durch Fortsetzen und Aufaddieren der einzelnen lokalen Funktionen, die durch die Zeilenvektoren gegeben sind:

$$V(c) := \sum_{i=1}^{F} \Theta_{\overline{\Omega}} P^i\left(c^i\right),$$

dabei bezeichnet $\Theta_{\overline{\Omega}}$ triviale Fortsetzung nach $\overline{\Omega}$ durch Null.

Eine auf diese Weise konstruierte Funktion soll *zulässig* heißen genau dann, wenn zwei Verbindungsbedingungen und eine Randbedingung erfüllt sind:

- Die lokalen Funktionen, aus denen die globale Funktion zusammengeflickt wurde, sind auf gemeinsamen Verbindungspunkten der jeweiligen Mengen X gleich:

$$P^i\left(c^i\right) \;=\; P^j\left(c^j\right) \quad \text{auf } X^i \cap X^j.$$

- Die Ableitung der lokalen Funktionen in Richtung der zugeordneten Normalenvektoren ist stetig beim Übergang über den Rand in einem Verbindungspunkt aus den Mengen Y. Dafür bezeichne η^i das Vektorfeld von Einheitsvektoren auf $\partial\omega_i$, welches normal zum Rand ist und nach außerhalb von ω_i weist. Dann soll gelten[1]:

$$\nabla P^i\left(c^i\right)\cdot\eta^i(y) \;=\; -\nabla P^j\left(c^j\right)\cdot\eta^j(y) \quad \text{für alle } y\in Y^i\cap Y^j.$$

- Die Funktionswerte in Verbindungspunkten, die auf $\partial\Omega$ liegen, sind gleich den Funktionswerten einer gegebenen Funktion[2] $B\in\mathcal{C}(\partial\Omega)$, die Randbedingungen eines Problems wiedergibt:

$$P^i\left(c^i\right) \;=\; B \quad \text{auf } X^i\cap\partial\Omega.$$

Schließlich ist der Funktionenraum, der für gegebene Randbedingungen die Diskretisierung definiert, der Raum aller zulässigen Funktionen:

$$\mathcal{V}_B := \left\{ V(c) \;:\; c\in\mathbb{R}^{F\times K} \text{ und } V(c) \text{ zulässig } \right\}.$$

Der Übersicht halben sind hier noch einmal die festgelegten natürlichen Zahlen mit ihrer Bedeutung festgehalten:

D	Die Dimension des Raumes, in der Ω liegt.
F	Die Anzahl der finiten Elemente in einer Zerlegung von Ω.
M	Die Anzahl der Verbindungspunkte pro Element sowohl für Funktionswerte als auch für Richtungsableitungen.
N	Die Anzahl der Kollokationspunkte pro Element.
K	Definiert als $K := M + N$, die Dimension des Funktionenraumes $\mathcal{P}$, der jedem Element zugeordnet ist.

0.2 Der Algorithmus von Doedel

Zunächst soll einmal vereinfachend vorausgesetzt werden, daß A linear ist. In dieser Ausgangssituation schlug Eusebius Doedel in [D] ein mögliches Gleichungssystem zur Bestimmung einer zulässigen Funktion vor, welche vermutlich näherungsweise die Differentialgleichung erfüllt. Man kann dafür zunächst nachprüfen, daß der Raum $\mathcal{V}_B$ genau die Dimension $N\cdot F$ hat. Daher liegt es nahe, die Erfüllung der Differentialgleichung in den Kollokationspunkten zu fordern:

$$\text{Finde } \tilde{u}_0\in\mathcal{V}_B, \text{ so daß für alle } z\in Z:\; A\tilde{u}_0(z) = F(z).$$

Das ergibt genau $N\cdot F$ Gleichungen, durch die hoffentlich ein Element in $\mathcal{V}_B$ eindeutig bestimmt ist. Ein wesentlicher Gewinn, den man durch die Wahl dieser Gleichungen hat, ist, daß das entstehende System extrem effizient gelöst werden kann: Falls die finiten Elemente durch fortgesetzte Zweiteilung des Ausgangsgebietes Ω entstehen, so kann der rekursive direkte Lösungsalgorithmus '*Nested dissection*', welcher in Kapitel 8 genau beschrieben wird, dazu verwendet werden. Außerdem besitzen die lokalen Funktionenräume eine recht hohe Dimension, wodurch eine hohe Genauigkeit erzielt werden kann, die beispielsweise für die Untersuchung von Bifurkationsszenarien wichtig ist.

Es war jedoch noch nicht geklärt worden, unter welchen Umständen

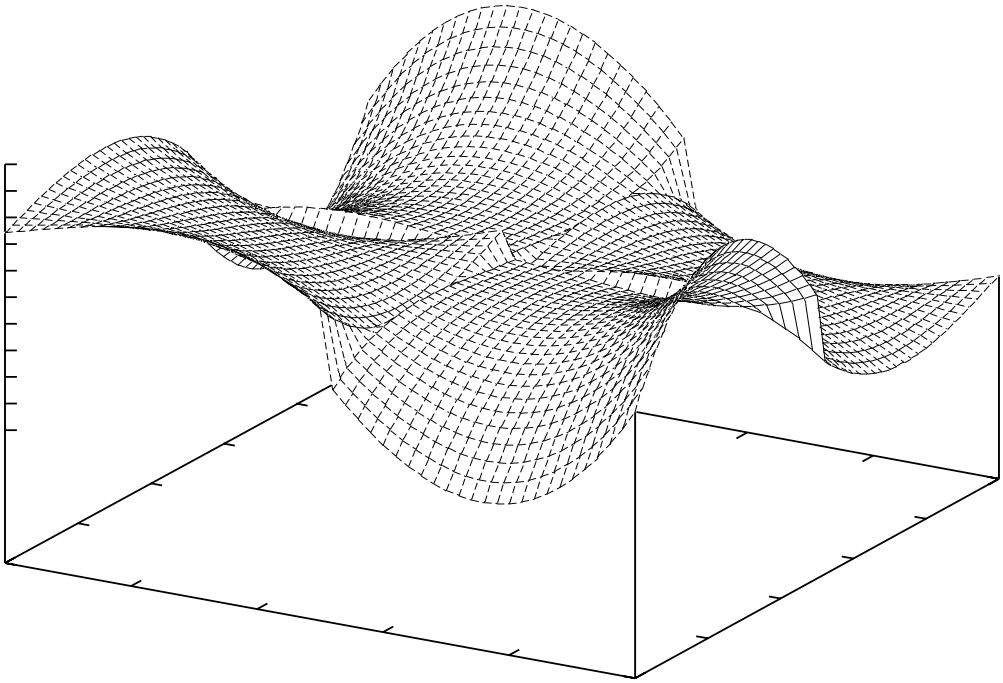
- Das Verfahren tatsächlich auf ein nichtsinguläres Gleichungssystem führt
- Die potentielle Näherungslösung $\tilde{u}_0$ gegen die exakte Lösung u_0 konvergiert, falls das Gebiet immer feiner unterteilt wird.

[1] Der Einfachheit der Notation zuliebe benutzen wir die in der Differentialgeometrie übliche Konvention, daß skalare Multiplikation mit einem Tangentialvektor im Punkt y Auswertung des Gradienten in y impliziert, d.h.

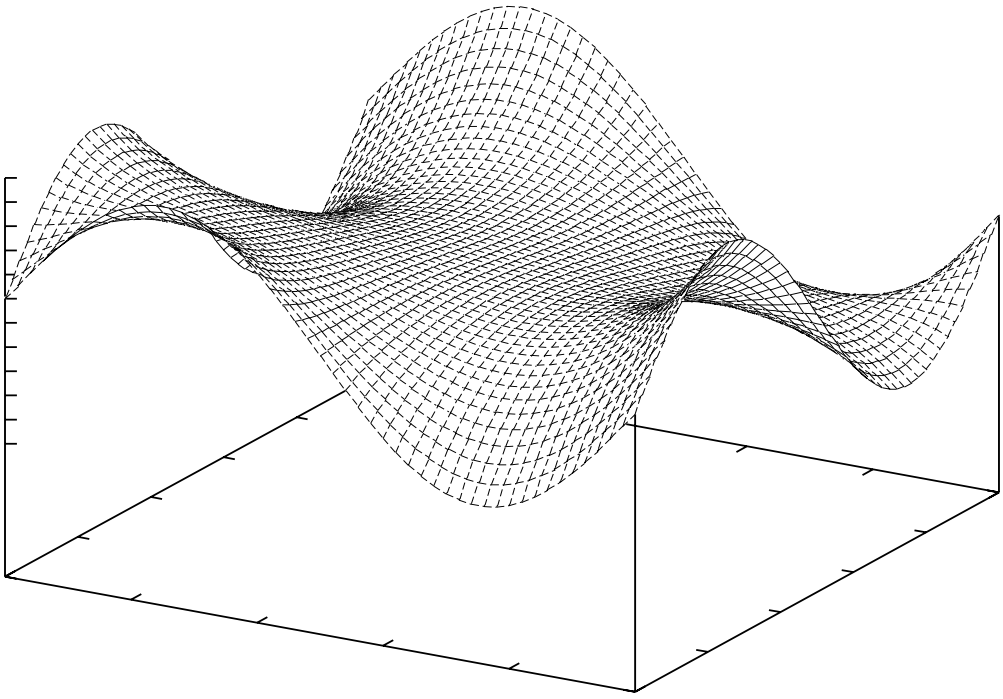
$$\nabla P(c)\cdot\eta(y) := \nabla P(c)\big|_y\cdot\eta(y)$$

[2] Zumeist wird $B\equiv 0$ sein

Typisches Verhalten von Lösungen, die mit dem Verfahren gewonnen werden:



Bei einem groben $2x2$-Gitter sind die unstetigen Übergänge noch deutlich erkennbar.



Nach Verfeinerung des Gitters werden die Lösungen zunehmend glatter.

0.3 Ausblick auf die folgenden Untersuchungen

Das Ziel der folgenden Kapitel in Teil I wird es sein, diese beiden Mißstände anzugehen und hinreichende Kriterien herauszuarbeiten, bei deren Erfüllung sowohl die Existenz und Eindeutigkeit der Kollokationslösungen als auch deren Konvergenz gegen die exakte Lösung eines Problems gegeben sein wird. Dabei werden vor allem Methoden aus der Arbeit [B] zu sogenannten *Variational Crimes*[3] von Klaus Böhmer zum Einsatz gelangen.

Dafür werden in Kapitel 1 zunächst abstrakte Kriterien bewiesen, unter denen die schwachen Lösungen in den Räumen V^h gegen die exakte Lösung konvergieren. Die Herangehensweise ist die Verallgemeinerung der in der Literatur[4] zu findenden Konvergenztheorie für konforme finite Elemente auf den nichtkonformen Fall.

Diese Kriterien liefern Bedingungen an den Operator, dessen zugehörige Bilinearform einige Eigenschaften aufweisen muß, von denen die Elliptizität die am schwierigsten nachzuweisende ist. Im folgenden Kapitel 2 wird daher zunächst gezeigt, daß die Operatoren einer speziellen Klasse partieller Differentialgleichungen, der sogenannten elliptischen Randwertprobleme, eben die geforderten Voraussetzungen erfüllen. Im Zuge dessen wird mit Hilfe einer Regularitätsaussage ein Beweis für die Stabilität der Diskretisierung im Falle einer elliptischen Bilinearform erbracht.

Als weiteres Kriterium ergibt sich die Existenz von Interpolations- und Glättungsoperatoren mit gewissen Eigenschaften, welche in den Kapiteln 3 und 4 unter weiteren Anforderungen an die finiten Elemente nachgewiesen werden. Diese Anforderungen werden allesamt von der Art sein, daß man sie für ein gegebenes Standardelement, das sogenannte *Referenzelement*, nachprüfen muß. Der eigentlich Nachweis kann dann natürlich im Vorfeld per Hand erfolgen, bequemer und für komplizierte Elemente vermutlich notwendig wird es jedoch sein, die Bedingungen durch das Programm prüfen zu lassen, welches die Berechnung durchführt. Da es nur um Tests geht, ob gewisse Matrizen invertierbar sind, wird man ihm diese Aufgabe relativ gefahrlos anvertrauen können.

Schließlich hat man bisher nur die Konvergenz der schwachen Lösungen, und es fehlt noch der Nachweis, daß auch die Kollokationslösungen gegen die exakte Lösung konvergieren. Darum kümmert sich Kapitel 5, welches gleichzeitig auch deren Existenz und Eindeutigkeit nachweist. Dabei wird eine Bedingung an die Kollokationspunkte in Form der Existenz einer geeigneten Quadraturformel ins Spiel kommen.

Die weiteren Bedingungen, die noch an die finiten Elemente und das behandelte Problem gestellt werden müssen, werden im Laufe des Textes an geeigneter Stelle eingeführt. Damit dennoch die Übersicht gewahrt bleibt, findet sich in Kapitel 6, welches den theoretischen Teil abschließt, noch eine Zusammenfassung aller Forderungen und der damit erreichten Resultate.

Teil II befaßt sich dann mit der Praxis der Methode und soll überprüfen, inwiefern sich die theoretisch erzielten Resultate in konkreten Rechnungen wiederspiegeln. Es wird in Kapitel 7 in aller Kürze das Programmpaket vorgestellt, das parallel zu dieser Arbeit entstanden ist, eine ausführliche Dokumentation liegt nur in elektronischer Form vor. Das anschließende Kapitel 8 beschreibt ausführlicher den eben nur skizzierten Lösungsalgorithmus aus [D] und seine Implementation im vorliegenden Programm. Es richtet sich daher vornehmlich an Leser, die sich mit der Programmierung befassen, und kann von anderen getrost übersprungen werden.

Zum Abschluß werden in Kapitel 9 mit diesem Programmsystem etliche Beispielrechnungen durchgeführt, die systematisch die verschiedenen Einflüsse variieren, welche laut der theoretischen Resultate für die Güte der Konvergenz entscheidend sein sollten. Daraus abgeleitete Empfehlungen für die Wahl des Referenzelementes runden die Ergebnisse ab.

In den Anhängen schließlich finden sich noch technische Hilfssätze, deren Beweis im Haupttext diesen zu arg zerfleddert hätte. Außerdem ist dies der Ort, an dem des Komforts wegen noch einmal einige wohlbekannte zentrale Theoreme zitiert werden, die für diese Arbeit benötigt wurden. Zusätzlich werden Quellen angegeben, wo man ihren Beweis finden kann.

[3] 'Verbrechen gegen die Variationsrechnung' - die in der englischen Fachliteratur übliche Bezeichnung für Diskretisierungsverfahren mit nichtkonformen finiten Elementen

[4] Hier vornehmlich bei W.Hackbusch in [H]

Teil I

Theorie

Kapitel 1

Das Variationsproblem

1.0 Diskretisierung und schwache Lösungen

Ziel dieses Kapitels ist es, auf recht abstrakter Ebene den Rahmen zu analysieren, der in dieser Arbeit Gegenstand der Untersuchung ist. Dabei wird zunächst präzisiert, was unter Diskretisierung und Näherungslösungen auf der Ebene von Bilinearformen verstanden werden soll, sowie Kriterien erarbeitet, unter denen Konvergenz der Näherungslösungen gegen die exakte Lösung gewährleistet ist. Entscheidend hierfür ist die sogenannte Stabilität, welche für den Spezialfall elliptischer Bilinearformen für das von uns untersuchte Szenario nachgewiesen wird.

Gegeben sei im folgenden ein Gebiet $\Omega \subset \mathbb{R}^d$ und abgeschlossener Unterraum $\mathrm{V} \sqsubset H^1(\Omega)$. Weiter soll eine stetige Bilinearform a auf $\mathrm{V} \times \mathrm{V}$ mit Stetigkeitskonstante α gegeben sein. Sei außerdem $f \in \mathrm{V}'$ eine stetige Linearform auf V mit Stetigkeitskonstante $\phi = \|f\|_{\mathrm{V}'}$. Unter gewissen Zusatzvoraussetzungen, die später noch geklärt werden, ist dann eine *exakte (schwache) Lösung* $u_0 \in \mathrm{V}$ eindeutig bestimmt durch

$$a(u_0, v) = f(v) \quad \text{für alle } v \in \mathrm{V}.$$

Eine *Diskretisierung* des Problems zu einer Indexmenge $H \subset (0, 1]$ mit Häufungspunkt 0 besteht nun aus den folgenden Zutaten:

- Einer Familie von Zerlegungen $(\omega_i^h)_{1 \le i \le F^h}$ des Gebietes Ω in finite Elemente.
- Einer Familie $(\mathrm{V}^h)_{h \in H}$ endlichdimensionaler Unterräume von $L^2(\Omega)$. Diese sollen die Eigenschaft haben, daß für jedes $v^h \in \mathrm{V}^h$ die Restriktionen $v^h\big|_{\omega_i^h}$ in $\mathcal{C}^\infty(\omega_i^h)$ liegen - man sagt in diesem Zusammenhang auch gerne, daß Funktionen aus V^h 'stückweise glatt' auf den finiten Elementen sind. Der Sinn der Sache ist, daß dadurch Normen auf V^h definiert werden können, die mit den üblichen Normen auf $H^1(\Omega)$ in gewisser Weise verträglich sind. Die Konstruktion läuft dadurch ab, daß die Integrale über ganz Ω durch die Summen der Integrale über finite Elemente ersetzt werden. Die neuen *diskreten* Normen werden durch ein h als Index kenntlich gemacht. Es ist also zum Beispiel

  $$\left\|v^h\right\|_{h,W_p^1} := \left(\sum_{i=1}^{F^h} \left\|v^h\right\|_{W_p^1(\omega_i^h)}^p\right)^{1/p}.$$

  In analoger Weise sind die weiteren Normen und Halbnormen definiert. Das innere Produkt $(\diamond, \diamond)_2$ von $L^2(\Omega)$ macht sowieso auch auf $\mathrm{V}^h \subset L^2(\Omega)$ noch Sinn.

  Abkürzend soll $\|\diamond\|_h := \|\diamond\|_{h,H^1}$ die von der üblichen Norm auf $H^1(\Omega)$ induzierte diskrete Norm bedeuten. Man beachte, daß alle diskreten Normen auch für Funktionen im Raum V wohldefiniert sind und dort mit den üblichen Normen übereinstimmen, da das Integral die Ränder der finiten Elemente als Nullmengen nicht sieht. Insbesondere liefern die diskreten Normen also auch wohldefinierte Normen auf dem vergrößerten Raum $\mathrm{H}^h := \mathrm{V} + \mathrm{V}^h$.
- Einer Familie von Bilinearformen $(a^h)_{h \in H}$, welche Fortsetzungen der ursprünglichen Bilinearform a auf die Räume H^h sein sollen. Zusätzlich soll die Eigenschaft der Stetigkeit, nun bezüglich der Norm $\|\diamond\|_h$, aber mit der gleichen Konstanten α, erhalten bleiben.

- Einer Familie von Linearformen $(f^h)_{h\in H}$, welche ebenfalls Fortsetzungen der ursprünglichen Linearform f auf die Räume H^h sein sollen. Wieder soll die Eigenschaft der Stetigkeit bezüglich der Norm $\|\diamond\|_h$ mit der gleichen Konstanten ϕ erhalten bleiben.

Die neuen (Bi-)linearformen induzieren nun unter gewissen noch zu untersuchenden Voraussetzungen für jedes $h \in H$ eine *schwache Näherungslösung* u_0^h, die eindeutig festgelegt ist durch

$$a^h(u_0^h, v^h) = f^h(v^h) \quad \text{für alle } v^h \in \mathcal{V}^h.$$

Zu einer sinnvollen Diskretisierung gehört selbstverständlich noch eine Reihe von weiteren Eigenschaften. Wesentlich ist, daß die V^h den Raum V zunehmend gut approximieren in dem Sinne, daß von h unabhängige Konstanten $C_{\mathrm{ac}}, C_{\mathrm{ip}} > 0$ existieren, so daß

$$\operatorname{dist}\left(u, \mathrm{V}^h\right) \leq h^{R-1} \cdot C_{\mathrm{ip}} \left|u\right|_{H^R(\Omega)} \text{ für alle } u \in \mathrm{V} \cap H^R(\Omega)$$
$$\text{und umgekehrt } \operatorname{dist}\left(\mathrm{V}, u^h\right) \leq h^G \cdot C_{\mathrm{ac}} \left\|u^h\right\|_h \text{ für alle } u^h \in \mathrm{V}^h,$$

dabei sind $R > 1$ und $G > 0$ geeignete Konstanten. Etwas schärfer formuliert soll sogar gelten: Für jedes $h \in H$ existiert ein *Interpolationsoperator* $I^h : \mathrm{V} \cap H^R(\Omega) \to \mathrm{V}^h$ mit

$$\left\|I^h u - u\right\|_h \leq h^{R-1} \cdot C_{\mathrm{ip}} \left|u\right|_{H^R(\Omega)} \text{ für alle } u \in \mathrm{V} \cap H^R(\Omega).$$

Im allgemeinen wird man leider nicht erreichen können, daß *jede* Funktion aus V interpoliert werden kann, da zumeist $R > 1$ ist. Dies liegt daran, daß die Interpolation überhaupt nur wohldefiniert ist, wenn die zu interpolierende Funktion gewisse Glattheitseigenschaften hat, da man z.B. lokal Funktionswerte kennen muß. Das Sobolev-Lemma ergibt dann notwendige und hinreichende Bedingungen an die Glattheit der Funktionen. Da insbesondere die Lösung u_0 in den späteren Beweisen interpoliert werden muß, stellt dies eine zusätzliche Regularitätsforderung an u_0 dar, von der man sich im Vorfeld überzeugen muß. Im allgemeinen wird es mit sich bringen, daß sowohl der Rand von Ω als auch die vorgegebene Funktion F und die Koeffizientenfunktionen des Differentialoperators genügend 'harmlos' sind. Entsprechende Kriterien finden sich z.B. in [H], Kapitel 9.1.

Des weiteren muß für jedes $h \in H$ ein *Glättungsoperator* $E^h : \mathrm{V}^h \to \mathrm{V}$ mit

$$\left\|E^h u^h - u^h\right\|_h \leq h^G \cdot C_{\mathrm{ac}} \left\|u^h\right\|_h \text{ für alle } u^h \in \mathrm{V}^h$$

existieren. In diesem Falle hat man zum Glück keine Einschränkungen an die Funktionen, die geglättet werden können, dies ist selbstverständlich auch notwendig. Der Mangel an Einschränkungen erklärt sich dadurch, daß Funktionen in V^h stückweise glatt sind und man sich daher um die Existenz von Funktions- und Ableitungswerten keine Sorgen machen muß.

Die Konvergenzaussage für die Räume folgt sofort aus der Existenz von Interpolations- und Glättungsoperator. Sie werden uns im folgenden erlauben, die Lösung u_0 durch Funktionen aus V^h mit $h \to 0$ immer besser zu approximieren, ebenso beliebige Funktionen aus V^h durch Funktionen aus V. Dabei hat man die Güte der Approximation aufgrund der Abschätzungen recht genau unter Kontrolle. In späteren Beweisen wird noch deutlich werden, auf welche Weise man Nutzen aus dieser Tatsache ziehen kann.

Unter geeigneten Umständen, die nun untersucht werden sollen, ist dann nicht nur die bloße Existenz, sondern auch die Konvergenz der Näherungslösungen gegen die exakte schwache Lösung gesichert. Leser, die sich in der Materie etwas auskennen, werden feststellen, daß die folgende Entwicklung der Struktur nach dem üblichen Zugang folgt, wie er z.B. in [H] ausgeführt wird. Man beachte jedoch, daß die meisten der hier bewiesenen Theoreme Verallgemeinerungen der von dort bereits bekannten darstellen: Das hier untersuchte Szenario stammt von nichtkonformen finiten Elementen, daher gilt im Gegensatz zum normalerweise untersuchten Fall $\mathrm{V}^h \not\subset \mathrm{V}$.

1.1 Bilinearformen und zugeordnete Operatoren

In diesem Abschnitt wird H stets ein Hilbertraum sein, dies ist für den folgenden Begriff wesentlich. Zu einer Bilinearform b auf $\mathrm{H} \times \mathrm{H}$ existiert dann nämlich der eindeutig bestimmte *zugeordnete Operator* $B \in \mathcal{L}(\mathrm{H}, \mathrm{H}')$, dessen definierende Gleichung lautet:

$$\langle Bu, v\rangle_{\mathrm{H}'\times\mathrm{H}} = b(u, v) \text{ für alle } u, v \in \mathrm{H}.$$

Zwischen Invertierbarkeit von B und der Lösbarkeit des Variationsproblems besteht naturgemäß eine enge Beziehung. Ebenso sind die folgenden Zahlen stark daran beteiligt:

1.1 Definition. Sei $\mathbb{E}\mathrm{H} := \{v \in \mathrm{H} \,:\, \|v\|_{\mathrm{H}} = 1\}$ die Einheitskugel in H. Weiter sei für eine stetige Bilinearform $b : \mathrm{H} \times \mathrm{H} \to \mathbb{C}$ mit Stetigkeitskonstante β

$$\omega_b := \inf_{u\in\mathbb{E}\mathrm{H}} \sup_{v\in\mathbb{E}\mathrm{H}} |b(u,v)| \le \beta < \infty \quad \text{und}$$
$$\overline{\omega}_b := \inf_{v\in\mathbb{E}\mathrm{H}} \sup_{u\in\mathbb{E}\mathrm{H}} |b(u,v)| \le \beta < \infty.$$

Das nächste Lemma taucht die Verbindung zwischen alldem in helles Licht:

1.2 Lemma. *Sei $B \in \mathcal{L}(H, H')$ der zur stetigen Bilinearform b gehörende Operator. Dann sind die folgenden beiden Aussagen äquivalent:*

(i) $\omega_b > 0$ *und* $\overline{\omega}_b > 0$

(ii) $B^{-1} \in \mathcal{L}(H', H)$ *existiert*

Falls eine der beiden Aussagen richtig ist, so gilt außerdem noch

$$\omega_b = \overline{\omega}_b = \left\|B^{-1}\right\|^{-1}_{H\leftarrow H'},$$

weiter existiert für alle $f \in H'$ ein $u_f \in H$ so daß

$$b(u_f, v) = f(v) \text{ für alle } v \in H, \text{ und } \|u_f\|_H \le \frac{1}{\omega_b}\|f\|_{H'}.$$

Beweis. Die einzelnen Implikationen werden der Reihe nach gezeigt:
(i) $\Longrightarrow$ *(ii):* $B^{-1} \in \mathcal{L}(\mathrm{H}, \mathrm{H}')$ existiere. Dann ergibt stures Ausrechnen:

$$\begin{aligned}
\omega_b &= \inf_{u\in\mathbb{E}\mathrm{H}} \sup_{v\in\mathbb{E}\mathrm{H}} |b(u,v)| = \inf_{0\neq u\in\mathrm{H}} \sup_{0\neq v\in\mathrm{H}} \frac{|b(u,v)|}{\|u\|_{\mathrm{H}}\|v\|_{\mathrm{H}}} = \inf_{0\neq u\in\mathrm{H}} \sup_{0\neq v\in\mathrm{H}} \frac{\left|\langle Bu, v\rangle_{\mathrm{H}'\times\mathrm{H}}\right|}{\|u\|_{\mathrm{H}}\|v\|_{\mathrm{H}}} \\
&= \inf_{0\neq u'\in\mathrm{H}'} \sup_{0\neq v\in\mathrm{H}} \frac{\left|\left\langle BB^{-1}u', v\right\rangle_{\mathrm{H}'\times\mathrm{H}}\right|}{\|B^{-1}u'\|_{\mathrm{H}}\|v\|_{\mathrm{H}}} \\
&= \inf_{0\neq u'\in\mathrm{H}'} \frac{1}{\|B^{-1}u'\|_{\mathrm{H}}} \sup_{0\neq v\in\mathrm{H}} \frac{\left|\langle u', v\rangle_{\mathrm{H}'\times\mathrm{H}}\right|}{\|v\|_{\mathrm{H}}} = \inf_{0\neq u'\in\mathrm{H}'} \frac{\|u'\|_{\mathrm{H}'}}{\|B^{-1}u'\|_{\mathrm{H}}} \\
&= \left(\sup_{0\neq u'\in\mathrm{H}'} \frac{\left\|B^{-1}u'\right\|_{\mathrm{H}}}{\|u'\|_{\mathrm{H}'}}\right)^{-1} = \frac{1}{\|B^{-1}\|_{\mathrm{H}\leftarrow\mathrm{H}'}} > 0.
\end{aligned}$$

Ganz analog folgt $\overline{\omega}_b = 1/\left\|B'^{-1}\right\|_{\mathrm{H}\leftarrow\mathrm{H}'}$, man beachte, daß B' der zur dualen Bilinearform b' mit

$$b'(u,v) := b(v,u)$$

gehörige Operator ist. Diese Eigenschaft sorgt dann dafür, daß Vertauschung von u und v in der Definition von $\overline{\omega}_b$ zur Ersetzung von b durch b' und damit B durch B' führt. Die Normen von B^{-1} und B'^{-1} stimmen aber überein, daher ist $\omega_b = \overline{\omega}_b$.

(ii) $\Longrightarrow$ *(i):* Um diese Implikation zu beweisen, müßte man etwas ausschweifiger werden und unter anderem noch den Riesz-Isomorphismus zwischen H und H′ mit etlichen Eigenschaften heranziehen. Da es sich um eine Standardaussage handelt, sei auf die Literatur verwiesen, man lese den Beweis z.B. in [H], Lemma 6.5.3 nach.

Die für unser weiteres Vorgehen entscheidenden Zusätze ergeben sich wie folgt: Zunächst ist die erste behauptete Gleichung eben schon bewiesen worden, und für ein beliebiges $f \in \mathrm{H}'$ gilt mit $u_f := B^{-1}f$:

$$b(u_f, v) = \langle Bu_f, v\rangle_{\mathrm{H}'\times\mathrm{H}} = \langle f, v\rangle_{\mathrm{H}'\times\mathrm{H}} = f(v) \text{ für alle } v \in \mathrm{H}.$$

□

In den weiteren Abschnitten beachte man, daß das vorherige Lemma anwendbar ist für die Bilinearform a auf V, da V als abgeschlossener Unterraum eines Hilbertraumes wieder ein Hilbertraum ist. Gleiches gilt für die Bilinearformen a^h auf den endlichdimensionalen Unterräumen V^h. In diesem Sinne denken wir uns, wenn von ω_{a^h} die Rede ist, a^h ab sofort immer als Abbildung $a^h : \mathrm{V}^h \times \mathrm{V}^h \to \mathbb{R}$, obwohl a^h auch auf dem größeren Raum H^h definiert ist und auf V mit a übereinstimmt - auch von dieser Tatsache wird reger Gebrauch gemacht.

1.2 Koerzive und elliptische Bilinearformen. Stabilität.

In diesem Abschnitt sollen zunächst wesentliche Eigenschaften definiert werden, die eine Bilinearform auszeichnen können. Ein Hauptziel ist es, den Zusammenhang von Koerzivität und Elliptizität zur Stabilität zu klären, dabei wird es sich ergeben, daß Stabilität sofort aus Koerzivität und unter zusätzlichen Voraussetzungen auch aus der schwächeren Eigenschaft der Elliptizität folgt.

1.3 Definition. Seien $\mathrm{H} \subset L^2(\Omega)$ ein normierter Vektorraum mit Norm $\|\diamond\|_{\mathrm{H}}$ und $b : \mathrm{H} \times \mathrm{H} \to \mathbb{C}$ eine Bilinearform, $\kappa > 0$ und $\mu \in \mathbb{R}$. b heißt dann *(κ, μ)-elliptisch* über $L^2(\Omega)$ genau dann, wenn

$$|b(v,v)| \geq \kappa \|v\|_{\mathrm{H}}^2 - \mu(v,v)_2 \text{ für alle } v \in \mathrm{V}.$$

b heißt *κ-koerziv*, wenn sogar $\mu = 0$ gewählt werden kann.

In Zukunft werden diese Eigenschaften natürlich vor allem für a und a^h interessant sein. Die Rolle von H spielt im Falle von a der Raum $(\mathrm{V}, \|\diamond\|_{H^1(\Omega)})$, im Falle von a^h im Sinne der Schlußbemerkung zum letzten Abschnitt die Räume $(\mathrm{V}^h, \|\diamond\|_h)$.

Die entscheidende Bedeutung der Koerzivität ist, daß sie auf einfache Weise sowohl Existenz aller schwachen Lösungen als auch Stabilität der Diskretisierung garantiert, falls a und alle a^h κ-koerziv mit der gleichen Konstante κ sind. Analoges gilt für Elliptizität unter gewissen zusätzlichen Voraussetzungen, was ungleich schwieriger zu zeigen ist. Beides wird in Kürze bewiesen, zunächst soll jedoch erst einmal definiert werden, was Stabilität eigentlich genau bedeuten soll.

1.4 Notation. Um im folgenden nicht immer zwei genau gleich aussehende Aussagen für a und a^h hinschreiben zu müssen, wird folgende Notation vereinbart: Ein (h) in einer Formel soll bedeuten, daß die Formel sowohl mit als auch ohne h gültig ist. Steht an mehreren Stellen in einer Formel ein (h), so muß entweder überall ein h oder nirgends ein h stehen. Wo Unklarheiten bestehen könnten, wird die Formel sicherheitshalber ausformuliert.

1.5 Definition. Die *Stabilitätsindizes* ω, $\overline{\omega}$ und ω^h der Diskretisierung sind die Zahlen

$$\omega^{(h)} := \omega_{a^{(h)}} = \inf_{u \in \mathbb{E}\mathrm{V}^{(h)}} \sup_{v \in \mathbb{E}\mathrm{V}^{(h)}} \left| a^{(h)}(u,v) \right|$$
$$\text{und} \quad \overline{\omega} := \overline{\omega}_a = \inf_{v \in \mathbb{E}\mathrm{V}} \sup_{u \in \mathbb{E}\mathrm{V}} |a(u,v)|$$

Die Diskretisierung heißt *ϵ-stabil* mit $\epsilon > 0$, falls

$$\omega \geq \epsilon, \overline{\omega} \geq \epsilon \text{ und } \omega^h \geq \epsilon \text{ für alle } h \in H.$$

Die entscheidende Bedeutung der Stabilität ist, daß dann alle schwachen Lösungen existieren und außerdem ihre Normen gleichmäßig beschränkt bleiben. Wir wollen dies etwas präzisieren:

1.6 Satz. *Bei ϵ-stabiler Diskretisierung existieren die schwachen Lösungen u_0 und u_0^h für alle $h \in H$. Weiter gilt:*

$$\|u_0\|_h \leq \frac{\phi}{\epsilon}$$
$$\textit{und } \left\|u_0^h\right\|_h \leq \frac{\phi}{\epsilon} \quad \textit{für alle } h \in H.$$

Beweis. Wegen $0 < \epsilon \leq \omega = \omega_a$ und $0 < \epsilon \leq \overline{\omega}_a$ ist die Aussage für u_0 eine unmittelbare Konsequenz von Lemma 1.2.

Um den Rest der Behauptung zu zeigen, wird bewiesen, daß $\overline{\omega}_{a^h} = \omega_{a^h}$ gilt. Der Grund dafür ist letztendlich, daß die Vektorräume V^h endlichdimensional sind. Die Bedingung $\omega^h \geq \epsilon$ ist nämlich zunächst äquivalent zu

$$\sup_{v^h \in \mathbb{E}\mathrm{V}^h} \left|a^h(u^h, v^h)\right| \geq \epsilon \left\|u^h\right\|_h \text{ für alle } u^h \in \mathrm{V}^h.$$

Links steht aber nun für den zu a^h gehörenden Operator $A^h \in \mathcal{L}(\mathrm{V}^h, \mathrm{V}^{h\prime})$ die Dualnorm von $A^h u^h$. Es ist also

$$\left\|A^h u^h\right\|_{\mathrm{V}^{h\prime}} \geq \epsilon \left\|u^h\right\|_h,$$

mithin ist A^h injektiv. Da $\dim \mathrm{V}^h = \dim \mathrm{V}^{h\prime}$, existiert $(A^h)^{-1} \in \mathcal{L}(\mathrm{V}^{h\prime}, \mathrm{V}^h)$. Die Behauptung für u_0^h folgt nun wiederum mit Lemma 1.2. □

Gleichmäßige Koerzivität impliziert Stabilität:

1.7 Lemma. *Sei $\kappa > 0$ so daß a und a^h für alle $h \in H$ κ-koerziv sind. Dann ist die Diskretisierung ϵ-stabil, wobei $\epsilon = 1/\kappa$ gewählt werden kann.*

Beweis. Wegen Koerzivität ist für festes $u = \mathbb{E}\mathrm{V}^{(h)}$:

$$\sup_{v \in \mathbb{E}\mathrm{V}^{(h)}} \left| a^{(h)}(u,v) \right| \geq \left| a^{(h)}(u,u) \right| \geq \frac{1}{\kappa},$$

also ist auch das Infimum über alle diese u größer oder gleich $1/\kappa$, womit sofort die Behauptung folgt. □

1.3 Stabilität impliziert Konvergenz

Der erste Satz liefert eine abstrakte Abschätzung für den Diskretisationsfehler im Falle der Stabilität der Diskretisierung. Anschließend wird gezeigt, daß unter der Voraussetzung der Existenz genügend guter Interpolations- und Glättungsoperatoren daraus bereits Konvergenz der Näherungslösungen gegen die exakte Lösung folgt. Insbesondere liegt also im Falle der Koerzivität aller beteiligten Bilinearformen sofort Konvergenz vor, da dies Stabilität der Diskretisierung implizierte.

1.8 Satz. *Die Diskretisierung sei ϵ-stabil. Dann gilt für alle $h \in H$:*

$$\left\| u_0 - u_0^h \right\|_h \leq \left(1 + \frac{\alpha}{\epsilon}\right) \inf_{v^h \in V^h} \left\| u_0 - v^h \right\|_h + \frac{1}{\epsilon} \sup_{w^h \in \mathbb{E}V^h} \left| a^h(u_0 - u_0^h, w^h) \right|$$

Beweis. Sei $v^h \in \mathrm{V}^h$ beliebig. Dann gilt mit der Definition der Stabilität und Stetigkeit von a^h:

$$\begin{aligned}
\left\| u_0 - u_0^h \right\|_h &\leq \left\| u_0 - v^h \right\|_h + \left\| v^h - u_0^h \right\|_h \\
&\leq \left\| u_0 - v^h \right\|_h + \frac{1}{\epsilon} \sup_{w^h \in \mathbb{E}\mathrm{V}^h} \left| a^h(v^h - u_0^h, w^h) \right| \\
&= \left\| u_0 - v^h \right\|_h + \frac{1}{\epsilon} \sup_{w^h \in \mathbb{E}\mathrm{V}^h} \left| a^h(v^h - u_0, w^h) + a^h(u_0 - u_0^h, w^h) \right| \\
&\leq \left(1 + \frac{\alpha}{\epsilon}\right) \left\| u_0 - v^h \right\|_h + \frac{1}{\epsilon} \sup_{w^h \in \mathbb{E}\mathrm{V}^h} \left| a^h(u_0 - u_0^h, w^h) \right|
\end{aligned}$$

Nach Bildung des Infimums über alle v^h folgt die Behauptung, da der rechte Summand nicht mehr von v^h abhängt. □

Die Terme auf der rechten Seite der Abschätzung müssen nun untersucht werden. Dies ist Gegenstand der beiden folgenden Lemmas. Der erste hängt ab von der Güte der Interpolation:

1.9 Lemma. *Unter Voraussetzung der Existenz der Interpolationsoperatoren gilt:*

$$\inf_{v^h \in V^h} \left\| u_0 - v^h \right\|_h \leq h^{R-1} \cdot C_{ip} \left| u_0 \right|_{H^R(\Omega)}.$$

Beweis. Man beobachte, daß $I^h u_0 \in \mathrm{V}^h$, und wende die vorausgesetzte Abschätzung für den Interpolationsfehler an. □

Der nächste Schritt ist die Abschätzung des Supremums-Terms in der Ungleichung aus Satz 1.8. Dies ist nun die Stelle, wo die Existenz der Glättungsoperatoren entscheidend einfließt.

1.10 Lemma. *Unter Voraussetzung der Existenz der Glättungsoperatoren gilt:*

$$\sup_{w^h \in \mathbb{E}V^h} \left| a^h(u_0 - u_0^h, w^h) \right| \leq h^G \cdot C_{ac} \left(\alpha \left\| u_0 \right\|_h + \phi \right).$$

Beweis. Sei $w^h \in \mathrm{V}^h$ mit $\left\| w^h \right\|_h = 1$. Dann gilt für alle $w \in \mathrm{V}$:

$$\begin{aligned}
\left| a^h(u_0 - u_0^h, w^h) \right| &= \left| a^h(u_0, w^h) - a^h(u_0^h, w^h) \right| && \text{(da } a^h \text{ bilinear)} \\
&= \left| a^h(u_0, w^h) - f^h(w^h) \right| && \text{(Definition von } u_0^h\text{)} \\
&= \left| a^h(u_0, w^h) - a^h(u_0, w) + a^h(u_0, w) \right. \\
&\qquad \left. - \left(f^h(w^h) - f^h(w) \right) - f^h(w) \right| && \text{(Fundamentaltrick der Analysis)} \\
&= \left| a^h(u_0, w^h - w) - f^h(w^h - w) \right| && \text{(Definition von } u_0\text{)} \\
&\leq \left(\alpha \left\| u_0 \right\|_h + \phi \right) \cdot \left\| w^h - w \right\|_h && \text{(Dreiecksungleichung, Stetigkeit)}
\end{aligned}$$

Wähle speziell $w := E^h w^h$, so folgt mit der vorausgesetzten Qualität der Glättungsoperatoren:

$$\begin{aligned}\left|a(u_0 - u_0^h, w^h)\right| &\leq (\alpha \, \|u_0\|_h + \phi) \cdot h^G \cdot C_{\mathrm{ac}} \, \left\|w^h\right\|_h \\ &= (\alpha \, \|u_0\|_h + \phi) \cdot h^G \cdot C_{\mathrm{ac}}\end{aligned}$$

und damit die Behauptung des Lemmas. □

Zusammenfassend erhalten wir also das erste Hauptresultat über die Konvergenz der Approximation:

1.11 Theorem. *Seien alle a^h κ-koerziv auf $\mathcal{V}^h$. Die exakte Lösung u_0 liege in $H^R(\Omega)$ mit $R \geq 2$. Dann gilt für alle $h \in H$:*

$$\left\|u_0 - u_0^h\right\|_h \;\leq\; h^{\min\{G, R-1\}} \cdot \Big[C_{ac}\,\kappa\;(\alpha\;\|u_0\|_h \;+\; \phi) \;+\; C_{ip}\;(1 + \kappa\,\alpha)\;|u_0|_{H^R(\Omega)}\Big]$$

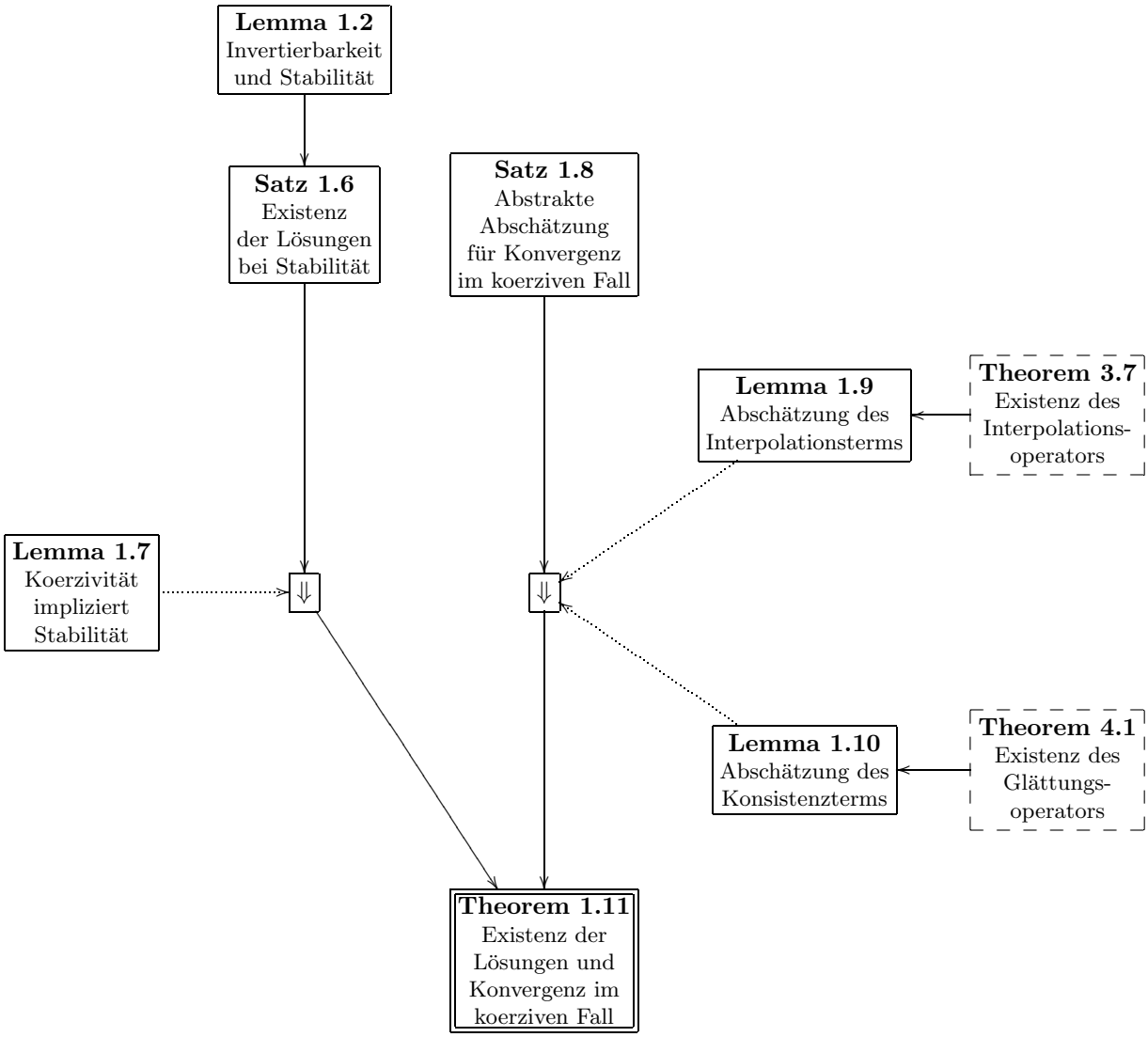
Beweis. Stabilität bezüglich $\epsilon = 1/\kappa$ folgt mit Lemma 1.7. Kombiniere nun Satz 1.8 mit Lemma 1.9 und 1.10. □

1.12 *Bemerkung.* Geht man den Beweisgang noch einmal durch, so stellt man fest, daß eine höhere Konvergenzordnung nur zu erreichen ist, falls man sowohl die Konvergenzordnung der Interpolation als auch der Glättung verbessert. Bei der Interpolation ist das noch relativ leicht, Theorem 3.7 wird die Aussage liefern, daß man lediglich die Dimension der lokalen Funktionenräume geeignet erhöhen muß. Allerdings geht das nur unter der Voraussetzung, daß die Lösung u_0 eine genügend hohe Regularität R aufweist, wobei R mindestens so groß ist wie die gewünschte Konvergenzordnung erhöht um eins.

Derzeit ist außerdem noch keine allgemeine Abschätzung zur Hand, mit der sichergestellt werden kann, daß man die Fehlerordnung der Glättung auf ähnlich einfache Art verbessern könnte. In Kapitel 4 wird sich weiter herausstellen, daß G sehr klein ist, im allgemeinen wird man nur $G = 1/2$ erreichen. Dies ist allerdings nicht weiter tragisch, wie sich bald herausstellen wird, da beim Übergang zum Kollokationsverfahren noch eine andere abstrakte Abschätzung hergeleitet wird. Verwendet man diese, so erhält man ein Resultat für die Konvergenzordnung, welches von der Ordnung der Glättung unabhängig ist. ◇

1.4 Beweisstruktur

Zur besseren Orientierung wird jedem Kapitel aus dem theoretischen Teil ein Graph beigefügt, der die Zusammenhänge in der Beweisstruktur zwischen den zugrundeliegenden, bzw. den im Verlauf des Kapitels verwendeten oder bewiesenen Sätzen illustriert.



Kapitel 2

Elliptische Randwertprobleme

In diesem Kapitel soll der Zusammenhang zwischen den Bilinearformen in der Variationsformulierung und den Differentialoperatoren herausgearbeitet werden. Insbesondere wird eine Klasse von Operatoren eingeführt, deren Bilinearformen die notwendigen Eigenschaften für die in den letzten Abschnitten hergeleitete Konvergenztheorie erfüllen. Dies sind die sogenannten elliptischen Operatoren, welche sehr viele wichtige Beispiele umfassen, auf ein paar davon wird im Laufe der Diskussion noch näher eingegangen.

Mit Hilfe einer Aussage über die Regularität von Lösungen wird außerdem das Stabilitätsresultat für koerzive Bilinearformen aus dem letzten Kapitel auf elliptische Bilinearformen übertragen.

2.0 Bilinearformen

2.1 Definition. Eine Matrix $(a_{ij})_{i,j=1,\dots,D}$ mit Koeffizienten $a_{ij} \in L^\infty(\Omega)$ heißt *(gleichmäßig) elliptisch*, falls eine Konstante $\kappa \geq 0$ existiert, so daß für alle $x \in \Omega$ und $\xi \in \mathbb{R}^D$:

$$\frac{1}{2} \sum_{i,j=1}^{D} a_{ij}(x)\, \xi_i \xi_j \;\geq \kappa \sum_{i=1}^{D} \xi_i^2 .$$

Im folgenden seien ebenfalls $b_1, \dots, b_d \in L^\infty(\Omega)$ und $c_0 \in L^\infty(\Omega)$. Diese definieren eine Bilinearform $a : H^1(\Omega) \times H^1(\Omega) \to \mathbb{R}$ durch

$$a(u,v) := \int_\Omega \sum_{i,j=1}^{D} a_{ij}\, \partial_i u \cdot \partial_j v \;+\; \sum_{k=1}^{D} b_k\, \partial_k u \cdot v \;+\; c_0\, u \cdot v$$

Um die Notation etwas übersichtlicher zu halten wird im folgenden vereinbart, daß doppelt auftretende Indizes von 1 bis D summiert werden. Damit schreibt es sich etwas kürzer:

$$a(u,v) = \int_\Omega a_{ij}\, \partial_i u \cdot \partial_j v \;+\; b_k\, \partial_k u \cdot v \;+\; c_0\, u \cdot v$$

Die Zerlegungen induzieren dann für alle $h \in H$ in naheliegender Weise Bilinearformen auf $\mathrm{V}^h \times \mathrm{V}^h$:

$$a^h(u^h, v^h) := \sum_{n=1}^{F^h} \int_{\Omega_n^h} a_{ij}\, \partial_i u^h \cdot \partial_j v^h \;+\; b_k\, \partial_k u^h \cdot v^h \;+\; c_0\, u^h \cdot v^h$$

Diese Darstellung liefert bei gleichmäßig elliptischer Koeffizientenmatrix (a_{ij}) eine Klasse von Bilinearformen, die alle Eigenschaften erfüllen, wie sie für die Konvergenztheorie gebraucht werden. Sowohl für a als auch die a^h hat man dann nämlich:

- Stetigkeit mit einer gemeinsamen Stetigkeitskonstanten,
- Elliptizität mit gemeinsamen Elliptizitätskonstanten, daraus folgend
- Stabilität der Diskretisierung, somit Existenz und Konvergenz der schwachen Lösungen.

Wir zeigen dies in einer Reihe von Lemmas:

2.2 Lemma. *a und alle a^h sind stetig mit der Stetigkeitskonstanten*

$$\alpha := \sum_{i,j=1}^{D} \|a_{ij}\|_{L^\infty(\Omega)} \; + \; \sum_{k=1}^{D} \|b_k\|_{L^\infty(\Omega)} \; + \; \|c_0\|_{L^\infty(\Omega)} \, .$$

Beweis. Die Behauptung folgt mit der Abschätzung

$$\begin{aligned}
\left|a^h(u^h, v^h)\right| &\leq \sum_{n=1}^{F^h} \left\|a_{ij}\, \partial_i u^h \cdot \partial_j v^h\right\|_{L^1(\Omega_n^h)} + \left\|b_k\, \partial_k u^h \cdot v^h\right\|_{L^1(\Omega_n^h)} + \left\|c_0\, u^h \cdot v^h\right\|_{L^1(\Omega_n^h)} \\
&\leq \sum_{n=1}^{F^h} \|a_{ij}\|_{L^\infty(\Omega_n^h)} \left\|\partial_i u^h\right\|_{L^2(\Omega_n^h)} \left\|\partial_j v^h\right\|_{L^2(\Omega_n^h)} \\
&\qquad + \|b_k\|_{L^\infty(\Omega_n^h)} \left\|\partial_k u^h\right\|_{L^2(\Omega_n^h)} \left\|v^h\right\|_{L^2(\Omega_n^h)} \\
&\qquad + \; \|c_0\|_{L^\infty(\Omega_n^h)} \left\|u^h\right\|_{L^2(\Omega_n^h)} \left\|v^h\right\|_{L^2(\Omega_n^h)} \qquad \text{(Hölder)} \\
&\leq \alpha \sum_{n=1}^{F^h} \left\|u^h\right\|_{H^1(\Omega_n^h)} \cdot \left\|v^h\right\|_{H^1(\Omega_n^h)} \\
&\leq \alpha \; \left\|u^h\right\|_h \cdot \left\|v^h\right\|_h \qquad \text{(Lemma A.4)}
\end{aligned}$$

für alle $u^h, v^h \in \mathrm{V}^h$, die in völlig analoger Weise auch für a gezeigt wird, nur daß die Summation und die letzte Zeile dann sogar entfällt. In der vorletzten Zeile der Ungleichungskette wurden die Supremumsnormen auf den einzelnen Elementen durch die Supremumsnorm auf ganz Ω, bzw. die L^2-Normen durch geeignete Sobolev-Normen nach oben abgeschätzt. □

2.3 Lemma. (Verallgemeinerte Gårding-Ungleichung). *a und alle a^h sind (κ, μ)-elliptisch . Dabei kommt das κ aus der Definition der gleichmäßigen Elliptizität der Koeffizientenmatrix (a_{ij}), die Konstante μ ist gegeben durch*

$$\begin{aligned}
\mu &:= \kappa + \frac{\beta^2}{4\kappa} - \gamma \quad , \text{ wobei} \\
\beta &:= \sum_{k=1}^{D} \|b_k\|_{L^\infty(\Omega)} \\
\gamma &:= \operatorname{ess\,inf} c_0
\end{aligned}$$

Insbesondere sind a und a^h sogar κ-koerziv , falls $\gamma \geq \kappa + \beta^2/(4\kappa)$.

Beweis. Wir zeigen die Aussage zunächst wieder für a^h. Für alle $v^h \in \mathrm{V}^h$ gilt aufgrund der Elliptizität von (a_{ij}) punktweise unter dem Integral angewandt mit $\xi = \nabla v(x)$:

$$\begin{aligned}
\sum_{n=1}^{F^h} \int_{\omega_n^h} a_{ij}\, \partial_i v^h \cdot \partial_j v^h &\geq 2\kappa \sum_{i=1}^{F^h} \int_{\omega_n^h} \left|\nabla v^h\right|_2^2 \qquad \text{(Elliptizität)} \\
&= 2\kappa \sum_{i=1}^{F^h} \left|v^h\right|_{H^1(\omega_n^h)}^2 \quad \text{(Definition von } |\diamond|_{H^1(\omega_n^h)}\text{)} \\
&= 2\kappa \left|v^h\right|_{h,H^1}^2 \qquad \text{(Definition von } |\diamond|_{h,H^1}\text{)}
\end{aligned}$$

Damit folgt für beliebiges $\mu \in \mathbb{R}$:

$$a^h(v^h, v^h) + \mu \left\|v^h\right\|_{h,L^2}^2 \; \geq \; 2\kappa \left|v^h\right|_{h,H^1}^2 \; + \; \sum_{n=1}^{F^h} \int_{\omega_n^h} b_k \partial_k v^h \cdot v^h \; + \; (c_0 + \mu) v^h \cdot v^h$$

Der mittlere Summand läßt sich folgendermaßen abschätzen:

$$\begin{aligned}
\left|\sum_{n=1}^{F^h}\int_{\omega_n^h} b_k\,\partial_k v^h\cdot v^h\right| &\le \|b_k\|_{L^\infty(\Omega)}\sum_{n=1}^{F^h}\left\|\partial_k v^h\right\|_{L^2(\omega_n^h)}\left\|v^h\right\|_{L^2(\omega_n^h)} && \text{(Hölder)}\\
&\le \sum_{k=1}^{D}\|b_k\|_{L^\infty(\Omega)}\sum_{n=1}^{F^h}\left|v^h\right|_{H^1(\omega_n^h)}\left\|v^h\right\|_{L^2(\omega_n^h)} && \text{(Definition von } |\diamond|_{H^1(\omega_n^h)})\\
&= \beta\sum_{n=1}^{F^h}\left|v^h\right|_{H^1(\omega_n^h)}\left\|v^h\right\|_{L^2(\omega_n^h)} && \text{(Definition von } \beta)\\
&\le \beta\left|v^h\right|_{h,H^1}\cdot\left\|v^h\right\|_{h,L^2} && \text{(Lemma A.4)}
\end{aligned}$$

Für den letzten Summanden findet man:

$$\begin{aligned}
\sum_{n=1}^{F^h}\int_{\omega_n^h}(c_0+\mu)v^h\cdot v^h &\ge (\gamma+\mu)\sum_{n=1}^{F^h}\int_{\omega_n^h}\left|v^h\right|^2 && \text{(Definition von } \gamma)\\
&= (\gamma+\mu)\left\|v^h\right\|_{h,L^2}^2 && \text{(Definition von } \|\diamond\|_{h,L^2})
\end{aligned}$$

Zusammen mit der ursprünglichen Ungleichung ergibt sich dann

$$a^h(v^h,v^h)+\mu\left\|v^h\right\|_{h,L^2}^2 \;\ge\; 2\kappa\left|v^h\right|_{h,H^1}^2 \;-\; \beta\left|v^h\right|_{h,H^1}\cdot\left\|v^h\right\|_{h,L^2} \;+\; (\gamma+\mu)\left\|v^h\right\|_{h,L^2}^2$$

Zum gewünschten Ergebnis kommt man nun durch trickreiche Anwendung der verallgemeinerten Ungleichung zwischen arithmetischem und geometrischen Mittel. Es gilt nämlich für beliebige $x,y\in\mathbb{R}$ und $\delta>0$:

$$xy\le\frac{\delta}{2}\,x^2+\frac{1}{2\delta}\,y^2.$$

Wendet man diese auf die gegebene Situation an mit $\delta=2\kappa/\beta$, $x=\left|v^h\right|_{h,H^1}$ und $y=\left\|v^h\right\|_{h,L^2}$ an, so ergibt sich:

$$\begin{aligned}
a^h(v^h,v^h)+\mu\left\|v^h\right\|_{h,L^2}^2 \;\ge\; 2\kappa\left|v^h\right|_{h,H^1}^2 \;&-\; \beta\left(\frac{2\kappa}{2\beta}\left|v^h\right|_{h,H^1}^2+\frac{\beta}{4\kappa}\left\|v^h\right\|_{h,L^2}^2\right)\\
&+\;(\gamma+\mu)\left\|v^h\right\|_{h,L^2}^2\\
=\;\kappa\left(\left|v^h\right|_{h,H^1}^2+\left\|v^h\right\|_{h,L^2}^2\right) \;&+\; \left(\mu+\gamma-\kappa-\frac{\beta^2}{4\kappa}\right)\left\|v^h\right\|_{h,L^2}^2\\
=\;\kappa\left\|v^h\right\|_{h,H^1}^2 \;&+\; \left(\mu+\gamma-\kappa-\frac{\beta^2}{4\kappa}\right)\left\|v^h\right\|_{h,L^2}^2
\end{aligned}$$

Das bedeutet aber gerade, daß a^h (κ,μ)-elliptisch ist, falls μ so gewählt wird, daß

$$\mu+\gamma-\kappa-\frac{\beta^2}{4\kappa}=0,$$

und das war die Behauptung.

Die Ungleichung für a folgt wie eben völlig analog, indem auf die Summation verzichtet wird - der mißtrauische Leser kann den Beweis für diesen Spezialfall aber auch bei [BS], Satz 5.6.8 nachschlagen. □

2.1 Ein Regularitätssatz

Um die Stabilitätsaussage, die man für koerzive Bilinearformen hatte, auf die allgemeineren elliptischen zu übertragen, braucht man an einer Stelle eine etwas höhere Regularität der Lösungen als $u_0\in H^1(\Omega)$. Dies ist nur unter weiteren Einschränkungen an das Ausgangsgebiet und die Koeffizientenfunktionen möglich. Das entsprechende Resultat soll hier nur zitiert werden, da der Beweis noch einige Vorarbeit mehr erfordern würde.

2.4 Satz. *Sei Ω beschränkt und konvex. Die Bilinearform*

$$a(u,v) = \int_\Omega a_{ij}\,\partial_i u \cdot \partial_j v \;+\; b_k\,\partial_k u \cdot v \;+\; c_0\,u \cdot v$$

besitze im Hauptteil Lipschitz-stetige Koeffizienten

$$a_{ij} \in \mathcal{C}^{0,1}(\overline{\Omega}) \quad \textit{für } 1 \le i,j \le D.$$

Sei weiter $f \in L^\infty(\Omega)$ und $u_0 \in V$ die schwache Lösung der Aufgabe

$$a(u_0,v) \;=\; \int_\Omega f \cdot v \quad \textit{für alle } v \in V.$$

Dann ist sogar $u_0 \in V \cap H^2(\Omega)$ mit

$$\|u_0\|_{H^2(\Omega)} \le C \cdot \left(\|f\|_{L^\infty(\Omega)} \;+\; \|u_0\|_h \right).$$

Die Konstante C hängt dabei nur vom Durchmesser von Ω ab.

Beweis. Wegen der verallgemeinerten Gårding-Ungleichung 2.3 ist a (κ,μ)-elliptisch . Die Aussage folgt dann mit Satz 9.1.22 aus [H]. □

2.2 Stabilität für elliptische Bilinearformen.

Wir haben in Lemma 1.7 gesehen, daß die wichtige Eigenschaft der Stabilität relativ leicht nachzuweisen ist, falls gleichmäßige Koerzivität vorliegt. Wesentlich schwieriger und nur unter weiteren Einschränkungen möglich ist es, ein ähnliches Resultat für die schwächere gleichmäßige Elliptizität zu erzielen. In diesem Abschnitt soll die Aufgabe angegangen werden, den Weg dorthin ebnet das folgende Theorem, welches eigentlich aus der Eigenwerttheorie von Bilinearformen stammt, uns aber hier gute Dienste leistet. Auf die eigentliche Theorie der Eigenwerte soll nicht näher eingegangen werden, um die Arbeit nicht mit für unsere Ziele unnötiger Theorie zu befrachten.

Der Beweis des Theorems ergibt sich durch filigranes Zusammenwirken des Glättungsoperators mit der bereits gezeigten Stabilität und Konvergenz im koerziven Fall. Leider ist er nicht ganz elementar und benötigt an der entscheidenden Stelle die eben zitierte Regularitätsaussage. Der Grund besteht in der Notwendigkeit, eine Lösung hinreichend gut zu interpolieren, von der man zunächst nur weiß, daß sie in $\mathrm{V} \subset H^1(\Omega)$ liegt. Allerdings ist das Analogon des Theorems in der konformen Theorie schon sehr schwierig[1] zu beweisen, und es wäre daher wohl vermessen zu hoffen, daß es ausgerechnet im nichtkonformen Fall einfacher sein sollte. So ist der folgende Beweis auch einer der kompliziertesten in der gesamten Arbeit.

2.5 Theorem. *Die Bilinearformen a und a^h seien allesamt (κ,μ)-elliptisch . Die Voraussetzungen für die H^2-Regularität der Lösungen aus Satz 2.4 seien erfüllt. Dann gibt es Zahlen $C > 0$ und $\eta(h) > 0$ mit $\lim_{h\to 0}\eta(h) = 0$, so daß gilt:*

$$\omega^h \ge C\omega - \eta(h) \textit{ für alle } h \in H.$$

Beweis. Nach Definition sind die Bilinearformen

$$a_\mu^{(h)}(u,v) := a^{(h)}(u,v) - \mu(u,v)_2$$

für alle $h \in H$ κ-koerziv , d.h. die Diskretisierung ist nach Lemma 1.7 bezüglich dieser Bilinearformen $\epsilon = 1/\kappa$-stabil. Sei nun zunächst $h \in H$ fest. Dann existieren für ein beliebiges $u^h \in \mathrm{V}^h$ mit $\left\|u^h\right\|_h = 1$

$$\begin{aligned} z \in \mathrm{V} \text{ als Lösung von } a_\mu(z,v) &= -\mu(E^h u^h, v)_2 =: g^h(v) \text{ für alle } v \in \mathrm{V} \\ \text{und } z^h \in \mathrm{V}^h \text{ als Lösung von } a_\mu^h(z^h,v^h) &= -\mu(u^h,v^h)_2 =: g(v^h) \text{ für alle } v^h \in \mathrm{V}^h, \end{aligned}$$

da $g, g^h \in \mathrm{V}' \cap \mathrm{V}^{h\prime}$ mit $\left\|g^h\right\|_{\mathrm{V}^{h\prime}} = \left\|E^h u^h\right\|_h$ und $\|g\|_{\mathrm{V}^{h\prime}} = \left\|u^h\right\|_h = 1$. Man erhält nun eine Reihe von Aussagen:

[1] vergleiche [H], Lemma 11.2.7

(*i*) Für alle $v \in \mathrm{V}$ gilt:

$$\begin{aligned} a_\mu(E^h u^h - z, v) &= a(E^h u^h - z, v) - \mu(E^h u^h - z, v)_2 && \text{(Definition } a_\mu\text{)} \\ &= a(E^h u^h, v) - a(z, v) - \mu(E^h u^h, v)_2 + \mu(z, v)_2 && \text{(Bilinearität)} \\ &= a(E^h u^h, v) - a_\mu(z, v) - \mu(E^h u^h, v)_2 && \text{(Definition } a_\mu\text{)} \\ &= a(E^h u^h, v) && \text{(Definition } z\text{)} \end{aligned}$$

(*ii*) Für alle $v^h \in \mathrm{V}^h$ gilt:

$$\begin{aligned} a_\mu^h(u^h - z^h, v^h) &= a^h(u^h - z^h, v^h) - \mu(u^h - z^h, v^h)_2 && \text{(Definition } a_\mu^h\text{)} \\ &= a^h(u^h, v^h) - a^h(z^h, v^h) - \mu(u^h, v^h)_2 + \mu(z^h, v^h)_2 && \text{(Bilinearität)} \\ &= a^h(u^h, v^h) - a_\mu^h(z^h, v^h) - \mu(u^h, v^h)_2 && \text{(Definition } a_\mu^h\text{)} \\ &= a^h(u^h, v^h) && \text{(Definition } z^h\text{)} \end{aligned}$$

(*iii*) Es ist

$$1 - h^G \cdot C_{\mathrm{ac}} \;\le\; \left\|E^h u^h\right\|_h \;\le\; 1 + h^G \cdot C_{\mathrm{ac}},$$

denn nach Dreiecksungleichung und Abschätzung für die Glättung gilt

$$\begin{aligned} 1 = \left\|u^h\right\|_h &= \left\|u^h - E^h u^h + E^h u^h\right\|_h \\ &\le h^G \cdot C_{\mathrm{ac}} \left\|u^h\right\|_h \;+\; \left\|E^h u^h\right\|_h \\ &= h^G \cdot C_{\mathrm{ac}} \;+\; \left\|E^h u^h\right\|_h \\ \text{und} \quad \left\|E^h u^h\right\|_h &= \left\|E^h u^h - u^h + u^h\right\|_h \\ &\le h^G \cdot C_{\mathrm{ac}} \left\|u^h\right\|_h \;+\; \left\|u^h\right\|_h \\ &= h^G \cdot C_{\mathrm{ac}} \;+\; 1. \end{aligned}$$

Insbesondere folgt damit $\left\|g^h\right\|_{\mathrm{V}^{h\prime}} = \left\|E^h u^h\right\|_h \le 1 + h^G C_{\mathrm{ac}}$ und

$$\begin{aligned} \omega(1 - h^G \cdot C_{\mathrm{ac}}) &\le \omega \left\|E^h u^h\right\|_h && \text{(gerade gezeigt)} \\ &\le \sup_{v \in \mathbb{E}\mathrm{V}} \left|a(E^h u^h, v)\right| && \text{(nach Definition von } \omega\text{)} \\ &= \sup_{v \in \mathbb{E}\mathrm{V}} \left|a_\mu(E^h u^h - z, v)\right| && \text{(nach } (i)\text{)} \\ &\le C_\alpha \left\|E^h u^h - z\right\|_h, \end{aligned}$$

denn mit a ist sicherlich auch a_μ stetig mit einer geeigneten Konstanten C_α.

(*iv*) Nötig ist auch eine Abschätzung für $\left\|z - z^h\right\|_h$. Sei dafür

$$\tilde{z} \in \mathrm{V} \text{ Lösung von } a_\mu(\tilde{z}, v) = g(v) \text{ für alle } v \in \mathrm{V},$$

eine solche Lösung existiert wiederum wegen Koerzivität von a_μ. Dann ist $z - \tilde{z}$ Lösung von

$$a_\mu(z - \tilde{z}, v) = (g^h - g)(v) \text{ für alle } v \in \mathrm{V},$$

und nach Satz 1.6 gilt:

$$\begin{aligned} \|z - \tilde{z}\|_h &\le \kappa \left\|g^h - g\right\|_{\mathrm{V}'} \\ &\le\; \kappa \left\|E^h u^h - u^h\right\|_h \\ &\le\; h^G \cdot C_{\mathrm{ac}}\, \kappa \left\|u^h\right\|_h \\ &=\; h^G \cdot C_{\mathrm{ac}}\, \kappa, \\ \text{sowie} \quad \|\tilde{z}\|_h &\le\; \kappa \|g\|_{\mathrm{V}'} \;=\; \kappa. \end{aligned}$$

Nach dem Regularitätssatz 2.4 gilt aber sogar

$$\tilde{z} \in H^2(\Omega) \quad \text{und} \quad \|\tilde{z}\|_{H^2(\Omega)} \;\le\; C_1 \cdot (\|g\|_{\mathrm{V}'} \;+\; \|\tilde{z}\|_h)\,.$$

Daher greift Theorem 1.11 mit $R = 2$ und man hat weiter:

$$\begin{aligned}\left\|\tilde{z} - z^h\right\|_h &\leq h^{\min\{1,G\}} \cdot \left[\kappa\, C_\alpha\, C_{\text{ac}}\, \|\tilde{z}\|_h + (1 + \kappa\, C_\alpha)\, C_{\text{ip}}\, \|\tilde{z}\|_{H^2(\Omega)} + \kappa\, \|g\|_{\text{V}'}\, C_{\text{ac}}\right] \\ &\leq h^{\min\{1,G\}} \cdot \left[\kappa^2\, C_\alpha\, C_{\text{ac}}\ +\ (1 + \kappa\, C_\alpha)\, C_{\text{ip}}\, C_1(1 + \kappa)\ +\ C_{\text{ac}}\, \kappa\right] \\ &=: h^{\min\{1,G\}} \cdot \tilde{C}.\end{aligned}$$

Insgesamt folgt

$$\begin{aligned}\left\|z - z^h\right\|_h &\leq \|z - \tilde{z}\|_h\ +\ \left\|\tilde{z} - z^h\right\|_h \\ &\leq h^{\min\{1,G\}} \cdot (C_{\text{ac}}\, \kappa\ +\ \tilde{C}).\end{aligned}$$

Nun ist alles beieinander. Mit den bisherigen Ergebnissen kann man nämlich endlich die gewünschte Abschätzung für ω^h durchführen:

$$\begin{aligned}&\sup_{v^h \in \mathbb{E}\text{V}^h} \left|a^h(u^h, v^h)\right| \\ &= \sup_{v^h \mathbb{E}\text{V}^h} \left|a^h_\mu(u^h - z^h, v^h)\right| && \text{(nach (ii))} \\ &\geq \frac{1}{\kappa} \left\|u^h - z^h\right\|_h && \text{(Koerzivität von } a^h_\mu\text{)} \\ &\geq \frac{1}{\kappa} \left\|E^h u^h - z\right\|_h\ -\ \frac{1}{\kappa} \left\|u^h - E^h u^h\right\|_h\ -\ \frac{1}{\kappa} \left\|z - z^h\right\|_h && \text{(Dreiecksungleichung)} \\ &\geq \frac{1}{\kappa} \left\|E^h u^h - z\right\|_h\ -\ h^{\min\{1,G\}} \cdot \frac{1}{\kappa}\, C_{\text{ac}}\ -\ h^{\min\{1,G\}} \cdot \frac{1}{\kappa} \left(C_{\text{ac}}\kappa\ +\ \tilde{C}\right) && \text{(Glättung und (iv))} \\ &\geq \frac{\omega}{C_\alpha\, \kappa} - h^{\min\{1,G\}} \cdot \frac{1}{\kappa} \left[\frac{\omega C_{\text{ac}}}{C_\alpha}\ +\ C_{\text{ac}}\ +\ C_{\text{ac}}\, \kappa\ +\ \tilde{C}\right]. && \text{(nach (iii))}\end{aligned}$$

Da $u^h \in \mathbb{E}\text{V}^h$ beliebig war, folgt die Aussage des Theorems mit

$$C := \frac{1}{C_\alpha\, \kappa} \text{ und } \eta(h) := h^{\min\{1,G\}} \cdot \frac{1}{\kappa} \left[\frac{\omega C_{\text{ac}}}{C_\alpha}\ +\ C_{\text{ac}}(1\ +\ \kappa)\ +\ \tilde{C}\right]$$

wegen $G > 0$. □

2.6 Korollar. *Alle a^h und a seien (κ, μ)-elliptisch . Der zur Bilinearform a gehörende Operator $A \in \mathcal{L}(V', V)$ sei invertierbar. Dann existiert ein $\epsilon > 0$ und $h_0 > 0$, so daß*

$$\omega^h \geq \epsilon \text{ für alle } h \in H,\ h \leq h_0.$$

Mit anderen Worten: Die Diskretisierung ist für hinreichend kleine h ϵ-stabil, d.h. alle schwachen Lösungen existieren und sind in der Norm beschränkt durch ϕ/ϵ. Insbesondere gilt dann für $h \leq h_0$:

$$\left\|u_0 - u_0^h\right\|_h\ \leq\ h^{\min\{R-1,G\}} \cdot \left[\frac{C_{ac}}{\epsilon}\, (\alpha\, \|u_0\|_h\ +\ \phi)\ +\ C_{ip} \left(1 + \frac{\alpha}{\epsilon}\right) |u_0|_{H^R(\Omega)}\right].$$

Beweis. Falls A invertierbar ist, so ist nach Lemma 1.2 $\overline{\omega} = \omega = \omega_a > 0$. Die erste Behauptung folgt dann mit der Abschätzung des Theorems und $\eta(h) \to 0$, die Abschätzung für den Fehler der Diskretisierung folgt durch einen analogen Schluß wie in Theorem 1.11. □

Das Ergebnis der ganzen Mühen ist nun, daß die im letzten Kapitel erarbeitete Konvergenztheorie voll auf elliptische Bilinearformen von der Form wie oben definiert anwendbar ist. Insbesondere gilt dies also für die elliptischen Randwertprobleme, partielle Differentialgleichungen mit dem zur Bilinearform a gehörenden Differentialoperator. Dieser Übergang wird im nächsten Abschnitt beschrieben.

2.3 Operatorform und Randfehler

Es soll nun hergeleitet werden, welche Gestalt die Differentialoperatoren besitzen und welcher Art die Randbedingungen sind, die man natürlicherweise zu einer Bilinearform der Gestalt

$$a(u,v) = \int_\Omega a_{ij}\,\partial_i u \cdot \partial_j v \;+\; b_k\,\partial_k u \cdot v \;+\; c_0\,u \cdot v$$

assoziiert. Partielle Integration liefert zunächst einmal

$$a(u,v) = \int_\Omega (\partial_j(a_{ij}\,\partial_i u) \;+\; b_k\,\partial_k u \cdot v \;+\; c_0\,u)\cdot v \;+\; \int_{\partial\Omega} \sum_{i,j=1}^{D} a_{ij}\,\eta_j\,\partial_i u \cdot v.$$

Der zu a gehörende Differentialoperator A ist mithin gegeben durch

$$A := \partial_j(a_{ij}\,\partial_i) \;+\; b_k\,\partial_k \;+\; c_0,$$

falls man die Randbedingungen derart wählt, daß das bei der partiellen Integration auftretende Randintegral für beliebige u und v aus V verschwindet. Funktionen $u \in \mathrm{V}$ müssen also auf jedem Teilstück von $\partial\Omega$ entweder identisch Null sein, oder die sogenannten natürlichen Randbedingungen

$$\sum_{i,j=1}^{D} a_{ij}\,\eta_j\,\partial_i u \;=\; 0$$

erfüllen. Jede klassische Lösung $u_0 \in \mathrm{V} \cap \mathcal{C}^2(\Omega)$ der Gleichung $Au_0 = F$ ist dann auch eine schwache Lösung, umgekehrt ist eine Lösung des Variationsproblems, welche sogar in $\mathcal{C}^2(\Omega)$ liegt, auch eine Lösung der partiellen Differentialgleichung mit den gegebenen Randbedingungen. Diese Gleichung nennt man bei Dirichlet- oder natürlichen Randbedingungen und einem Operator mit gleichmäßig elliptischer Koeffizientenmatrix auch ein *elliptisches Randwertproblem* oder eine *elliptische Differentialgleichung*.

Geht man nun zu den Bilinearformen a^h über, so hat man damit zu kämpfen, daß Funktionen aus V^h auf den Rändern der finiten Elemente nicht stetig sind. Insbesondere ist also der klassische Operator nur im Inneren der finiten Elemente überhaupt definiert, und bei der partiellen Integration muß man noch Fehler berücksichtigen, die durch Integrale auf den Elementrändern gegeben sind. Es gilt für Funktionen $u^h, v^h \in \mathrm{V}^h$:

$$\begin{aligned} a^h(u^h,v^h) &= \sum_{n=1}^{F^h} \int_{\omega_n^h} a_{ij}\,\partial_i u^h \cdot \partial_j v^h \;+\; b_k\,\partial_k u^h \cdot v^h \;+\; c_0\,u^h \cdot v^h \\ &= \sum_{n=1}^{F^h} \int_{\omega_n^h} Au^h \cdot v^h \;+\; \int_{\partial\omega_n^h} \sum_{j=1}^{D} a_{ij}\,\eta_j\,\partial_i u_n^h \cdot v_n^h, \end{aligned}$$

man erinnere sich daran, daß u_n^h die lokale Funktion aus $\mathcal{C}^\infty(\overline{\omega_n^h})$ bezeichnet, durch welche u^h dort definiert ist. Wegen der Unstetigkeit der Funktionen u^h und v^h auf den Elementrändern und der Tatsache, daß diese wegen $\mathrm{V}^h \not\subset \mathrm{V}$ im allgemeinen nicht die Randbedingungen erfüllen, ergibt die Summe über die Randintegrale nicht Null wie im Falle konformer finiter Elemente.

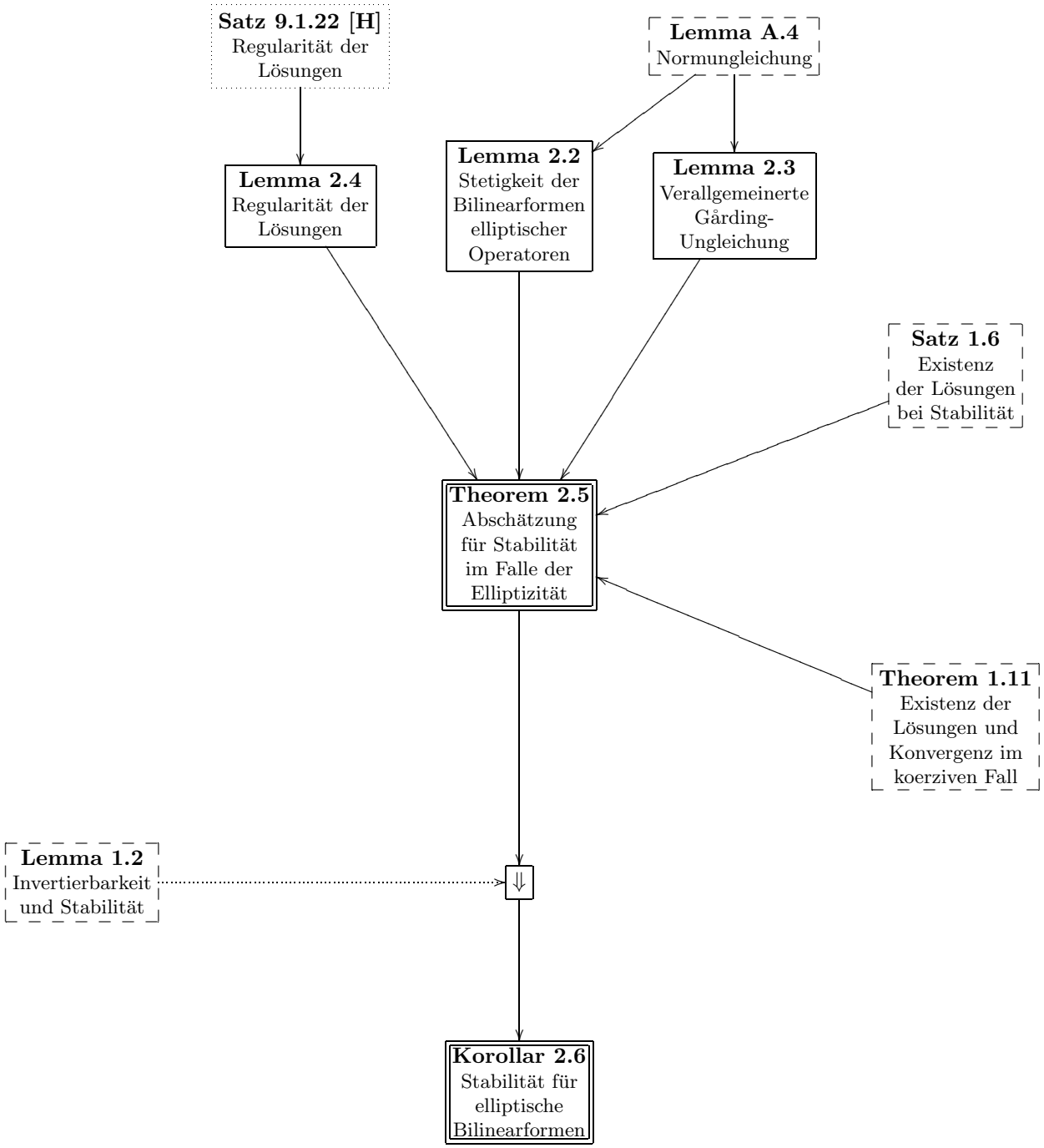
2.7 Definition. Der Fehler beim Übergang von a^h zur Operatordarstellung sei bezeichnet mit

$$\begin{aligned} \rho^h(u^h,v^h) :&= a^h(u^h,v^h) - \sum_{n=1}^{F^h} \int_{\omega_n^h} Au^h \cdot v^h \\ &= \sum_{n=1}^{F^h} \int_{\partial\omega_n^h} \sum_{j=1}^{D} a_{ij}\,\eta_j\,\partial_i u_n^h \cdot v_n^h. \end{aligned}$$

Die Abschätzung dieses Fehlers wird essentiell notwendig beim Übergang zur Kollokation, hängt aber stark von der Art der finiten Elemente ab. Die Diskussion wird erst im Kapitel über die Glättungsoperatoren speziell für den in der Einführung beschriebenen Typ durchgeführt.

2.4 Beweisstruktur

Zur besseren Orientierung wird jedem Kapitel aus dem theoretischen Teil ein Graph beigefügt, der die Zusammenhänge in der Beweisstruktur zwischen den zugrundeliegenden, bzw. den im Verlauf des Kapitels verwendeten oder bewiesenen Sätzen illustriert.



Kapitel 3

Die Interpolationsoperatoren

Bisher waren die Räume V^h noch recht abstrakte Objekte mit gewissen notwendigen Eigenschaften. Es soll nun konkreter die Situation $\mathrm{V}^h := \mathcal{V}_0^h$ untersucht werden, d.h. die V^h entstehen durch Diskretisierung mit nichtkonformen finiten Elementen, so wie es in Abschnitt 0.1 geschildert worden ist.

Das erste Ziel dieses Kapitels ist es, eine Familie stetiger Operatoren

$$I^h : \mathrm{V} \cap H^R(\Omega) \to \mathrm{V}^h$$

mit der für Konvergenz erforderlichen Approximationseigenschaft

$$\left\|u - I^h u\right\|_h \;\leq\; h^{R-1} \cdot C_{\mathrm{ip}}\, |u|_{H^R(\Omega)} \text{ für alle } h \in H \text{ und } u \in V \cap H^R(\Omega)$$

zu konstruieren. Dabei ist R eine geeignete natürliche Zahl, welche die zusätzliche Forderung mit sich bringt, daß die exakte Lösung u_0 der Variationsaufgabe sogar in $\mathrm{V} \cap H^R(\Omega)$ liegt, also eine gewisse Regularität aufweist. Dies ergibt sich aus den Beweisen in den vorherigen Kapiteln, in denen insbesondere die Lösung u_0 durch Funktionen in V^h interpoliert werden mußte.

Im folgenden wird zunächst ein festes $h \in H$ fixiert und die entsprechenden Indizes unterdrückt, um die Notation übersichtlicher zu halten.

Die Konstruktion wird nur für den speziellen Fall $2M = K \Leftrightarrow N = M$ explizit durchgeführt, die Methode ist jedoch mit etwas mühseligerer Schreibweise leicht auf den Fall $M \leq N$ verallgemeinerbar, an geeigneter Stelle wird dies durch einen Hinweis näher erläutert. In dieser Verallgemeinerung ist die Einschränkung nicht allzu schwerwiegend: Der Lösungsalgorithmus wird wesentlich genauer, wenn mehr Kollokationspunkte im Spiel sind, und ab einer gewissen Untergrenze ist die Bedingung dann auf natürliche Weise erfüllt, ohne daß die Elementgeometrie allzu seltsam erscheint - wie man sich leicht vorstellen kann, liegen bei einer homogenen Verteilung der Punkte über die Fläche des Elementes automatisch mehr Punkte im Inneren als auf dem Rand.

Anschließend wird dann ein Kriterium bewiesen, mit dessen Erfüllung die Existenz des Interpolationsoperators sichergestellt ist. Jenes kann zur Programmlaufzeit durch eine einzelne LR-Zerlegung einer $K \times K$-Matrix pro Referenzelement überprüft werden, jedoch natürlich auch im Vorfeld, falls die genaue Geometrie und die Basisfunktionen bekannt sind. Durchgeführt wird ein exakter mathematischer Beweis nur für einen einzigen Elementtyp (den unter der Nebenbedingung einfachsten denkbaren), da die Rechnungen per Hand schnell ausgesprochen mühsam werden - es macht keinen Spaß, größere Matrizen auf einem Blatt Papier zu invertieren. In der Praxis wird diese Aufgabe wohl das Programm übernehmen, welches die Berechnungen durchführt, und bei der Definition des Referenzelements nachprüfen, ob dieses alle erforderlichen Eigenschaften aufweist - es werden später noch mehr hinzukommen.

Der Nachweis der Konvergenzeigenschaft ist nach geeigneter Umformulierung des Szenarios eine direkte Anwendung der Interpolationstheorie für konforme finite Elemente, das entsprechende Resultat wird dann auch lediglich aus [BS] zitiert.

3.0 Lokale Konstruktion

Wir arbeiten zunächst rein lokal auf einem ausgewählten Element i und werden dessen Index in den nächsten Absätzen unterdrücken. Man behalte aber bitte im Hinterkopf, daß die folgenden Definitionen vom Element abhängen, sofern nicht ausdrücklich anderweitiges behauptet wird.

Der erste Schritt ist die Definition eines lokalen Interpolationsoperators, der als Eingabedaten einen Vektor $u \in \mathbb{R}^M$ und einen Vektor $v \in \mathbb{R}^M$ erhält und diese linear auf einen Vektor $c \in \mathbb{R}^K$ transformiert, dessen zugehörige lokale Funktion $P(c)$ die folgenden Eigenschaften hat:

- Die Funktionswerte in den Verbindungspunkten sind durch u gegeben:

$$P(c)(x_m) = u_m \quad \text{für } 1 \leq m \leq M$$

- Die Werte der Richtungsableitungen in den Verbindungspunkten legt v fest:

$$\nabla P(c) \cdot \eta(y_m) = v_m \quad \text{für } 1 \leq m \leq M$$

Diese zwei Bedingungen führen zu einem System von $2M$ Gleichungen für K Unbekannte, daher die Bedingung[1] $2M = K$.

Wir beschreiben das entstehende Gleichungssystem durch die Matrizen, die auch E.Doedel in seinen Texten zur Abkürzung verwendet:

$$\Phi := \begin{bmatrix} \phi_1(x_1) & \dots & \phi_1(x_M) \\ \vdots & \ddots & \vdots \\ \phi_K(x_1) & \dots & \phi_K(x_M) \end{bmatrix}$$

$$R_\Phi := \begin{bmatrix} \nabla\phi_1 \cdot \eta(y_1) & \dots & \nabla\phi_1 \cdot \eta(y_M) \\ \vdots & \ddots & \vdots \\ \nabla\phi_K \cdot \eta(y_1) & \dots & \nabla\phi_K \cdot \eta(y_M) \end{bmatrix}$$

Mit diesen können die Beziehungen nun umgeschrieben werden zu

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \Phi^* \\ R_\Phi^* \end{bmatrix} \cdot c$$

und nahe liegt die folgende

3.1 Definition. Der *lokale Interpolationsoperator* auf dem finiten Element i ist die lineare Abbildung $I^i_{\text{loc}} : \mathbb{R}^{2M} \to \mathbb{R}^K$, die durch die folgende Matrix gegeben ist:

$$I^i_{\text{loc}} := \begin{bmatrix} \Phi^* \\ R_\Phi^* \end{bmatrix}^{-1}$$

Damit der Operator wohldefiniert ist, muß die Matrix natürlich invertierbar sein. Zunächst soll angenommen werden, daß dies der Fall ist und die Diskussion wird auf den Abschnitt 3.2 verschoben.

3.1 Globale Konstruktion

Falls alle lokalen Interpolationsoperatoren existieren, ist der Rest einfach: Man nehme einfach die Daten der zu interpolierenden Funktion als Eingabedaten für die lokalen Interpolationsoperatoren und klebe die entstehenden lokalen Funktionen zu einer globalen zusammen.

3.2 Definition. Für eine gegebene Funktion $g \in \mathcal{C}^1(\overline{\Omega})$ und für jedes finite Element $1 \leq i \leq F$ seien die Vektoren $u^i(g)$ und $v^i(g)$ definiert durch

$$u^i(g) := \begin{bmatrix} g(x^i_1) \\ \vdots \\ g(x^i_M) \end{bmatrix}, \quad v^i(g) := \begin{bmatrix} \nabla g \cdot \eta^i(y^i_1) \\ \vdots \\ \nabla g \cdot \eta^i(y^i_M) \end{bmatrix},$$

die Vektoren $c^i(g)$ seien aus ihnen durch lokale Interpolation entstanden:

$$c^i(g) := I^i_{\text{loc}} \begin{bmatrix} u^i(g) \\ v^i(g) \end{bmatrix}.$$

[1] Ist $N > M \implies K > 2M$, so sind es zuwenig Gleichungen, in diesem Fall kann man einfach nach Belieben zusätzliche Gleichheit von Funktionswerten z.B. auf Kollokationspunkten fordern und muss die im folgenden auftauchenden Matrizen geeignet erweitern.

Dann ist der *globale Interpolationsoperator* $I : \mathcal{C}^1(\overline{\Omega}) \to \mathcal{V}_0$ gegeben durch

$$I(g) := \sum_{i=1}^{F} \Theta_{\overline{\Omega}} P\left(c^i(g)\right)$$

Da die Funktionswerte und Richtungsableitungen der lokalen Funktionen in den Verbindungspunkten auf die entsprechenden Werte von g festgelegt sind, ist die interpolierende Funktion automatisch eine zulässige Funktion. Der Operator ist also, falls die lokalen Operatoren existieren, nach Konstruktion wohldefiniert. Er muß allerdings noch fortgesetzt werden zu einem Operator auf ganz V.

Dies wird sogleich in Abschnitt 3.3 im Zuge der Umstellung der Notation auf die in [BS] verwendete erfolgen, indem die obige Definition in eine etwas allgemeinere eingebettet wird. In diesem Sinne dienten die obigen Betrachtungen mehr der Veranschaulichung der Lage, allerdings war es sowieso notwendig, einmal die Matrizen zu notieren, von deren Invertierbarkeit man sich überzeugen muß.

3.2 Existenz

Dieser Abschnitt befaßt sich mit der Existenz des lokalen Interpolationsoperators. Diese hängt natürlich stark von der Geometrie des Elements und der Wahl der Basisfunktionen ab, so daß sie lediglich von Fall zu Fall diskutiert werden kann. Allerdings wird der Begriff der affinen Äquivalenz von Elementen eingeführt werden, welcher die Dinge etwas vereinfacht dadurch, daß nur einige wenige Referenzelemente untersucht werden müssen. Schließlich wird noch für eine bestimmte quadratische Elementgeometrie und eine Basis von Polynomen die Existenz konkret bewiesen.

Hier ist auch die geeignete Stelle, die Lage in einer leicht veränderten Notation zu betrachten. Bisher haben wir sie von E.Doedel übernommen, jedoch ist ein etwas anderer Standpunkt ebenfalls nützlich: Die Formulierung als finites Element im Sinne von S.Brenner und L.Scott aus [BS]. Auf diese Weise können wir alle Resultate bezüglich des Interpolationsfehlers direkt von dort übernehmen.

Um dieses Ziel zu erreichen, muß zunächst zu jedem finiten Element ein weiteres Objekt assoziiert werden, die sogenannte *nodale Basis* $\mathcal{N} = \{N_1, \ldots, N_K\}$, welche eine Basis des Dualraums der lokalen Funktionen sein muß:

$$\mathcal{P}' = \langle N_1, \ldots, N_K \rangle_{\mathbb{R}} .$$

Damit die Beschreibung unserer finiten Elemente der von Brenner/Scott verwendeten gleicht, müssen die *nodalen Variablen* als Auswertung von Funktions- und Richtungsableitungswerten in den entsprechenden Verbindungspunkten definiert werden:

$$\begin{aligned} N_k(g) &:= g(x_m) \quad \text{für } 1 \le m \le M \\ N_{k+M}(g) &:= \nabla g \cdot \eta(y_m) \quad \text{für } 1 \le m \le M \end{aligned}$$

Erfreut bemerken wir, daß die den lokalen Interpolationsoperator beschreibende Matrix nun eine wesentlich symmetrischere Gestalt hat:

$$\begin{bmatrix} \Phi^* \\ R^*_\Phi \end{bmatrix} = \begin{bmatrix} N_1(\phi_1) & \ldots & N_1(\phi_K) \\ \vdots & \ddots & \vdots \\ N_K(\phi_1) & \ldots & N_K(\phi_K) \end{bmatrix}$$

Da nun alle Information über Verbindungspunkte und Ableitungsrichtungen in die nodalen Variablen eincodiert ist, werden wir ein finites Element ab sofort mit $(\omega, Z, \mathcal{P}, \mathcal{N})$ bezeichnen.

3.3 Definition. Sei $(\omega, Z, \mathcal{P}, \mathcal{N})$ ein finites Element und $\varphi : \mathbb{R}^D \to \mathbb{R}^D$ eine affine Abbildung, $\varphi(x) = Ax+b$ mit nichtsingulärer Matrix A. Dann heißt $(\omega, Z, \mathcal{P}, \mathcal{N})$ *affin äquivalent* zu $(\tilde{\omega}, \tilde{Z}, \tilde{\mathcal{P}}, \tilde{\mathcal{N}})$ genau dann, wenn

(*i*) $\varphi(\omega) = \tilde{\omega}$ und $\varphi(Z) = \tilde{Z}$

(*ii*) $\varphi^* \tilde{\mathcal{P}} = \mathcal{P}$

(*iii*) $\varphi_* \mathcal{N} = \tilde{\mathcal{N}}$

In diesem Fall schreiben wir $(\omega, Z, \mathcal{P}, \mathcal{N}) \cong_\varphi (\tilde{\omega}, \tilde{Z}, \tilde{\mathcal{P}}, \tilde{\mathcal{N}})$.

3.4 *Bemerkung.* Man erinnere sich, daß der *pull-back* φ^* definiert ist durch $\varphi^*(\tilde{\phi}) := \tilde{\phi} \circ \varphi$, der *push-forward* φ_* durch $(\varphi_* N)(\tilde{\phi}) := N(\varphi^* \tilde{\phi})$.

Dank affiner Äquivalenz ist es nun ausreichend, die Existenz des lokalen Interpolationsoperators nur noch für einige Referenzelemente zu beweisen:

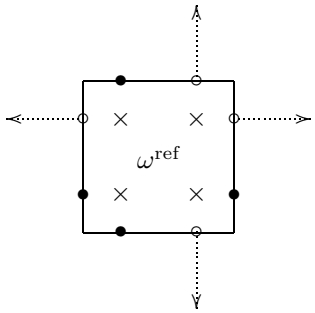
3.5 Lemma. *Falls der lokale Interpolationsoperator existiert für $(\omega, Z, \mathcal{P}, \mathcal{N})$, dann existiert er auch für alle zu $(\omega, Z, \mathcal{P}, \mathcal{N})$ affin äquivalenten Elemente. In der Tat sind die Matrizen für lokale Interpolation sogar auf allen diesen finiten Elementen gleich.*

Beweis. Sei $(\omega, Z, \mathcal{P}, \mathcal{N}) \cong_\varphi (\tilde{\omega}, \tilde{Z}, \tilde{\mathcal{P}}, \tilde{\mathcal{N}})$. Dann ist

$$\begin{aligned}
\begin{bmatrix} \tilde{N}_1(\tilde{\phi}_1) & \dots & \tilde{N}_1(\tilde{\phi}_K) \\ \vdots & \ddots & \vdots \\ \tilde{N}_K(\tilde{\phi}_1) & \dots & \tilde{N}_K(\tilde{\phi}_K) \end{bmatrix} &= \begin{bmatrix} \varphi_* N_1(\tilde{\phi}_1) & \dots & \varphi_* N_1(\tilde{\phi}_K) \\ \vdots & \ddots & \vdots \\ \varphi_* N_K(\tilde{\phi}_1) & \dots & \varphi_* N_K(\tilde{\phi}_K) \end{bmatrix} && \text{(Bedingung (iii))} \\
&= \begin{bmatrix} N_1(\varphi^* \tilde{\phi}_1) & \dots & N_1(\varphi^* \tilde{\phi}_K) \\ \vdots & \ddots & \vdots \\ N_K(\varphi^* \tilde{\phi}_1) & \dots & N_K(\varphi^* \tilde{\phi}_K) \end{bmatrix} && \text{(Definition } \varphi_*) \\
&= \begin{bmatrix} N_1(\phi_1) & \dots & N_1(\phi_K) \\ \vdots & \ddots & \vdots \\ N_K(\phi_1) & \dots & N_K(\phi_K) \end{bmatrix}, && \text{(Bedingung (ii))}
\end{aligned}$$

die lokalen Interpolationsoperatoren sind also tatsächlich gleich. □

3.6 Beispiel. Die Existenz des lokalen Interpolationsoperators soll nun für das im folgenden skizzierte Referenzelement gezeigt werden. Dabei ist $\omega^{\text{ref}} = (0,1)^2$ das Einheitsquadrat, ausgefüllte Punkte bezeichnen Verbindungspunkte X^{ref} für Funktionswerte, Ringe bezeichnen Verbindungspunkte Y^{ref} für Richtungsableitungs, und Kreuze markieren Kollokationspunke, die jedoch in diesem Kontext unwichtig sind. Alle eingezeichneten Vektoren haben die Länge eins, Punkte sind von links unten beginnend im Gegenuhrzeigersinn numeriert.



Das Referenzelement

Der Grund, warum das Element so asymmetrisch konstruiert wurde, ist in einem ganz grundsätzlichen Problem zu suchen, welches bei der Wahl von Polynomen als Basisfunktionen stets aufzutauchen scheint: Die Tendenz, singuläre Matrizen zu produzieren, falls die Elementgeometrie oder die Basisfunktionen sehr symmetrisch gewählt worden sind. Aus dem gleichen Grund ist auch der von uns gewählte lokale Funktionenraum leicht asymmetrisch in den Polynomen höheren Grades. Derzeit ist noch kein sauberer Algorithmus bekannt, der eine Konstruktion von 'funktionierenden' Basisfunktionen erlaubt.

In diesem Fall ist nun $M = N = 4$ und $K = 8$, der Raum der lokalen Funktionen ist also achtdimensional. Wir verwenden $\mathcal{P} = \langle 1, x, y, x^2, y^2, xy, x^3, xy^3 \rangle_{\mathbb{R}}$, und werden durch stures Ausrechnen beweisen, daß der lokale Interpolationsoperator wohldefiniert ist. Sei dafür $h = 1/4$ der Abstand von Null zum nächsten Verbindungspunkt, dann lautet die zu invertierende Matrix

$$\begin{bmatrix} N_1(\phi_1) & \dots & N_1(\phi_8) \\ \vdots & \ddots & \vdots \\ N_8(\phi_1) & \dots & N_8(\phi_8) \end{bmatrix} = \begin{bmatrix} 1 & h & 0 & h^2 & 0 & 0 & h^3 & 0 \\ 1 & 1 & h & 1 & h^2 & h & 1 & h^3 \\ 1 & h & 1 & h^2 & 1 & h & h^3 & h \\ 1 & 0 & h & 0 & h^2 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & h-1 & 0 & 0 \\ 0 & 1 & 0 & 2 & 0 & 1-h & 3 & (1-h)^3 \\ 0 & 0 & 1 & 0 & 2 & 1-h & 0 & 3-3h \\ 0 & -1 & 0 & 0 & 0 & h-1 & 0 & (h-1)^3 \end{bmatrix}$$

welche dank einer Determinante von 111/1024 auch tatsächlich invertierbar ist, wie der dazu geneigte masochistisch veranlagte Leser nachrechnen kann. ◇

3.3 Konvergenz der Interpolation

Es soll natürlich nicht 'das Rad neu erfunden' werden, aus diesem Grund haben wir uns ja bereits auf den Beschreibungsrahmen von S.Brenner und L.Scott zurückgezogen, was uns nun erlauben wird, die muskulösen Theoreme, die dort über Konvergenz bewiesen wurden, direkt auf unseren Fall anzuwenden.

Als zusätzliche Voraussetzung wird im folgenden verlangt, daß als Basis von $\mathcal{P}$ von nun an die zu $\mathcal{N}$ duale Basis verwendet wird, so daß ab sofort gilt:

$$N_k(\phi_j) = \delta_{kj}.$$

Dies ist nur eine technische Forderung für die Beweise in diesem Abschnitt und keine echte Einschränkung, da die interpolierende Funktion als solche von der speziellen Wahl der Basis natürlich unabhängig ist. Weiterhin bemerke man, daß die Existenz dieser speziellen Basis garantiert ist, falls man nur irgendeine Basis gefunden hat, in der der lokale Interpolationsoperator wohldefiniert ist. Um dies einzusehen, prüfe man nach, daß in diesem Fall die neue Basis $\{\hat{\phi}_1, \ldots, \hat{\phi}_K\}$, definiert durch

$$\hat{\phi}_k := P\,(I_{\mathrm{loc}} e_k)\,,$$

nach Konstruktion dual zu $\mathcal{N}$ ist und so die verlangte Eigenschaft besitzt.

Das erfreuliche an der neuen Basis ist nun, daß der lokale Interpolationsoperator eine sehr einfache Gestalt annimmt:

$$\begin{bmatrix} \Phi^* \\ R^*_\Phi \end{bmatrix} = \begin{bmatrix} N_1(\phi_1) & \ldots & N_1(\phi_K) \\ \vdots & \ddots & \vdots \\ N_K(\phi_1) & \ldots & N_K(\phi_K) \end{bmatrix} = \mathbf{1}_K,$$

in der Tat ist kaum etwas einfacheres vorstellbar als das nun vorliegende $I_{\mathrm{loc}} = \mathbf{1}_K$. Als Konsequenz gilt

$$c(g) = \begin{bmatrix} u(g) \\ v(g) \end{bmatrix} = \begin{bmatrix} N_1(g) \\ \vdots \\ N_K(g) \end{bmatrix},$$

$$\text{folglich ist } I_{\mathrm{loc}}(g) = \sum_{k=1}^{K} N_k(g)\phi_k 1_{\omega_i},$$

und der lokale Interpolationsoperator, wie wir ihn definiert haben, stimmt mit demjenigen aus [BS] überein, man vergleiche mit der Definition (3.3.1) dort. Die angenehme Folgerung ist zuerst einmal, daß alle Aussagen über die Qualität der Approximation sofort für unser Szenario verfügbar sind.

Das Resultat, an dem wir speziell interessiert sind, ist gültig, falls Ω polyedrisch ist und eine Familie von Zerlegungen $(\omega_i^h)_{1\le i\le F^h}$ zu reellen Zahlen $h \in H \subset (0,1]$ vorliegt, welche die folgenden Eigenschaften aufweist:

- Sie ist *nicht ausgeartet*, was hier bedeuten soll, daß eine Konstante $\rho > 0$ existiert so daß für alle $h \in H$ and $1 \le i \le F^h$:
$$\operatorname{diam} B_{\omega_i^h} \ge \rho \operatorname{diam} \omega_i^h,$$
wobei $B_{\omega_i^h}$ die größte Kugel ganz enthalten in ω_i^h ist, bezüglich der ω_i^h sternförmig ist.
- Die Zerlegungen werden schnell genug feiner, wenn h kleiner wird, in dem exakten Sinne daß
$$\max\{\operatorname{diam} \omega_i^h \;:\; 1 \le i \le F^h\} \le h \operatorname{diam} \Omega \quad \text{für alle } h \in H.$$

Seien weiterhin alle finiten Elemente in jeder Zerlegung affin äquivalent zu einem Referenzelement $(\omega^{\mathrm{ref}}, \mathcal{P}^{\mathrm{ref}}, \mathcal{N}^{\mathrm{ref}})$, für welches mit einer feste natürliche Zahl R mit $R > D/p + 1$ gilt:

- ω^{ref} ist sternförmig bezüglich einer geeigneten Kugel $B_{\omega^{\mathrm{ref}}} \subset \omega^{\mathrm{ref}}$,
- $\mathbb{P}_{R-1} \subset \mathcal{P}^{\mathrm{ref}} \subset W^R_\infty(\omega^{\mathrm{ref}})$, wobei $\mathbb{P}_{R-1}$ der Raum aller Polynome in D Variablen von Grad höchstens gleich $R-1$ ist, und
- $\mathcal{N}^{\mathrm{ref}} \subset \mathcal{C}^1(\overline{\omega}^{\mathrm{ref}})'$.

Eine solche Familie von Zerlegungen und Elementen wird *zulässig* genannt, im folgenden wird stillschweigend angenommen daß dies der Fall ist, wenn irgendwo der Index h auftaucht. Dann gilt das folgende Theorem über den Interpolationsfehler des Operators I^h zur Zerlegung (ω_i^h):

3.7 Satz. *Falls alle oben angeführten Voraussetzungen erfüllt sind, so existiert eine Konstante $C_{ip} > 0$, abhängig nur vom Referenzelement und den Konstanten D, R und ρ, so daß für alle $h \in H$, $0 \le s \le R$ und für alle $w \in W_p^R(\Omega)$:*

$$\left(\sum_{1 \le i \le F^h} \left\| w - I^h w \right\|^p_{W_p^s(\omega_i^h)} \right)^{\frac{1}{p}} \le C_{ip}\, h^{R-s}\ |w|_{W_p^R(\Omega)}\,.$$

Beweis. Dies ist genau Satz (4.4.20) aus [BS], leicht umformuliert, um unserer Notation gerecht zu werden. □

In dem uns interessierenden Fall müssen wir $s = 1$ und $p = 2$ setzen, da $\mathrm{V} \sqsubset H^1(\Omega)$. Das gewünschte Resultat ist also erzielt, falls $R > D/2+1 \ge 2$ gewählt werden kann, die zu interpolierende Lösung u_0 sollte also zumindest im Raum $H^3(\Omega)$ liegen, und bei höheren Dimensionen als 2 sogar noch glatter sein. Dies ist eine sehr starke Forderung an die Regularität des Problems und nur dann der Fall, wenn sowohl der Rand von Ω als auch die vorgegebene Funktion F und die Koeffizientenfunktionen des Differentialoperators genügend 'harmlos' sind. Entsprechende Kriterien finden sich z.B. in [H], Kapitel 9.1.

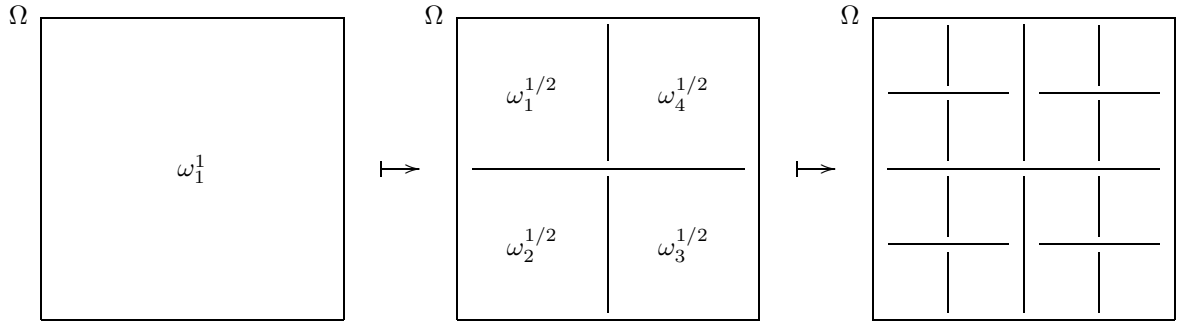
3.8 Beispiel. Zum Abschluß soll noch eine zulässige Folge von Zerlegungen für Intervalle in $\mathbb{R}^2$ mit $h \to 0$ konstruiert werden. Dies geschieht explizit nur für die die Klasse quadratischer Elemente, für die die Existenz des lokalen Interpolationsoperators in 3.6 bewiesen wurde. Das verwendete Schema funktioniert aber genauso für beliebige Elemente mit dieser Eigenschaft und einem Rechteck als Definitionsbereich.

Technisch wird Ω rekursiv geviertelt, diese rekursive Unterteilung ist auch eine Grundvoraussetzung für den Lösungsalgorithmus *Nested Dissection*, der in 8 beschrieben wird.

Als Referenzelement findet wie eingangs erwähnt das Element aus 3.6 Verwendung, mit $\omega^{\mathrm{ref}} = (0,1)^2$ und $\mathcal{P}^{\mathrm{ref}} = \langle 1, x, y, x^2, y^2, xy, x^3, xy^3 \rangle_{\mathbb{R}}$. Der minimale akzeptable Wert für R ist 3, da $l + D/2 = 1+1 = 2$, also muß der volle Raum von Polynomen zweiten Grades

$$\mathbb{P}_2 = \langle 1, x, y, x^2, xy, y^2 \rangle_{\mathbb{R}}$$

eine Teilmenge von $\mathcal{P}^{\mathrm{ref}}$ sein. Dies ist zum Glück der Fall, daher gilt Theorem 3.7, vorausgesetzt wir können eine Folge von Zerlegungen konstruieren, die nicht ausgeartet ist, deren Elemente schnell genug kleiner werden und die alle affin äquivalent zu unserem Referenzelement sind. Das hört sich nach viel Arbeit an, ist aber ganz leicht: Wir starten mit $h = 1$ entsprechend dem Ausgangsquadrat Ω und halbieren h in jedem Schritt, dabei wird jedes finite Element der aktuellen Zerlegung einmal quer und einmal längs unterteilt.



Dann ist offenbar $\operatorname{diam} B_{\omega_i^h} = \operatorname{diam} \omega_i^h$ für alle i und h, daher ist die Zerlegung nicht ausgeartet mit $\rho = 1$. Der Durchmesser aller finiten Elemente ω_i^h ist $h \cdot \operatorname{diam} \Omega$, und die Zerlegung wird schnell genug feiner in dem Sinne, der gefordert war. Die Eigenschaft der affinen Äquivalenz soll nach Konstruktion erfüllt sein: Von der Existenz einer affinen Abbildung, die ω^{ref} auf ω_i^h abbildet, überzeugt man sich durch scharfes Hinsehen. Die nodalen Variablen und Punkte von ω_i^h definiere man dann einfach so, daß die Eigenschaften für affine Äquivalenz gegeben sind. ◇

Eine wichtige Folgerung, die in [BS] wie schon die Aussagen zur Interpolation mit Hilfe der Theorie der Approximation durch Polynome in Sobolev-Räumen bewiesen worden ist, war die sogenannte *inverse Abschätzung*. Diese setzt verschiedene Normen auf den diskretisierenden Räumen V^h zueinander in Bezug.

Das entsprechende Theorem soll hier zur weiteren Verwendung in der Arbeit zitiert werden.

3.9 Theorem. *Seien $0 \leq m \leq l$ und $1 \leq p \leq \infty, 1 \leq q \leq \infty$ beliebig. Dann existiert eine von h unabhängige Konstante C, so daß für alle $u^h \in V^h$:*

$$\left\|u^h\right\|_{h,W_p^l} \leq h^{m-l+\min\left(0,\frac{D}{p}-\frac{D}{q}\right)} \cdot C \left\|u^h\right\|_{h,W_q^m}$$

Beweis. Siehe [BS], Theorem 4.5.11. □

3.4 Interpolation auf Kollokationspunkten

Bisher ist der Interpolationsoperator so definiert worden, daß er die Werte der nodalen Variablen auf den Rändern als Eingabedaten bekommt. Dies führt auf Gleichungssysteme, welche sich rein lokal lösen lassen. Schwieriger in den Griff zu bekommen ist die Aufgabe, eine Interpolation durchzuführen, wenn eine Funktion in $\mathcal{V}^h$ mit vorgegebenen Werten auf den Kollokationspunkten zu finden ist, denn dabei sind die gemeinsamen Werte der lokalen Funktionen in den Verbindungspunkten selbst Unbekannte, man hat also ein im Allgemeinen sehr großes und unüberschaubares globales Gleichungssystem zu lösen. Diese Aufgabe muß jedoch angegangen werden, zum einen muß die Existenz von Funktionen mit bestimmten Eigenschaften für den Übergang zum Kollokationsverfahren sichergestellt werden, zum anderen möchte man natürlich wissen, ob das entstehende globale Gleichungssystem, das beim Algorithmus von Doedel entsteht, überhaupt lösbar ist. Beide Aufgaben sind gewissermaßen äquivalent, wie sich gleich zeigen wird.

Im Endeffekt werden wir dann zwei verschiedene Möglichkeiten haben, den Interpolationsoperator zu wählen und es stellt sich natürlich die Frage, ob es nicht übertriebene Mehrarbeit darstellt, die Existenz von beiden zu sichern. Einerseits würde es jedoch noch sehr viel mehr theoretische Arbeit erfordern, für die Interpolation auf Kollokationspunkten die Konvergenz zu beweisen, die ja beim bereits untersuchten Operator fast umsonst aus der Theorie der finiten Elemente übernommen worden konnte. Andererseits ist jedoch wie bereits erwähnt die Interpolation von Funktionen mit vorgegebenen Werten auf den Kollokationspunkten für den Übergang zum Kollokationsverfahren absolut notwendig. Im Endeffekt scheint es also unumgänglich, tatsächlich beide Möglichkeiten parallel zu untersuchen.

Schreiten wir nun zur Tat. Es soll untersucht werden, welche Forderungen an das Referenzelement gestellt werden müssen, damit zu einer vorgegebenen Funktion $g \in \mathcal{C}(\Omega)$ und einem linearen Operator $A : \mathrm{V}^h \to \mathrm{V}^h$ eindeutig eine Funktion $\phi \in \mathcal{V}^h$ existiert, so daß

$$g(z) = A\phi(z) \text{ für alle } z \in Z^h.$$

Wegen $\dim \mathcal{V}^h = F^h \cdot M = \#Z^h$ ist durchaus zu hoffen, daß dies unter gewissen Voraussetzungen der Fall sein wird. Der exakte Beweis, wann dieser Fall eintritt, wird im folgenden mit Hilfe einer Induktion über die finiten Elemente einer Zerlegung für festes $h \in H$ erbracht werden.

Für den Induktionsanfang wird sicherlich eine erste Forderung sein müssen, daß die Interpolationsaufgabe für das Referenzelement lösbar ist. Diese und die weiteren notwendige Eigenschaften sollen nun präzisiert werden. Dabei werden zunächst einige Matrizen eingeführt, die die Formulierung der Resultate erleichtern sollen.

Das Referenzelement besitze S Seiten mit jeweils L Verbindungspunkten für Funktionswerte u und Normalenableitungen v, d.h. $N = S \cdot L$. Zu einer Seite korrespondierende Objekte werden durch ein hochgestelltes $s \in \{1, \ldots, S\}$ kenntlich gemacht. Es gelten dann für eine lokale Funktion

$$\phi = c_1\phi_1 + \cdots + c_K\phi_K$$

und für alle $1 \leq s \leq S$ die folgenden Beziehungen:

$$u^s = \begin{bmatrix} \phi_1(x_1^s) & \ldots & \phi_K(x_1^s) \\ \vdots & \ddots & \vdots \\ \phi_1(x_L^s) & \ldots & \phi_K(x_L^s) \end{bmatrix} \cdot c =: \Phi^{*s} \cdot$$

$$v^s = \begin{bmatrix} \nabla\phi_1 \cdot \eta(y_1^s) & \ldots & \nabla\phi_K \cdot \eta(y_1^s) \\ \vdots & \ddots & \vdots \\ \nabla\phi_1 \cdot \eta(y_N^s) & \ldots & \nabla\phi_K \cdot \eta(y_N^s) \end{bmatrix} \cdot c =: R_\Phi^{*s} \cdot c$$

und

$$G = \begin{bmatrix} A\phi_1(z_1) & \dots & A\phi_K(z_1) \\ \vdots & \ddots & \vdots \\ A\phi_1(z_M) & \dots & A\phi_K(z_M) \end{bmatrix} \cdot c =: \Psi \cdot c,$$

dabei ist $G = \begin{bmatrix} g(z_1) & \dots & g(z_M) \end{bmatrix}^T$ der Vektor von Funktionswerten in den Kollokationspunkten.

Induktionsanfang. Die Forderung ist, daß das Kollokationsproblem stets lösbar ist, wenn auf den Seiten eines einzelnen finiten Elementes entweder Funktionswerte oder Ableitungswerte vorgegeben werden - dabei sollen diese auch gemischt auftreten dürfen, nur auf jeder Seite muß man sich auf eines von beidem festlegen. Etwas genauer heißt dies in mathematischer Notation unter Verwendung der eben definierten Matrizen:

- Für jede beliebige Wahl von $\Upsilon^s \in \{\Phi^{*s}, R_\Phi^{*s}\}$ und $w^s \in \mathbb{R}^L$ für alle $1 \le s \le S$ soll das Gleichungssystem

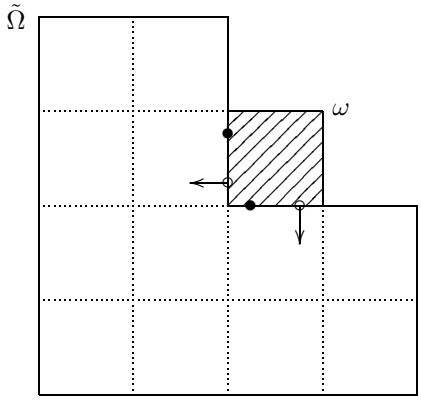
$$\begin{bmatrix} w^1 \\ \vdots \\ w^S \\ G \end{bmatrix} = \begin{bmatrix} \Upsilon^1 \\ \vdots \\ \Upsilon^S \\ \Psi \end{bmatrix} \cdot c =: \Upsilon \cdot c \tag{3.1}$$

  eindeutig lösbar sein, d.h die Matrix Υ invertierbar.

Induktionsschluß. Ist Lösbarkeit des lokalen Falles gegeben, so kann man sich induktiv an das gesamte Gebiet Ω herantasten. Es muß dafür untersucht werden, wie es um die Existenz einer interpolierenden Funktion steht, wenn man an ein Gebiet, auf dem die Interpolationsaufgabe lösbar ist, ein weiteres finites Element 'anklebt'. Das Gebiet Ω und die Diskretisierung müssen dafür die folgende Verträglichkeitsbedingung erfüllen:

- Ist $\tilde{\Omega}$ eine Vereinigung abgeschlossenen finiten Elementen, so ist eine *zulässige Erweiterung von* $\tilde{\Omega}$ eine Vereinigung $\tilde{\Omega} \cup \omega$ mit einem finiten Element ω derart, daß der Durchschnitt $\overline{\tilde{\Omega}} \cap \overline{\omega}$ genau aus einer Anzahl σ von Seiten von ω besteht. Wegen der Konvexität der finiten Elemente gibt es dann σ verschiedene Elemente in $\tilde{\Omega}$, welche mit ω genau eine Seite gemeinsam haben.

  Ab sofort wird gefordert, daß für alle $h \in H$ das Gesamtgebiet $\Omega = \bigcup_{i=1}^{F^h} \omega_i^h$ durch eine Folge von zulässigen Erweiterungen aus ω_1^h entsteht.



- Punkte in $X_{\tilde{\Omega}} \cap X_\omega$ für Berechnung von u^*
- ∘ Punkte in $Y_{\tilde{\Omega}} \cap Y_\omega$ für Berechnung von v^*
- ω, Lebensbereich der lokalen Funktion ϕ

Es muß jetzt lediglich noch untersucht werden, ob die Kollokationsaufgabe auf einer zulässigen Erweiterung $\tilde{\Omega} \cup \omega$ mit beliebigen vorgegebenen gemischten Randbedingungen lösbar ist, falls sie auf $\tilde{\Omega}$ lösbar ist. Dann folgt die Lösbarkeit für das Gesamtgebiet Ω durch Induktion nach der Anzahl der Elemente.

Sei also $\tilde{\Omega} \cup \omega$ eine zulässige Erweiterung von $\tilde{\Omega}$. Die Seiten von ω seien so numeriert, daß die gemeinsamen σ Seiten in der Numerierung die ersten sind. Sei $u^* \in \mathbb{R}^{\sigma L}$ der Vektor der Funktionswerte und $v^* \in \mathbb{R}^{\sigma L}$ der Vektor der Ableitunswerte auf den gemeinsamen Seiten. Sei $w \in \mathbb{R}^{(S-\sigma)L}$ der Vektor der übrigen vorgegebenen Daten (gemischte Ableitungs- und Funktionswerte) auf den restlichen Seiten von ω. Nach Voraussetzung existiert dann eine invertierbare Matrix Υ, so daß für die gesuchte lokale Funktion $\phi = c_1^*\phi_1 + \cdots + c_K^*\phi_K$ auf ω gilt:

$$\Upsilon c^* = \begin{bmatrix} u^* \\ w \\ G \end{bmatrix} \Leftrightarrow c^* = \Upsilon^{-1} \begin{bmatrix} u^* \\ w \\ G \end{bmatrix}, \tag{3.2}$$

wobei G wieder der Vektor der Funktionswerte in den Kollokationspunkten von ω ist. Weiterhin hat man

$$v^* = \begin{bmatrix} R_\Phi^{*1} \\ \vdots \\ R_\Phi^{*\sigma} \end{bmatrix} \cdot c^* =: \Xi \cdot c^*. \tag{3.3}$$

Nach Induktionsannahme ist das Interpolationsproblem auf dem Gebiet $\tilde{\Omega}$ für beliebige vorgegebene Randbedingungen eindeutig lösbar, daher kommt eine dritte Gleichung ins Spiel:

$$v^* = \Phi_{\tilde{\Omega}} u^* + k_{\tilde{\Omega}},$$

d.h. v^* und u^* hängen aufgrund der auf $\tilde{\Omega}$ gültigen Bedingungen affin linear voneinander ab, mit invertierbarer Matrix $\Phi_{\tilde{\Omega}}$ und einem festen Vektor $k_{\tilde{\Omega}}$. Aus dem Induktionsschritt wird sofort folgen, daß das dann auch für das erweiterte Gebiet richtig ist. Zur Vereinfachung kann weiter angenommen werden, daß die Basis so gewählt ist, daß $\Phi_{\tilde{\Omega}} = \mathbf{1}_{(\sigma L)}$ gilt.

Es müssen nun u^*, v^*, c^* gefunden werden, so daß die drei Gleichungen erfüllt sind. Sei dafür $(\Xi\Upsilon^{-1})^L$ die Matrix gebildet aus den $\sigma \cdot L$ linken Spalten von $\Xi\Upsilon^{-1}$. Dann folgt durch Einsetzen von (3.2) in (3.3) die Bedingung:

$$\mathbf{1}_{(\sigma L)} u^* + k_{\tilde{\Omega}} = \Xi \cdot c^* = (\Xi\Upsilon^{-1})^L u^* + (\Xi\Upsilon^{-1}) \begin{bmatrix} 0 \\ w \\ G \end{bmatrix}$$

$$\Leftrightarrow \left(\mathbf{1}_{(\sigma L)} - (\Xi\Upsilon^{-1})^L\right) u^* = (\Xi\Upsilon^{-1}) \begin{bmatrix} 0 \\ w \\ G \end{bmatrix} - k_{\tilde{\Omega}}.$$

Die Matrix links ist invertierbar, falls 1 kein Eigenwert der Matrix $(\Xi\Upsilon^{-1})^L$ ist. Diese Eigenschaft ist von der Wahl der Basis unabhängig, so daß wir als letztes die folgende lokale Bedingung stellen wollen:

- Bei beliebiger Wahl der Art der Randbedingungen soll niemals die Zahl 1 ein Eigenwert der Matrix $(\Xi\Upsilon^{-1})^L$ sein.

Ist dieses Kriterium erfüllt, so erhält man also bei beliebig vorgegebenen Randbedingungen stets eine Lösung für u^*, durch Rückwärtseinsetzen in Gleichung (3.3) also auch für v^* und schließlich aus Gleichung (3.2) für c^*. Die lokale Funktion, die sich daraus ergibt, erfüllt dann die Interpolationsbedingungen, womit der Induktionsschluß erbracht ist.

Mithin haben wir das folgende Theorem, welches zum Glück lediglich leicht nachprüfbare lokale Forderungen an die Lösbarkeit von Gleichungssystemen stellt und das Ausgangsproblem damit wesentlich reduziert hat.

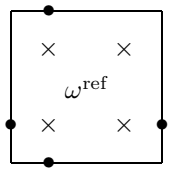
3.10 Theorem. *Die drei mit '•' markierten Bedingungen seien für einen vorgegebenen Operator $A : V^h \to V^h$ erfüllt. Für jede vorgegebene Funktion $g \in \mathcal{C}(\Omega)$ existiert dann eine Funktion $\phi \in \mathcal{V}^h$, so daß*

$$A\phi(z^h) = g(z^h) \text{ für alle } z \in Z^h.$$

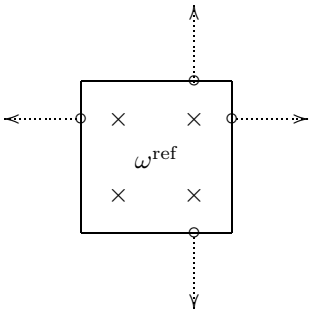
Sind die Bedingungen speziell für $A = \mathrm{id}_{V^h}$ erfüllt, dann gibt es also für alle $h \in H$ und beliebige vorgegebene Funktionswerte auf den Kollokationspunkten eine Funktion in $\mathcal{V}^h$, die diese Werte annimmt.

3.5 Beschränktheit

In diesem Abschnitt soll gezeigt werden, daß speziell der Interpolationsoperator auf den Kollokationsstellen beschränkt ist. Diese Eigenschaft ist notwendig, damit die Stabilität beim späteren Übergang von der schwachen Formulierung zu den Kollokationsgleichungen erhalten bleibt. Um das Ziel zu erreichen, werden zunächst einige weitere elementare Eigenschaften der lokalen Interpolation bewiesen. Zwei verschiedene Arten der Interpolation werden hier eine Rolle spielen: Der erste Operator I_u^h interpoliert Funktionswerte auf Kollokationsstellen und Verbindungspunkten, der zweite Operator I_v^h die Werte der Normalenableitungen auf Verbindungspunkten und die Funktionswerte auf den Kollokationsstellen.



I_u^h interpoliert Funktionswerte u auf Verbindungspunkten und Funktionswerte z auf Kollokationspunkten



I_v^h interpoliert die Werte v der Normalenableitungen auf Verbindungspunkten und Funktionswerte z auf Kollokationspunkten

Beide Operatoren sind nach der im letzten Abschnitt angeführten Voraussetzung 3.1 wohldefiniert und approximieren die exakte Funktion mit kleiner werdendem h beliebig gut. Sie bilden jedoch *nicht* in den Raum V^h ab, da entweder die Verbindungsbedingungen für Funktionswerte oder Normalenableitungen im allgemeinen verletzt sein werden. Es gilt jedoch

$$I_u^h : H^R(\Omega) \to \tilde{\mathrm{V}}^h \text{ und } I_v^h : H^R(\Omega) \to \tilde{\mathrm{V}}^h,$$

wobei jeweils

$$\tilde{\mathrm{V}}^h := \{u \in L^\infty \,:\, u\big|_{\omega_i^h} \in \mathcal{P}_i^h \text{ für alle } 1 \le i \le F^h \,\}.$$

Das erste Lemma sagt aus, daß beide Operatoren sich lokal unabhängig von h abschätzen lassen durch eine Konstante multipliziert mit dem Maximum aller Eingabedaten.

3.11 Lemma. *Sei $w = (u, z) \in \mathbb{R}^{M+N}$, bzw. $w = (v, z) \in \mathbb{R}^{M+N}$ ein Vektor von Eingabedaten für die lokalen Interpolationsoperatoren. Dann gilt für die interpolierenden lokalen Funktionen:*

$$\left\|(I_u^h w)_i\right\|_h \le C \left\|w\right\|_\infty, \text{ bzw. } \left\|(I_v^h w)_i\right\|_h \le \tilde{C} \left\|w\right\|_\infty.$$

Die Konstanten C und $\tilde{C}$ sind von h, w und i unabhängig.

Beweis. Der Beweis kann für beide Fälle gleichzeitig durchgeführt werden. In beiden hat man nämlich lokal auf dem finiten Element ω_i^h eine Matrix $I_{\text{loc,i}}^h$, die den Eingabedaten w die Koeffizienten $c_i = I_{\text{loc,i}}^h w$ der interpolierenden lokalen Funktion zuordnet. Für diese gilt dann nach Definition mit der affinen Transformation $\varphi := \varphi_i^h$, die ω_i^h auf ω^{ref} abbildet:

$$\begin{aligned}
\left\|P_i^h\left(I_{\text{loc,i}}^h w\right)\right\|_{H^1(\omega_i^h)} &= \left\|\sum_{k=1}^K (I_{\text{loc,i}}^h w)_k \, (\varphi^* \phi_k)\right\|_{H^1(\omega_i^h)} && \text{(Definition } P_i^h\text{)} \\
&\le \sum_{k=1}^K \left|(I_{\text{loc,i}}^h w)_k\right| \cdot \left\|\varphi^* \phi_k\right\|_{H^1(\omega_i^h)} && \text{(Dreiecksungleichung)} \\
&\le \left\|I_{\text{loc,i}}^h w\right\|_\infty \sum_{k=1}^K \left\|\phi_k\right\|_{H^1(\omega^{\text{ref}})} && \text{(Lemma A.3)} \\
&\le \left\|I_{\text{loc,i}}^h\right\|_\infty \left\|w\right\|_\infty \sum_{k=1}^K \left\|\phi_k\right\|_{H^1(\omega^{\text{ref}})} && (I_{\text{loc,i}}^h \text{ stetig}) \\
&= C \cdot \left\|I_{\text{loc,i}}^h\right\|_\infty \left\|w\right\|_\infty
\end{aligned}$$

mit

$$C := \sum_{k=1}^{\hat{K}} \|\phi_k\|_{H^1(\omega^{\mathrm{ref}})} .$$

Da nach Lemma 3.5 die Matrix $I^h_{\mathrm{loc,i}}$ für alle zu ω^{ref} affin äquivalenten finiten Elemente gleich ist, folgt die Behauptung. □

Das zweite Lemma besagt, daß bei gewissen interpolierenden Funktionen die lokalen Werte der Normalenableitungen beschränkt sind durch die lokalen Funktionswerte und umgekehrt.

3.12 Lemma. *Für eine lokale Funktion aus $\mathcal{P}^h_i$, die auf allen Kollokationsstellen den Wert Null annimmt, ist der Zusammenhang zwischen den Funktionswerten u auf den Verbindungspunkten und den Werten v der Normalenableitungen auf den Verbindungspunkten gegeben durch*

$$v \;=\; \Pi \cdot u,$$

wobei $\Pi \in \mathbb{R}^{M\times M}$ eine von h und i unabhängige Matrix ist. Insbesondere gilt

$$\|v\|_\infty \;\le\; C\, \|u\|_\infty$$

mit einer von h und i unabhängigen Konstanten $C > 0$.

Beweis. Dies wurde bereits bewiesen: Der lineare Zusammenhang folgt aus den Gleichungen (3.2) und (3.3), die Unabhängigkeit der Matrix vom finiten Element steht in Lemma 3.5. □

Das dritte Lemma schließlich enthält die gesamte Technik, die man braucht, um das globale Interpolationsproblem auf den Kollokationsstellen - das ja letztendlich die Lösung eines sehr komplizierten globalen Gleichungssystems erforderte, über dessen Verhalten man nur sehr wenig wußte - auf Approximationen zu reduzieren, die im wesentlichen lokal sind.

Im folgenden bezeichnet für eine Funktion $g \in \tilde{\mathrm{V}}^h$ der Vektor $[u]_g$ die Werte der Sprünge in den Funktionswerten auf den Verbindungspunkten, etwas formaler:

$$[u]_g \;:=\; \big(g_i(x) - g_j(x) \text{ für } x \in \omega^h_i \cap \omega^h_j\big)_{x\in X^h} .$$

Analog ist $[v]_g$ der Sprung in den Normalenableitungen:

$$[v]_g \;:=\; \big(\nabla g_i \cdot \eta(y) - \nabla g_j \cdot \eta(y) \text{ für } y \in \omega^h_i \cap \omega^h_j\big)_{y\in X^h} .$$

Da I^h_u die Funktionswerte exakt interpoliert, hat man also stets

$$[u]_{I^h_u(g)} = 0, \text{ analog ebenso } [v]_{I^h_v(g)} = 0.$$

Es wird jedoch im allgemeinen

$$[u]_{I^h_v(g)} \neq 0, \text{ sowie } [v]_{I^h_u(g)} \neq 0$$

sein. Eine Funktion $g \in \tilde{\mathrm{V}}^h$ liegt in V^h genau dann, wenn sowohl $[u]_g$ als auch $[v]_g$ Null ist.

3.13 Lemma. *Sei $h \in H$ fest und $[v] \in \mathbb{R}^{M\cdot F^h}$ ein Vektor von vorgegebenen Fehlern in den Normalenableitungen. Dann existiert eine* Korrekturfunktion *$\tilde{g} = \tilde{g}([v]) \in L^\infty(\Omega)$ mit $\tilde{g}\big|_{\omega^h_i} \in \mathcal{P}^h_i$, so daß gilt:*

(i) $\tilde{g}$ hat auf allen Kollokationsstellen Z^h den Funktionswert 0.

(ii) $\tilde{g}$ verringert den Fehler in den Normalenableitungen für hinreichend kleine h:

$$\|[v] \,+\, [v]_{\tilde{g}}\|_\infty \;\le\; h \cdot C\, \|[v]\|_\infty .$$

(iii) $\tilde{g}$ macht keine Fehler bei den Funktionswerten: $[u]_{\tilde{g}} = 0$.

(iv) $\|\tilde{g}\|_h \;\le\; \tilde{C} \cdot \|[v]\|_\infty$.

Die Konstanten C und $\tilde{C}$ hängen dabei weder von h noch von $[v]$ ab.

Beweis. Der Beweis erfolgt in mehreren Schritten. Zunächst wird der Fehler in den Normalenableitungen zu Null gemacht, dabei handelt man sich allerdings einen Fehler in den Funktionswerten ein. Dieser wird dann nachkorrigiert, wodurch man im Endeffekt den Fehler in den Normalenableitungen reduziert hat.

Schritt 1: Es existiert eine Funktion $\phi \in L^\infty(\Omega)$ mit $\phi\big|_{\omega_i^h} \in W^1_\infty(\omega_i^h)$, so daß

$$|\phi|_{h,W^1_\infty} \;\leq\; \|[v]\|_\infty\,,\;\; \|\phi\|_{L^\infty(\Omega)} \;\leq\; h\cdot C\,\|[v]\|_\infty\,,\; \phi\big|_{Z^h} = 0 \text{ und } [v]_\phi \;=\; [v].$$

Die Konstruktion dieser Funktion ist in Lemma A.7 ausgearbeitet.

Schritt 2: Wähle als erste Korrektur die Interpolierende $g_1 := I_v^h(\phi)$. Es gilt dann nach der Abschätzung für den Interpolationsfehler nach [BS], Theorem 3.7:

$$\begin{aligned} \big\|\phi \,-\, I_v^h(\phi)\big\|_{L^\infty(\Omega)} \;&\leq\; h\cdot C\,|\phi|_{h,W^1_\infty} \\ &\leq\; h\cdot C\,\|[v]\|_\infty \\ \Longrightarrow \qquad \big\|I_v^h(\phi)\big\|_{L^\infty(\Omega)} \;&\leq\; h\cdot C\,\|[v]\|_\infty\,. \end{aligned}$$

Da der Sprung in den Funktionswerten einer Funktion niemals größer sein kann als das Doppelte der L^∞-Norm, gilt auch für den Fehler in den Funktionswerten

$$\|[u]_{g_1}\|_\infty \;\leq\; \big\|I_v^h(\phi)\big\|_{L^\infty(\Omega)} \;\leq\; h\cdot C\,\|[v]\|_\infty\,.$$

Schritt 3: Es existiert eine Funktion $\tilde\phi \in L^\infty(\Omega)$ mit $\tilde\phi\big|_{\omega_i^h} \in \mathcal{C}^\infty(\omega_i^h)$, so daß

$$\tilde\phi\big|_{Z^h} = 0 \text{ und } [u]_{\tilde\phi} \;=\; [u]_{g_1}.$$

Die Konstruktion dieser Funktion wird in Lemma A.6 vorgeführt.

Schritt 4: Korrektur von $[u]_{g_1}$. Als zweite Korrektur setzt man nun $g_2 := I_u^h(\tilde\phi)$ an. Wegen Lemma 3.12 kann man dann wie folgt abschätzen:

$$\|[v]_{g_2}\|_\infty \;\leq\; C\,\|[u]_{g_1}\|_\infty \;\leq\; h\cdot C\,\|[v]\|_\infty\,.$$

Die Funktion $\tilde g := -g_1 + g_2$ hat also nach Konstruktion die gewünschten Eigenschaften (i), (ii) und (iii): (i) ist klar, (ii) folgt wegen

$$\begin{aligned} \|[v] \,+\, [v]_{-g_1+g_2}\|_\infty \;&=\; \|[v] \,+\, [v]_{-g_1} \,+\, [v]_{g_2}\|_\infty \\ &\leq\; \|[v] \,+\, [v]_{-g_1}\|_\infty \;+\; \|[v]_{g_2}\|_\infty \\ &=\; \|[v]_{g_2}\|_\infty \\ &\leq\; h\cdot C\,\|[v]\|_\infty\,, \end{aligned}$$

und Eigenschaft (iii) wegen

$$\begin{aligned} [u]_{\tilde g} \;&=\; [u]_{-g_1+g_2} \;=\; [u]_{-g_1} \,+\, [u]_{\tilde\phi} \\ &=\; [u]_{-g_1} \,+\, [u]_{g_1} \;=\; 0. \end{aligned}$$

Schließlich hat man auch Eigenschaft (iv) nach Lemma 3.11, da die Norm $\|\diamond\|_h$ einer interpolierenden Funktion durch das Maximum aller Eingabedaten beschränkt ist. □

Mit Hilfe dieses Lemmas ist nun der Beweis unseres Hauptanliegens nur noch etwas Rechentechnik:

3.14 Theorem. *Es existiert ein $0 < h_0 \leq 1$ mit folgender Eigenschaft: Falls $h_0 \geq h \in H$ und $z \in \mathbb{R}^{N\cdot F^h}$ ein vorgegebener Vektor von Funktionswerten auf den zugehörigen Kollokationsstellen ist, dann existiert eine Funktion $g \in V^h$, die diese Werte dort annimmt. Für ihre Norm gilt*

$$\|g\|_h \;\leq\; C\cdot\|z\|_\infty$$

mit einer von h und z unabhängigen Konstanten C.

Beweis. Man interpoliere zunächst die vorgegebenen Werte z auf den Kollokationsstellen durch die Funktion $g_0 := I_u^h(z, 0)$, wobei die Funktionswerte auf den Verbindungspunkten überall Null gesetzt werden. Dabei macht man nach Lemma 3.12 bei den Ableitungswerten einen Fehler $[v_0]$ mit

$$\|[v_0]\|_\infty \leq C \cdot \|z\|_\infty .$$

Definiere nun mit Lemma (3.13) induktiv eine Folge $(\tilde{g}_j)$ von Korrekturen durch

$$\tilde{g}_{j+1} \;:=\; \tilde{g}([v_j]) \text{ und } [v_{j+1}] \;:=\; [v_j] + [v]_{\tilde{g}_{j+1}}.$$

Es gilt dann nach (3.13, ii)

$$\|[v_j]\|_\infty \;\leq\; h \cdot C \;\|[v_{j-1}]\|_\infty \;\leq\; \ldots \;\leq\; (h \cdot C)^j \;\|[v_0]\|_\infty , \tag{3.4}$$

damit folgt nach (3.13, iv)

$$\|\tilde{g}_{j+1}\|_h \;\leq\; \tilde{C} \cdot \|[v_j]\|_\infty \;\leq\; (h \cdot C)^j \,\tilde{C} \;\|v_0\|_\infty .$$

Durch Addition gelangt man zu einer Folge von Funktionen

$$g_k \;:=\; g_0 \;+\; \sum_{j=1}^{k} \tilde{g}_j.$$

Da C von h unabhängig ist, kann $h_0 \ll 1$ gewählt werden, so daß $h_0 \cdot C =: q < 1$. Dann ist die Reihe absolut konvergent in V^h, und es existiert

$$g := \lim_{k \to \infty} g_k,$$

außerdem ist die Norm von g unabhängig von h beschränkt. Dies sieht man folgendermaßen ein: Für $\tilde{h} < h_0$ ist

$$\begin{aligned}
\|g\|_h \;&\leq\; \|g_0\|_h \;+\; \sum_{j \geq 1} \|\tilde{g}_j\|_h \\
&\leq\; C' \;\|z\|_\infty \left(1 \;+\; \sum_{j \geq 0} (\tilde{h} \cdot C)^j \right) \\
&\leq\; C' \;\|z\|_\infty \left(1 \;+\; \sum_{j \geq 0} q^j \right) \\
&=\; C' \;\|z\|_\infty \left(1 \;+\; \frac{1}{1-q} \right),
\end{aligned}$$

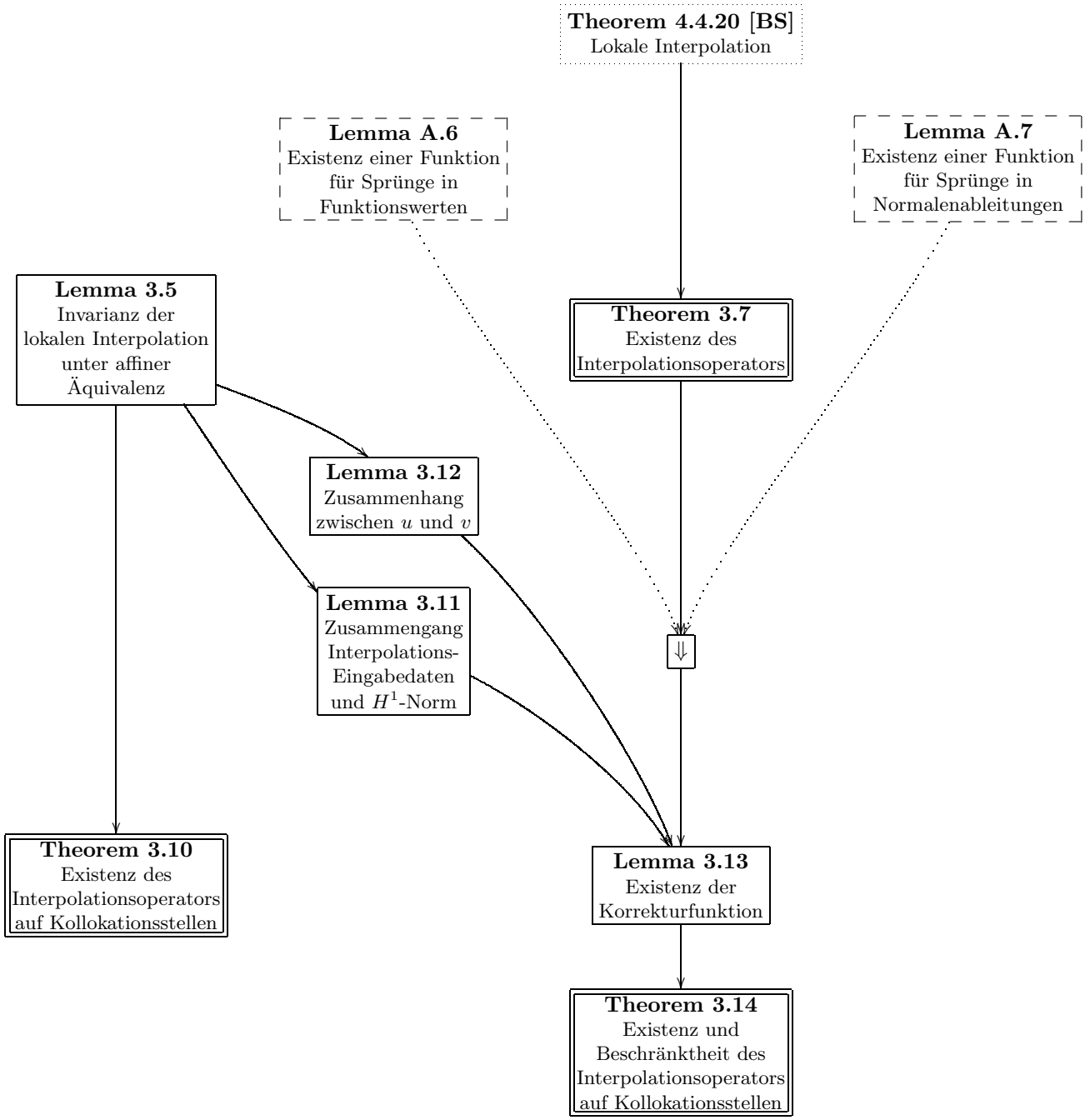
alle auftretenden Konstanten sind unabhängig von h.

Nach (3.13, iii) ist $[u]_g = 0$. Hinzu kommt die Abschätzung (3.4) für den Fehler beim Übergang der Normalenableitungen, daher liegt g in V^h. Nach Konstruktion erfüllt g wegen (3.13, i) außerdem die Forderung an die Funktionswerte in den Kollokationsstellen. □

3.15 *Bemerkung.* Das vorliegende Theorem ist auch ein neuerlicher Beweis dafür, daß der Interpolationsoperator auf den Kollokationspunkten existiert. Er ist allerdings weniger konstruktiv als der aus Theorem 3.10, und auch nicht für allgemeinere Operatoren als die Identität brauchbar. Ein großer Vorteil ist allerdings, daß von den Voraussetzungen lediglich die erste zur Lösbarkeit des lokalen Gleichungssystems (3.1) bestehen bleibt.

3.6 Beweisstruktur

Zur besseren Orientierung wird jedem Kapitel aus dem theoretischen Teil ein Graph beigefügt, der die Zusammenhänge in der Beweisstruktur zwischen den zugrundeliegenden, bzw. den im Verlauf des Kapitels verwendeten oder bewiesenen Sätzen illustriert.



Kapitel 4

Glättung

Dieses Kapitel ist im wesentlichen der Existenz der Glättungsoperatoren $E^h : \mathrm{V}^h \to \mathrm{V}$ mit den geforderten Eigenschaften gewidmet. Das entsprechende Hauptresultat lautet:

4.1 Satz. *Es existiert eine Familie von Operatoren $E^h : V^h \to V$ und eine von h unabhängige Konstante C_{ac}, so daß für alle $h \in H$ und $v^h \in V^h$:*

$$\left\| v^h - E^h v^h \right\|_h \leq h^{1/2} \cdot C_{ac} \left\| v^h \right\|_h$$

Die E^h haben weiterhin die Eigenschaft, daß die jeweiligen Interpolationsoperatoren I^h linksseitige Inverse zu ihnen sind, d.h es gilt für alle $h \in H$:

$$I^h \circ E^h = id_{V^h}$$

Man hat damit also eine Aussage über die Güte der Approximation einer gegebenen Funktion in V^h durch Funktionen in V. Dabei ist leider festzustellen, daß diese nicht mehr wie bei der Interpolation von Eigenschaften des Referenzelementes und der exakten Lösung abhängt, sondern lediglich linear mit h besser wird. Es wird also keine Methode an die Hand gegeben, um wie im Falle der Interpolation die Konvergenzordnung durch geeignete Veränderung der Parameter der Diskretisierung zu verbessern, wie zum Beispiel durch die Wahl von mehr Verbindungspunkten. Dies ist jedoch nicht weiter tragisch: Für das Hauptresultat bezüglich der Konvergenz der Kollokationslösungen wird der Glättungsoperator lediglich eingesetzt, um Stabilität nachzuweisen. Die Konvergenzordnung der Kollokationslösungen gegen die exakte Lösung hängt jedoch von der Glättung nicht ab.

Ein weiteres Ergebnis, welches zum Kontext der Glättung gehört, ist die Analyse der Fehler, die in der Gestalt von Randintegralen beim Übergang zur Operatordarstellung für die Bilinearformen a^h auftreten.

4.0 Konstruktion des Operators und Fehlerabschätzung

Wir wollen zunächst die Konstruktion angehen, die wie schon beim Interpolationsoperator lokal angegangen wird. In der Tat ist der Glättungsoperator sogar ein spezieller Interpolationsoperator, allerdings auf einem *erweiterten Referenzelement*. Dieses wird dann konform sein in dem Sinne, daß die davon induzierten interpolierenden Funktionen in V liegen. Erreicht wird dies durch Vergrößerung des lokalen Funktionenraumes und das Hinzufügen von nodalen Variablen, wie sie in folgendem Lemma konstruiert werden.

4.2 Lemma. *Es existiert ein erweitertes Referenzelement $\left(\omega^{ref}, Z, \hat{\mathcal{P}}, \hat{\mathcal{N}}\right)$ mit $\mathcal{N}^{ref} \subset \hat{\mathcal{N}}$ und $\mathcal{P}^{ref} \subset \hat{\mathcal{P}}$, so daß für die induzierten Interpolationsoperatoren $\hat{I^h}$ gilt:*

$$\text{bild}\, \hat{I^h} \subset V$$

Beweis. (Skizze). Sei $\hat{L}$ die Anzahl von Verbindungspunkten für Funktionswerte auf jeder Seite des erweiterten Referenzelements und $\hat{k}$ der maximale Grad der Polynome aus $\hat{\mathcal{P}}$ eingeschränkt auf eine Seite. Ein einzelner Erweiterungsschritt bestehe darin, einen Verbindungspunkt auf jeder Seite des finiten Elements und eine passende Anzahl linear unabhängiger Polynome von minimal möglichem Grad zu $\hat{\mathcal{P}}$ hinzuzufügen. Dies muß und kann derart geschehen, daß die Bedingung für die Existenz der lokalen Interpolation, $\det \hat{I}_{\text{loc}} \neq 0$, erfüllt bleibt.

Da nun der Grad der Polynome langsamer wächst als die Anzahl der Punkte pro Seite, ist nach einer endlichen Anzahl von Erweiterungsschritten $\hat{L} \geq \hat{k} + 1$ erreicht. Das bedeutet aber, daß die lokalen Funktionen aneinandergrenzender finiter Elemente auf den Elementrändern in mehr als $\hat{k}$ Punkten übereinstimmen. Da es sich dort um Polynome von Grad $\hat{k}$ handelt, müssen sie also auf dem ganzen Rand gleich sein, mithin hat man Konformität erreicht.

Diese Hinweise sollten ausreichen, um eine recht gute Vorstellung vom genauen Beweisgang zu erhalten. Der exakte Algorithmus, um das erweiterte Referenzelement zu konstruieren, ist zusammen mit dem wegen seiner Länge hier nicht ausformulierten Beweis seiner Korrektheit zu finden bei [B], Algorithmus 2.25. □

Dies definiert zunächst nur einen Operator auf V, der jedoch fortgesetzt werden kann zu einem Operator auf H^h, und der dann mithin auch auf V^h definiert ist. Der Grundgedanke ist, daß das Ergebnis der Interpolation nur abhängig ist vom Vektor der Eingabedaten, den man auch für unstetige Funktionen geeignet definieren kann. Man lege die 'Funktionswerte' für eine zu interpolierende Funktion folgendermaßen fest: Sei $\hat{x} \in \hat{X}^h$ ein Punkt auf einem Elementrand. Dann gibt es eine endliche maximale Menge von Indizes $\{i_1(\hat{x}) \ldots, i_{G(\hat{x})}(\hat{x})\}$, so daß $\hat{x} \in \omega^h_{i_g(\hat{x})}$ für alle g. Definiere nun den Eingabewert für den Interpolationsoperator $\hat{I^h}$ auf diesem Punkt $\hat{x}$ als den Mittelwert aller angrenzenden lokalen Funktionen der zu interpolierenden Funktion v^h:

$$\hat{u}(\hat{x}) := \frac{1}{G(\hat{x})} \sum_{g=1}^{G(\hat{x})} v^h_{i_g(\hat{x})},$$

dabei ist v^h_i die zum finiten Element i gehörige lokale Funktion zu v^h. Auf diese Weise ist dann der Interpolationsoperator auch für Funktionen $v^h \in \mathrm{V}^h$ wohldefiniert, man beachte, daß lediglich nodale Variablen für Funktionswerte hinzukommen, so daß man sich um die Definition neuer Werte für Normalenableitungen keine Gedanken machen muß.

4.3 Definition. Für jedes $h \in H$ ist der Glättungsoperator E^h definiert als der auf obige Weise auf V^h fortgesetzte Interpolationsoperator $\hat{I^h}$. Dieser wird induziert durch die zum erweiterten Referenzelement aus Lemma 4.2 affin äquivalenten Elemente zur Zerlegung $(\Omega_h)_{h \in H}$.

Nach Konstruktion gilt dann zumindest schon einmal $I^h \circ \hat{I^h} = \mathrm{id}_{\mathrm{V}^h}$, denn $\hat{I^h}(v^h)$ hat auf den alten Verbindungspunkten die gleichen Funktions- und Ableitungswerte wie v^h selber, da v^h in diesen Punkten stetig war. Interpolation liefert dann wieder v^h. Ebenso erbt E^h natürlich die Eigenschaft $\mathrm{bild}(E^h) \subset \mathrm{V}$ von $\hat{I^h}$.

Es bleibt dann nur noch die Abschätzung aus Satz 4.1 zu beweisen. Der lange Beweis soll in dieser Arbeit nicht zitiert werden, er findet sich in Abschnitt 2.6 '*Anti-Crime-Transformation from V^h to V*' von [B], Theorem 2.27.

4.1 Analyse der Randintegrale

Ebenfalls in den Kontext der Glättung paßt eine genauere Untersuchung der Fehler, die beim Übergang zur Operatordarstellung auftreten. Diese waren gegeben durch

$$\rho^h(u^h, v^h) = \sum_{n=1}^{F^h} \int_{\partial\omega^h_n} \sum_{i,j=1}^{d} a_{ij}\, \eta_j\, \partial_i u^h_n \cdot v^h_n$$

und müssen für den Konvergenz- und Stabilitätsbeweis gewisse Abschätzungen erfüllen, die nun bewiesen werden sollen.

Die erste Abschätzung ist recht schwach und für den Fall eines beliebigen $u^h \in \mathrm{V}^h$. Man braucht sie lediglich zum Nachweis der Stabilität, für die Konvergenzqualität wird der nächste Satz herangezogen, der sie dann auf den Fall $u^h = u^h_0$ der exakten Lösung und optimal verteilte Punkte spezialisieren wird.

4.4 Satz. *Die lokalen Funktionenräume und Verbindungspunkte seien derart gewählt, daß die Voraussetzungen für Lemma B.1 erfüllt*[1] *sind. Dann existiert eine Konstante $C_\rho > 0$, so daß für alle $h \in H$ und Funktionen $u^h, v^h \in V^h$ gilt:*

$$\left|\rho^h(u^h, v^h)\right| \;\leq\; h \cdot C_\rho \left\|u^h\right\|_h \left\|v^h\right\|_h .$$

[1] Siehe dazu auch Bemerkung B.2

Beweis. Seien $u^h, v^h \in \mathrm{V}^h$. Dann ist

$$\begin{aligned}
\left|\rho^h(u^h, v^h)\right| &= \left| \sum_{n=1}^{F^h} \int_{\partial\omega_n^h} \sum_{i,j=1}^{d} a_{ij}\, \eta_j\, \partial_i u_n^h \cdot v_n^h \, ds \right| \\
&\leq \max_{i,j=1}^{d} \|a_{ij}\|_{L^\infty(\Omega)} \left| \sum_{n=1}^{F^h} \sum_{i,j=1}^{d} \int_{\partial\omega_n^h} \eta_j\, \partial_i u_n^h \cdot v_n^h \, ds \right| \\
&\leq \max_{i,j=1}^{d} \|a_{ij}\|_{L^\infty(\Omega)} \sum_{e\in\Omega^h} \sum_{i,j=1}^{d} \left| \int_e \eta_j \, \left[\partial_i u^h \cdot v^h\right] ds \right| .
\end{aligned}$$

Dabei läuft die Summe $e \in \Omega^h$ über alle verschiedenen Seiten von finiten Elementen in der Zerlegung Ω^h. In der letzten Zeile meint $(\eta_1, \ldots, \eta_d)$ einen Normaleneinheitsvektor zur Seite e und $\left[\partial_i u^h \cdot v^h\right]$ den Sprung der in eckigen Klammern stehenden Funktion beim Übergang über die Seite in Richtung dieses Vektors. Dabei wird diese außerhalb von Ω als konstant Null angesehen. Der Ausdruck kommt dadurch zustande, daß jede Seite in der vorherigen Summe über alle finiten Elemente genau zweimal auftauchte - für jedes angrenzende Element einmal, wobei die zu den beiden Elementen gehörenden Normalenvektoren verschiedene Vorzeichen in allen Komponenten hatten.

Sei nun e gemeinsame Seite von ω_m^h und ω_n^h. Man kann dann weiter abschätzen:

$$\begin{aligned}
\left| \int_e \left[\partial_i u^h \cdot v^h\right] ds \right| &= \left| \int_e \partial_i u_m^h \cdot v_m^h \, - \, \partial_i u_n^h \cdot v_n^h \, ds \right| \\
&= \left| \int_e \partial_i u_m^h \cdot \left(v_m^h - v_n^h \right) \, + \, v_n^h \cdot \left(\partial_i u_m^h - \partial_i u_n^h \right) ds \right| \\
&\leq \left| \int_e \partial_i u_m^h \cdot \left[v^h\right] ds \right| \, + \, \left| \int_e v_n^h \cdot \left[\partial_i u^h\right] ds \right| .
\end{aligned}$$

Sowohl der Sprung von v^h als auch der von $\partial_i u^h$ hat in allen Verbindungspunkten, die auf e liegen, Nullstellen. Lemma B.1 kann daher in Verbindung mit dem zweiten Teil von Bemerkung B.2 angewendet werden. Zusammen mit dem üblichen Homogenitätsargument, das in Lemma A.3 explizit formuliert ist, liefert es

$$\begin{aligned}
\left| \int_e \partial_i u_m^h \cdot \left[v^h\right] ds \right| &\leq h \cdot C \, \left|u_m^h\right|_{H^1(\omega_m^h)} \cdot \left(\left|v_n^h\right|_{H^1(\omega_n^h)} \, + \, \left|v_m^h\right|_{H^1(\omega_m^h)} \right) \\
\left| \int_e v_n^h \cdot \left[\partial_i u^h\right] ds \right| &\leq h \cdot C \, \left|v_n^h\right|_{H^1(\omega_m^h)} \cdot \left(\left|u_n^h\right|_{H^1(\omega_n^h)} \, + \, \left|u_m^h\right|_{H^1(\omega_m^h)} \right) , \\
\text{also insgesamt } \left| \int_e \left[\partial_i u^h \cdot v^h\right] ds \right| &\leq h \cdot C \sum_{j,k\in\{n,m\}} \left|u_j^h\right|_{H^1(\omega_j^h)} \left|v_k^h\right|_{H^1(\omega_k^h)} .
\end{aligned}$$

Summiert man über alle Seiten, so ist die Anzahl der Male, die jedes finite Element auftaucht, endlich und unabhängig von h, da die Zerlegung nicht ausgeartet ist. Die Behauptung folgt daher mit Lemma A.4. □

Es folgt nun der angekündigte Spezialfall des Satzes für die exakte Lösung u_0^h, der bei einer höheren Zahl von Verbindungspunkten in optimaler Lage und genügend glatter Lösung eine wesentlich stärkere Konvergenzaussage liefert.

4.5 Satz. *Das Referenzelement besitze auf jeder Seite L Verbindungspunkte, die wie die Punkte der Gauss-Legendre-Quadratur verteilt seien. Weiter liege die exakte Lösung u_0^h in $H^{L+2}(\Omega)$. Dann existiert eine Konstante $C'_\rho > 0$, so daß für alle $h \in H$ und Funktionen $v^h \in V^h$ gilt:*

$$\left|\rho^h(u_0^h, v^h)\right| \, \leq \, h^L \cdot C'_\rho \, \left\|u_0^h\right\|_{H^{L+2}(\Omega)} \left\|v^h\right\|_h .$$

Beweis. Wir setzen an wie eben und müssen dann nur noch das Integral

$$\left| \int_e \left[\partial_i u_0^h \cdot v^h\right] \right|$$

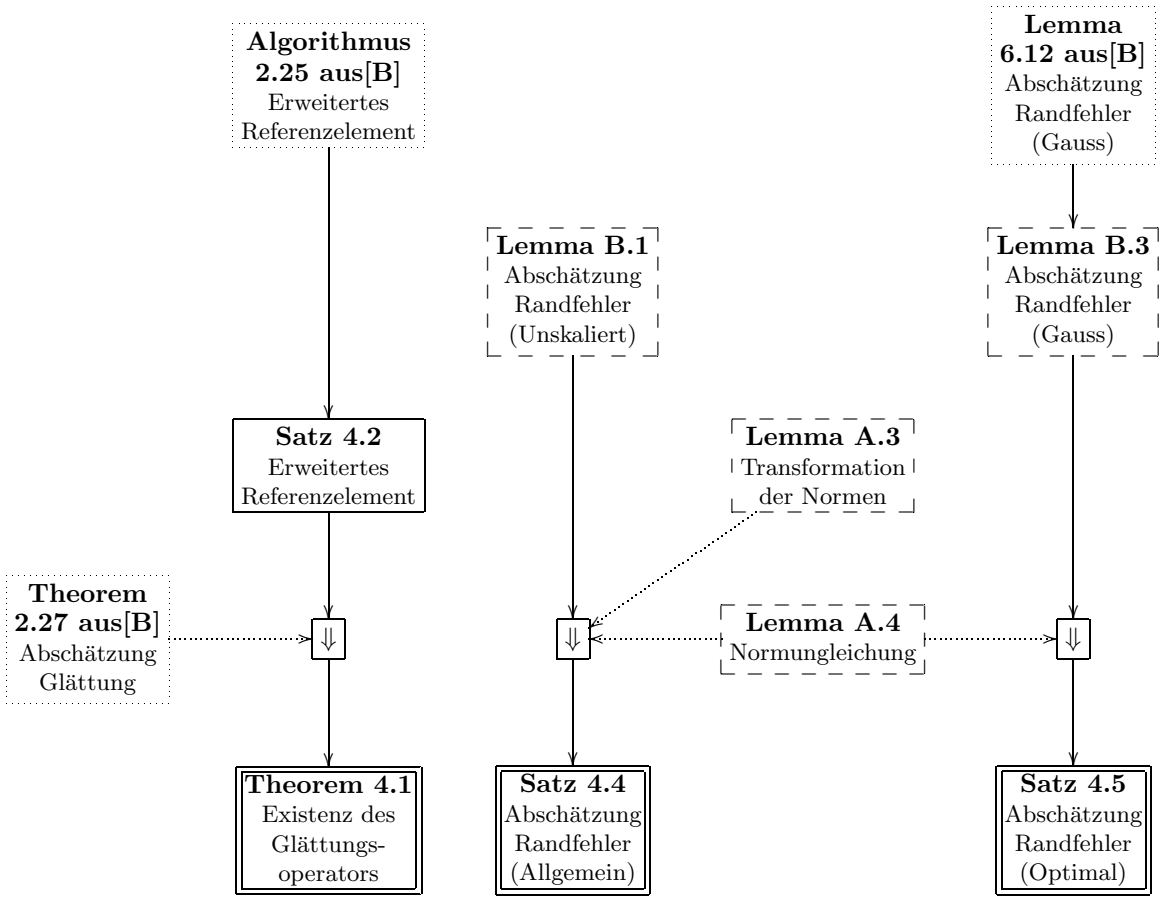
geeignet abschätzen. Da der Sprung von $\partial_i u_0^h$ auf e konstant Null ist, kann man $\partial_i u_0^h$ aus den eckigen Klammern herausziehen und erhält mit Lemma B.3:

$$\begin{aligned}\left|\int_e \partial_i u_0^h \ [v^h]\right| &\leq h^L \cdot C'_\rho \ \left|\partial_i u_0^h\right|_{H^{L+1}(\Omega)} \left(\left\|v^h\right\|_{H^1(\omega_m^h)} + \left\|v^h\right\|_{H^1(\omega_n^h)} \right) \\ &\leq h^L \cdot C'_\rho \ \left\|u_0^h\right\|_{H^{L+2}(\Omega)} \left(\left\|v^h\right\|_{H^1(\omega_m^h)} + \left\|v^h\right\|_{H^1(\omega_n^h)} \right).\end{aligned}$$

Die Behauptung folgt dann nach Summation mit der gleichen Technik wie im letzten Satz unter Zuhilfenahme von Lemma A.4. □

4.2 Beweisstruktur

Zur besseren Orientierung wird jedem Kapitel aus dem theoretischen Teil ein Graph beigefügt, der die Zusammenhänge in der Beweisstruktur zwischen den zugrundeliegenden, bzw. den im Verlauf des Kapitels verwendeten oder bewiesenen Sätzen illustriert.



Kapitel 5

Kollokation

Eine Konvergenz- und Stabilitätsaussage für das Kollokationsverfahren, das Hauptresultat dieser Arbeit, wird hier bewiesen. Dafür wird zunächst ein zum Kollokationsverfahren äquivalentes Variationsverfahren angegeben, und sodann mit Hilfe einer ähnlichen abstrakten Abschätzung wie in Kapitel 1 und der Resultate aus den Folgekapiteln Konvergenz für das Variationsproblem gezeigt. Stabilität für das Gleichungssystem der Kollokation folgt aus der Existenz des beschränkten Interpolationsoperators in den Kollokationsstellen. Mit Hilfe der spezialisierten Abschätzung für die Randfehler erzielt man im Falle einer Gauss-Legendre-Verteilung der Verbindungspunkte eine hohe Ordnung der Konvergenz.

5.0 Formulierung als Variationsaufgabe

Bisher wurde bewiesen, daß bei einer zulässigen Familie von Diskretisierungen die Lösungen u_0^h der diskretisierten Versionen des Variationsproblems gegen die exakte Lösung u_0 des Variationsproblems konvergieren, wobei der Fehler höchstens proportional zu h ist. Dabei waren u_0^h und u_0 wie folgt charakterisiert:

$$\begin{aligned} a(u_0, v) &= f(v) \quad \text{für alle } v \in \mathrm{V} \\ \text{und } a^h(u_0^h, v^h) &= f^h(v^h) \quad \text{für alle } v^h \in \mathrm{V}^h. \end{aligned}$$

Die nächste Problematik, die nun konsequenterweise angegangen werden muß, ist die Frage, wie sich das im Lösungsalgorithmus verwendete Kollokationsverfahren in diesen Rahmen einfügt. Es wäre natürlich höchst optimistisch zu glauben, daß es für jedes h bereits die Lösung u_0^h liefert, tatsächlich ist das auch leider nicht der Fall. Wir bezeichnen im folgenden mit $\tilde{u}_0^h$ die *Lösung des Kollokationsproblems*: Das eindeutig bestimmte

$$\tilde{u}_0^h \in \mathrm{V}^h \text{ mit } A\tilde{u}_0^h(z) = F(z) \text{ für alle } z \in Z^h.$$

Dabei ist A der zur Bilinearform a gehörende Differentialoperator und $F \in \mathcal{C}(\Omega)$ die Funktion, die die Linearform f erzeugt. Man beachte, daß für F Funktionswerte im Inneren von Ω definiert sein müssen, daher ist es nicht mehr möglich, z.B. beliebiges $F \in L^\infty(\Omega)$ zuzulassen.

Die entscheidende Frage lautet nun: Konvergiert auch $\tilde{u}_0^h$ gegen u_0, und wie gut? Alternativ würde es wegen der bisherigen Ergebnisse auch ausreichen, daß der Abstand von $\tilde{u}_0^h$ zu u_0^h beliebig klein wird, die Dreiecksungleichung liefert dann den Rest. Dafür soll zunächst das Kollokationsproblem in ein Variationsproblem mit modifizierter Bilinearform überführt werden.

Man wähle dafür für jedes $h \in H$ eine Basis $\{\psi_z\}_{z \in Z^h}$ von V^h mit der Eigenschaft

$$\psi_z(\tilde{z}) \;=\; \delta_z^{\tilde{z}} \text{ für alle } z, \tilde{z} \in Z^h,$$

dies ist insbesondere dann möglich, falls die Bedingungen für die Interpolation auf Kollokationsstellen in Theorem 3.14 erfüllt sind.

Jede Funktion $v^h \in \mathrm{V}^h$ läßt sich dann in diese Basis entwickeln gemäß

$$v^h = \sum_{z \in Z^h} v^h(z)\psi_z.$$

5.1 *Bemerkung.* Der aufmerksame Leser wird sich fragen, *was* da eigentlich interpoliert wird. Zu seiner Beruhigung sei an Lemma A.5 erinnert: Es gibt eine Konstante C, so daß für beliebig vorgegebenes $x \in \mathbb{R}^D$ und $\epsilon > 0$ eine glatte Funktionen $\theta_{x,\epsilon} : \mathbb{R}^D \to [0,1]$ existiert mit

(*i*) $\theta_{x,\epsilon}(x) = 1$,

(*ii*) $\operatorname{supp}\theta_{x,\epsilon} \subset B_\epsilon(x)$ und

(*iii*) $\|\theta_{x,\epsilon}\|_{H^1(\mathbb{R}^D)} \leq C$.

Man kann nun $x = z$ und $\epsilon > 0$ so wählen, daß

$$\theta_{z,\epsilon}(\tilde{z}) \;=\; \delta_z^{\tilde{z}} \text{ für alle } z, \tilde{z} \in Z^h,$$

dies ist möglich, weil die Kollokationsstellen diskret in Ω liegen. Interpoliert man dann $\theta_{z,\epsilon}$ mit dem Interpolationsoperator auf Kollokationsstellen, so hat man die Funktion ψ_z mit den gewünschten Eigenschaften. Da die Interpolation nach Theorem 3.14 durch gleichmäßig beschränkte Operatoren gegeben ist, so sind wegen (*iii*) die Funktionen ψ_z stets beschränkt in der Norm $\|\diamond\|_h$. Der folgende Übergang (5.1) von der schwachen Formulierung zu den Kollokationsgleichungen führt wegen der in Kürze bewiesenen Stabilität und der gleichmäßig beschränkten Basistransformation dann stets auf gleichmäßig gut konditionierte Gleichungssysteme. ◇

Für den Übergang zur Kollokation wähle man nun weiter eine Quadraturformel

$$Q^h(v^h) := \sum_{i=1}^{F^h} Q_i^h(v^h) := \sum_{i=1}^{F^h} \sum_{z \in Z_i^h} v^h(z) q_z^h$$

für Funktionen $v^h \in \mathrm{V}^h$, deren Koeffizienten q_z^h seien möglichst so gewählt, daß der Genauigkeitsgrad der Formel maximiert wird. Für den Konvergenzbeweis wird ausreichen, daß sich der Quadraturfehler für Produkte von Funktionen global abschätzen läßt gemäß

$$\left| \sum_{i=1}^{F^h} Q_i^h(v^h w^h) - \int_{\omega_i^h} v^h w^h \right| \;\leq\; h \cdot C_\mathrm{q} \left\|v^h\right\|_h \left\|w^h\right\|_h \text{ für alle } v^h, w^h \in \mathrm{V}^h$$

mit einer von h unabhängigen Konstanten C_q. Folgerung B.5 aus Lemma B.4 garantiert, daß diese Abschätzung stets gültig ist.

Über die Quadratur wird dann eine Näherung $\tilde{a}^h$ für die Bilinearform a^h und eine Näherung $\tilde{f}^h$ für die Linearform f^h konstruiert. Es gilt nach 2.7

$$a^h(u_0^h, v^h) = \rho^h(u_0^h, v^h) + \sum_{i=1}^{F^h} \int_{\omega_i^h} A u_0^h \cdot v^h$$
$$\text{und } f^h(v^h) = \sum_{i=1}^{F^h} \int_{\omega_i^h} F \cdot v^h,$$

Dabei ist ρ^h der in Abschnitt 2.3 eingeführte Korrekturterm, der beim Übergang zur Operatorform der Gleichung im Raum V^h leider auftritt: Wegen der Unstetigkeit auf den Rändern der ω_i^h heben sich die Randterme, die bei der partiellen Integration ins Spiel kommen, im allgemeinen nicht weg. Eine genauere Untersuchung wurde im letzten Kapitel durchgeführt und ergab in Satz 4.4 unter den Voraussetzungen von Lemma B.1, daß mit einer Konstanten $C_\rho > 0$ gilt:

$$\rho^h(u^h, v^h) \leq h\, C_\rho \left\|u^h\right\|_h \left\|v^h\right\|_h \text{ für alle } u^h, v^h \in \mathrm{V}^h.$$

Die Voraussetzungen dafür wollen wir von nun an als erfüllt annehmen. Es liegen nun die folgenden Definitionen nahe:

$$\tilde{a}^h(u_0^h, v^h) := Q^h(A u_0^h \cdot v^h)$$
$$\text{und } \tilde{f}^h(v^h) := Q^h(F \cdot v^h).$$

Das Kollokationsproblem ist dann äquivalent zum Variationsproblem bezüglich dieser Bilinearform.

Dies sieht man folgendermaßen ein:

$$
\begin{aligned}
& \tilde{a}^h(\tilde{u}_0^h, v^h) = \tilde{f}^h(v^h) \text{ für alle } v^h \in \mathrm{V}^h \\
\Leftrightarrow\ & \tilde{a}^h(\tilde{u}_0^h, \psi_z^h) = \tilde{f}^h(\psi_z^h) \text{ für alle } z \in Z^h && \text{(da } \{\psi_z^h\}_{z\in Z^h} \text{ Basis von } \mathrm{V}^h) \\
\Leftrightarrow\ & Q^h(A\tilde{u}_0^h \cdot \psi_z^h) = Q^h(F \cdot \psi_z^h) \text{ für alle } z \in Z^h && \text{(Definition } \tilde{a}^h, \tilde{f}^h) \\
\Leftrightarrow\ & \sum_{\tilde{z}\in Z^h} A\tilde{u}_0^h(\tilde{z})\psi_z^h(\tilde{z})q_{\tilde{z}}^h = \sum_{\tilde{z}\in Z^h} F(\tilde{z})\psi_z^h(\tilde{z})q_{\tilde{z}}^h \text{ für alle } z \in Z^h && \text{(Definition } Q^h) \qquad (5.1) \\
\Leftrightarrow\ & \sum_{\tilde{z}\in Z^h} A\tilde{u}_0^h(\tilde{z})\delta_z^{\tilde{z}}q_{\tilde{z}}^h = \sum_{\tilde{z}\in Z^h} F(\tilde{z})\delta_z^{\tilde{z}}q_{\tilde{z}}^h \text{ für alle } z \in Z^h && \text{(Definition } \psi_z^h) \\
\Leftrightarrow\ & A\tilde{u}_0^h(z) = F(z) \text{ für alle } z \in Z^h. && \text{(Definition } \delta_z^{\tilde{z}})
\end{aligned}
$$

5.1 Stabilität und Konvergenz

Eben wurde verifiziert, daß die Näherungslösung $\tilde{u}_0^h$ der Kollokation charakterisiert wird durch

$$\tilde{a}^h(\tilde{u}_0^h, v^h) = \tilde{f}^h(v^h) \text{ für alle } v^h \in \mathrm{V}^h,$$

und wir können nun zeigen, daß in Lemma 1.10 ein Korrekturterm einfließt, der den Einfluß der Modifikation der Bilinearform auf die Lösung mißt. Entscheidend für die Gültigkeit eines abstrakten Konvergenzkriteriums ist wiederum die Stabilität, dieses mal für die Bilinearformen $\tilde{a}^h$.

5.2 Definition. Die Kollokation heißt $\tilde{\epsilon}$-stabil mit einer Konstanten $\tilde{\epsilon} > 0$, falls für alle $h \in H$:

$$\tilde{\omega}^h := \inf_{u^h\in\mathbb{E}\mathrm{V}^h} \sup_{v^h\in\mathbb{E}\mathrm{V}^h} \left|\tilde{a}^h(u^h, v^h)\right| \geq \tilde{\epsilon}.$$

5.3 Satz. *Die Kollokation sei $\tilde{\epsilon}$-stabil, und alle Bilinearformen $\tilde{a}^h$ seien stetig mit gemeinsamer Stetigkeitskonstante $\tilde{\alpha}$. Dann gilt folgende Fehlerabschätzung:*

$$
\begin{aligned}
\left\|\tilde{u}_0^h - u_0\right\|_h \leq \frac{1}{\tilde{\epsilon}} & \sup_{v^h\in\mathbb{E}V^h} \left|\tilde{a}^h(u_0, v^h) - a^h(u_0, v^h)\right| \\
& + \frac{1}{\tilde{\epsilon}} \sup_{v^h\in\mathbb{E}V^h} \left|\tilde{f}^h(v^h) - f^h(v^h)\right| \\
& + \left(1 + \frac{\tilde{\alpha}}{\tilde{\epsilon}}\right) \inf_{u^h\in\mathbb{E}V^h} \left\|u_0 - u^h\right\|_h .
\end{aligned}
$$

Beweis. Nach Voraussetzung gilt $\tilde{\omega}^h \geq \tilde{\epsilon}$. Damit folgt für beliebiges $u^h \in \mathrm{V}^h$:

$$
\begin{aligned}
\tilde{\epsilon}\left\|u^h - \tilde{u}_0^h\right\|_h &\leq \sup_{v^h\in\mathbb{E}\mathrm{V}^h} \left|\tilde{a}^h(u^h - \tilde{u}_0^h, v^h)\right| && \text{(Stabilität)} \\
&\leq \sup_{v^h\in\mathbb{E}\mathrm{V}^h} \left|\tilde{a}^h(u^h - u_0 + u_0 - \tilde{u}_0^h, v^h)\right| && \text{(Fundamentaltrick)} \\
&\leq \sup_{v^h\in\mathbb{E}\mathrm{V}^h} \left|\tilde{a}^h(u_0, v^h) - \tilde{f}^h(v^h)\right| \\
&\qquad + \sup_{v^h\in\mathbb{E}\mathrm{V}^h} \left|\tilde{a}^h(u^h - u_0)\right| && \text{(Dreiecksungl., Definition } \tilde{u}_0^h) \\
&\leq \sup_{v^h\in\mathbb{E}\mathrm{V}^h} \left|\tilde{a}^h(u_0, v^h) - a^h(u_0, v^h) + f^h(v^h) - \tilde{f}^h(v^h)\right| \\
&\qquad + \sup_{v^h\in\mathbb{E}\mathrm{V}^h} \left|\tilde{a}^h(u^h - u_0)\right| && \text{(Definition } u_0) \\
&\leq \sup_{v^h\in\mathbb{E}\mathrm{V}^h} \left|a^h(u_0, v^h) - \tilde{a}^h(u_0, v^h)\right| \\
&\qquad + \sup_{v^h\in\mathbb{E}\mathrm{V}^h} \left|\tilde{f}^h(v^h) - f^h(v^h)\right| + \tilde{\alpha}\left\|u^h - u_0\right\|_h . && \text{(Dreiecksungl., Stetigkeit)}
\end{aligned}
$$

Zusammen mit der Dreiecksungleichung liefert diese Abschätzung

$$
\begin{aligned}
\tilde{\epsilon}\left\|u_0 - \tilde{u}_0^h\right\|_h &\leq \tilde{\epsilon}\left\|u_0 - u^h\right\|_h + \tilde{\epsilon}\left\|u^h - \tilde{u}_0^h\right\|_h \\
&\leq \sup_{v^h\in\mathbb{E}\mathrm{V}^h} \left|a^h(u_0, v^h) - \tilde{a}^h(u_0, v^h)\right| \\
&\qquad + \sup_{v^h\in\mathbb{E}\mathrm{V}^h} \left|\tilde{f}^h(v^h) - f^h(v^h)\right| + (\tilde{\epsilon} + \tilde{\alpha})\left\|u^h - u_0\right\|_h .
\end{aligned}
$$

u^h war aber beliebig gewählt, also steht die Behauptung nach Bildung des Infimums über alle $u^h \in \mathbb{E}\mathrm{V}^h$ und Division durch $\tilde{\epsilon}$ da. □

Es bleibt noch die Aufgabe, geeignete Kriterien für die $\tilde{\epsilon}$-Stabilität bereitzustellen und die einzelnen Terme abzuschätzen. Für den von u_0 unabhängigen Teil gewinnt man eine Aussage als leichte Konsequenz aus der Genauigkeitsforderung an die Quadratur:

5.4 Lemma. *Mit der geforderten Genauigkeit der Quadraturformel gilt:*

$$\sup_{v^h \in \mathbb{E}V^h} \left| \tilde{f}^h(v^h) - f^h(v^h) \right| \leq \; h \cdot C_q \, \phi$$

Beweis. Einsetzen der Definitionen ergibt für beliebiges $v^h \in \mathrm{V}^h$:

$$\begin{aligned} \left| \tilde{f}^h(v^h) - f^h(v^h) \right| &= \left| \sum_{i=1}^{F^h} \int_{\omega_i^h} F \cdot v^h \; - \; Q_i^h(F \cdot v^h) \right| \\ &\leq \; h \cdot C_{\mathrm{q}} \, \|F\|_h \cdot \left\|v^h\right\|_h \qquad \text{(Quadraturabschätzung)} \end{aligned}$$

□

Für den zweiten Korrekturterm ist der Weg etwas komplizierter, da man noch die Fehler bei den Randintegralen ins Spiel bringen muß. Man findet jedoch eine ähnliche Abschätzung, die auch bei der Überprüfung der $\tilde{\epsilon}$-Stabilität noch entscheidend einfließen wird:

5.5 Lemma. *Für beliebiges $u^h \in V^h$ gilt:*

$$\sup_{v^h \in \mathbb{E}V^h} \left| \tilde{a}^h(u^h, v^h) - a^h(u^h, v^h) \right| \leq \; h \, (C_\rho + C_q \, C_\rho + C_q \, \alpha) \left\|u^h\right\|_h .$$

Außerdem sind alle $\tilde{a}^h$ stetig mit einer gemeinsamen Stetigkeitskonstanten $\tilde{\alpha}$.

Beweis. Man hat zunächst wegen der Dreiecksungleichung, der Genauigkeitsforderung an die Quadraturformel und Satz 4.4 für festes $v^h \in \mathrm{V}^h$:

$$\begin{aligned} \left| \tilde{a}^h(u^h, v^h) - a^h(u^h, v^h) \right| \leq \; & \left| \rho^h(u^h, v^h) \right| \; + \; \left| \sum_{i=1}^{F^h} Q_i^h(Au^h \cdot v^h) - \int_{\omega_i^h} Au^h \cdot v^h \right| \\ & \leq \; h \cdot \left(C_\rho \, \left\|u^h\right\|_h \left\|v^h\right\|_h \; + \; C_{\mathrm{q}} \, \left\|Au^h\right\|_h \left\|v^h\right\|_h \right) \\ & = \; h \cdot \left(C_\rho \, \left\|u^h\right\|_h \; + \; C_{\mathrm{q}} \, \left\|Au^h\right\|_h \right) \end{aligned}$$

Weiter gilt nach dem Satz von Riesz[1]

$$\begin{aligned} \left\|Au^h\right\|_h &= \sup_{v^h \in \mathbb{E}\mathrm{V}^h} \left| (Au^h, v^h)_2 \right| \\ &= \sup_{v^h \in \mathbb{E}\mathrm{V}^h} \left| \sum_{i=1}^{F^h} \int_{\omega_i^h} Au^h \cdot v^h \right| \\ &= \sup_{v^h \in \mathbb{E}\mathrm{V}^h} \left| \rho^h(u^h, v^h) \; + \; a^h(u^h, v^h) \right| \qquad \text{(Definition } \rho^h\text{)} \\ &\leq (C_\rho + \alpha) \, \left\|u^h\right\|_h \left\|v^h\right\|_h . \qquad \text{(Stetigkeit, Definition } C_\rho\text{)} \end{aligned}$$

Die beiden Abschätzungen zusammen ergeben für den Fall $v^h \in \mathbb{E}\mathrm{V}^h$

$$\left| \tilde{a}^h(u^h, v^h) - a^h(u^h, v^h) \right| \leq \; h \, (C_\rho + C_{\mathrm{q}} \, C_\rho + C_{\mathrm{q}} \, \alpha) \left\|u^h\right\|_h ,$$

[1] Genauer ist diese Formel richtig, weil sowohl $H^m(\Omega) \sqsubset L^2(\Omega) \sqsubset H^{-m}(\Omega)$, als auch $(\mathrm{V}^h, \|\diamond\|_h) \sqsubset (\mathrm{V}^h, \|\diamond\|_{h,L^2}) \sqsubset (\mathrm{V}^{h\prime}, \|\diamond\|_h{}')$ sogenannte *Gelfand-Dreier* bilden, und man daher anstelle der Paarung $\langle \diamond, \diamond \rangle_{\mathrm{V}^{h\prime} \times \mathrm{V}^h}$, bzw. $\langle \diamond, \diamond \rangle_{H^{-m}(\Omega) \times H^m(\Omega)}$ auch das L^2-Skalarprodukt $(\diamond, \diamond)_2$ verwenden kann. Man lese dies und die Ableitung der Formel z.B. in [H], Abschnitt 6.3.3 nach.

damit folgt die erste Behauptung. Mit der Dreiecksungleichung hat man außerdem

$$\begin{aligned}\left|\tilde{a}^h(u^h,v^h)\right| &\le \left|a^h(u^h,v^h)\right| + h\,(C_\rho + C_{\mathrm{q}}\,C_\rho + C_{\mathrm{q}}\,\alpha)\left\|u^h\right\|_h\left\|v^h\right\|_h \\ &\le (\alpha + C_\rho + C_{\mathrm{q}}\,C_\rho + C_{\mathrm{q}}\,\alpha)\left\|u^h\right\|_h\left\|v^h\right\|_h \\ &=: \tilde{\alpha}\left\|u^h\right\|_h\left\|v^h\right\|_h,\end{aligned}$$

also die Gleichstetigkeit der $\tilde{a}^h$. □

Schließlich ist die Kollokation bei ϵ-stabiler Diskretisierung automatisch $\tilde{\epsilon}$-stabil für geeignetes $\tilde{\epsilon}$ und genügend kleine h.

5.6 Lemma. *Die Diskretisierung sei ϵ-stabil. Dann existiert ein $\tilde{h} \in (0,1]$, so daß für alle $h \le \tilde{h}$:*

$$\tilde{\omega}^h \ge \tilde{\epsilon},$$

wobei ein beliebiges $0 < \tilde{\epsilon} < \epsilon$ fest gewählt werden kann. Mit anderen Worten ist die Kollokation für hinreichend kleine h $\tilde{\epsilon}$-stabil.

Beweis. Sei $u^h \in \mathbb{E}\mathrm{V}^h$. Dann gilt:

$$\begin{aligned}&\sup_{v^h\in\mathbb{E}\mathrm{V}^h}\left|\tilde{a}^h(u^h,v^h)\right| \\ &\ge \sup_{v^h\in\mathbb{E}\mathrm{V}^h}\Big(\left|a^h(u^h,v^h)\right| - \left|\tilde{a}^h(u^h,v^h) - a^h(u^h,v^h)\right|\Big) && \text{(Dreiecksungleichung)} \\ &\ge \sup_{v^h\in\mathbb{E}\mathrm{V}^h}\Big(\left|a^h(u^h,v^h)\right| - C\cdot h\left\|u^h\right\|_h\Big) && \text{(Lemma 5.5)} \\ &\ge \epsilon - C\cdot h\left\|u^h\right\|_h && (\epsilon\text{-Stabilität}) \\ &= \epsilon - C\cdot h.\end{aligned}$$

Damit ist aber auch das Infimum über alle $u^h \in \mathbb{E}\mathrm{V}^h$ größer oder gleich der rechten Seite, womit wegen $C\cdot h \to 0$ die Behauptung folgt. □

Man hat damit das Kollokationsverfahren endlich im Griff, insbesondere ist unter allen bisher gemachten Voraussetzungen die Konvergenz gesichert. Baut man nämlich die Aussagen aus diesem Kapitel zusammen, so erhält man das folgende Theorem, das allerdings noch etwas schwächlich ist, weil es zunächst nur lineare Konvergenz sichert. Dies tut es allerdings in relativ großer Allgemeinheit, und es wird sogleich für den Fall speziellerer Verteilungen der Verbindungspunkte verbessert werden.

5.7 Theorem. *Sei*

$$a(u,v) = \int_\Omega a_{ij}\,\partial_i u\cdot\partial_j v \;+\; b_k\,\partial_k u\cdot v \;+\; c_0\,u\cdot v$$

die Bilinearform eines elliptischen Randwertproblems und $\{(\omega_i^h)_{1\le i\le F^h}\}_{h\in H}$ eine zulässige Familie von Zerlegungen von Ω in finite Elemente. Alle weiteren Bedingungen für die Anwendung von Theorem 2.6, Satz 2.4 über die Regularität von Lösungen und Satz 4.4 für die Abschätzung der Randfehler seien erfüllt. Dann gibt es ein $\tilde{h} > 0$ und ein $\tilde{\epsilon} > 0$ und eine von h unabhängige Konstante $C > 0$, so daß für $0 < h \le \tilde{h}$:

(*i*) *Die durch a und die finiten Elemente induzierte Kollokationslösung $\tilde{u}_0^h$ existiert und erfüllt $\left\|\tilde{u}_0^h\right\|_h \le \phi/\tilde{\epsilon}$.*

(*ii*) *Der Fehler der Diskretisierung läßt sich abschätzen durch*

$$\left\|\tilde{u}_0^h - u_0\right\|_h \;\le\; C\,h.$$

Beweis. Die Aussage (*i*) folgt wegen der in Lemma 5.6 soeben bewiesenen Stabilität aus Satz 1.6. Für die zweite Aussage kombiniere man Satz 5.3 mit den Abschätzungen 5.4 und 5.5 und der Aussage 1.9 über den Interpolationsfehler. □

Eine wesentlich verbesserte Konvergenzaussage und das Hauptresultat dieser Arbeit erhält man, wenn die Verbindungspunkte optimal für die Quadratur als Gauss-Legendre-Punkte gewählt sind.

5.8 Theorem. *Sei*

$$a(u,v) = \int_\Omega a_{ij}\,\partial_i u \cdot \partial_j v \;+\; b_k\,\partial_k u \cdot v \;+\; c_0\,u \cdot v$$

die Bilinearform eines elliptischen Randwertproblems und $\{(\omega_i^h)_{1\le i\le F^h}\}_{h\in H}$ *eine zulässige Familie von Zerlegungen von* Ω *in finite Elemente. Alle weiteren Bedingungen für die Anwendung von Theorem 2.6 und Satz 2.4 über die Regularität von Lösungen seien erfüllt. Das Referenzelement besitze* L *Verbindungspunkte auf jeder Seite, die wie die Punkte der Gauss-Legendre-Quadratur verteilt sind. Weiter sollen globale Quadraturformeln* Q^h *für die Kollokationsstellen existieren, welche für ein* $Q \ge 1$ *die Fehlerabschätzungen*

$$\left|\int_\Omega u^h \cdot v^h \;-\; Q^h(u^h \cdot v^h)\right| \;\le\; C_q\,h^Q \cdot \left\|u^h\right\|_h \left\|v^h\right\|_h$$

mit einer von h *unabhängigen Konstanten* C_q *für alle* $v^h \in V^h$ *und* $u^h \in \{Au_0^h, F\}$ *erfüllen. Dann gibt es ein* $\tilde h > 0$ *und ein* $\tilde\epsilon > 0$ *und eine von* h *unabhängige Konstante* $C > 0$*, so daß für* $0 < h \le \tilde h$*:*

(*i*) *Die durch* a *und die finiten Elemente induzierte Kollokationslösung* $\tilde u_0^h$ *existiert und erfüllt* $\left\|\tilde u_0^h\right\|_h \le \phi/\tilde\epsilon$*.*

(*ii*) *Der Fehler der Diskretisierung läßt sich abschätzen durch*

$$\left\|\tilde u_0^h - u_0\right\|_h \;\le\; C\,h^{\min\{L,R-1,Q\}},$$

falls die exakte Lösung u_0 *in* $H^R(\Omega)$ *liegt und alle Polynome von Grad* $R-1$ *im Raum der lokalen Funktionen enthalten sind.*

Beweis. Wegen der stärkeren Voraussetzungen kann man zunächst 5.4 verbessern zu

$$\sup_{v^h\in\mathbb{E}\mathrm{V}^h} \left|\tilde f^h(v^h) - f^h(v^h)\right| \le\; h^Q \cdot C_{\mathrm{q}}\,\phi.$$

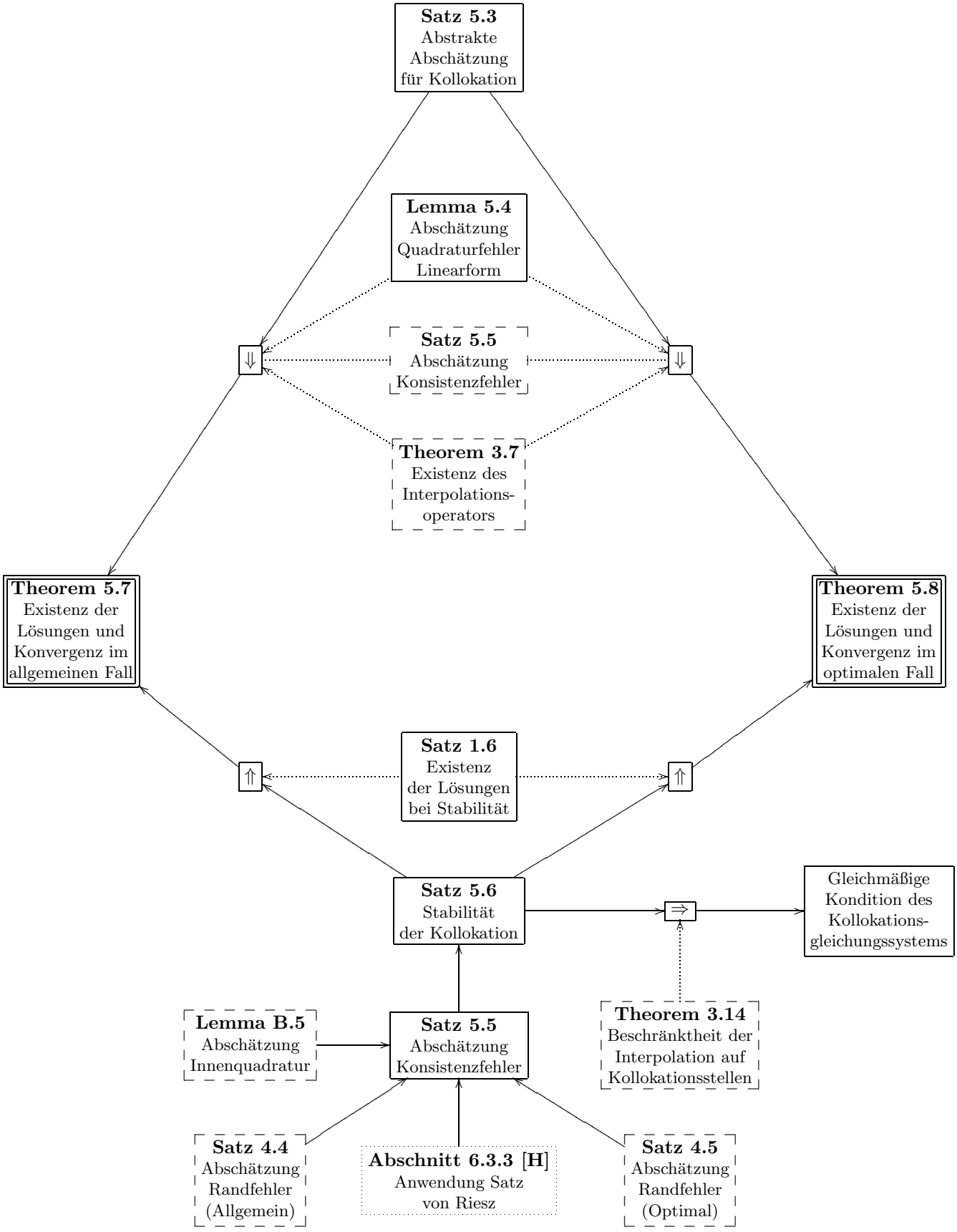
Mit der Randfehlerabschätzung 4.5 für die exakte Lösung erhält man außerdem eine spezialisierte Fassung von 5.5:

$$\sup_{v^h\in\mathbb{E}\mathrm{V}^h} \left|\tilde a^h(u_0^h, v^h) - a^h(u_0^h, v^h)\right| \le\; h^{\min\{Q,L\}}\,(C'_\rho + C_{\mathrm{q}}\,C'_\rho + C_{\mathrm{q}}\,\alpha)\left\|u_0^h\right\|_h.$$

Durch Kombination dieser Aussagen mit Lemma 1.9 liefert die abstrakte Abschätzung 5.3 nun die Behauptung. □

5.2 Beweisstruktur

Zur besseren Orientierung wird jedem Kapitel aus dem theoretischen Teil ein Graph beigefügt, der die Zusammenhänge in der Beweisstruktur zwischen den zugrundeliegenden, bzw. den im Verlauf des Kapitels verwendeten oder bewiesenen Sätzen illustriert.



Kapitel 6

Zusammenfassung

In diesem Kapitel sollen die über die bisherigen theoretischen Untersuchungen verstreuten Voraussetzungen und Resultate noch einmal im Sinne eines Überblicks wiederholt werden.

6.0 Forderungen an die Geometrie

Ziel der Bestrebungen war es, ein Randwertproblem der Gestalt

$$Au = F \text{ auf } \Omega \text{ und } u\big|_{\partial\Omega} = B$$

für ein beschränktes Gebiet Ω, einen vorgegebenen partiellen Differentialoperator A, eine Funktion $F \in L^\infty(\Omega)$ und $B \in L^2(\partial\Omega)$ zu lösen. Damit die hier vorgestellte Diskretisierungsmethode durchführbar ist, muß dieses Gebiet notwendigerweise polygonal sein, damit eine Familie von Zerlegungen $\left\{\left(\omega_i^h \;:\; 1 \le i \le F^h\right)\right\}_{h\in H}$ von Ω in finite Elemente existieren kann. Diese Zerlegungen müssen für alle $h \in H$ die folgenden Eigenschaften aufweisen:

- Ω ist disjunkte Vereinigung über die finiten Elemente:

$$\bigcup_{i=1}^{F^h} \overline{\omega_i^h} = \overline{\Omega} \text{ und } \omega_i^h \cap \omega_j^h = \emptyset \text{ für } i \neq j.$$

- Die finiten Elemente sind Polygone, und für $i \neq j$ ist der Durchschnitt $\sigma_{ij}^h := \overline{\omega_i^h} \cap \overline{\omega_j^h}$ entweder leer oder eine gemeinsame Seite.
- Die Zerlegung ist *nicht ausgeartet*.
- Die Zerlegungen werden schnell genug feiner, wenn h kleiner wird, in dem exakten Sinne daß

$$\max\{\operatorname{diam}\omega_i^h \;:\; 1 \le i \le F^h\} \le h \operatorname{diam}\Omega.$$

Sämtliche finiten Elemente sollen durch affine Äquivalenz von einem Referenzelement $(\omega^{\mathrm{ref}}, Z^{\mathrm{ref}}, \mathcal{P}^{\mathrm{ref}}, \mathcal{N}^{\mathrm{ref}})$ induziert werden. Dabei erfülle die affine Transformation ϕ_i^h, welche das Referenzelement auf ω_i^h abbildet, die Bedingung $\det \phi_i^h \le h^D \cdot \operatorname{diam}\Omega$. An die Geometrie des Referenzelements wird ebenfalls eine Reihe von Bedingungen gestellt:

- $\mathbb{P}_{R-1} \subset \mathcal{P}^{\mathrm{ref}} \subset \mathbb{P}[D]$, wobei $\mathbb{P}_{R-1}$ der Raum aller Polynome in D Variablen von Grad höchstens gleich $R-1$ ist und $R > D/2+1$ eine feste natürliche Zahl.
- $\mathcal{N}^{\mathrm{ref}} \sqsubset \mathcal{C}^1(\overline{\omega}^{\mathrm{ref}})'$ ist Basis des Dualraums von $\mathcal{P}^{\mathrm{ref}}$. Dabei bestehen die nodalen Variablen aus $\mathcal{N}^{\mathrm{ref}}$ in der Auswertung der lokalen Funktion auf einer Menge von Punkten $X^{\mathrm{ref}} \subset \partial\omega^{\mathrm{ref}}$, sowie der Auswertung der Richtungsableitungen normal zum Rand in einer Menge von Punkten $Y^{\mathrm{ref}} \subset \partial\omega^{\mathrm{ref}}$.
- Es gilt $M := \#X^{\mathrm{ref}} = \#Y^{\mathrm{ref}}$ und $N := \#Z^{\mathrm{ref}} \ge M$. Außerdem sei die Dimension K von $\mathcal{P}^{\mathrm{ref}}$ gleich $M+N$.
- Auf jeder Seite des Referenzelements liegt mindestens ein Punkt aus X^{ref} und ein Punkt aus Y^{ref}. Auf jeder Seite sollen gleich viele Verbindungspunkte von jedem der beiden Typen liegen.

- Für ein finites Element ω_i^h werden durch affine Äquivalenz die Verbindungspunkte $X_i^h := \phi_i^h X^{\mathrm{ref}}$ und $Y_i^h := \phi_i^h Y^{\mathrm{ref}}$ induziert. Ist nun für zwei verschiedene finite Elemente ω_i^h und ω_j^h die gemeinsame Seite $\sigma_{ij}^h \neq \emptyset$, so müssen die Verbindungspunkte der beiden Elemente auf der gemeinsamen Seite übereinstimmen:
$$X_i^h \cap \sigma_{ij}^h = X_j^h \cap \sigma_{ij}^h \text{ und } Y_i^h \cap \sigma_{ij}^h = Y_j^h \cap \sigma_{ij}^h.$$

Unter diesen Voraussetzungen induziert die Menge der Zerlegungen einen wohldefinierten, endlichdimensionalen Raum $\mathrm{V}^h \sqsubset L^2(\Omega)$ von Funktionen, der durch folgende Eigenschaften charakterisiert ist:

- Die Restriktionen der Funktionen aus V^h auf finite Elemente ω_i^h liegen in $\varphi_i^{h*}\mathcal{P}^{\mathrm{ref}}$, d.h. sie sind dort glatt. Auf den Rändern können im allgemeinen Unstetigkeiten auftreten.
- Nodale Variablen, die zu gemeinsamen Verbindungspunkten von finiten Elementen gehören, liefern für Funktionen aus V^h den gleichen Wert.
- Nodale Variablen, die zu Funktionsauswertungen in Punkten aus $\partial\Omega$ gehören, liefern den Funktionswert von B in diesem Punkt. Die nodalen Variablen für Auswertung von Richtungsableitungen auf $\partial\Omega$ sind hingegen frei.

Es ist dann gewährleistet, daß die Interpolation und Glättung wohldefiniert ist und die für Konvergenz erforderlichen Abschätzungen erfüllt. Dies waren die Theoreme 3.7 und 4.1, deren Aussagen wie folgt lauteten:

Es existieren von h unabhängige Konstanten $C_{ip} > 0$ und $C_{ac} > 0$, sowie Operatoren

$$I^h : V \cap H^R(\Omega) \to V^h \text{ und } E^h : V^h \to V,$$

so daß gilt:

(i) Die Interpolation erfüllt für alle $u \in H^R(\Omega)$ die Abschätzung

$$\left\|u - I^h u\right\|_h \;\leq\; h^{R-1} \cdot C_{ip}\, |u|_{H^R(\Omega)}\,.$$

(ii) Die Glättung erfüllt für alle $u^h \in V^h$ die Abschätzung

$$\left\|u^h - E^h u^h\right\|_h \;\leq\; h^{1/2} \cdot C_{ac}\, \left\|u^h\right\|_h\,.$$

Dabei ist $V \sqsubset H^1(\Omega)$ der Raum von Funktionen, auf dem die Variationsformulierung des Ausgangsproblems definiert ist.

6.1 Forderungen an die Bilinearformen

Eine weitere Voraussetzung dafür, daß die im ersten Kapitel hergeleitete Konvergenztheorie für den nichtkonformen Fall anwendbar ist, sind eine Reihe von Bedingungen, die an die zum Operator A gehörende Bilinearform a gestellt werden müssen.

Grundlegend ist die Forderung an die Bilinearform a und die auf H^h fortgesetzten Bilinearformen a^h nach der Stetigkeit mit einer gemeinsamen Stetigkeitskonstanten α. Gleiches muß für die Linearformen f und f^h gelten, mit einer Stetigkeitskonstanten ϕ.

Das Hauptresultat aus Kapitel 1 sagte dann aus, daß im Falle der gleichmäßigen κ-Koerzivität aller Bilinearformen die Diskretisierung stabil ist und die exakte schwache Lösung existiert. Im Falle einer genügend hohen Glattheit der exakten Lösung tritt daher Konvergenz der schwachen Näherungslösungen ein.

Die Diskretisierung sei stabil, hinreichend dafür ist κ-Koerzivität von a und allen a^h. Dann existieren die exakte Lösung u^h und alle schwachen Näherungslösungen u_0^h der Variationsaufgabe. Falls sogar u_0 in $H^R(\Omega)$ mit $R \geq 2$ liegt, konvergieren sie in dem Sinne, daß für alle $h \in H$

$$\left\| u_0 - u_0^h \right\|_h \;\leq\; h^{\min\{R-1,G\}} \cdot C$$

mit einer von h unabhängigen Konstanten C.

Die Bedingung $u_0 \in H^2(\Omega)$ ist natürlich im allgemeinen schwer nachzuprüfen, sie war auch notwendige Voraussetzung dafür, daß bereits die schwächere (κ, μ)-Elliptizität aller Bilinearformen für die Stabilität ausreicht. Satz 2.4 lieferte hinreichende Bedingungen an die Bilinearform a und das Ausgangsgebiet dafür, daß die Lösung zumindest H^2-regulär ist. Weitere nützliche Aussagen darüber finden sich übrigens in [H], Kapitel 9, 'Regularität der Lösung'.

Die Bilinearform eines elliptischen Randwertproblems,

$$a(u,v) = \int_\Omega a_{ij}\, \partial_i u \cdot \partial_j v \;+\; b_k\, \partial_k u \cdot v \;+\; c_0\, u \cdot v,$$

mit gleichmäßig elliptischer Koeffizientenmatrix $(a_{ij}) \in L^\infty(\Omega)^{D\times D}$ ist (κ, μ)-elliptisch und stetig mit einer Konstanten α. Sie induziert α-stetige Bilinearformen a^h, die ebenfalls (κ, μ)-elliptisch sind.
Falls die Lösungen der Variationsaufgabe für beliebiges $f \in L^\infty(\Omega)$ stets H^2-regulär sind, so folgt aus der gleichmäßigen (κ, μ)-Elliptizität aller Bilinearformen und der Invertierbarkeit des zur Bilinearform a gehörende Operator $A \in \mathcal{L}(V', V)$ die Stabilität der Diskretisierung. Insbesondere hat man eine analoge Konvergenzaussage wie für den koerziven Fall.

6.2 Forderungen an das Referenzelement

Von Eigenschaften des Referenzelementes hängt im wesentlichen ab, ob die Existenz und Approximationseigenschaft des Interpolationsoperators gesichert ist. Die Forderungen an geometrische Eigenschaften, die für den Interpolationsoperator nach Brenner/Scott notwendig waren, sind bereits in Abschnitt 6.0 zusammengestellt.

Damit jedoch der Übergang zum Kollokationsverfahrens formuliert werden kann, ist es weiter vonnöten, daß Funktionen in V^h existieren, die in den Kollokationsstellen beliebig vorgegebene Werte annehmen. Dafür muß die Interpolation auf Kollokationspunkten gemäß Abschnitt 3.4 wohldefiniert sein. Die erforderliche Voraussetzung für Theorem 3.14 war:

- Für jede beliebige Wahl von $\Upsilon^s \in \{\Phi^{*s}, R_\Phi^{*s}\}$ für alle $1 \leq s \leq S$ soll die Matrix

$$\Upsilon := \begin{bmatrix} \Upsilon^1 \\ \vdots \\ \Upsilon^S \\ \Psi \end{bmatrix}$$

  invertierbar sein.

Die Definitionen der Matrizen sollen hier nicht wiederholt werden, sie hängen allesamt ausschließlich vom Referenzelement ab, und die geforderten Eigenschaften können leicht durch das Programm verifiziert werden, welches mit den Rechnungen beauftragt ist. Man beachte, daß es nicht nötig ist, daß die stärkeren Voraussetzungen für Theorem 3.10 ebenfalls erfüllt sind, auch wenn dieses dann noch einen alternativen Beweis für die Existenz des Interpolationsoperators auf Kollokationsstellen liefert.

6.3 Konvergenzresultat

Um das Gleichungssystem für die Kollokation aufzustellen, muß nun noch angenommen werden, daß Funktionswerte für die Funktion F im Inneren aller finiten Elemente wohldefiniert sind. Das heißt, daß die Zerlegungen so gewählt sein müssen, daß eventuelle Unstetigkeitsstellen von F stets auf Elementrändern zu liegen kommen.

Für den Übergang zur Variationsformulierung der Kollokation müssen die Kollokationspunkte außerdem derart gelegen sein, daß für jedes $h \in H$ eine Quadraturformel

$$Q^h(v^h) = \sum_{i=1}^{F^h} Q_i^h(u^h v^h) = \sum_{i=1}^{F^h} \sum_{z \in Z_i^h} v^h(z) q_z^h$$

existiert, welche mit einer festen Zahl $Q \geq 1$ und einer von h unabhängigen Konstanten $C_{\mathrm{q}} > 0$ der Abschätzung

$$\left| \sum_{i=1}^{F^h} Q_i^h(u^h v^h) \; - \; \int_{\omega_i^h} u^h v^h \right| \; \leq \; h^Q \cdot C_{\mathrm{q}} \; \left\| u^h \right\|_h \; \left\| v^h \right\|_h \text{ für alle } u^h, v^h \in \mathrm{V}^h$$

genügt. Möchte man für gute Konvergenz $Q > 1$ erreichen, so wird dies im allgemeinen eine weitere Forderung an das Referenzelement ergeben. Die allgemeine Gültigkeit der Formel für $Q = 1$, die man für lineare Konvergenz und Stabilität benötigt, ist in Lemma B.5 gezeigt worden.

Es gilt dann das folgende erste Hauptresultat dieser Arbeit, welches die Existenz und eine etwas schwächliche lineare Konvergenz der Kollokationslösungen garantiert. Für dieses erste Resultat müssen außerdem die Voraussetzungen für die Abschätzung der Randfehler aus Satz 4.4 erfüllt sein.

Es existiert ein $\tilde{h} > 0$ und ein $\tilde{\epsilon} > 0$ und eine von h unabhängige Konstante $C > 0$, so daß für $0 < h \leq \tilde{h}$:

(*i*) *Die Kollokationslösung $\tilde{u}_0^h \in V^h$ mit*

$$A\tilde{u}_0^h(z^h) \; = \; F(z^h) \text{ für alle } z^h \in Z^h$$

existiert, ist eindeutig bestimmt und erfüllt $\left\| \tilde{u}_0^h \right\|_h \leq \phi/\tilde{\epsilon}$. Insbesondere ist das Gleichungssystem für die Kollokation lösbar und nach Bemerkung 5.1 unabhängig von h gleichmäßig gut konditioniert.

(*ii*) *Der Fehler der Diskretisierung läßt sich abschätzen durch*

$$\left\| \tilde{u}_0^h - u_0 \right\|_h \; \leq \; C\, h.$$

Durch optimale Wahl der Lage der Verbindungspunkte und gleichzeitige Erhöhung der Dimension des Raumes lokaler Funktionen kann man bei genügend hoher Glattheit der Lösung eine höhere Konvergenzordnung erzielen. Die Voraussetzungen für die Abschätzung der Randfehler sind dann automatisch erfüllt. Dies ist das zweite Hauptresultat.

Das Referenzelement besitze auf jeder Seite L Verbindungspunkte, die wie die Stützstellen der Gauss-Legendre-Quadraturformeln verteilt seien. Dann läßt sich obige Konvergenzaussage verbessern zu

$$\left\| \tilde{u}_0^h - u_0 \right\|_h \; \leq \; C\, h^{\min\{L, R-1, Q\}}.$$

6.4 Ansatzpunkte für weitere Untersuchungen

Es soll hier nicht verschwiegen werden, daß es trotz des schon einmal sehr erfreulichen Resultates ein paar Dinge gibt, die noch einer genaueren Untersuchung harren. Zum einen ist die Geometrie der Elemente in manchen der hier verwendeten eher technischen Lemmas recht restriktiv gewesen: Diverse Male wurde direkt davon Gebrauch gemacht, daß es sich bei den finiten Elementen wie im Original beschrieben um Quadrate im $\mathbb{R}^2$ handelt. Zwar lassen sich die entsprechenden Aussagen offensichtlich auf Dreiecke verallgemeinern, dies erfordert jedoch noch einigen Aufwand, will man es detailliert ausarbeiten. Auch eine entsprechende Verallgemeinerung auf höhere Dimensionen wäre wünschenswert.

Ebenfalls ein noch recht unbekanntes Terrain stellen die mehrdimensionalen Quadraturformeln dar, die man für eine gute Konvergenz braucht. Es gibt zwar große Datenbanken für Spezialfälle, z.B. [Q], und allgemeinere Aussagen, falls man Tensorprodukte bekannter eindimensionaler Formeln wählt. Die numerischen Resultate, die im nächsten Teil dieser Arbeit angeführt werden, sind jedoch in dieser Beziehung recht überraschend: Die Verteilung der Kollokationsstellen scheint für die Konvergenzordnung im wesentlichen egal zu sein, diese wird quasi ausschließlich durch die Anzahl der Verbindungspunkte und die Qualität der Interpolation bestimmt. Es scheint also so zu sein, als gäbe es für die meisten Verteilungen der Kollokationsstellen Quadraturformeln, die bereits so gut sind, daß der konvergenzlimitierende Faktor durch die Randfehler oder eine nicht hinreichend glatte exakte Lösung gegeben ist - einmal ganz abgesehen davon, daß es natürlich auch noch eine völlig andere Beweismethode geben könnte, die von der Existenz von Quadraturformeln gar keinen Gebrauch machen muß. Resultate für den Quadraturfehler in der erforderlichen Allgemeinheit scheinen noch nicht bekannt zu sein und sind sicherlich so schwierig zu finden, daß diese Aufgabe in dieser Arbeit leider nicht mehr angegangen werden kann.

Wir wollen daher mit dem bisher Erreichten für dieses Mal zufrieden sein, uns entspannt zurücklehnen und die Früchte unserer Arbeit genießen, indem wir im nächsten Teil das Verfahren in Aktion erleben und ihm dabei zuschauen, wie es sich bei der Lösung partieller Differentialgleichungen in der Praxis schlägt.

Teil II

Praxis

Kapitel 7

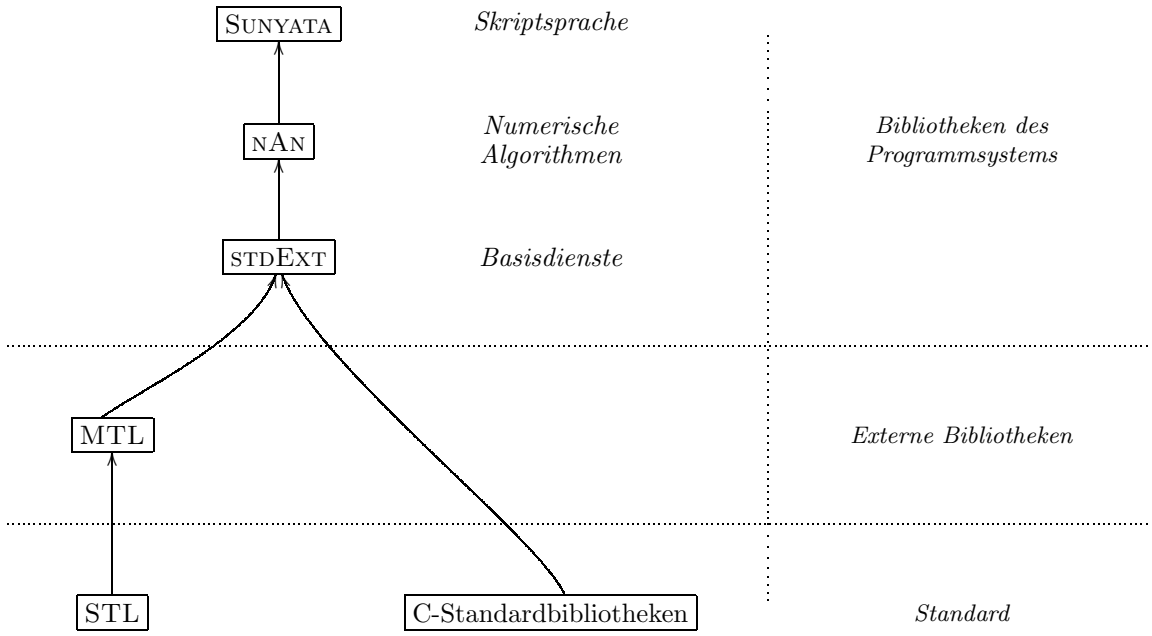
Das Programmpaket

In diesem Kapitel soll die Software kurz vorgestellt werden, mit der die numerischen Beispielrechnungen durchgeführt wurden, die die im letzten Teil hergeleiteten Resultate exemplarisch verifizieren sollen. Es handelt sich um eine Bibliothek für Algorithmen der numerischen Analysis, die über eine einfache Interpretersprache zugänglich gemacht werden. Im Mittelpunkt stehen dabei Lösungsverfahren für Differentialgleichungen. Viele der Teilsysteme sind noch im Stadium der Planung, aber die Module, welche für das hier vorgestellten Verfahren benötigt werden, sind in der Version für quadratische Gitter im $\mathbb{R}^2$ und allgemeine elliptische Randwertprobleme mit einer Gleichung fertig implementiert.

Da es sich um ein recht umfangreiches Projekt handelt, wird vom Gesamtsystem nur die allgemeine Struktur erläutert. Lediglich die Implementation des von E.Doedel in [D] beschriebenen Algorithmus, welcher in direktem Zusammenhang mit dem Thema dieser Arbeit steht, wird später noch detaillierter dargelegt. Eine ausführliche Dokumentation für Entwickler liegt in elektronischer Form vor und ist nicht Gegenstand dieses Textes. Sie ist zusammen mit Quelltexten und fertig übersetzten, ausführbaren Versionen des Skriptinterpreters für Linux oder Windows verfügbar unter *http://www.bastian-goldluecke.de/projekte*.

7.0 Konzept und Architektur

Das Programmsystem ist vollständig in klassischem[1] C++ programmiert derart, daß der gleiche Quelltext sowohl unter Linux als auch unter Windows mit Microsoft Visual C++ compiliert werden kann. Es ist hierarchisch in drei Schichten unterteilt, welche drei verschiedene statische Bibliotheken bilden.



[1] Darunter versteht der Autor die moderate Verwendung von Templates im Gegensatz zur Praxis der in Mode gekommenen modernen Template-Bibliotheken, die naturgemäß davon wimmeln.

- stdExt. Dies ist die Basisbibliothek, welche die Eigenheiten der verschiedenen Betriebssysteme kapselt, indem die benötigten Dienste in entsprechenden Klassen versteckt werden, auf die die darüberliegenden Schichten in einheitlicher Form zugreifen können. Außerdem wird die C++- Standardbibliothek um nützliche Containerklassen erweitert. Im folgenden wird diese Schicht keine weitere Erwähnung mehr finden, da sie für die Mathematik nicht relevant ist.

- nAn. In dieser Schicht werden alle Strukturen und Algorithmen implementiert, welche für numerische Berechnungen verwendet werden. Für Methoden aus der linearen Algebra wird dabei zumeist auf die MTL[2] zurückgegriffen, insbesondere sind die Matrix- und Vektorklassen von nAn direkt aus dieser Bibliothek spezialisiert. Da sich hier die meisten für uns relevanten Dinge abspielen, wird auf diese Schicht im nächsten Abschnitt noch ausführlicher eingegangen werden.

- Sunyata. Dies ist die letzte Schicht und gleichzeitig der Name der Skriptsprache[3], welche die Algorithmen aus nAn verfügbar macht. Das Ziel war dabei, einen schnellen und einfachen Einstieg in die Verfahren der Bibliothek zu ermöglichen, ohne komplexen Code schreiben oder das Programm neu compilieren zu müssen. Des weiteren wird die Durchführung vieler verschiedenartiger numerischer Experimente durch die Verwendung von Skripten stark erleichtert. Die Verwendung der Sprache wird in Kapitel 9 näher erläutert, in welchem mit ihrer Hilfe zahlreiche Beispiele durchgerechnet werden. Ansonsten soll hier auf die -wenn auch sehr interessante- Programmierung des Interpreters nicht weiter eingegangen werden, da dies in den Bereich der Informatik gehört und den Umfang dieser Arbeit sprengen würde. Interessierte Leser seien wieder auf die Dokumentation im Internet verwiesen.

Besonderer Wert wurde darauf gelegt, daß sich die Objekte, die man in der Mathematik findet, direkt in der Klassenhierarchie der Bibliothek wiederfinden. Da es sich in erster Linie um ein System zum Experimentieren handelt, wurde außerdem stets leichter Lesbarkeit und Ausbaufähigkeit des Codes der Vorzug gegenüber geringfügigen Performancegewinnen gegeben. Insbesondere der Aspekt der Erweiterbarkeit stand sehr im Mittelpunkt, neue Algorithmen und verbesserte Versionen der alten sollten sich möglichst nahtlos in die bestehende Struktur integrieren. Erreicht wird dies unter anderem durch ein zentrales Konzept für grundlegende mathematische Objekte, welches die wesentlichen Operationen auf diesen durch virtuelle Funktionen verfügbar macht, wodurch Spezialisierungen und besondere Implementationen erleichtert werden. Im folgenden wird diese Klassenhierarchie kurz vorgestellt.

7.1 Struktur des Moduls nAn

Hierbei handelt es sich was Mathematik betrifft um das zentrale Teilsystem. Es stellt Klassenhierarchien mathematischer Objekte zur Verfügung, sowie einen Satz von Algorithmen, die mit diesen operieren. Zunächst werden die wichtigsten Klassen kurz vorgestellt, ohne auf Details einzugehen, all dies ist ausführlich in der Online-Dokumentation zu finden.

- *Vektoren und Matrizen* werden abgebildet durch die Klassen *CRealVector* und *CRealMatrix*. Beides sind direkte Spezialisierungen der entsprechenden dicht besetzten Versionen aus der MTL. Methoden der linearen Algebra sind aus historischen Gründen zum großen Teil in globalen Funktionen gekapselt.

- *Gebiete im* $\mathbb{R}^n$ repräsentiert die abstrakte Basisklasse *CRealDomain*. Für diverse Fälle existieren Spezialisierungen, insbesondere für mehrdimensionale Intervalle. Diese Klasse beherrscht die naheliegenden Auskünfte zum Beispiel in der Art, ob bestimmte Vektoren im dargestellten Gebiet liegen, sie wird unter anderem als Definitionsbereich für Funktionsobjekte oder finite Elemente verwendet.

- *Funktionen endlichdimensionaler Vektorräume* sind der Zuständigkeitsbereich der Klasse *CRealFunction*. Naturgemäß ist dies eines der zentralen Objekte der Bibliothek, jeder Algorithmus, welcher in irgendeiner Weise mit Funktionen zu tun hat, wie beispielsweise das Newton-Verfahren, bekommt Objekte dieser Klasse als Parameter mitgegeben. Umgekehrt geben Algorithmen, welche Funktionen

[2]Hinter diesem Akronym verbirgt sich die Matrix Template Library, für weitere Informationen konsultiere man *http://www.mtl.com*.

[3]Jene ist vom Autor in einem seiner philosophischeren Momente getauft worden: Sunyata ist aus dem Sanskrit entlehnt und bedeutet soviel wie 'Die Leerheit aller Worte und Begriffe' - was auch immer das in diesem Zusammenhang heißen mag, so weit sind die Überlegungen dann nicht mehr fortgeschritten. nAn ist übrigens weniger tiefsinnig und eine heitere Anspielung auf NaN, was in der Informatik üblicherweise für 'Not a Number' steht. In vielerlei Deutungen ist also nAn gewissermaßen das Gegenteil, es könnte zum Beispiel für 'numerische Analysis' stehen.

als Rückgabewerte liefern, Objekte dieser Klasse zurück. Die Handhabung ist äußerst flexibel dadurch, daß die Funktionsauswertung in der virtuellen Funktion *Evaluate* geschieht, die nach Belieben überschrieben werden kann. *CRealFunction* bietet davon ausgehend Default-Implementationen, um numerische Ableitungen zu bilden, letztere können aber vom Anwender auch mit eigenen Methoden angepaßt werden. Außerdem ist es möglich, Objekte dieses Typs direkt aus bestimmten Funktionspointern des Compilers zu erzeugen, wodurch die die etwas umständiche Definition einer neuen Klasse pro Funktion vermieden werden kann.

Besonders mächtig wird das Konzept jedoch erst in Verbindung mit dem Modul SUNYATA, welches durch seinen Interpreter zusätzlich eine abgeleitete Klasse zur Verfügung stellt, in der Funktionen auch symbolisch durch einen mathematischen Ausdruck definiert werden können - über die Funktion kann also zur Laufzeit des Programms entschieden werden. Diese Art von Funktionen beherrschen auch das symbolische Differenzieren, effizient werden sie dadurch, daß die symbolische Form zuerst in einen schnell ausführbaren, internen Pseudocode compiliert wird. Auf diese Möglichkeit der Definition wird später in Verbindung mit der Skriptsprache noch etwas genauer eingegangen.

- *Endlichdimensionale Räume von Funktionen* sind vor allen für die Implementation von finiten Elementen gedacht. Sie verwalten ein Basissystem von Funktionen, zumeist ein Tensorprodukt, und interpretieren davon ausgehend Vektoren als Funktionen in dem dadurch definierten Vektorraum. Als Operationen beherrschen sie dann effizientes Berechnen von Funktions- oder Ableitungswerten der dadurch gegebenen Funktionen, im Falle von Fourierbasen auch in Form der schnellen Fouriertransformation.

- *Finite Elemente und Operatoren* implementieren die Methode der finite Elemente. Die Klasse *COperator* definiert einen Operator A auf einem Funktionenraum, er benutzt Objekte der Klasse *CFiniteElement*, um eine Diskretisierung dieser Räume durchzuführen und davon ausgehend zum Beispiel die Gleichung $Au = F$ zu lösen. Im nächsten Kapitel wird speziell die Implementation der Methode der Doedelschen finiten Elemente geschildert.

Zum Modul gehören weiterhin eine Reihe von Funktionen ohne spezielle Klassenzugehörigkeit, die grundlegende Algorithmen der Numerischen Analysis implementieren. Dazu gehören Dinge wie das Newton-Verfahren, Runge-Kutta-Methoden für gewöhnliche Differentialgleichungen, Bezier-Splines und dergleichen mehr. Da es sich um Standardverfahren handelt, wird hier nicht weiter darauf eingegangen. Zur Definition ihrer Schnittstelle verwenden sie allesamt die oben beschriebenen Klassen, wodurch die Bibliothek ein homogenes Erscheinungsbild präsentiert.

Kapitel 8

Der Lösungsalgorithmus

Ziel dieses Kapitels ist die Vorstellung eines effizienten Lösungsalgorithmus für die im ersten Teil untersuchten Diskretisierungsmethoden. Es handelt sich um das direkte, rekursive Verfahren *Nested Dissection*, und ist bereits in [D] von E.Doedel beschrieben worden. Dies soll hier der Vollständigkeit halber im Hinblick auf die speziellen Bezeichnungen der Arbeit noch einmal geschehen, anschließend wird die Implementation im hier vorliegenden Programmsystem etwas genauer diskutiert. Das Ziel ist dabei, eventuellen Interessenten einen schnellen Einstieg in das Studium der vorliegenden Online-Dokumentation und der Quelltexte zu ermöglichen, in denen der Algorithmus implementiert ist. Mit Hilfe dieser Übersicht sollte es hoffentlich gut möglich sein nachzuvollziehen, was das Programm an welcher Stelle tut. Bei der Beschreibung der Implementation werden außerdem noch mögliche Optimierungen erläutert, welche Teile der Berechnung um ein Vielfaches beschleunigen können.

8.0 Mathematische Beschreibung

Der Algorithmus *Nested Dissection* dient dazu, eine zulässige $\tilde{u}_0^h$ Funktion im Raum V^h zu finden, die das Gleichungssystem

$$A\tilde{u}_0^h(z) = F(z) \text{ für alle } z \in Z^h$$

für die Kollokation löst. Die Forderung $\tilde{u}_0^h \in \mathrm{V}^h$ gibt dabei zusätzliche Gleichungen für die verschiedenen Bedingungen, zur Erinnerung:

- Übereinstimmung der Funktionswerte der lokalen Funktionen auf benachbarten finiten Elementen ω_i^h und ω_j^h auf $X_i^h \cap X_j^h$,
- Übereinstimmung der Normalenableitungen der lokalen Funktionen auf benachbarten finiten Elementen ω_i^h und ω_j^h auf $Y_i^h \cap Y_j^h$,
- Lokale Funktionen auf finiten Elementen ω_i^h mit $X_i^h \cap \partial\Omega \neq \emptyset$ müssen in den Punkten aus $X_i^h \cap \partial\Omega$ die auf $\partial\Omega$ vorgegebenen Randbedingungen erfüllen.

Insgesamt ergeben sich dadurch für jedes finite Element N Kollokationsgleichungen und M Gleichungen durch Nebenbedingungen, wodurch dann eine lokale Funktion eindeutig bestimmt wird. Die Funktionswerte u auf den Punkten X^h tauchen neben den Koeffizienten c der lokalen Funktionen als Unbekannte in den Gleichungen auf, ebenso die Werte v der Normalenableitungen. Auf jedem finiten Element hat man einen affin linearen Zusammenhang zwischen den Unbekannten u und v. Die Idee ist nun, über die Baumstruktur der Zerlegung von Ω rekursiv von den Blättern beginnend zu einem Zusammenhang zwischen den Unbekannten $u_{\partial\Omega}$ und $v_{\partial\Omega}$ auf dem Rand von Ω zu gelangen. Auf diese Weise hat man dann die Unbekannten im Inneren von Ω komplett aus den Gleichungen eliminiert und das Gleichungssystem stark verkleinert.

Es wird von nun an ein festes $h \in H$ fixiert und der entsprechende Index im folgenden unterdrückt. Der erste Schritt ist die Beobachtung, daß unter gewissen Umständen auf jedem finiten Element ω_i die Koeffizienten der gesuchten lokalen Funktion durch die Funktionswerte ausgedrückt werden können. Dies führt dazu, daß man die Koeffizienten aus dem Gleichungssystem eliminieren kann.

Es sei wieder A der (lineare) Differentialoperator, dann besteht wie im Kapitel über die lokale Interpolation ein Zusammenhang zwischen den Koeffizienten $c \in \mathbb{R}^K$, den Werten $u \in \mathbb{R}^N$ der lokalen

Funktion auf den Verbindungspunkten und den Werten $f \in \mathbb{R}^N$ der vorgegebenen Funktion F auf den Kollokationspunkten. Dieser wird durch folgendes Gleichungssystem gegeben:

$$\begin{bmatrix} \Phi^* \\ \Psi_A \end{bmatrix} \cdot c := \begin{bmatrix} \phi_1(x_1) & \dots & \phi_K(x_1) \\ \vdots & \ddots & \vdots \\ \phi_1(x_M) & \dots & \phi_K(x_M) \\ A\phi_1(z_1) & \dots & A\phi_K(z_1) \\ \vdots & \ddots & \vdots \\ A\phi_1(z_N) & \dots & A\phi_K(z_N) \end{bmatrix} \cdot c = \begin{bmatrix} u_1 \\ \vdots \\ u_M \\ f_1 \\ \vdots \\ f_N \end{bmatrix}. \tag{8.1}$$

Ist die Matrix auf der linken Seite invertierbar, was eine notwendige Bedingung für das Funktionieren des Algorithmus ist, so kann man das System nach c auflösen und ausnutzen, daß auch die Werte v der Normalenableitungen der lokalen Funktion von c abhängen. Dies ergibt dann folgenden Zusammenhang:

$$\begin{aligned} v &= R_\Phi^* \cdot c := \begin{bmatrix} \nabla\phi_1 \cdot \eta(y_1) & \dots & \nabla\phi_K \cdot \eta(y_1) \\ \vdots & \ddots & \vdots \\ \nabla\phi_1 \cdot \eta(y_N) & \dots & \nabla\phi_K \cdot \eta(y_N) \end{bmatrix} \cdot c \\ &= R_\Phi^* \cdot \begin{bmatrix} \Phi^* \\ \Psi_A \end{bmatrix}^{-1} \begin{bmatrix} u \\ f \end{bmatrix} \\ &=: \Pi \begin{bmatrix} u \\ f \end{bmatrix}. \end{aligned}$$

Zerlegt man die Matrix $\Pi \in \mathbb{R}^{N \times M}$ in die N linken Spalten Π^L und die M rechten Spalten Π^R, so gelangt man zu der affin linearen Beziehung

$$\begin{aligned} v &= \Pi^L u + \Pi^R f \\ &=: Au + b \end{aligned} \tag{8.2}$$

mit der im allgemeinen vom finiten Element abhängigen Matrix $A \in \mathbb{R}^{N \times N}$ und dem Vektor $b \in \mathbb{R}^N$.

Für alle Elemente zusammengenommen, inklusive der Verbindungs- und Randbedingungen, ergeben diese Gleichungen ein großes globales System für die Werte u und v, welches alle Informationen über das Problem enthält. Sind daraus die Werte u berechnet worden, so kann man über die lokale Beziehung

$$c = \begin{bmatrix} \Phi^* \\ \Psi_A \end{bmatrix}^{-1} \begin{bmatrix} u \\ f \end{bmatrix} \tag{8.3}$$

die Koeffizienten c für jedes finite Element bestimmen.

Als nächstes wird beschrieben, wie das globale Gleichungssystem effizient gelöst werden kann. Auf jedem finiten Element besteht ein affiner Zusammenhang zwischen Funktions- und Ableitungswerten auf den Verbindungspunkten. Man kann dann bei zwei benachbarten Elementen i und j die gemeinsamen Werte u_{ij} und v_{ij} auf dem Innenrand eliminieren[1]. Für die folgende Untersuchung bezeichne weiterhin u_i und v_i die Daten auf $\partial\omega_i \setminus \overline{\omega}_j$, beziehungsweise u_j und v_j die Daten auf $\partial\omega_j \setminus \overline{\omega}_i$. Auf dem gemeinsamen Rand liegen L Verbindungspunkte.

Dann gelten bei geeigneter Sortierung der Verbindungspunkte und Zerlegung der Matrizen Beziehungen der Form

$$v_i = A_i^{ol} u_i + A_i^{or} u_{ij} + b_i^o \qquad v_j = A_j^{ol} u_j + A_j^{or} u_{ij} + b_j^o \tag{8.4}$$

$$v_{ij} = A_i^{ul} u_i + A_i^{ur} u_{ij} + b_i^u \quad v_{ij} = A_j^{ul} u_j + A_j^{ur} u_{ij} + b_j^u. \tag{8.5}$$

Die Bedeutung der gar nicht so kryptischen hochgestellten Indizes ist dabei:

o - die oberen $N - L$ Zeilen,
u - die unteren L Zeilen,
l - die linken $N - L$ Spalten und
r - die rechten L Spalten.

[1] Der Einfachheit halber wird davon ausgegangen, daß der Normalenvektor, mit dem die Ableitung berechnet wird, für beide finiten Elemente gleich orientiert ist, so daß die Vorzeichen übereinstimmen

In diesem System kann man die beiden unteren Beziehungen (8.5) gleichsetzen und gelangt so zu einer neuen Darstellung für die u_{ij}:

$$\begin{aligned} u_{ij} &= B^{-1}\left(A_j^{ul}u_j - A_i^{ul}u_i + b_j^u - b_i^u\right), \\ \text{mit } B &:= A_i^{ur} - A_j^{ur}. \end{aligned} \tag{8.6}$$

Substitution in die Gleichungen (8.4) liefert dann

$$\begin{aligned} v_i &= A_i^{ol}u_i + A_i^{or}B^{-1}\left(A_j^{ul}u_j - A_i^{ul}u_i + b_j^u - b_i^u\right) + b_i^o \\ v_j &= A_j^{ol}u_j + A_j^{or}B^{-1}\left(A_j^{ul}u_j - A_i^{ul}u_i + b_j^u - b_i^u\right) + b_j^o. \end{aligned}$$

Dies ist wieder ein System der Form

$$\begin{aligned} v_i &= \hat{A}_{11}u_i + \hat{A}_{12}u_j + \hat{r}_1 \\ v_j &= \hat{A}_{21}u_i + \hat{A}_{22}u_j + \hat{r}_2, \end{aligned} \tag{8.7}$$

mit den Matrizen, bzw. Vektoren

$$\begin{aligned} \hat{A}_{11} &= A_i^{ol} - A_i^{or}B^{-1}A_i^{ul} & \hat{A}_{12} &= A_i^{or}B^{-1}A_j^{ul} & r_1 &= B^{-1}\left(b_j^u - b_i^u\right) + b_i^o \\ \hat{A}_{21} &= -A_j^{or}B^{-1}A_i^{ul} & \hat{A}_{22} &= A_j^{ol} + A_j^{or}B^{-1}A_j^{ul} & r_2 &= B^{-1}\left(b_j^u - b_i^u\right) + b_j^o. \end{aligned}$$

Das Ergebnis ist also ein neuer affiner Zusammenhang zwischen den Funktionswerten und den Werten der Normalenableitungen auf dem Rand der *Vereinigung* der beiden finiten Elemente $\overline{\omega}_i \cup \overline{\omega}_j$. Ist nun die Zerlegung des Ausgangsgebietes Ω in finite Elemente durch rekursive Zweiteilung entstanden, so gelangt man über diesen Prozeß also zu einem affinen Zusammenhang

$$v_{\partial\Omega} = A_\Omega u_{\partial\Omega} + b_\Omega \tag{8.8}$$

der Funktionswerte und Normalenableitungen auf dem Rand des Gesamtgebietes. Die Werte $u_{\partial\Omega}$ sind nun aber durch die Randbedingungen vorgegeben, durch dieses Gleichungssystem sind also die Werte $v_{\partial\Omega}$ eindeutig festgelegt. Mittels Rückwärtseinsetzen in die Gleichungen (8.6) und (8.5) auf den Teilgebieten ergeben sich dann schrittweise die Funktions- und Ableitungswerte auf den Innenrändern.

Aus der Herleitung ist klar, daß die eigentliche Geometrie der finiten Elemente keine Rolle spielte - es könnten statt Quadraten auch Dreiecke oder Siebenecke oder gänzlich unregelmäßige Gebilde sein. Wesentlich ist lediglich die Eigenschaft, daß die finiten Elemente durch rekursive Zweiteilung des Ausgangsgebietes entstehen und auf jedem einzelnen finiten Element die Normalenableitungen durch die Funktionswerte ausgedrückt werden können.

Zum Abschluß noch ein paar Bemerkungen:

- Die Bedingung, daß die Normalenableitungen explizit durch die Funktionswerte darstellbar sind, ist für manche Probleme der Bifurkationsnumerik schon zu stark, tatsächlich gibt es auch eine modifizierte Version des Algorithmus, in der lediglich ein gewisser Zusammenhang der Form $Av + Bu = f$ bekannt sein muß. Diese Version ist in dem kleinen Testsystem jedoch noch nicht integriert.
- Der Algorithmus ist außer für Dirichlet-Randbedingungen auch für solche Randbedingungen geeignet, die eine gewisse Normalenableitung auf $\partial\Omega$ vorschreiben - insbesondere also die natürlichen Randbedingungen für elliptische Randwertprobleme. Dafür muß lediglich das letzte System (8.8) geeignet umgestellt und gelöst werden. Der Rechenaufwand vergrößert sich allerdings signifikant, da nochmals die sehr große Matrix A_Ω invertiert werden muß.

8.1 Newton-Verfahren

Die bisherige Herleitung scheint anzudeuten, daß das Verfahren nur für lineare Differentialoperatoren geeignet ist. Mit Hilfe des zum Standard gehörenden Newtonschen Iterationsverfahrens können jedoch auch nichtlineare Gleichungen damit behandelt werden.

Dafür seien c, u und v die Daten der Näherungslösung in einem Iterationsschritt, N ein nichtlinearer Operator und A seine Linearisierung in c. Zur Präzisierung: Üblicherweise ist N elliptisch mit einer *Inhomogenität*, d.h. zum Beispiel

$$N(u) = \Delta u + g(u).$$

Dann ist A gegeben durch

$$Au \;=\; \Delta u \,+\, Dg(c)u.$$

Die Residuen der aktuellen Näherung sind die Vektoren

$$\begin{aligned} r_c \;&:=\; [N\phi(z_m) \,-\, f_m]_{1\le m\le M}, \\ r_u \;&:=\; \Phi\cdot c \,-\, u, \\ \text{und } r_v \;&:=\; R_\Phi\cdot c \,-\, v, \\ \text{wobei } \phi &:= \sum_{k=1}^{K} c_k\phi_k \end{aligned} \tag{8.9}$$

die zur Lösung gehörende lokale Funktion sein soll. Man beachte, daß man zur Berechnung des Fehlers auf den Kollokationspunkten den vollen nichtlinearen Operator auswertet. Das linearisierte Gleichungssystem aus dem letzten Abschnitt wird nun für Korrekturen δc, δu und δv angesetzt mit dem Ziel, durch die Korrektur zu einer neuen Näherung $c+\delta c$, $u+\delta u$ und $v+\delta v$ zu gelangen, welche die Residuen zu Null macht und damit die exakte Lösung besser approximiert. Das zu lösende modifizierte Gleichungssystem in diesem Iterationsschritt lautet also

$$\begin{aligned} \Psi_A\cdot\delta c \;&=\; f \,+\, \delta u \,+\, r_c \\ \delta u \,-\, \Phi\cdot\delta c \;&=\; -r_u \\ \delta v \,-\, R_\Phi\cdot\delta c \;&=\; -r_v, \end{aligned}$$

wobei A der nach obigem Verfahren linearisierte Operator ist. Löst man dieses wie oben auf, so gelangt man wieder zu einer affin linearen Beziehung

$$\begin{aligned} \delta v \;&=\; A\,\delta u \,+\, b, \\ \text{mit } A \;&=\; \Pi^L(u+r_u) \text{ und } b \;=\; -\Pi^R(f+r_c) - r_v. \end{aligned} \tag{8.10}$$

Die weiteren Lösungsschritte zur Bestimmung von δc, δu und δv ergeben sich dann analog obiger Herleitung.

8.2 Implementation

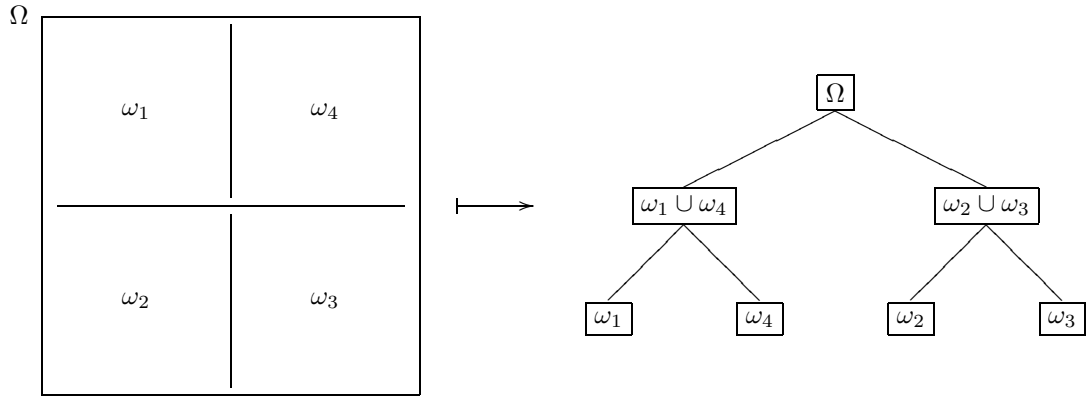
Getreu der Philosophie der Bibliothek ist der beschriebene Algorithmus schnörkellos und mit hohem Wiedererkennungswert in Code umgesetzt worden. Lediglich an einigen wenigen Stellen wurde Gebrauch von möglichen Optimierungen gemacht, diese werden aber gleich noch im Detail beschrieben. Die folgenden Ausführungen sollen dabei helfen, die zum Kern des Algorithmus gehörenden Stellen im Programmtext zu identifizieren und die Struktur seines Ablaufs im groben zu identifizieren, um ein genaueres Studium zu erleichtern. Liegt dies nicht im Interesse des Lesers, so kann er den Rest dieses Kapitels getrost überspringen.

Die Implementation aller Objekte, auf die im folgenden Bezug genommen wird, findet sich komplett in der Quelldatei *doedel.cpp*.

8.2.1 Datenstruktur

Die zentrale Datenstruktur, auf der operiert wird, ist ein Baum von Objekten der Klasse *CDoedelFiniteElement*. Anders als der Name vielleicht vermuten läßt, repräsentiert diese nicht notwendigerweise ein Blatt in der Hierarchie, also ein finites Element im eigentlichen Sinne, sondern wird in höheren Ebenen auch die Vereinigung mehrerer finiter Elemente verwalten. Insbesondere stellt das an der Wurzel des Baumes stehende Objekt dieser Klasse das gesamte Gebiet Ω dar.

8.1 Beispiel. Die folgende Graphik illustriert den Zusammenhang zwischen einer Zerlegung des Gebietes und der zugehörigen Baumstruktur:



Jeder der Knoten des Baumes wird dabei durch ein Objekt der Klasse *CDoedelFiniteElement* repräsentiert. ◇

Die relevanten Daten, die ein Objekt der Klasse *CDoedelFiniteElement* hält, sind:

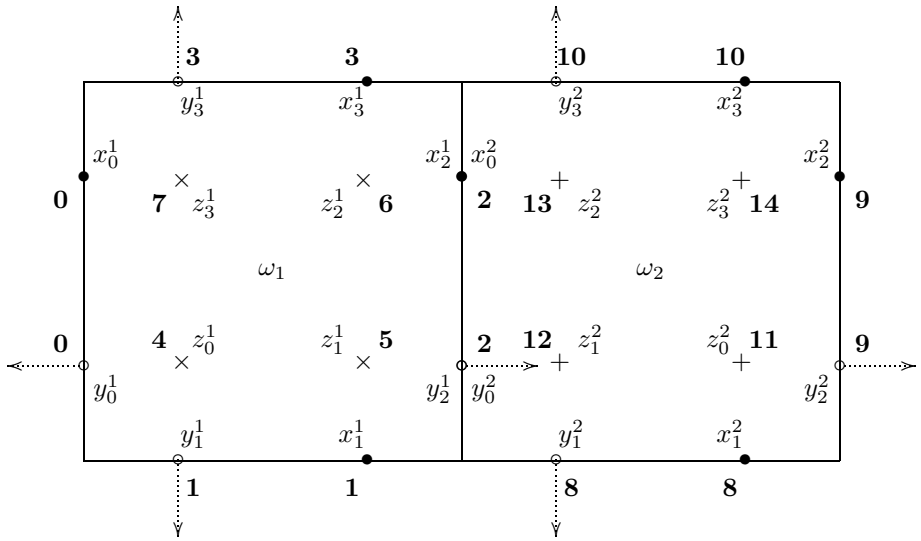
- *Informationen über die Position der Verbindungs- und Kollokationsstellen.* Letztere werden nur in Blättern abgelegt, da sie auf höheren Ebenen nicht mehr benötigt werden. Zusätzlich wird bei zusammengesetzten Elementen noch vermerkt, welche der Verbindungspunkte zu welchem Kindelement gehören, diese Daten braucht man zur Berechnung der Matrizen in (8.7). Punkte sind aus Gründen der Speichereffizienz als Indizes in ein globales Array aller Punkte abgelegt.
- *Einen Verweis auf ein Referenzelement*, in welchem unter anderem abgelegt ist, wie der lokale Funktionenraum aussieht. Dieser wird während der Berechnungen auf das Gebiet des finiten Elementes skaliert, falls es sich um ein Blatt in der Hierarchie handelt, ein solches muß daher auch seinen Definitionsbereich speichern.
- *Matrizen für die Berechnung.* Dies sind einerseits diejenigen, welche ausschließlich von der Elementgeometrie abhängen und daher nur einmal im Vorfeld angelegt werden müssen - hierunter fällt z.B. die Matrix Φ. Die meisten der Matrizen hängen allerdings vom Operator oder dem Problem ab und werden während eines jeden Durchlaufs des Lösungsalgorithmus neu initialisiert.
- *Daten über die Struktur der Zerlegung.* Dazu gehören Verweise auf die Wurzel des Baumes sowie über eventuell vorhandene Kindelemente.

Hier sind noch einmal die Variablen der Klasse *CDoedelFiniteElement* im einzelnen aufgeführt, zusammen mit ihrer Bedeutung und ihrem Verwendungszweck. Es wird helfen, sie zu kennen, wenn man die gleich folgende Beschreibung für die Vorgehensweise des Algorithmus verstehen möchte.

Name	Beschreibung
m_pReferenceElement	Referenzelement
m_pInterval	Definitionsbereich als mehrdimensionales Intervall
m_AP	Matrix A aus (8.10)
m_RHS	Vektor b aus (8.10)
m_rU	Residuenvektor r_u aus (8.9)
m_rV	Residuenvektor r_v aus (8.9)
m_rC	Residuenvektor r_c aus (8.9)
m_BP_LItoGI	Tabelle, die dem lokalen Index eines Randpunktes seinen globalen Index zuordnet
m_BP_GItoLI	Tabelle, die dem globalen Index eines Randpunktes seinen lokalen Index zuordnet
m_OwnBP	Array der lokalen Indizes der Randpunkte, die ausschließlich zu diesem finiten Element gehören
m_CommonBP	Array der lokalen Indizes der Randpunkte, die sich dieses finite Element mit seinem Nachbarelement teilt
m_pParent	In der Hierarchie höherliegendes Element
m_Childs	Array der Kindelemente

Für den globalen Index sind die Randpunkte aller finiten Elemente fortlaufend numeriert, der lokale Index läuft nur über die Randpunkte eines einzelnen finiten Elementes.

8.2 Beispiel. Man betrachte die Situation in der folgenden Grafik für den einfachsten interessanten Fall, daß Ω aus zwei finiten Elementen zusammengesetzt ist. Bei Punkten u_i und v_i bezeichnet das i den lokalen Index bezogen auf das Element, in dem der Punkt liegt. Die globalen Indizes sind jeweils **fett** neben den entsprechenden Punkt gedruckt. Man beachte, daß Punkten u_i und v_i mit gleichem lokalen Index nur ein globaler Index zugeordnet wird. Der Grund dafür ist, daß die Daten für Funktionswerte und Normalenableitungen in zwei verschiedenen globalen Arrays abgelegt werden.



Die daraus aufgebauten lokalen Tabellen lauten dann wie folgt:

Element 1	*Element 2*
$m\_BP\_LItoGI\,[0] = 0$ $m\_BP\_LItoGI\,[1] = 1$ $m\_BP\_LItoGI\,[2] = 2$ $m\_BP\_LItoGI\,[3] = 3$	$m\_BP\_LItoGI\,[0] = 2$ $m\_BP\_LItoGI\,[1] = 8$ $m\_BP\_LItoGI\,[2] = 9$ $m\_BP\_LItoGI\,[3] = 10$
$m\_BP\_LItoGI\,[4] = 4$ $m\_BP\_LItoGI\,[5] = 5$ $m\_BP\_LItoGI\,[6] = 6$ $m\_BP\_LItoGI\,[7] = 7$	$m\_BP\_LItoGI\,[4] = 11$ $m\_BP\_LItoGI\,[5] = 12$ $m\_BP\_LItoGI\,[6] = 13$ $m\_BP\_LItoGI\,[7] = 14$
$m\_OwnBP\,[0] = 0$ $m\_OwnBP\,[1] = 1$ $m\_OwnBP\,[2] = 3$	$m\_OwnBP\,[0] = 1$ $m\_OwnBP\,[1] = 2$ $m\_OwnBP\,[2] = 3$
$m\_CommonBP\,[0] = 2$	$m\_CommonBP\,[0] = 0$

In den Arrays *m_BP_GItoLI* steckt die gleiche Information wie in *m_BP_LItoGI* , nur daß jeweils die linke und rechte Seite vertauscht sind.

Offensichtlich könnte man die Datenstrukturen für diesen Fall noch optimieren, da es nicht wirklich notwendig ist, auch die Kollokationsstellen global zu speichern - im Prinzip würde es reichen, daß sie lokal bekannt sind. Da der speichertechnische Mehraufwand jedoch nicht wesentlich ist, wurde im Hinblick auf zukünftige denkbare Erweiterungen mehr Information abgelegt, als eigentlich nötig. ◇

8.2.2 Algorithmus

Die Implementation des Algorithmus wird nun anhand des Programmablaufs erklärt. Der erste Schritt ist die Konstruktion eines Objektes der Klasse *CDoedelOperator*. Dieser repräsentiert im wesentlichen eine Beschreibung des zu lösenden Problems, daher gehören zu den Konstruktionsparametern natürlicherweise

- der elliptische Differentialoperator A,
- die Funktion F für die rechte Seite der Gleichung,
- die gewünschten Randbedingungen für das Problem sowie
- das Ausgangsgebiet Ω.

Anschließend ruft man zur Definition der gewünschten Struktur der Diskretisierung die Methode

- *CDoedelOperator::CreateGrid*, welche als Parameter die gewünschte Zerlegungstiefe und das Referenzelemente übergeben bekommt. Zur Zeit werden nur quadratische Ausgangsgebiete untersucht, daher erübrigt sich ein tiefsinniges Verfahren, um die Zerlegung durchzuführen. Stattdessen wird Ω einfach die angegebene Anzahl von Rekursionsstufen halbiert, abwechselnd waagerecht und senkrecht. Sinnvollerweise sollte die Tiefe daher gerade sein. Dabei wird der Baum der Elemente erzeugt, der im obigen Abschnitt Datenstrukturen beschrieben wurde. Hierfür zuständig ist die rekursive Methode *CreateDomainSubdivision* von *CDoedelFiniteElement*.

Nach dieser Vorinitialisierung kann der Algorithmus für das Lösen der Gleichung gerufen werden. Dieser steckt in

- *CDoedelOperator::Solve*, welche keine weiteren Parameter mehr benötigt. Sie implementiert das Newton-Verfahren aus 8.1. Die Kenntnis ihrer lokalen Variablen wird im folgenden nützlich sein:

Name	Beschreibung
U	Funktionswerte u der aktuellen Näherung
V	Normalenableitungen v der aktuellen Näherung
C	Koeffizienten c der aktuellen Näherung
dU	Korrektur δu der Funktionswerte
dV	Korrektur δv der Normalenableitungen
dC	Korrektur δc der Koeffizienten
BoundaryPoints	Indizes der Punkte auf Elementrändern
BP_Index	Tabelle, die lokalen Indizes auf Elementen den globalen Index im Array BP zuordnet
BP	Globales Array aller Randpunkte

Die Sortierung der Randpunkte in den Vektoren U, V, dU und dV entspricht dabei der Reihenfolge der Punkte im Array BP. In den Vektoren C und dC sind die Koeffizienten aller lokalen Funktionen in der Reihenfolge der finiten Elemente abgelegt.

Der erste Schritt ist nun die Initialisierung des globalen Arrays BP von Randpunkten. Dies geschieht durch Iteration über alle Blattelemente, wobei doppelt auftretende Punkte eliminiert werden. Damit die finiten Elemente wissen, an welcher Position in diesem Array die zu ihnen gehörenden Punkte stehen, wird parallel die entsprechende Tabelle *BP_Index* erstellt. Dies ist später wichtig, um die richtigen Stellen in den Vektoren U und V lokalisieren zu können.

8.3 Beispiel. In der Situation von Beispiel 8.2 besteht das Array *BoundaryPoints* aus der Liste der globalen Indizes aller Randpunkte in der Reihenfolge ihres Auftretens. Das Array enthält also die Einträge

BoundaryPoints

0	1	2	3	8	9	10

In BP sind die Koordinaten der Randpunkte in genau dieser Reihenfolge abgelegt. *BP_Index* enthält die umgekehrte Zuordnung wie *BoundaryPoints*. ◇

Wenn dieser Index bekannt ist, können die Matrizen Φ und R_Φ berechnet werden, welche ausschließlich von der Elementgeometrie und dem lokalen Raum von Funktionen abhängen. Eine genauere Untersuchung offenbart sogar, daß sie für Blattelemente identisch sind, welche vom gleichen Referenzelement induziert wurden. In unserem Fall sind dies alle, so daß die Matrizen in der Tat nur ein einziges Mal bestimmt werden müssen. Dies erledigt die rekursive Methode *CDoedelFiniteElement::PreIterationSetup*, welche die im ersten Blattelement erzeugten Matrizen in alle anderen Blattelemente kopiert. Sie legt außerdem für jedes Element die Tabellen *m_BP_LItoGI* und *m_BP_LItoGI* an.

Die allgemeine, bis hierhin unkritische Initialisierung ist nun abgeschlossen und der erste Iterationsschritt des Newton-Verfahrens kann starten. Zunächst werden dafür in der rekursiven Methode *CDoedelFiniteElement::SetupIteration* die von der aktuellen Lösung und dem Differentialoperator abhängigen Matrizen in (8.10) berechnet. Dazu gehört insbesondere die Matrix Π, für deren Berechnung eine LU-Zerlegung der Matrix $\begin{bmatrix}\Phi & \Psi_A\end{bmatrix}^T$ durchgeführt werden muß. Diese ist in der Tat nicht stets invertierbar, sondern kann bei ungünstiger Wahl der Kollokationsstellen und einem nicht damit verträglichen Operator schon einmal singulär sein. Der Algorithmus bricht in diesem Fall mit einer entsprechenden Fehlermeldung ab. Hier bietet sich auch eine einfache und im Erfolgsfalle sehr einträgliche Methode der Optimierung an: Ist der Operator A linear im Funktionsargument, d.h. von der Form

$$A(u)(x,y) = \sum_{i,j=0}^{2} a_{ij}(x,y)\,\partial_i\partial_j u(x,y) + f(x,y),$$

so ist die Matrix Ψ_A ausschließlich vom Referenzelement abhängig. In diesem Fall muß nur ein einziges Mal eine Matrixinversion durchgeführt werden, das Ergebnis wird dann in jedes Blattelement übertragen.

Die errechneten Daten werden nun in der eigentlichen Kernprozedur ausgewertet:

– *CDoedelFiniteElement::SolveUVNestedDissection* bestimmt zunächst in der rekursiven Methode *CDoedelFiniteElement::NestedDissection* mit Hilfe der bereits bekannten Beziehungen (8.10) und den Formeln (8.7) die affin lineare Abhängigkeit zwischen den Funktionswerten und Normalenableitungen auf den zusammengesetzten Elementen.

  Anschließend wird das Gleichungssystem (8.8) auf Ω gelöst, allerdings nur dann, wenn keine reinen Dirichlet-Randbedingungen vorliegen. Ansonsten wird die bekannte Matrix A_Ω einfach dazu verwendet, um δv_Ω aus δu_Ω auszurechnen.

  Nun sind sowohl δu_Ω als auch δv_Ω bekannt, und in der rekursiven Methode *CDoedelFiniteElement::Backsubstitution* werden mittels der Beziehungen (8.5) und (8.6) die Änderungen der Funktions- und Ableitungswerte auf den Innenrändern berechnet.

An dieser Stelle sind nun alle Änderungen δu und δv auf sämlichen Rändern von finiten Elementen bekannt. *CDoedelFiniteElement::CalculateCoefficientDelta* berechnet daraus über die Gleichung (8.3) die Änderungen δc der Koeffizienten für die lokalen Funktionen.

War nun die Änderung der Koeffizienten größer als ein gewisser Schwellenwert, so wird ein weiterer Iterationsschritt durchgeführt, andernfalls bricht der Algorithmus an dieser Stelle mit einer Erfolgsmeldung ab.

Kapitel 9

Numerische Resultate

In diesem Kapitel wird die Anfertigung von numerischen Beispielrechnungen mit Hilfe der Skriptsprache SUNYATA kurz umrissen. Anschließend werden die Resultate für einige ausgewählte Gleichungen diskutiert, insbesondere wird angesprochen, welche Auswirkungen die verschiedenen Wahlmöglichkeiten für das Referenzelement auf die Genauigkeit der Lösung haben. Dabei werden einige Faustregeln erarbeitet, die bei der optimalen Auswahl helfen können, und mit Hilfe der Aussagen aus dem theoretischen Teil begründet.

9.0 Die Skriptsprache

Die Beispiele für die Diplomarbeit sind allesamt mit Hilfe der Interpretersprache SUNYATA angefertigt worden. Ihre Hauptaufgabe besteht darin, eine Schnittstelle zu Funktionen und Objekten aus der Bibliothek NAN bereitzustellen. Ist eine geeignete Implementation vorhanden, so kann auf diese ohne viel Aufwand in einem Skript zurückgegriffen werden, mitsamt der Möglichkeit, wichtige Methoden der Objekte anzuwenden.

Die wohl wichtigste Option der Sprache, welche sie für die Mathematik besonders geeignet macht, ist jedoch die Definition symbolischer Funktionen. Diese arbeiten dank der internen Übersetzung in Pseudocode sehr effizient und können direkt an die Algorithmen der Bibliothek übergeben werden.

Die wesentlichen Aspekte kann man am besten an Hand eines Beispielskripts erläutern, welches die Differentialgleichung aus Abschnitt (9.3) löst. Zeilen, die Kommentare enthalten, beginnen mit einen '!'.

Der erste Teil des Skriptes definiert einige Konstanten, die den Lösungsalgorithmus betreffen: Die Geometrie des Referenzelementes, die Anzahl der Unterteilungen von Omega und die maximale Zahl von Iterationen des Newton-Verfahrens, die im Falle eines nichtlinearen Operators durchgeführt werden.

```
! Beispielskript fuer SUNYATA.

! KONSTANTEN
! Anzahl der Unterteilungen von Omega
!   (= Anzahl der Ebenen des Zerlegungsbaumes ohne die Wurzel)
nTeilungen = 10
! Maximale Anzahl von Iterationen
nIterMax   = 10

! Anzahl der Verbindungspunkte pro Seite
nVP        = 2
! Anzahl der Kollokationspunkte pro Seite
nKP        = 3
```

Anschließend wird das Gesamtgebiet Ω definiert, hier als das Einheitsquadrat $[0,1] \times [0,1]$.

```
! SCHRITT 1: Definition von Omega
! Minimum des Intervalls. Man beachte die Syntax fuer einen
! zweidimensionalen Vektor, bzw. eine Liste, die in geschweiften
! Klammern eingeschlossen wird.
Min = { 0.0,  0.0 }
! Maximum des Intervalls.
```

```
Max = { 1.0,  1.0 }
! Konstruktion eines Objektes vom Typ 'RealInterval'
! mit den Parametern 'Min' und 'Max'. Die Parameterlisten
! fuer Funktionsaufrufe wie dem Konstruktor hier werden
! in eckige Klammern eingeschlossen.
Omega = RealInterval[ Min, Max ]
```

Der nächste Schritt ist der trickreichste: Es werden die Funktionen definiert, welche die Randbedingungen und den Differentialoperator beschreiben. Die Randbedingungen sind hier stets vom Dirichlet-Typ und werden daher einfach durch eine zweiparametrige Funktion $B : \partial\Omega \to \mathbb{R},\ (x, y) \mapsto B(x, y)$ dargestellt. Die Syntax für die Definition einer Funktion in SUNYATA ist dabei

$$'\&' + \text{Anzahl Parameter} + '\{' + \text{Funktionsrumpf} + '\}'.$$

Im Funktionsrumpf werden die einzelnen Parameter mit $'\#n'$ angesprochen, wobei n die Nummer des Parameters bezeichnet.

Der Differentialoperator ist von der Form

$$A(u) = \Delta u + F(u, \diamond),$$

gelöst wird die Differentialgleichung

$$\Delta u + F(u, \diamond) = 0.$$

Die Funktion F muß hier definiert werden. Das Zeichen $'\diamond'$ steht dabei natürlich für den Ort (x, y), daher hängt F von drei Parametern ab. Optional kann man F auch zweiparametrig wählen, in diesem Fall ist der Operator der lineare Laplace-Operator mit einer rein ortsabhängigen Störung. Das Programm ist dann in der Lage, geeignete Optimierungen durchzuführen.

```
! SCHRITT 3: Funktion fuer Randbedingungen
B = &2{ #1^3*log[#1] + #2^3*log[#2] }

! SCHRITT 4: Inhomogenitaet des Laplace-Operators
F = &3{ - #1 - (6*#2-#2^3)*log[#2] - (6*#3-#3^3)*log[#3] - 5*(#2+#3)}
```

Zu guter letzt wird noch das Operatorobjekt mit Hilfe der bisher definierten Parameter konstruiert und die Methode *Solve* aufgerufen, welche den Lösungsalgorithmus durchführt. Dies korrespondiert direkt mit dem Aufruf der entsprechenden C++-Methode *CDoedelOperator::Solve*.

```
! SCHRITT 5: Initialisierung des Operators
TestOp = DoedelOperator[ Omega, F, B ]

! SCHRITT 6: Aufruf des Loesungsalgorithmus
TestOp.Solve[ nIterMax, nTeilungen, nVP, nKP, 0.0, 1,0 ]
```

Die letzten drei Zahlparameter von *Solve* bedeuten dabei in der Reihenfolge der Verwendung:

- Eine optionale zusätzliche Störung der Kollokationsstellen. Falls die Matrix $\begin{bmatrix}\Phi & \Psi_A\end{bmatrix}$ singulär oder sehr schlecht konditioniert ist, kann die Angabe einer kleinen Störung im Bereich von 0.01 bis etwa 0.1, welche die Kollokationsstellen zufällig etwas verschiebt, hier eventuell Abhilfe schaffen.
- Die Verteilung der Verbindungspunkte. Mögliche Angaben sind hier die Zahlenwerte
  0. Äquidistante Verteilung.
  1. Stützstellen der Gauss-Legendre Quadratur.
  2. Zufällige Verteilung.
- Die Verteilung der Kollokationsstellen mit den gleicher Bedeutung der Parameterwerte.

In Verbindung mit den übrigen kommentierten Beispielen, die zusammen mit SUNYATA aus dem Internet bezogen werden können, sollte diese kurze Einführung ausreichen, um eigene Operatoren und Referenzelemente zu definieren. Ansonsten wird für eine ausführliche Beschreibung der Skriptsprache wie üblich auf die Programmdokumentation verwiesen.

9.1 Ausführung von Skripten und Ausgabe

Hat man nach mühevoller Arbeit ein Skript verfaßt, so möchte man es natürlich auch ausführen lassen. Hierzu speichere man es zunächst mit der Dateiendung *.sya* ab und rufe sodann den Kommandozeileninterpreter auf. Dieser meldet sich mit ein paar Informationen zu Version und Lizenzvereinbarungen[1] und erwartet anschließend eine Eingabe. Zur Ausführung einer Skriptdatei dient das Kommando

```
Execute["Script"]
```

die Dateiendung wird dabei weggelassen. War das Skript fehlerfrei, so wird nun die Differentialgleichung gelöst und dabei eine Reihe von Ausgabedateien erzeugt. Die Daten der Lösung werden in einem Format gespeichert, das vom bekannten Ausgabetool GnuPlot[2] direkt weiterverarbeitet werden kann. Im einzelnen finden sich die folgenden Files:

- *Solve.log.* Hier werden nützliche Informationen über einzelne Schritte des Lösungsprozesses zur späteren Analyse abgelegt.
- *Grid.plt.* Dieses kurze Skript kann direkt von GnuPlot ausgeführt werden und zeichnet eine grafische Darstellung des diskretisierten Ausgangsgebietes zusammen mit den Verbindungs- und Kollokationsstellen.
- *Solution.dat.* Die Funktionswerte der Näherungslösung werden in dieser Datei abgelegt. Das Format ist derart gewählt, daß die Lösung durch den Befehl *splot* von GnuPlot als dreidimensionale Grafik dargestellt werden kann. Die Abbildungen aus den nächsten Abschnitten sind zum großen Teil auf diese Weise entstanden.
- *Solution_error.dat.* Im gleichen Format wie in *Solution.dat* werden hier die Werte des absoluten Fehlers der Näherungslösung ausgegeben, falls die exakte Lösung des Problems bekannt ist.

9.2 Helmholtz-Gleichung mit Lösung in $\mathcal{C}^\infty(\overline{\Omega})$

Als erstes einfaches Beispiel wurde die elliptische Helmholtz-Gleichung auf $\Omega = [0,1]^2$ gewählt. In der hier untersuchten Form lautet sie

$$\Delta u \; - \; u \; = \; F,$$

mit einer Funktion $F : \overline{\Omega} \to \mathbb{R}$. Die Randbedingungen und die Funktion F werden dabei derart konstruiert, daß eine exakte analytische Lösung bekannt ist.

Das Ziel ist die Illustration der Genauigkeit der Approximation der Lösung, wobei der Einfluß verschiedener Faktoren untersucht wird. Dazu gehören insbesondere

- Die Anzahl der Kollokations- und Verbindungspunkte, und
- die Lage dieser Punkte auf dem Rand, bzw. im Inneren des Referenzelementes. Dabei werden als mögliche verschiedene Verteilungen äquidistante Punkte, die Stellen der Gauss-Legendre-Quadratur und zufällige Positionen untersucht.
- Die Anzahl der Unterteilungen des Ausgangsgebietes, d.h. der Parameter h der Diskretisierung,
- der Grad der Glattheit der exakten Lösung, welche nach den Untersuchungen in Kapitel 3 auch die Genauigkeit der Interpolation beeinflußt.

Die Auswahl fiel unter anderem deshalb auf die Helmholtz-Gleichung, weil dadurch Vergleiche mit einem Differenzenverfahren sehr einfach möglich sind, da eine Implementation für exakt diesen Operator auf einer Webseite abrufbar ist[3].

Als erstes wird eine glatte Lösung untersucht. Die Funktion F ist gegeben durch

$$\begin{aligned} F(x,y) \; := \; \exp(x^2-y^2) \cdot [2x^2 &- 2x - 2y - xy - 3x^2y + 4x^3y \\ &- 4x^4y + 2y^2 + 5xy^2 - x^2y^2 - 4x^3y^2 \\ &+ 4x^4y^2 + 4xy^3 - 4x^2y^3 - 4xy^4 + 4x^2y^4], \end{aligned}$$

[1] Das gesamte Paket wird unter der GNU General Public License als freie Software ('OpenSource') vertrieben.
[2] www.gnuplot.org
[3] numawww.mathematik.tu-darmstadt.de/numerik/pdgl/helmholtz.html

diese ist gerade so gemacht, daß sie zu den Randbedingungen

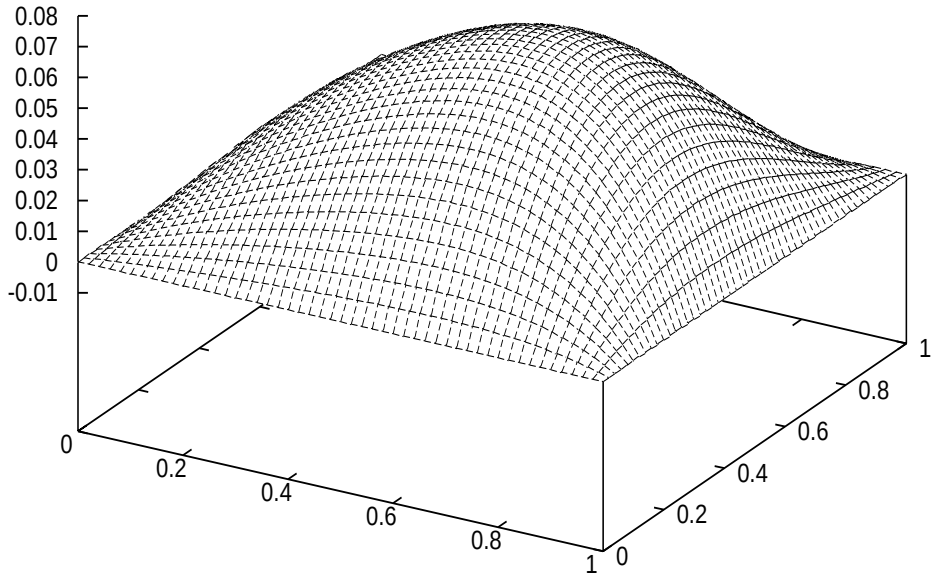
$$B(x,y) \;:=\; x(1-x)\cdot y(1-y)\cdot \exp(x^2-y^2)$$

'paßt' und die exakte Lösung der dadurch definierten Differentialgleichung einfach ebenfalls

$$u_0(x,y) \;:=\; x(1-x)\cdot y(1-y)\cdot \exp(x^2-y^2)$$

lautet. Dies ist offenbar eine glatte Funktion auf $\overline{\Omega}$.

Abbildung 1: Glatte Lösung



$$u_0(x,y) \;:=\; x(1-x)\cdot y(1-y)\cdot \exp(x^2-y^2)$$

Es folgt nun eine Tabelle, in der für verschiedene Geometrien (L,N) des Referenzelementes der maximale Fehler Δ_h der Näherungslösung auf den Verbindungspunkten in Abhängigkeit von der Feinheit des für die Diskretisierung verwendeten Gitters aufgetragen ist. L bedeutet die Anzahl der Verbindungspunkte auf jeder Seite des Referenzelementes, N die Anzahl der Kollokationspunkte. Insgesamt hat ω^{ref} also $M = 4\cdot L$ Verbindungsstellen, die Dimension des lokalen Raums von Funktionen beträgt dann $K = N + M = 4\cdot L + N$ auf jedem der F finiten Elemente. Zu lösen ist daher jeweils ein System für insgesamt $F\cdot(4\cdot L + N)$ Gleichungen.

Der Parameter h errechnet sich aus der angegebenen Größe $F = n \times n$ des Gitters gemäß $h = 1/n$. Ebenfalls eingetragen ist eine geschätzte Konvergenzordnung k, die mit den Daten verträglich ist, d.h. der Fehler ist näherungsweise $\mathcal{O}(h^k)$. Dabei wurde k bestimmt nach der folgenden Faustregel: Bilde jeweils

$$\log_2\left(\frac{\Delta_h}{\Delta_{h/2}}\right)$$

für alle N und nimm den Durchschnitt, auf Halbe abgerundet, als Schätzung. k_T ist die von Theorem 5.8 vorhergesagte minimale Konvergenzordnung $k_T := \min(L, R-1)$ ohne Berücksichtigung der Ordnung Q der Quadraturformeln, für die kein exaktes Resultat bekannt ist.

Hier ist zunächst derjenige Fall dargestellt, welcher sich zumeist als am günstigsten herausgestellt hat: Eine Gauss-Legendre-Verteilung der Verbindungspunkte gepaart mit einer äquidistanten Verteilung der Kollokationsstellen.

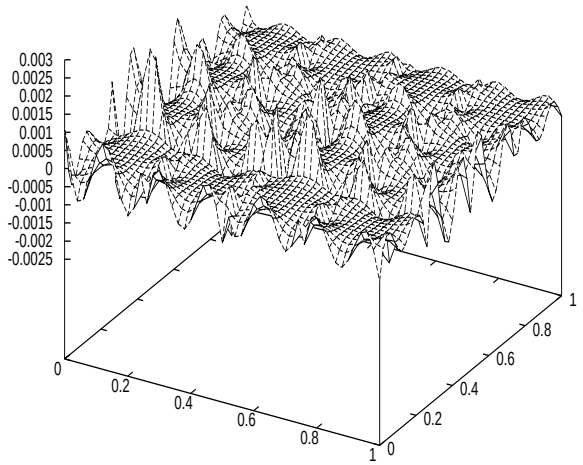
Verbindungspunkte: Gauss-Legendre
Kollokationspunkte: Äquidistant

R.E.	*Maximaler Fehler bei einem Gitter von*							
(L,N)	2×2	4×4	8×8	16×16	32×32	64×64	k	k_T
$(2,9)$	4.924e-03	5.060e-05	2.684e-06	1.262e-07	8.007e-09	2.542e-10	4.5	2
$(3,16)$	1.007e-03	6.263e-02	1.416e-04	2.727e-04	4.245e-06	6.806e-08	n/a	3
$(3,36)$	3.361e-05	5.946e-07	7.686e-09	1.068e-10	2.591e-12	2.169e-14	5.5	3
$(4,36)$	2.125e-03	4.127e-05	4.047e-06	9.231e-08	2.430e-09	3.400e-11	5	4
$(4,49)$	1.854e-06	1.973e-08	8.118e-11	4.128e-13	2.307e-15	2.456e-16	6.5	4

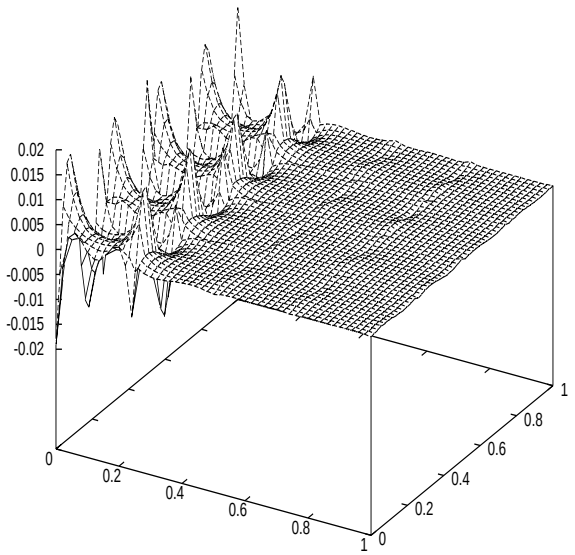
Es springen sofort zwei Tatsachen ins Auge:

- Bei den Elementen $(3,16)$ und $(4,36)$ ist die Konvergenz schlechter als beim jeweils kleineren Typ mit weniger Verbindungspunkten. Der Grund dafür ist vermutlich im Verfahren *Nested Dissection* zu suchen: Bei bestimmten ungünstigen Konstellationen von Verbindungs- und Kollokationsstellen sind die lokalen Gleichungssysteme (8.1) oder die Matrix B aus (8.6) sehr schlecht konditioniert und produzieren große Fehler. Dies ist zum Beispiel bei diesen beiden Elementen häufig der Fall, und wird in späteren Tabellen noch augenfälliger: Der Fehler schnellt dann plötzlich um mehrere Größenordnungen nach oben, obwohl h verkleinert wird! Der Algorithmus versucht aber, derart schlecht konditionierte Systeme zu erkennen, und gibt in solchen Fällen eine Warnung aus. Abbildung 2 zeigt das typische Verhalten des Fehlers in einer solchen Situation, wo ein einzelnes finites Element den Gesamtfehler vervielfacht. Als Konsequenz kann man im folgenden die Ergebnisse dieser beiden Elementtypen eigentlich ignorieren, da sie nicht sehr zuverlässig sind.
- Bei den anderen Elementen, bei denen der Algorithmus tadellos funktioniert, stellt sich heraus, daß die in der Praxis erzielte Konvergenzordnung wesentlich besser ist, als von der Theorie vorausgesagt, und eher im Bereich um $L+5/2$ zu liegen scheint. Entweder kann also die Abschätzung 4.5 über den Fehler bei den Randintegralen noch wesentlich verbessert werden, oder es gibt noch einen anderen ganz anderen Beweisgang, der zum Ziel führt. Es liegt aber wegen der späteren Experimente mit äquidistant verteilten Randpunkten sehr nahe, daß der Quadraturfehler auf dem Rand in irgendeiner Weise dabei eine Rolle spielt.

Abbildung 2: Verhalten des Fehlers



Im Normalfall: Gleichmäßige Verteilung über die finiten Elemente



Bei Entartung durch singuläre lokale Gleichungssysteme: In einzelnen finiten Elementen konzentriert

Geht man bei den Kollokationsstellen ebenfalls zu einer Gauss-Legendre-Verteilung über, so verbessert sich weder die Konvergenzordnung noch der absolute Fehler signifikant. Die erforderliche Mindestanzahl für die Kollokationsstellen scheint offenbar zu garantieren, daß auch bei äquidistanter Verteilung die entstehende Quadraturformel für das Gesamtgebiet Ω schon derart gut ist, daß der limitierende Faktor die Randfehler sind.

Verbindungspunkte: Gauss-Legendre
Kollokationspunkte: Gauss-Legendre

R.E.	*Maximaler Fehler bei einem Gitter von*							
(L, N)	2×2	4×4	8×8	16×16	32×32	64×64	k	k_T
$(2, 9)$	4.178e-04	4.209e-05	2.540e-06	1.177e-07	4.680e-09	1.406e-09	4.5	2
$(3, 16)$	3.313e-02	1.243e-03	2.820e-04	5.832e-06	3.879e-07	1.047e-07	3	3
$(3, 36)$	5.767e-05	9.940e-07	6.631e-09	8.566e-11	1.358e-12	3.471e-14	5.5	3
$(4, 36)$	2.064e-03	1.778e-04	9.247e-07	2.657e-09	3.345e-03	8.841e-13	n/a	4
$(4, 49)$	1.648e-06	1.412e-08	9.566e-11	4.721e-13	3.816e-15	5.482e-15	6.5	4

Sogar eine völlig zufällige Verteilung der Kollokationsstellen im Referenzelement bei gleichbleibender Anzahl ändert nichts Wesentliches am Resultat, solange die lokalen Gleichungssysteme gut konditioniert bleiben. Allerdings werden die Resultate ungleichmäßiger, so daß einer wohldefinierten Verteilung schon der Vorzug gegeben werden sollte.

Verbindungspunkte: Gauss-Legendre
Kollokationspunkte: Zufällig

R.E.	*Maximaler Fehler bei einem Gitter von*							
(L, N)	2×2	4×4	8×8	16×16	32×32	64×64	k	k_T
$(2, 9)$	1.253e-01	4.056e-04	7.773e-06	6.128e-06	6.055e-07	6.305e-07	4	2
$(3, 16)$	2.193e-03	5.373e-04	2.092e-05	2.152e-04	3.280e-08	4.959e-01	n/a	3
$(3, 36)$	3.812e-05	8.488e-07	1.525e-08	1.157e-09	4.008e-13	1.192e-14	5.5	3
$(4, 36)$	7.761e-04	2.973e-04	5.880e-07	7.000e-09	9.005e-03	2.153e+05	n/a	4
$(4, 49)$	2.203e-06	6.614e-07	1.193e-08	2.807e-10	6.121e-14	3.924e-14	6	4

Ändert man aber im Gegenzug die Randpunkte auf eine äquidistante Verteilung, so führt dies stets zu einer Verschlechterung der Konvergenz, die teilweise recht deutlich ausfällt. Die Fehler in den Randintegralen scheinen bei einer genügend glatten Lösung die entscheidende limitierende Rolle zu spielen. Für diesen Fall wurde keine optimierte Version von 4.4 ausgearbeitet, so daß die Theorie auch nur $k \geq 1$ voraussagt. Wesentlich besser als diese mickrige Vorhersage sind die Resultate zwar, trotzdem ist die Verwendung von äquidistant verteilten Randpunkten nicht empfehlenswert und wurde in dieser Arbeit auch deshalb nicht theoretisch weiterverfolgt.

Verbindungspunkte: Äquidistant
Kollokationspunkte: Gauss-Legendre

R.E.	*Maximaler Fehler bei einem Gitter von*						
(L, N)	2×2	4×4	8×8	16×16	32×32	64×64	k
$(2, 9)$	4.601e-03	1.212e-03	2.938e-04	7.309e-05	1.824e-05	4.556e-06	2
$(3, 16)$	4.112e-02	5.933e-03	1.056e-03	3.522e-04	2.763e-06	1.153e-07	3.5
$(3, 36)$	3.007e-03	5.541e-05	7.684e-07	1.079e-08	1.619e-10	1.637e-11	5
$(4, 36)$	2.983e-02	1.272e-03	2.325e-05	8.716e-07	2.328e-08	3.521e-10	5
$(4, 49)$	4.430e-05	7.315e-07	9.721e-09	1.474e-10	2.284e-12	4.425e-14	5.5

9.3 Helmholtz-Gleichung mit Lösung in $\mathcal{C}^2(\overline{\Omega})$

In diesem Abschnitt soll noch einmal derselbe Operator untersucht werden, jedoch dieses Mal mit einer Inhomogenität F, welche eine Lösung liefert, die nur in $\mathcal{C}^2(\overline{\Omega})$ liegt.

Hier ist F definiert als

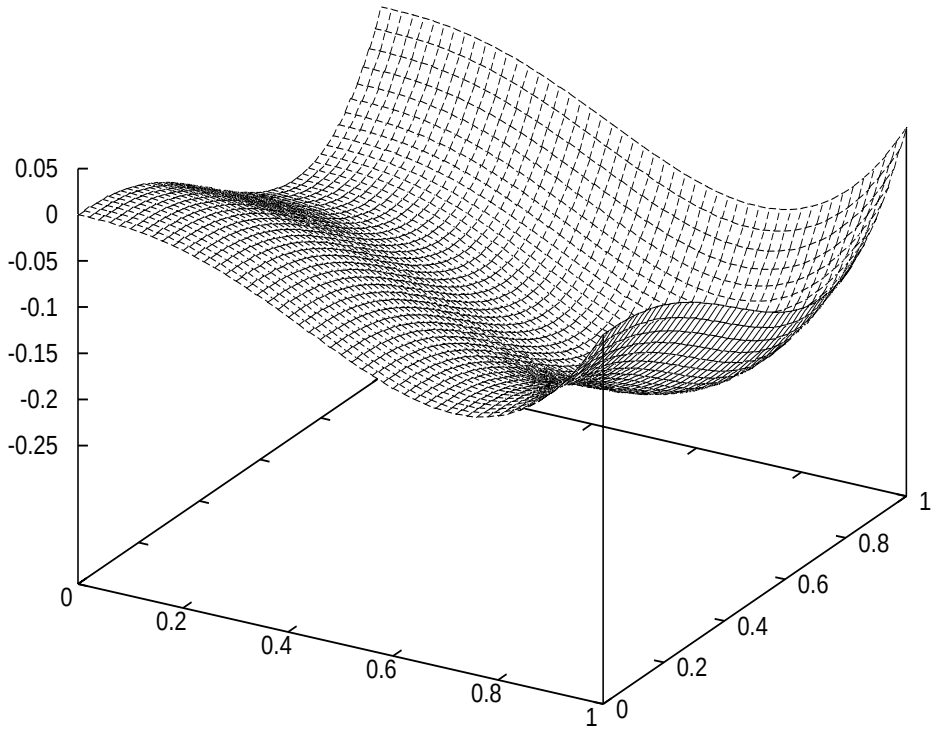
$$F(x,y) \;:=\; (6x - x^3) \cdot \log(x) \;+\; (6y - y^3) \cdot \log(y) \;+\; 5(x+y),$$

Randbedingungen und exakte Lösung lauten

$$B(x,y) \;:=\; u_0(x,y) \;:=\; x^3 \cdot \log(x) \;+\; y^3 \cdot \log(y).$$

Die Funktion B als Abbildung $\partial\Omega \to \mathbb{R}$, beziehungsweise u_0 eingeschränkt auf $\partial\Omega$ liegt sogar in $\mathcal{C}^3(\partial\Omega)$.

Abbildung 3: Lösung in $\mathcal{C}^2(\overline{\Omega})$



$$u_0(x,y) \;:=\; x^3 \cdot \log(x) \;+\; y^3 \cdot \log(y)$$

Die folgende Tabelle enthält wieder die Daten der Fehler auf den Verbindungspunkten bei der gleichen Auswahl von Referenzelementen und Unterteilungen. Wegen der Sobolev-Ungleichung A.1 liegt u_0 jetzt nicht mehr in $H^5(\Omega)$. Sicherlich ist aber in $u_0 \in H^3(\Omega)$, daher wurde $R = 3$ für die theoretische Vorhersage k_T von k angesetzt.

Verbindungspunkte: Gauss-Legendre
Kollokationspunkte: Äquidistant

R.E.	*Maximaler Fehler bei einem Gitter von*							
(L, N)	2×2	4×4	8×8	16×16	32×32	64×64	k	k_T
$(2, 9)$	4.790e-04	1.122e-04	1.776e-05	2.521e-06	3.444e-07	4.030e-08	2.5	2
$(3, 16)$	1.254e-01	1.193e-01	2.918e-04	1.144e-01	1.300e-04	1.052e-06	n/a	2
$(3, 36)$	9.471e-04	8.345e-05	8.041e-05	1.087e-06	8.535e-07	3.082e-08	2.5	2
$(4, 36)$	2.617e-02	1.079e-03	1.014e-03	1.195e-04	3.824e-03	8.685e-07	n/a	2
$(4, 49)$	1.646e-04	6.384e-05	3.487e-06	1.568e-07	5.832e-08	2.271e-09	2.5	2

Ein Vergleich mit den Daten der ersten Tabelle führt zu dem erwarteten Resultat: Die schlechte Interpolation ist hier der limitierende Faktor für die Qualität der Näherungslösung. Die Konvergenzordnung ist auch bei einer Erhöhung der Anzahl der Verbindungspunkte oder Kollokationsstellen durchweg maximale nur etwa 2.5 und liegt daher zwar leicht über dem von der Theorie vorausgesagten Ergebnis, ist aber qualitativ damit konsistent. Vermutlich ist die Lösung doch noch ein wenig glatter als pessimistischerweise zunächst angenommen.

Die anderen Fälle verhalten sich hier analog wie im ersten Abschnitt. Auffällig aber erwartet ist lediglich, daß es nun beim Übergang zu einer äquidistanten Verteilung der Randpunkte nicht zu einer vergleichbaren Verschlechterung der Konvergenz kommt. Die naheliegende Erklärung ist, daß die mangelhafte Interpolation bereits den Boden des Abgrundes darstellt und die Lage auch durch ungüngstige Wahl der Punkte nicht mehr wesentlich verschlimmert werden kann.

9.4 Fazit: Auswahl des Referenzelementes

Aus den oben aufgelisteten Resultaten, die repräsentativ für diverse am Computer durchgeführte Experimente sind, lassen sich verschiedene Richtlinien ableiten, wie man das Referenzelement günstig wählen kann.

- Über die Lage der Verbindungspunkte braucht wohl nicht diskutiert zu werden, eine andere Wahl als die Gauss-Legendre-Punkte wäre verschenkte Genauigkeit.

- Die Lage der Kollokationsstellen kann hingegen eher in Frage gestellt werden. Eine von vorneherein äquidistante Verteilung macht insofern Sinn, als daß dann die Kollokationsstellen gleichmäßig im Referenzelement verteilt sind, kann aber auch leicht einmal ein singuläres lokales Gleichungssystem für die Funktionskoeffizienten liefern. Eine zufällige Verteilung behebt dieses Manko zwar, jedoch kann man eventuell Pech haben, und die Kollokationsstellen konzentrieren sich in einem bestimmten Bereich des Referenzelementes, was auch wenig wünschenswert ist. Bewährt hat sich deswegen insbesondere eine zunächst äquidistante Verteilung, die einer kleinen zufälligen Störung unterworfen wird.

- Eine Erhöhung der Anzahl der Kollokationsstellen bei gleichbleibender Anzahl an Verbindungspunkten erhöht die gesamte Rechenzeit nur unwesentlich. Andererseits ist die Genauigkeit der verwendeten Datentypen nicht mehr ausreichend, wenn der Grad der lokalen Polynome zu hoch wird. Die in den Tabellen aufgeführten Elementtypen stellen jeweils die Grenze dar, danach werden die lokalen Gleichungssysteme stets singulär, weil aufgrund von Rundungsfehlern die Terme höherer Ordnung in den lokalen Polynomen nicht mehr korrekt ausgewertet werden können.

  Hat man sich daher für eine bestimmte Anzahl an Verbindungspunkten entschieden, so sollte man die maximal mögliche Anzahl an Kollokationsstellen wählen, die laut der Tabellen noch funktionieren sollte. Übrigens fand sich kein Referenzelement mit mehr als vier Verbindungspunkten auf jeder Seite, für das der Algorithmus durchlief, ebenfalls weil dann wegen der geforderten Mindestdimension der Grad der Polynome zu hoch wird.

Bezogen auf das vorliegende Programmsystem führt dies zu folgender Empfehlung:

(*i*) *Man wähle das Element* $(2, 9)$*, wenn ein schnelles vorläufiges Ergebnis gewünscht ist, es liefert bei sehr kurzen Rechenzeiten ein vernünftiges Maß an Genauigkeit.*

(*ii*) *Das Referenzelement mit dem besten 'Preis-Leistungs-Verhältnis' ist das Element* $(3, 36)$*, welches für Rechnungen mit hoher Genauigkeit bevorzugt werden sollte.*

(*iii*) *Sein großer Bruder* $(4, 49)$ *liefert zwar zumeist leicht bessere Ergebnisse, allerdings bei einem unverhältnismäßig starken Anstieg der Rechenzeit. Es sollte daher das Element der Wahl sein, wenn genauestmögliche Daten gefordert sind und Rechenzeit nur eine Nebenrolle spielt.*

Teil III

Anhänge

Anhang A

Sobolev-Räume

Natürlich soll in diesem Anhang beileibe keine Einführung in die Theorie der Sobolev-Räume gegeben werden. Das Ziel ist vielmehr, daß die im Verlauf der Arbeit herangezogenen Theoreme zusammen mit einer Quelle, wo man ihren Beweis nachschlagen kann, zitiert werden. Lediglich die Aussagen und Abschätzungen, die nicht zum absoluten Standard gerechnet werden können, werden kurz bewiesen.

A.0 Einbettungs- und Dichtheitssätze

Hier werden die wesentlichen Theoreme darüber zusammengetragen, welche Funktionenräume wie in welchen Sobolev-Räumen liegen, bzw. wie die Sobolev-Räume untereinander in Beziehung stehen. Die erste Aussage ist die berühmte *Sobolev-Ungleichung*, welche es erlaubt, Funktionen in $W_p^k(\Omega)$ mit genügend großem k als stetig oder sogar eine gewisse Anzahl m von Malen als stetig differenzierbar aufzufassen.

A.1 Theorem. *Sei Ω ein d-dimensionales Gebiet mit Lipschitz-stetigem Rand. Seien $m, k \in \mathbb{N}$ mit $0 \le m < k$ und $0 \le p \le \infty$, so daß*

$$\begin{aligned} k - m &\ge d \qquad \text{falls } p = 1 \\ k - m &> d/p \quad \text{falls } p > 1. \end{aligned}$$

Dann existiert eine Konstante $C > 0$, so daß für alle $u \in W_p^k(\Omega)$:

$$\|u\|_{W_\infty^m(\Omega)} \;\le\; C\, \|u\|_{W_p^k(\Omega)}\,.$$

Insbesondere liegt eine $\mathcal{C}^m$-Funktion in der L^p-Äquivalenzklasse von u.

Beweis. Siehe [BS], Korollar 1.4.7. □

Das zweite wichtige Theorem, genannt *Spurtheorem*, erlaubt es, Funktionen aus $W_p^1(\Omega)$ auch auf $\partial\Omega$ als wohldefinierte L^p-Funktionen anzusehen.

A.2 Theorem. *Sei Ω ein Gebiet mit Lipschitz-stetigem Rand und $1 \le p \le \infty$. Dann existiert eine Konstante $C > 0$, so daß für alle $u \in W_p^1(\Omega)$:*

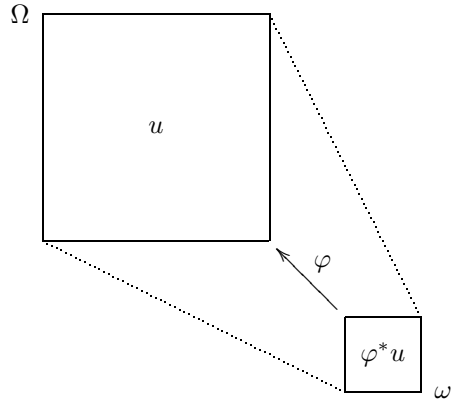
$$\|u\|_{L^p(\partial\Omega)} \;\le\; C\, \|u\|_{L^p(\Omega)}^{1-1/p}\, \|u\|_{W_p^1(\Omega)}^{1/p}\,.$$

Beweis. Siehe [BS], Theorem 1.6.6. □

A.1 Affine Transformationen

Dieser Abschnitt untersucht das Verhalten von Sobolev-Normen unter affinen Transformationen. Brauchbar sind die Aussagen vor allem beim Übergang vom Referenzelement einer Zerlegung zu einem konkreten skalierten Element und umgekehrt.

Was passiert mit Normen von Funktionen bei Skalierung?



A.3 Lemma. *Sei $\Omega \subset \mathbb{R}^d$ ein Gebiet, $0 < h \leq 1$ und $\varphi : \omega \to \Omega$ eine affine Abbildung, deren linearer Teil eine Skalierung um den Faktor $1/h$ darstellt, d.h. $\varphi(x) = M \cdot x + b$ mit*

$$M \;=\; \begin{bmatrix} \frac{1}{h} & 0 & \dots & 0 \\ 0 & \frac{1}{h} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{h} \end{bmatrix} \in \mathbb{R}^{d\times d}.$$

Weiter sei $u \in W_p^s(\Omega)$ und $\alpha \in \mathbb{N}^d$ ein Multi-Index mit $|\alpha| \leq s$. Dann gilt

$$\left(\int_\omega |D^\alpha(\varphi^* u)|^p\right)^{\frac{1}{p}} \;=\; h^{\frac{d}{p}-|\alpha|}\left(\int_\Omega |D^\alpha u|^p\right)^{\frac{1}{p}},$$

$$\textit{insbesondere ist } \|\varphi^* u\|_{W_p^s(\omega)} \;\leq\; h^{\frac{d}{p}-s}\cdot\|u\|_{W_p^s(\Omega)}.$$

Beweis. Durch Ausrechnen der linken Seite:

$$\begin{aligned}
\int_\omega |D^\alpha(\varphi^* u)|^p \;&=\; \int_\omega \frac{1}{h^{|\alpha|p}}\,|D^\alpha u\circ\varphi|^p && \text{(Kettenregel)}\\
&=\; \frac{1}{\det M}\cdot\frac{1}{h^{|\alpha|p}}\int_\omega |D^\alpha u|^p\circ\varphi\cdot|\det(D\varphi)| && \text{(Erweitern)}\\
&=\; h^{d-|\alpha|p}\int_\Omega |D^\alpha u|. && \text{(Trafo-Formel)}
\end{aligned}$$

Die Transformationsformel schlage man z.B. in [Ba] nach (Satz 19.4). Obige Formel ist damit bewiesen, der Zusatz ergibt sich direkt aus der Definition der Sobolev-Normen im Hinblick auf $h \leq 1$, da die Abschätzung dann am schwächsten wird, wenn $|\alpha|$ maximal ist. □

A.2 Aussagen für diskrete Normen

In diesem Abschnitt steht im Vordergrund, inwieweit sich die wohlbekannten Gleichungen und Ungleichungen für die Normen auf Sobolev-Räumen auf die diskretisierenden Räume V^h übertragen. Untersucht werden also die durch Zerlegung des Gebietes Ω induzierten Normen, wie sie im Kapitel 1 eingeführt worden sind. Dies geschieht hier im Anhang, um die Argumentationsstruktur des Haupttextes nicht ständig durch kleine Nebenrechnungen unterbrechen zu müssen.

Das erste Lemma dient mehr dazu, die häufig verwendete Aussage einmal explizit festzuhalten, es folgt unmittelbar aus der Hölderschen Ungleichung für Summen und der Definition der Normen auf H^h.

A.4 Lemma. *Für beliebige Funktionen $u^h, v^h \in H^h$, natürliche Zahlen $l, m \in \mathbb{N}$ sowie $1 \leq p, q \leq \infty$ mit $\frac{1}{p}+\frac{1}{q}=1$ gilt:*

$$\sum_{n=1}^{F^h} \left|u^h\right|_{W_p^l(\Omega_n^h)} \left|v^h\right|_{W_q^m(\Omega_h^h)} \;\leq\; \left|u^h\right|_{h,W_p^l}\cdot\left|v^h\right|_{h,W_q^m},$$

$$\textit{bzw. } \sum_{n=1}^{F^h} \left\|u^h\right\|_{W_p^l(\Omega_n^h)} \left\|v^h\right\|_{W_q^m(\Omega_h^h)} \;\leq\; \left\|u^h\right\|_{h,W_p^l}\cdot\left\|v^h\right\|_{h,W_q^m}$$

Beweis. Laut der Hölderschen Ungleichung gilt für beliebige $x, y \in \mathbb{R}^{F^h}$:

$$\sum_{n=1}^{F^h} |x_n \cdot y_n| \;\leq\; \left(\sum_{n=1}^{F^h} |x_n|^p\right)^{1/p} \cdot \left(\sum_{n=1}^{F^h} |y_n|^q\right)^{1/q}.$$

Die Aussage folgt dann unmittelbar durch Einsetzen der entsprechenden Normen. □

A.3 Existenz bestimmter Funktionen

Häufig werden in der Arbeit Funktionen mit ganz bestimmten Eigenschaften benötigt. Deren Existenz wird hier nachgewiesen.

Für die Kollokation benötigt man zunächst Funktionen, die an einer bestimmten Stelle 1 sind und beliebig kleinen Träger besitzen, deren H^1-Norm aber trotzdem beschränkt bleibt.

A.5 Lemma. *Es gibt eine Konstante C, so daß für beliebig vorgegebenes $x \in \mathbb{R}^D$ und $\epsilon > 0$ eine glatte Funktionen $\theta_{x,\epsilon} : \mathbb{R}^D \to [0,1]$ existiert mit*

(i) $\theta_{x,\epsilon}(x) = 1$,

(ii) $\operatorname{supp} \theta_{x,\epsilon} \subset B_\epsilon(x)$ *und*

(iii) $\|\theta_{x,\epsilon}\|_{H^1(\mathbb{R}^D)} \leq C$.

Beweis. Sei $x \in \Omega$, $\epsilon > 0$. Die Funktion $\theta_{x,\epsilon}$ mit den gewünschten Eigenschaften soll konstruiert werden. Wähle dafür zunächst $\delta > 0$, so daß das Intervall $[x-\delta, x+\delta]^D \subset B_\epsilon(x)$, und sodann eine der sattsam bekannten reellen 'Hubbelfunktionen' $\vartheta \in \mathcal{C}^\infty(\mathbb{R}, [0,1])$ mit

$$\vartheta(0) = 1 \text{ und } \operatorname{supp} \vartheta \subset (-1,1).$$

Definiere $\theta \in \mathcal{C}^\infty(\mathbb{R}^D, [0,1])$ durch

$$\theta(\tilde{z}) := \prod_{d=1}^{D} \vartheta(\tilde{z}_d).$$

Dann gilt $\theta(0) = 1$ und $\operatorname{supp} \theta \in (-1,1)^D$, außerdem

$$\begin{aligned} C \;:=\; \|\theta\|_{H^1(\mathbb{R}^D)} \;&\leq\; \prod_{d=1}^{D} \|\vartheta\|_{H^1(\mathbb{R}^D)} \qquad \text{(Fubini)} \\ &=\; \|\vartheta\|_{H^1(\mathbb{R}^D)}^D . \end{aligned}$$

Die Funktion

$$\theta_{x,\epsilon} \;:=\; \theta\left(\frac{\diamond - x}{\delta}\right)$$

hat dann offensichtlich die Eigenschaften (i) und (ii). (iii) sieht man folgendermaßen ein:

$$\begin{aligned} \|\theta_{x,\epsilon}\|_{H^1(\mathbb{R}^D)} \;&:=\; \left\|\theta\left(\frac{\diamond - x}{\delta}\right)\right\|_{H^1(\mathbb{R}^D)} \\ &\leq\; \|\theta(\diamond - x)\|_{H^1(\mathbb{R}^D)} \qquad \text{(Lemma A.3)} \\ &=\; \|\theta\|_{H^1(\mathbb{R}^D)} \qquad \text{(Translationsinvarianz)} \\ &=\; C. \end{aligned}$$

□

Eine Konsequenz ist, daß man stets bestimmte vorgegebene Sprünge in den Funktionswerten beim Übergang über die Verbindungspunkte garantieren kann. Dies wird, wie auch das danach folgende Lemma, für die Konstruktion des beschränkten Interpolationsoperators auf Kollokationsstellen in Theorem 3.14 benötigt.

A.6 Lemma. *Für beliebig vorgegebenes $\epsilon > 0$ und Werte $[u]$ für die Sprünge der Funktionswerte beim Übergang über die Verbindungspunkte existiert eine Funktion $\phi \in L^\infty(\Omega)$ mit $\phi\big|_{\omega_i^h} \in \mathcal{C}^\infty(\omega_i^h)$, so daß*

$$\tilde{\phi}\big|_{Z^h} = 0 \text{ und } [u]_\phi \;=\; [u].$$

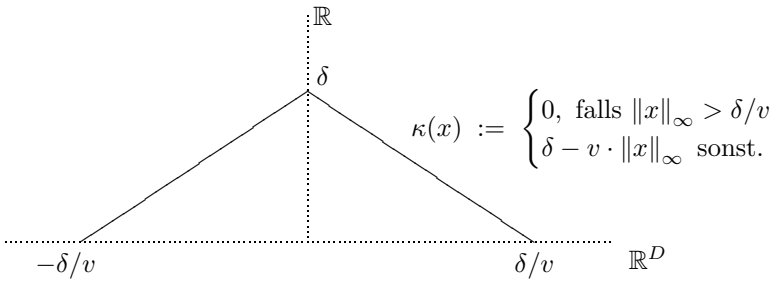
Beweis. Diese Aussage folgt unmittelbar aus dem letzten Lemma, da man sich über die Norm der entstehenden Funktion keine Gedanken machen muß. □

Die Konstruktion entsprechender Funktionen für Sprünge in den Normalenableitungen ist ein wenig komplizierter, da für den Beweis von Theorem 3.14 stärkere Eigenschaften benötigt werden.

A.7 Lemma. *Für beliebig vorgegebenes $\epsilon > 0$ und Werte $[v]$ für die Sprünge der Normalenableitungen beim Übergang über die Verbindungspunkte existiert eine Funktion $\phi \in L^\infty(\Omega)$ mit $\phi\big|_{\omega_i^h} \in W^1_\infty(\omega_i^h)$, so daß*

$$|\phi|_{h,W^1_\infty} \;\le\; \|[v]\|_\infty\,, \;\; \|\phi\|_{L^\infty(\Omega)} \;\le\; \epsilon,\; \phi\big|_{Z^h} = 0 \text{ und } [v]_\phi \;=\; [v].$$

Beweis. Sei $\delta > 0$ und $v \ge 0$ beliebig. Dann beschreibt der Graph der Funktion $\kappa = \kappa_{\delta,v} : \mathbb{R}^D \to \mathbb{R}$ einen Kegel mit Spitze in $(0, \delta)$ über der Kugel $B_{\delta/v}(0)$:



Es gilt $\|\kappa\|_{L^\infty(\mathbb{R}^D)} = \delta$ und $|\kappa|_{h,W^1_\infty} \;=\; v$, außerdem für jedes $\eta \in \mathbb{R}^D$ mit $\|\eta\|_\infty = 1$:

$$\nabla\kappa(0) \cdot \eta = -v.$$

Indem man nun an den Verbindungspunkte Translationen von $\kappa_{\delta,v}$ für geeignete Wahl von δ und v zusammenklebt, kann man die gesuchte Funktion konstruieren. Da die Norm $\|\diamond\|_{L^\infty(\mathbb{R}^D)}$ sowie der Träger von κ durch Verkleinerung von δ beliebig klein gewählt werden kann, ohne daß sich die Ableitungen ändern, folgen alle Behauptungen des Lemmas. □

A.4 Approximation durch Taylorpolynome

Entscheidend für die Interpolationstheorie ist die Approximation von Funktionen in Sobolev-Räumen durch sogenannte gemittelte Taylorpolynome. Es soll natürlich nicht die komplette Theorie hier angerissen werden, sondern nur eines der wichtigsten Theoreme zitiert werden, welches in dieser Arbeit Verwendung findet. Dieses gibt Aufschluß über die Abschätzung des Fehlers bei der Approximation einer Funktion durch ihr Taylorpolynom. Es bezeichnet

$$\rho_{\max} \;:=\; \sup\{\, \rho > 0 \;:\; \Omega \text{ ist sternförmig bezüglich einer Kugel mit Radius } \rho \,\}.$$

A.8 Lemma. (Bramble-Hilbert). *Sei B eine Kugel mit Radius ρ in Ω, so daß Ω bezüglich B sternförmig und $\rho > (1/2)\rho_{max}$ ist. Sei $T^m u$ das über B gemittelte Taylorpolynom der Ordnung m von $u \in W^m_p(\Omega)$, wobei $p \ge 1$. Dann gilt:*

$$|u \,-\, T^m u|_{W^k_p(\Omega)} \;\le\; C_{m,n,\gamma}\,(\operatorname{diam}\Omega)^{m-k}\;|u|_{W^m_p(\Omega)}\,.$$

Beweis. Siehe [BS], Lemma (4.3.8). □

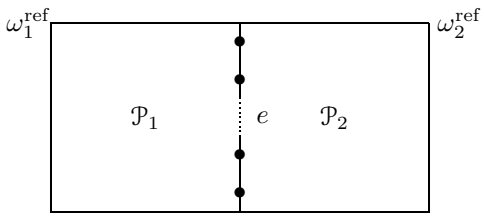
Anhang B

Quadraturfehler

B.0 Eindimensional

Für die Untersuchung der Fehler bei der partiellen Integration ist es erforderlich, eine Abschätzung des Integrals einer lokalen Funktion längs des Randes eines finiten Elementes durch eine Sobolev-Norm der Funktion im inneren der angrenzenden Elemente herzuleiten. Grundlegend dafür ist die folgende technische Abschätzung, die sich Aussagen über Fehler der numerischen Quadratur, das Bramble-Hilbert-Lemma und der Natur der Sache gemäß natürlich das Spurtheorem zunutze macht.

Zwei Referenzelemente mit gemeinsamer Seite



- Punkte in $X^1 \cap X^2$, insgesamt L Stück

B.1 Lemma. *Seien ω_1^{ref} und ω_2^{ref} zwei benachbarte Kopien eines Referenzelementes mit gemeinsamer Seite e, auf der L Verbindungspunkte liegen. Die zugehörigen endlichdimensionalen lokalen Funktionenräume sollen nur aus Polynomen p bestehen derart, daß ihre Restriktion $p\big|_e$ durch eine Quadraturformel, die die L Verbindungspunkte verwendet, exakt integriert wird. Dann existiert eine Konstante C, so daß für beliebige Funktionen $v, u \in \mathcal{P}_1 \oplus \mathcal{P}_2$ mit $v, u\big|_{X_1 \cap X_2} = 0$ und $i = 1, 2$:*

$$\left| \int_e u_i \, [v] \, ds \right| \;\leq\; C \cdot |u_i|_{H^1(\omega_i^{ref})} \left(|v_1|_{H^1(\omega_1^{ref})} \;+\; |v_2|_{H^1(\omega_2^{ref})} \right).$$

Beweis. Nach Voraussetzung ist $[v]$ auf e ein Polynom mit L Nullstellen in den L Verbindungspunkten. Nach Voraussetzung existiert für diese L Punkte eine Quadraturformel, die $[v]$ exakt integriert, daher ist also

$$\int_e [v] \, ds \;=\; 0.$$

Damit folgt für beliebige Konstanten $c_1, c_2 \in \mathbb{R}$:

$$\begin{aligned}
\left| \int_e u_i \, [v] \, ds \right| \;&=\; \left| \int_e (u_i - c_1) \, [v] \, ds \right| \\
&= \left| \int_e (u_i - c_1) \, [v - c_2] \, ds \right| \\
&\leq \|u_i - c_1\|_{L^2(e)} \; \|[v - c_2]\|_{L^2(e)} && \text{(Cauchy-Schwarz)} \\
&\leq C \cdot \|u_i - c_1\|_{H^1(\omega_i^{\text{ref}})} \left(\|v_1 - c_2\|_{H^1(\omega_1^{\text{ref}})} \;+\; \|v_2 - c_2\|_{H^1(\omega_2^{\text{ref}})} \right) && \text{(Spurtheorem A.2)}
\end{aligned}$$

Die Behauptung folgt dann mit dem Bramble-Hilbert-Lemma A.8, indem man für die Konstanten speziell die Taylorpolynome 0-ten Grades der entsprechenden Funktionen wählt. □

B.2 *Bemerkung.* Es stellt sich natürlich sofort die Frage, wie genau denn nun die lokalen Funktionenräume aussehen dürfen, damit das Lemma angewendet werden kann. Dies kann man nur bei genauerer Kenntnis der Gestalt des Referenzelementes beantworten. Geht man einmal von dem einfachen Fall aus, daß das Referenzelement das Einheitsquadrat ist, und die Polynome p aus Monomen $x^a \cdot y^b$ zusammengesetzt sind, so entnimmt man den grundlegenden Aussagen über die numerische Quadratur:

- Sind die Verbindungspunkte äquidistant verteilt, so ist eine zumindest hinreichende Bedingung, daß

$$a, b \leq \begin{cases} L-1 & \text{falls } L \text{ gerade,} \\ L & \text{falls } L \text{ ungerade.} \end{cases}$$

  Siehe dazu auch [S], Diskussion zu den Sätzen 8.4 und 8.5, bzw. [IK].

- Optimaler ist es, falls die Verbindungspunkte gemäß der Gauss-Legendre-Quadratur als die (geeignet skalierten) Nullstellen der Legendre-Polynome gewählt sind. In diesem Fall hat man wesentlich mehr Freiheit, da $a, b \leq 2L-1$ ausreichend ist - bei einer natürlichen Wahl der lokalen Polynome, die zumeist versucht, den Grad zu minimieren, ist die Bedingung im allgemeinen automatisch erfüllt.

Von ähnlichem Kaliber ist die folgende Abschätzung, welche für den Spezialfall dienlich ist, daß u eine genügend glatte Funktion ist. Diese Rolle spielt die exakte Lösung, und mit Hilfe der folgenden Aussage wird dann auch ein besonders wesentlicher Schritt für die Bestimmung der Konvergenzqualität eingeleitet.

B.3 Lemma. *Seien ω_i^h und ω_j^h zwei benachbarte finite Elemente mit gemeinsamer Seite e. Das gemeinsame Referenzelement besitze L Punkte pro Seite, die gemäß den Punkten der Gauss-Legendre-Quadratur verteilt seien. Dann existiert eine von h, i und j unabhängige Konstante C, so daß für $1 \leq k \leq d$ und beliebige Funktionen $u \in H^{L+2}(\Omega)$ und $v^h \in V^h$:*

$$\left| \int_e \partial_k u \, \left[v^h\right] ds \right| \;\leq\; C \cdot h^L \cdot |\partial_k u|_{H^{L+1}(\Omega)} \left(\left\| v_i^h \right\|_h \,+\, \left\| v_j^h \right\|_h \right).$$

Beweis. Der Beweis erfordert eine ähnliche, aber etwas elaboriertere Technik wie der des letzten Lemmas. Er ist ausgeführt in [B], Lemma 6.12. □

B.1 Mehrdimensional

In diesem Abschnitt werden hinreichende Bedingungen für die Existenz genügend guter Quadraturformeln für ganz Ω angegeben, die man für gute Konvergenz benötigt.

B.4 Lemma. *Für das Referenzelement existiere eine Quadraturformel für die Kollokationsstellen, die Polynome vom Grad Q exakt integriert. Dann gibt es eine von h unabhängige Konstante $C > 0$, so daß für alle Funktionen $u, v \in L^\infty(\Omega)$, deren Restriktionen auf finite Elemente sogar in $W_\infty^Q(\omega_i^h)$ liegen, gilt:*

$$\left| \int_\Omega u \cdot v \;-\; Q^h(u \cdot v) \right| \;\leq\; h^{Q+2} \cdot C \, \|u\|_{h,W_\infty^Q} \, \|v\|_{h,W_\infty^Q} .$$

Beweis. Da die Quadraturformeln für die finiten Elemente durch Skalierung aus der Formel für das Referenzelement entstehen, gilt mit einer von h und i unabhängigen Konstanten C' lokal[1]:

$$\left| Q_i^h u \right| \;\leq\; C' \operatorname{meas}\left(\omega_i^h\right) \|u\|_{L^\infty(\omega_i^h)} \text{ für alle } u \in \mathcal{C}(\omega_i^h).$$

Nach Voraussetzung werden nun insbesondere die gemittelten Taylorpolynome $T^Q(u \cdot v)$ auf den ein-

[1] Vergleiche [BS], 8.x.16

zelnen finiten Elementen, die ja von Grad Q sind, exakt integriert. Damit folgt:

$$
\begin{aligned}
&\left| \int_\Omega u \cdot v \; - \; Q^h(u \cdot v) \right| \\
&= \sum_{i=1}^{F^h} \left| \int_{\omega_i^h} u \cdot v \; - \; Q_i^h(u \cdot v) \right| \\
&= \sum_{i=1}^{F^h} \left| \int_{\omega_i^h} u \cdot v - \int_{\omega_i^h} T_i^Q(u \cdot v) \; + \; Q_i^h T_i^Q(u \cdot v) \; - \; Q_i^h(u \cdot v) \right| && \text{(Fundamentaltrick)} \\
&= \sum_{i=1}^{F^h} \left| \int_{\omega_i^h} \left(u \cdot v \; - \; T_i^Q(u \cdot v) \right) \; + \; Q_i^h \left(T_i^Q(u \cdot v) \; - \; u \cdot v \right) \right| && \text{(Linearität)} \\
&\leq \sum_{i=1}^{F^h} (1 + C') \cdot \operatorname{meas}\left(\omega_i^h\right) \left\| u \cdot v - T_i^Q(u \cdot v) \right\|_{L^\infty(\omega_i^h)} && \text{(Integralabschätzung, Voraussetzung)} \\
&\leq h^Q \cdot C \operatorname{meas}\left(\omega_i^h\right) \; |u \cdot v|_{h, W_\infty^Q} && \text{(Bramble-Hilbert-Lemma A.8)} \\
&\leq h^{Q+2} \cdot C \, \|u\|_{h, W_\infty^Q} \; \|v\|_{h, W_\infty^Q} \, .
\end{aligned}
$$

□

B.5 Folgerung. *Es existieren stets Quadraturformeln Q^h, so daß speziell für alle $u, v \in V^h$ gilt:*

$$
\left| \int_\Omega u \cdot v \; - \; Q^h(u \cdot v) \right| \; \leq \; h \cdot C \; \|u\|_h \; \|v\|_h \, .
$$

Beweis. Da das Referenzelement wegen $N \geq M$ mindestens drei Kollokationsstellen aufweisen muß, existiert zumindest eine Quadraturformel, die lineare Funktionen exakt integriert. Die Abschätzung aus Lemma B.4 ist daher mit $Q = 1$ richtig. Die Behauptung folgt nun durch Anwendung der inversen Abschätzung 3.9 auf u und v, wobei jeweils ein h verlorengeht. □

Anhang C

Alle experimentellen Daten

Auf den folgenden Seiten findet sich noch einmal eine ausführlichere Zusammenstellung aller erhobenen Daten. Die Überschriften der einzelnen Spalten bedeuten dabei:

L	Anzahl der Verbindungspunkte auf jeder Seite des (quadratischen) Referenzelementes
N	Die gesamte Zahl der Kollokationspunkte im Inneren des Referenzelementes
n×n	Die Gesamtzahl der finiten Elemente, es ist dann $h = 1/n$.
nVP	Die gesamte Anzahl an Verbindungspunkten einschließlich solcher, die auf dem Rand von Ω liegen
nKP	Die gesamte Anzahl an Kollokationsstellen.
$Zeit$	Die zum Lösen des Gleichungssystems mit Hilfe des Algorithmus 'Nested Dissection' aufgewandte Zeit in Sekunden
Max-F-VP	Der maximale Fehler auf den Verbindungspunkten.
$\emptyset$-F-VP	Der durchschnittliche Fehler auf den Verbindungspunkten.
Max-F-KP	Der maximale Fehler auf den Kollokationspunkten.
$\emptyset$-F-KP	Der durchschnittliche Fehler auf den Kollokationspunkten.

Der durchschnittliche Fehler ist aufgenommen, weil durch einen Vergleich mit dem maximalen Fehler recht gut erkennbar wird, ob ein 'Ausreißer' auf ein schlecht konditioniertes Gleichungssystem 8.1 zurückgeführt werden kann: Ist der Fehler im allgemeinen niedrig, und nur auf einem finiten Element wesentlich höher, so führt dies zu einer großen Diskrepanz zwischen durchschnittlichem und maximalen Fehler, und die Vermutung liegt nahe, daß das lokale Gleichungssystem auf einem oder einigen wenigen Elementen schlecht konditioniert war.

Das entstehende Gleichungssystem hatte nKP Unbekannte entsprechend der Dimension des entstehenden Funktionenraumes V^h. Hinzu kommen die zusätzlichen $2 \times nVP$ Unbekannten für die Funktions- und Ableitungswerte auf den Verbindungsstellen, abzüglich der Anzahl der Punkte auf $\partial\Omega$, die in der Tabelle nicht aufgeführt sind, es sind insgesamt $4n \cdot L$. Die Zeitangaben sind natürlich stark abhängig vom verwendeten Rechner und können daher nur als grobe Richtlinie dienen. Zum Einsatz kam hier ein 1.2GHz Athlon mit 256 MB RAM.

C.0 Helmholtz-Gleichung mit Lösung in $\mathcal{C}^\infty(\overline{\Omega})$

Fall I

Verbindungspunkte: Gauss-Legendre
Kollokationspunkte: Äquidistant

$V\times K$	n×n	nVP	nKP	Zeit	Max-F-VP	∅-F-VP	Max-F-KP	∅-F-KP
2×9	2×2	24	36	0	4.925e-004	7.593e-005	4.923e-003	1.090e-003
2×9	4×4	80	144	0	5.060e-005	8.741e-006	5.599e-004	7.686e-005
2×9	8×8	288	576	0	2.684e-006	6.409e-007	4.451e-005	5.048e-006
2×9	16×16	1088	2304	1	1.262e-007	3.258e-008	3.208e-006	3.076e-007
2×9	32×32	4224	9216	5	8.007e-009	1.719e-009	2.141e-007	1.873e-008
2×9	64×64	16640	36864	36	3.812e-010	8.806e-011	1.372e-008	1.143e-009
3×16	2×2	36	64	0	3.510e-002	3.793e-003	6.650e-002	1.783e-002
3×16	4×4	120	256	0	1.970e-003	2.143e-004	2.810e-003	5.156e-004
3×16	8×8	432	1024	0	1.425e-004	1.996e-005	2.570e-004	4.081e-005
3×16	16×16	1632	4096	2	6.335e-006	1.600e-006	7.987e-006	1.845e-006
3×16	32×32	6336	16384	15	7.532e-007	2.172e-007	7.935e-007	2.256e-007
3×16	64×64	24960	65536	117	4.243e-008	1.384e-008	4.297e-008	1.406e-008
3×36	2×2	36	144	1	5.403e-005	5.548e-006	1.225e-003	2.780e-004
3×36	4×4	120	576	1	7.983e-007	1.094e-007	6.111e-005	5.358e-006
3×36	8×8	432	2304	0	7.475e-009	1.820e-009	1.580e-006	8.936e-008
3×36	16×16	1632	9216	2	1.084e-010	3.249e-011	3.179e-008	1.440e-009
3×36	32×32	6336	36864	16	1.137e-012	3.644e-013	5.577e-010	2.250e-011
3×36	64×64	24960	147456	125	2.551e-014	8.871e-015	9.053e-012	3.539e-013
4×36	2×2	48	144	0	1.551e-001	1.431e-002	8.773e-001	1.435e-001
4×36	4×4	160	576	1	2.604e-004	3.176e-005	1.418e-003	2.310e-004
4×36	8×8	576	2304	1	3.687e-006	9.915e-007	7.148e-006	1.241e-006
4×36	16×16	2176	9216	5	1.934e-007	2.742e-008	2.271e-006	4.257e-007
4×36	32×32	8448	36864	34	2.076e-009	6.945e-010	2.152e-009	7.185e-010
4×36	64×64	33280	147456	359	8.736e-010	3.479e-010	8.885e-010	3.534e-010
4×49	2×2	48	196	1	1.765e-006	2.462e-007	2.165e-005	2.255e-006
4×49	4×4	160	784	0	1.500e-008	1.776e-009	1.512e-007	1.152e-008
4×49	8×8	576	3136	1	9.155e-011	1.011e-011	8.773e-010	4.384e-011
4×49	16×16	2176	12544	5	4.018e-013	5.264e-014	4.388e-012	1.973e-013
4×49	32×32	8448	50176	36	4.594e-015	1.699e-015	1.823e-014	2.198e-015
4×49	64×64	33280	200704	362	5.440e-015	2.169e-015	5.419e-015	2.204e-015

Fall II

Verbindungspunkte: Äquidistant
Kollokationspunkte: Äquidistant

$L\times N$	n×n	nVP	nKP	Zeit	Max-F-VP	∅-F-VP	Max-F-KP	∅-F-KP
2×9	2×2	24	36	0	4.026e-003	8.342e-004	4.564e-003	1.244e-003
2×9	4×4	80	144	1	1.174e-003	3.279e-004	1.228e-003	3.933e-004
2×9	8×8	288	576	0	2.915e-004	9.600e-005	2.927e-004	1.068e-004
2×9	16×16	1088	2304	0	7.290e-005	2.571e-005	7.304e-005	2.724e-005
2×9	32×32	4224	9216	5	1.823e-005	6.646e-006	1.824e-005	6.848e-006
2×9	64×64	16640	36864	37	4.555e-006	1.688e-006	4.556e-006	1.714e-006
3×16	2×2	36	64	0	6.493e-002	8.815e-003	1.187e-001	3.285e-002
3×16	4×4	120	256	0	9.532e-002	1.597e-002	1.051e-001	2.234e-002
3×16	8×8	432	1024	0	6.561e-004	1.398e-004	6.090e-004	1.549e-004
3×16	16×16	1632	4096	3	2.350e-004	4.078e-005	3.520e-004	6.505e-005
3×16	32×32	6336	16384	15	9.096e-007	2.182e-007	9.248e-007	2.251e-007
3×16	64×64	24960	65536	117	1.625e-007	3.788e-008	1.646e-007	3.849e-008
3×36	2×2	36	144	0	2.484e-003	3.341e-004	3.364e-003	4.933e-004
3×36	4×4	120	576	0	5.684e-005	9.823e-006	1.122e-004	1.295e-005
3×36	8×8	432	2304	0	7.743e-007	1.884e-007	2.259e-006	2.309e-007
3×36	16×16	1632	9216	2	1.115e-008	3.214e-009	3.495e-008	3.774e-009
3×36	32×32	6336	36864	16	1.615e-010	4.947e-011	5.702e-010	5.720e-011
3×36	64×64	24960	147456	123	6.302e-012	2.160e-012	8.923e-012	2.284e-012
4×36	2×2	48	144	0	2.302e-002	2.464e-003	2.811e-002	3.712e-003
4×36	4×4	160	576	0	1.105e+001	1.051e+000	1.625e+001	1.339e+000
4×36	8×8	576	2304	1	6.694e-005	1.086e-005	7.246e-005	9.335e-006
4×36	16×16	2176	9216	5	3.275e-007	1.070e-007	3.388e-007	1.129e-007
4×36	32×32	8448	36864	37	6.138e-006	2.111e-006	5.997e-006	2.172e-006
4×36	64×64	33280	147456	374	3.800e-010	1.353e-010	3.800e-010	1.374e-010
4×49	2×2	48	196	0	4.658e-005	6.413e-006	6.328e-005	8.866e-006
4×49	4×4	160	784	0	7.293e-007	1.580e-007	7.904e-007	1.883e-007
4×49	8×8	576	3136	1	9.701e-009	2.898e-009	1.021e-008	3.212e-009
4×49	16×16	2176	12544	5	1.470e-010	4.828e-011	1.490e-010	5.112e-011
4×49	32×32	8448	50176	37	2.286e-012	7.812e-013	2.290e-012	8.049e-013
4×49	64×64	33280	200704	383	3.514e-014	1.227e-014	3.515e-014	1.246e-014

Fall III

Verbindungspunkte: Gauss-Legendre
Kollokationspunkte: Gauss-Legendre

$L \times N$	n×n	nVP	nKP	Zeit	Max-F-VP	∅-F-VP	Max-F-KP	∅-F-KP
2×9	2×2	24	36	0	4.178e-004	8.982e-005	5.831e-003	1.491e-003
2×9	4×4	80	144	0	4.209e-005	5.392e-006	6.524e-004	9.884e-005
2×9	8×8	288	576	0	2.540e-006	3.945e-007	5.214e-005	6.345e-006
2×9	16×16	1088	2304	1	1.177e-007	1.949e-008	3.731e-006	3.947e-007
2×9	32×32	4224	9216	5	4.680e-009	1.548e-009	2.433e-007	2.440e-008
2×9	64×64	16640	36864	37	1.406e-009	4.739e-010	1.590e-008	1.522e-009
3×16	2×2	36	64	0	3.313e-002	4.165e-003	3.047e-001	7.500e-002
3×16	4×4	120	256	0	1.243e-003	1.271e-004	4.152e-003	4.719e-004
3×16	8×8	432	1024	1	2.820e-004	3.339e-005	8.020e-004	1.132e-004
3×16	16×16	1632	4096	2	5.832e-006	1.540e-006	1.260e-005	1.793e-006
3×16	32×32	6336	16384	15	3.879e-007	7.742e-008	9.415e-007	9.755e-008
3×16	64×64	24960	65536	116	1.047e-007	3.495e-008	1.072e-007	3.551e-008
3×36	2×2	36	144	0	5.767e-005	4.480e-006	1.632e-003	4.414e-004
3×36	4×4	120	576	0	9.940e-007	6.467e-008	7.165e-005	8.376e-006
3×36	8×8	432	2304	0	6.631e-009	1.622e-009	2.134e-006	1.433e-007
3×36	16×16	1632	9216	2	8.566e-011	2.529e-011	4.220e-008	2.285e-009
3×36	32×32	6336	36864	16	1.358e-012	4.419e-013	7.053e-010	3.541e-011
3×36	64×64	24960	147456	125	3.471e-014	1.218e-014	1.175e-011	5.600e-013
4×36	2×2	48	144	0	2.064e-003	3.003e-004	2.797e-002	3.919e-003
4×36	4×4	160	576	0	1.778e-004	2.198e-005	2.869e-003	3.854e-004
4×36	8×8	576	2304	1	9.247e-007	1.751e-007	1.050e-005	1.449e-006
4×36	16×16	2176	9216	5	2.657e-009	3.731e-010	5.175e-008	3.852e-009
4×36	32×32	8448	36864	37	3.345e-003	4.475e-005	3.033e-002	3.516e-004
4×36	64×64	33280	147456	376	8.841e-013	2.696e-013	1.308e-012	2.897e-013
4×49	2×2	48	196	1	1.648e-006	2.243e-007	2.856e-005	3.418e-006
4×49	4×4	160	784	0	1.412e-008	1.803e-009	2.207e-007	1.628e-008
4×49	8×8	576	3136	1	9.566e-011	9.803e-012	1.133e-009	6.016e-011
4×49	16×16	2176	12544	5	4.721e-013	4.559e-014	5.383e-012	2.373e-013
4×49	32×32	8448	50176	37	3.816e-015	1.448e-015	2.258e-014	1.989e-015
4×49	64×64	33280	200704	391	5.482e-015	2.213e-015	5.496e-015	2.248e-015

Fall IV

Verbindungspunkte: Äquidistant
Kollokationspunkte: Gauss-Legendre

$L \times N$	n×n	nVP	nKP	Zeit	Max-F-VP	∅-F-VP	Max-F-KP	∅-F-KP
2×9	2×2	24	36	0	4.601e-003	9.282e-004	7.612e-003	1.895e-003
2×9	4×4	80	144	0	1.212e-003	3.381e-004	1.611e-003	4.364e-004
2×9	8×8	288	576	0	2.938e-004	9.675e-005	3.117e-004	1.091e-004
2×9	16×16	1088	2304	0	7.309e-005	2.578e-005	7.415e-005	2.741e-005
2×9	32×32	4224	9216	5	1.824e-005	6.649e-006	1.830e-005	6.857e-006
2×9	64×64	16640	36864	37	4.556e-006	1.688e-006	4.560e-006	1.715e-006
3×16	2×2	36	64	0	4.112e-002	5.739e-003	1.459e-001	2.743e-002
3×16	4×4	120	256	0	5.933e-003	1.030e-003	1.111e-002	2.058e-003
3×16	8×8	432	1024	0	1.056e-003	2.535e-004	1.702e-003	3.601e-004
3×16	16×16	1632	4096	2	3.522e-004	5.319e-005	1.038e-003	1.258e-004
3×16	32×32	6336	16384	15	2.763e-006	6.478e-007	3.049e-006	6.775e-007
3×16	64×64	24960	65536	117	1.153e-007	2.986e-008	1.174e-007	3.036e-008
3×36	2×2	36	144	0	3.007e-003	3.977e-004	7.068e-003	8.031e-004
3×36	4×4	120	576	0	5.541e-005	1.017e-005	2.061e-004	1.592e-005
3×36	8×8	432	2304	1	7.684e-007	1.872e-007	4.545e-006	2.639e-007
3×36	16×16	1632	9216	2	1.079e-008	3.143e-009	7.911e-008	4.207e-009
3×36	32×32	6336	36864	16	1.619e-010	4.990e-011	1.452e-009	6.555e-011
3×36	64×64	24960	147456	131	1.637e-011	6.124e-012	2.369e-011	6.337e-012
4×36	2×2	48	144	0	2.983e-002	2.766e-003	8.852e-002	9.705e-003
4×36	4×4	160	576	1	1.272e-003	1.745e-004	4.623e-003	3.146e-004
4×36	8×8	576	2304	1	2.325e-005	4.986e-006	4.479e-005	6.536e-006
4×36	16×16	2176	9216	5	8.716e-007	1.803e-007	3.111e-006	2.551e-007
4×36	32×32	8448	36864	36	2.328e-008	9.132e-009	2.428e-008	9.420e-009
4×36	64×64	33280	147456	393	3.521e-010	1.133e-010	3.530e-010	1.151e-010
4×49	2×2	48	196	0	4.430e-005	6.139e-006	8.942e-005	1.046e-005
4×49	4×4	160	784	1	7.315e-007	1.585e-007	1.164e-006	2.019e-007
4×49	8×8	576	3136	1	9.721e-009	2.903e-009	1.133e-008	3.274e-009
4×49	16×16	2176	12544	5	1.474e-010	4.838e-011	1.524e-010	5.143e-011
4×49	32×32	8448	50176	35	2.284e-012	7.805e-013	2.305e-012	8.049e-013

Fall V

Verbindungspunkte: Gauss-Legendre
Kollokationspunkte: Zufällig

$L \times N$	n×n	nVP	nKP	Zeit	Max-F-VP	∅-F-VP	Max-F-KP	∅-F-KP
2×9	2×2	24	36	0	1.253e-001	1.892e-002	9.422e-002	3.177e-002
2×9	4×4	80	144	0	4.056e-004	8.052e-005	6.465e-004	1.781e-004
2×9	8×8	288	576	0	7.773e-006	1.347e-006	3.635e-005	5.681e-006
2×9	16×16	1088	2304	0	6.128e-006	1.850e-006	6.232e-006	1.935e-006
2×9	32×32	4224	9216	4	6.055e-007	2.044e-007	6.834e-007	2.297e-007
2×9	64×64	16640	36864	33	6.305e-007	2.155e-007	6.348e-007	2.203e-007
3×16	2×2	36	64	0	2.193e-003	1.545e-004	1.284e-003	2.076e-004
3×16	4×4	120	256	0	5.373e-004	7.592e-005	4.251e-004	9.639e-005
3×16	8×8	432	1024	0	2.092e-005	5.767e-006	2.295e-005	6.713e-006
3×16	16×16	1632	4096	1	2.152e-004	8.340e-005	2.140e-004	8.860e-005
3×16	32×32	6336	16384	13	3.280e-008	1.003e-008	3.301e-008	1.021e-008
3×16	64×64	24960	65536	106	4.959e-001	3.837e-002	4.864e-001	3.887e-002
3×36	2×2	36	144	0	3.812e-005	3.923e-006	1.163e-003	1.273e-004
3×36	4×4	120	576	0	8.488e-007	1.561e-007	4.652e-005	3.489e-006
3×36	8×8	432	2304	0	1.525e-008	4.154e-009	1.483e-006	7.410e-008
3×36	16×16	1632	9216	2	1.157e-009	3.720e-010	3.370e-008	1.370e-009
3×36	32×32	6336	36864	14	1.234e-012	4.008e-013	6.564e-010	1.916e-011
3×36	64×64	24960	147456	110	1.192e-014	3.538e-015	9.032e-012	2.939e-013
4×36	2×2	48	144	0	7.761e-004	1.126e-004	5.359e-004	1.167e-004
4×36	4×4	160	576	0	2.973e-004	2.905e-005	2.515e-004	3.187e-005
4×36	8×8	576	2304	0	5.880e-007	1.472e-007	5.288e-007	1.628e-007
4×36	16×16	2176	9216	4	7.000e-009	1.719e-009	6.455e-009	1.798e-009
4×36	32×32	8448	36864	32	9.005e-003	1.122e-004	7.649e-003	1.022e-004
4×49	2×2	48	196	0	2.203e-006	2.348e-007	1.399e-005	2.066e-006
4×49	4×4	160	784	0	6.614e-007	1.265e-007	6.715e-007	1.502e-007
4×49	8×8	576	3136	0	1.193e-008	2.996e-009	1.127e-008	3.287e-009
4×49	16×16	2176	12544	4	2.807e-010	7.929e-011	2.770e-010	8.368e-011
4×49	32×32	8448	50176	32	6.121e-014	1.297e-014	6.395e-014	1.322e-014
4×49	64×64	33280	200704	307	3.924e-014	1.175e-014	3.913e-014	1.191e-014

Fall VI

Verbindungspunkte: Äquidistant
Kollokationspunkte: Zufällig

$L \times N$	n×n	nVP	nKP	Zeit	Max-F-VP	∅-F-VP	Max-F-KP	∅-F-KP
2×9	2×2	24	36	0	8.639e-002	1.248e-002	6.160e-002	2.028e-002
2×9	4×4	80	144	0	1.044e-003	2.923e-004	1.298e-003	3.664e-004
2×9	8×8	288	576	0	2.942e-004	9.663e-005	3.163e-004	1.070e-004
2×9	16×16	1088	2304	0	7.320e-005	2.580e-005	7.378e-005	2.745e-005
2×9	32×32	4224	9216	4	1.800e-005	6.558e-006	1.802e-005	6.757e-006
2×9	64×64	16640	36864	33	4.174e-006	1.540e-006	4.172e-006	1.563e-006
3×16	2×2	36	64	0	3.024e-003	3.052e-004	1.371e-003	3.544e-004
3×16	4×4	120	256	0	4.702e-004	8.859e-005	3.901e-004	1.143e-004
3×16	8×8	432	1024	0	4.678e-005	1.379e-005	4.649e-005	1.571e-005
3×16	16×16	1632	4096	1	1.248e-004	4.563e-005	1.242e-004	4.849e-005
3×16	32×32	6336	16384	13	1.162e-008	2.969e-009	1.095e-008	3.052e-009
3×16	64×64	24960	65536	106	2.232e-006	7.407e-007	2.231e-006	7.522e-007
3×36	2×2	36	144	0	1.377e-003	1.841e-004	1.706e-003	2.559e-004
3×36	4×4	120	576	0	5.271e-005	7.790e-006	6.215e-005	9.012e-006
3×36	8×8	432	2304	0	7.438e-007	1.779e-007	1.795e-006	2.057e-007
3×36	16×16	1632	9216	2	9.952e-009	3.145e-009	1.761e-008	3.396e-009
3×36	32×32	6336	36864	14	1.632e-010	5.014e-011	5.305e-010	5.530e-011
3×36	64×64	24960	147456	109	1.233e-011	4.610e-012	1.252e-011	4.727e-012
4×36	2×2	48	144	0	5.111e-004	2.233e-005	3.292e-004	2.964e-005
4×36	4×4	160	576	0	5.817e-006	1.313e-006	4.668e-006	1.556e-006
4×36	8×8	576	2304	0	5.659e-007	1.455e-007	5.249e-007	1.630e-007
4×36	16×16	2176	9216	4	2.329e-009	6.823e-010	2.370e-009	7.214e-010
4×36	32×32	8448	36864	31	1.939e-009	6.461e-010	1.936e-009	6.660e-010
4×36	64×64	33280	147456	308	3.173e+013	2.510e+011	3.107e+013	2.449e+011
4×49	2×2	48	196	0	4.341e-005	6.119e-006	5.047e-005	8.188e-006
4×49	4×4	160	784	0	6.370e-007	1.413e-007	5.867e-007	1.674e-007
4×49	8×8	576	3136	0	7.296e-009	1.430e-009	6.839e-009	1.560e-009
4×49	16×16	2176	12544	4	1.758e-010	5.726e-011	1.737e-010	6.059e-011
4×49	32×32	8448	50176	32	2.203e-012	7.589e-013	2.203e-012	7.817e-013
4×49	64×64	33280	200704	308	1.972e-014	4.748e-015	1.968e-014	4.817e-015

C.1 Helmholtz-Gleichung mit Lösung in $\mathcal{C}^2(\overline{\Omega})$

Fall I

Verbindungspunkte: Gauss-Legendre
Kollokationspunkte: Äquidistant

$L \times N$	n×n	nVP	nKP	Zeit	Max-F-VP	∅-F-VP	Max-F-KP	∅-F-KP
2×9	2×2	24	36	0	4.790e-004	1.056e-004	4.864e-003	1.858e-003
2×9	4×4	80	144	0	1.122e-004	3.179e-005	6.649e-004	1.721e-004
2×9	8×8	288	576	0	1.776e-005	6.481e-006	8.664e-005	1.739e-005
2×9	16×16	1088	2304	0	2.521e-006	1.072e-006	1.105e-005	1.872e-006
2×9	32×32	4224	9216	5	3.444e-007	1.573e-007	1.387e-006	2.147e-007
2×9	64×64	16640	36864	37	4.322e-008	2.046e-008	1.720e-007	2.444e-008
3×16	2×2	36	64	1	4.121e-002	4.631e-003	9.267e-002	2.708e-002
3×16	4×4	120	256	0	2.557e-003	2.033e-004	6.639e-003	9.142e-004
3×16	8×8	432	1024	0	5.170e-004	3.199e-005	1.208e-003	9.506e-005
3×16	16×16	1632	4096	2	6.045e-005	6.625e-006	1.218e-004	9.336e-006
3×16	32×32	6336	16384	15	1.238e-005	1.139e-006	2.897e-005	1.457e-006
3×16	64×64	24960	65536	116	6.549e-007	3.764e-008	1.544e-006	4.789e-008
3×36	2×2	36	144	0	6.295e-004	7.944e-005	2.353e-003	3.262e-004
3×36	4×4	120	576	0	2.808e-004	1.568e-005	4.526e-004	4.956e-005
3×36	8×8	432	2304	0	4.334e-005	2.212e-006	5.029e-004	3.811e-005
3×36	16×16	1632	9216	2	3.158e-006	1.995e-007	4.465e-004	2.328e-005
3×36	32×32	6336	36864	16	1.989e-006	1.561e-007	1.799e-004	2.719e-006
3×36	64×64	24960	147456	123	1.879e-008	2.018e-009	2.584e-005	2.328e-007
4×36	2×2	48	144	0	1.178e+000	1.089e-001	6.614e+000	1.050e+000
4×36	4×4	160	576	1	1.386e-002	2.366e-003	6.004e-002	1.521e-002
4×36	8×8	576	2304	0	7.165e-004	2.583e-004	1.520e-003	3.767e-004
4×36	16×16	2176	9216	5	9.786e-004	1.123e-004	9.459e-003	1.607e-003
4×36	32×32	8448	36864	35	2.981e-005	9.123e-006	3.043e-005	9.506e-006
4×36	64×64	33280	147456	353	7.710e-006	5.786e-007	2.450e-005	1.435e-006
4×49	2×2	48	196	0	9.671e-005	1.242e-005	5.233e-004	7.101e-005
4×49	4×4	160	784	0	9.956e-006	1.027e-006	4.870e-005	4.355e-006
4×49	8×8	576	3136	1	7.273e-006	1.356e-006	1.069e-005	1.584e-006
4×49	16×16	2176	12544	5	6.260e-007	1.298e-007	9.456e-007	1.463e-007
4×49	32×32	8448	50176	36	3.819e-008	4.236e-009	1.124e-007	5.045e-009
4×49	64×64	33280	200704	357	1.203e-008	3.290e-009	1.897e-008	3.362e-009

Fall II

Verbindungspunkte: Äquidistant
Kollokationspunkte: Äquidistant

$L \times N$	n×n	nVP	nKP	Zeit	Max-F-VP	∅-F-VP	Max-F-KP	∅-F-KP
2×9	2×2	24	36	0	6.360e-003	1.332e-003	6.968e-003	2.011e-003
2×9	4×4	80	144	0	1.702e-003	5.611e-004	1.784e-003	6.743e-004
2×9	8×8	288	576	0	4.349e-004	1.715e-004	4.374e-004	1.904e-004
2×9	16×16	1088	2304	1	1.095e-004	4.683e-005	1.096e-004	4.956e-005
2×9	32×32	4224	9216	5	2.739e-005	1.218e-005	2.739e-005	1.255e-005
2×9	64×64	16640	36864	36	6.860e-006	3.109e-006	6.861e-006	3.156e-006
3×16	2×2	36	64	0	1.837e-002	2.091e-003	7.389e-002	1.999e-002
3×16	4×4	120	256	0	8.324e-002	1.543e-002	6.234e-002	2.066e-002
3×16	8×8	432	1024	0	4.665e-004	1.284e-004	1.078e-003	1.845e-004
3×16	16×16	1632	4096	3	1.832e-003	5.708e-004	1.785e-003	6.312e-004
3×16	32×32	6336	16384	15	2.796e-006	2.524e-007	9.303e-006	3.719e-007
3×16	64×64	24960	65536	116	1.117e-006	1.577e-007	3.003e-006	1.781e-007
3×36	2×2	36	144	0	1.947e-003	3.162e-004	2.450e-003	4.830e-004
3×36	4×4	120	576	0	5.761e-004	9.693e-005	7.115e-004	1.210e-004
3×36	8×8	432	2304	0	1.933e-004	2.824e-005	2.887e-004	3.179e-005
3×36	16×16	1632	9216	3	4.302e-005	3.970e-006	9.015e-005	5.349e-006
3×36	32×32	6336	36864	16	2.062e-005	7.299e-007	4.983e-005	1.023e-006
3×36	64×64	24960	147456	120	1.767e-006	3.656e-008	8.545e-006	5.924e-008
4×36	2×2	48	144	0	1.440e-001	1.501e-002	1.869e-001	2.478e-002
4×36	4×4	160	576	0	7.675e+002	7.302e+001	1.128e+003	9.287e+001
4×36	8×8	576	2304	1	2.548e-002	5.546e-003	2.652e-002	5.128e-003
4×36	16×16	2176	9216	5	7.641e-004	2.120e-004	8.223e-004	2.233e-004
4×36	32×32	8448	36864	36	1.990e-003	7.584e-004	1.942e-003	7.751e-004
4×36	64×64	33280	147456	362	2.768e-006	1.882e-007	3.018e-006	1.874e-007
4×49	2×2	48	196	0	1.112e-003	1.466e-004	1.701e-003	2.038e-004
4×49	4×4	160	784	1	3.125e-004	5.956e-005	3.418e-004	7.101e-005
4×49	8×8	576	3136	0	3.467e-005	7.323e-006	3.954e-005	8.041e-006
4×49	16×16	2176	12544	5	6.157e-006	1.222e-006	7.733e-006	1.284e-006
4×49	32×32	8448	50176	37	6.941e-007	1.566e-007	7.706e-007	1.607e-007
4×49	64×64	33280	200704	361	9.270e-008	2.355e-008	9.839e-008	2.386e-008

Fall III

Verbindungspunkte: Gauss-Legendre
Kollokationspunkte: Gauss-Legendre

$L \times N$	n×n	nVP	nKP	Zeit	Max-F-VP	∅-F-VP	Max-F-KP	∅-F-KP
2×9	2×2	24	36	0	3.840e-004	5.258e-005	5.672e-003	2.620e-003
2×9	4×4	80	144	0	5.165e-005	6.150e-006	7.152e-004	2.015e-004
2×9	8×8	288	576	0	8.103e-006	1.318e-006	9.327e-005	1.535e-005
2×9	16×16	1088	2304	1	8.553e-007	6.262e-008	1.125e-005	1.079e-006
2×9	32×32	4224	9216	5	1.229e-007	2.860e-008	1.440e-006	8.789e-008
2×9	64×64	16640	36864	36	1.788e-008	7.131e-009	1.842e-007	1.104e-008
3×16	2×2	36	64	0	3.569e-002	4.142e-003	2.326e-001	5.258e-002
3×16	4×4	120	256	0	1.527e-003	1.572e-004	8.662e-003	1.029e-003
3×16	8×8	432	1024	1	2.369e-004	1.362e-005	1.770e-003	1.071e-004
3×16	16×16	1632	4096	2	4.739e-005	4.807e-006	2.557e-004	1.101e-005
3×16	32×32	6336	16384	15	1.436e-005	6.772e-007	1.090e-004	2.236e-006
3×16	64×64	24960	65536	117	1.042e-006	2.090e-007	3.813e-006	2.290e-007
3×36	2×2	36	144	0	5.261e-003	4.229e-004	1.443e-002	2.361e-003
3×36	4×4	120	576	1	3.785e-004	5.431e-005	1.426e-003	2.175e-004
3×36	8×8	432	2304	0	1.096e-004	1.035e-005	6.778e-004	5.503e-005
3×36	16×16	1632	9216	2	1.740e-005	2.148e-006	8.484e-004	3.171e-005
3×36	32×32	6336	36864	16	4.968e-007	7.014e-008	2.486e-004	8.394e-006
3×36	64×64	24960	147456	123	2.071e-007	4.150e-008	3.159e-005	5.973e-007
4×36	2×2	48	144	0	1.582e-002	2.523e-003	3.526e-001	6.803e-002
4×36	4×4	160	576	0	5.308e-003	1.291e-003	1.594e-001	3.673e-002
4×36	8×8	576	2304	1	6.617e-004	1.584e-004	7.713e-003	1.914e-003
4×36	16×16	2176	9216	5	8.918e-006	9.850e-007	6.617e-005	6.223e-006
4×36	32×32	8448	36864	37	9.248e-001	1.227e-002	8.384e+000	9.850e-002
4×36	64×64	33280	147456	374	3.054e-007	1.146e-008	4.305e-006	6.009e-008
4×49	2×2	48	196	1	7.971e-005	9.761e-006	1.250e-003	1.717e-004
4×49	4×4	160	784	0	1.253e-005	1.832e-006	1.428e-004	1.218e-005
4×49	8×8	576	3136	1	1.703e-006	2.647e-007	1.917e-005	9.286e-007
4×49	16×16	2176	12544	5	1.759e-007	2.485e-008	1.632e-006	5.696e-008
4×49	32×32	8448	50176	36	1.789e-008	1.190e-009	2.251e-007	3.381e-009
4×49	64×64	33280	200704	368	4.047e-009	8.234e-010	3.760e-008	9.958e-010

Fall IV

Verbindungspunkte: Äquidistant
Kollokationspunkte: Gauss-Legendre

$L \times N$	n×n	nVP	nKP	Zeit	Max-F-VP	∅-F-VP	Max-F-KP	∅-F-KP
2×9	2×2	24	36	0	7.152e-003	1.473e-003	1.102e-002	3.165e-003
2×9	4×4	80	144	0	1.838e-003	6.018e-004	2.414e-003	7.875e-004
2×9	8×8	288	576	0	4.520e-004	1.788e-004	4.716e-004	2.024e-004
2×9	16×16	1088	2304	1	1.118e-004	4.795e-005	1.128e-004	5.101e-005
2×9	32×32	4224	9216	5	2.770e-005	1.235e-005	2.775e-005	1.274e-005
2×9	64×64	16640	36864	37	6.896e-006	3.130e-006	6.900e-006	3.179e-006
3×16	2×2	36	64	0	2.935e-002	3.382e-003	1.105e-001	2.584e-002
3×16	4×4	120	256	0	4.662e-003	5.783e-004	2.209e-002	2.601e-003
3×16	8×8	432	1024	0	9.665e-004	1.714e-004	4.446e-003	4.351e-004
3×16	16×16	1632	4096	2	4.317e-003	1.436e-003	6.715e-003	1.691e-003
3×16	32×32	6336	16384	15	1.819e-005	1.021e-006	8.055e-005	2.088e-006
3×16	64×64	24960	65536	116	9.896e-007	1.135e-007	3.393e-006	1.315e-007
3×36	2×2	36	144	0	4.653e-003	8.093e-004	2.148e-002	1.857e-003
3×36	4×4	120	576	0	6.877e-003	1.487e-003	1.757e-002	1.994e-003
3×36	8×8	432	2304	0	2.617e-004	6.961e-005	5.048e-004	8.117e-005
3×36	16×16	1632	9216	3	2.357e-004	9.729e-006	1.339e-003	3.088e-005
3×36	32×32	6336	36864	16	2.307e-005	4.952e-007	2.152e-004	1.652e-006
3×36	64×64	24960	147456	127	1.032e-004	1.398e-006	6.212e-004	3.718e-006
4×36	2×2	48	144	0	3.075e-001	2.819e-002	7.512e-001	1.023e-001
4×36	4×4	160	576	0	6.411e-002	1.054e-002	2.556e-001	2.412e-002
4×36	8×8	576	2304	1	7.145e-003	1.510e-003	2.194e-002	2.400e-003
4×36	16×16	2176	9216	4	2.109e-003	5.233e-004	8.551e-003	7.805e-004
4×36	32×32	8448	36864	36	2.360e-004	8.836e-005	2.912e-004	9.184e-005
4×36	64×64	33280	147456	382	3.574e-006	2.287e-007	8.299e-006	2.667e-007
4×49	2×2	48	196	0	2.128e-003	2.864e-004	3.897e-003	4.997e-004
4×49	4×4	160	784	0	4.716e-004	8.441e-005	8.461e-004	1.100e-004
4×49	8×8	576	3136	1	6.796e-005	1.342e-005	1.102e-004	1.531e-005
4×49	16×16	2176	12544	5	9.816e-006	2.089e-006	1.624e-005	2.232e-006
4×49	32×32	8448	50176	36	1.195e-006	2.454e-007	1.773e-006	2.535e-007
4×49	64×64	33280	200704	383	1.500e-007	3.356e-008	2.469e-007	3.410e-008

Fall V

Verbindungspunkte: Gauss-Legendre
Kollokationspunkte: Zufällig

$L \times N$	n×n	nVP	nKP	Zeit	Max-F-VP	∅-F-VP	Max-F-KP	∅-F-KP
2×9	2×2	24	36	0	3.028e-001	4.926e-002	2.122e-001	7.967e-002
2×9	4×4	80	144	0	4.343e-004	9.162e-005	6.897e-004	1.855e-004
2×9	8×8	288	576	0	8.594e-005	2.258e-005	1.418e-004	3.760e-005
2×9	16×16	1088	2304	0	3.550e-004	8.515e-005	3.466e-004	9.156e-005
2×9	32×32	4224	9216	4	2.554e-006	1.058e-006	2.556e-006	1.044e-006
2×9	64×64	16640	36864	33	6.171e-006	2.524e-006	6.164e-006	2.561e-006
3×16	2×2	36	64	0	7.868e-002	1.068e-002	6.376e-002	1.761e-002
3×16	4×4	120	256	0	4.841e-004	8.601e-005	4.641e-004	1.244e-004
3×16	8×8	432	1024	0	4.113e-004	8.776e-005	3.729e-004	9.971e-005
3×16	16×16	1632	4096	1	4.751e-002	2.852e-003	4.528e-002	3.082e-003
3×16	32×32	6336	16384	13	2.803e-006	5.683e-007	2.577e-006	5.836e-007
3×16	64×64	24960	65536	109	1.039e+004	3.825e+002	1.029e+004	3.860e+002
3×36	2×2	36	144	0	4.233e-003	3.814e-004	4.035e-002	7.242e-003
3×36	4×4	120	576	0	1.365e-002	2.499e-003	1.468e-002	3.037e-003
3×36	8×8	432	2304	0	4.509e-005	2.630e-006	2.112e-003	1.170e-004
3×36	16×16	1632	9216	2	3.815e-005	7.736e-006	1.109e-003	3.072e-005
3×36	32×32	6336	36864	14	1.156e-007	2.774e-008	4.866e-004	5.881e-006
3×36	64×64	24960	147456	107	1.126e-007	2.192e-008	9.772e-004	7.355e-006
4×36	2×2	48	144	0	2.058e-002	3.302e-003	1.516e-002	4.044e-003
4×36	4×4	160	576	0	2.743e-002	4.069e-003	2.343e-002	4.588e-003
4×36	8×8	576	2304	0	7.208e-005	1.849e-005	5.869e-005	2.041e-005
4×36	16×16	2176	9216	4	8.916e-005	2.238e-005	8.653e-005	2.354e-005
4×36	32×32	8448	36864	32	1.799e+001	4.875e-001	1.475e+001	4.237e-001
4×36	64×64	33280	147456	310	6.230e+010	3.102e+008	3.684e+010	1.734e+008
4×49	2×2	48	196	0	7.894e-004	7.047e-005	9.108e-004	1.971e-004
4×49	4×4	160	784	0	1.634e-003	3.481e-004	1.595e-003	4.310e-004
4×49	8×8	576	3136	0	2.336e-004	6.250e-005	2.250e-004	7.123e-005
4×49	16×16	2176	12544	4	3.548e-004	8.176e-005	3.112e-004	8.644e-005
4×49	32×32	8448	50176	32	6.652e-006	1.351e-006	6.030e-006	1.387e-006
4×49	64×64	33280	200704	306	9.586e-007	2.564e-007	8.895e-007	2.597e-007

Fall VI

Verbindungspunkte: Äquidistant
Kollokationspunkte: Zufällig

$L \times N$	n×n	nVP	nKP	Zeit	Max-F-VP	∅-F-VP	Max-F-KP	∅-F-KP
2×9	2×2	24	36	0	2.115e-001	3.274e-002	1.442e-001	5.160e-002
2×9	4×4	80	144	0	1.393e-003	5.027e-004	1.660e-003	6.200e-004
2×9	8×8	288	576	0	3.907e-004	1.571e-004	4.039e-004	1.733e-004
2×9	16×16	1088	2304	0	2.243e-004	4.660e-005	2.164e-004	4.970e-005
2×9	32×32	4224	9216	4	2.821e-005	1.265e-005	2.823e-005	1.303e-005
2×9	64×64	16640	36864	32	1.036e-005	4.850e-006	1.036e-005	4.924e-006
3×16	2×2	36	64	0	4.569e-002	5.974e-003	3.673e-002	1.049e-002
3×16	4×4	120	256	0	2.919e-004	5.496e-005	3.374e-004	7.937e-005
3×16	8×8	432	1024	0	2.621e-004	6.491e-005	2.594e-004	7.446e-005
3×16	16×16	1632	4096	1	1.155e-002	1.177e-003	1.127e-002	1.260e-003
3×16	32×32	6336	16384	13	9.145e-007	2.876e-007	8.435e-007	2.954e-007
3×16	64×64	24960	65536	106	8.039e-005	2.414e-005	8.034e-005	2.451e-005
3×36	2×2	36	144	0	1.915e-002	2.637e-003	5.713e-002	5.226e-003
3×36	4×4	120	576	0	3.126e-003	5.016e-004	3.748e-003	6.131e-004
3×36	8×8	432	2304	0	1.037e-003	5.127e-005	3.076e-003	7.587e-005
3×36	16×16	1632	9216	2	4.370e-004	1.174e-005	6.665e-004	1.243e-005
3×36	32×32	6336	36864	14	1.931e-004	2.514e-006	4.790e-004	4.615e-006
3×36	64×64	24960	147456	109	4.605e-004	3.088e-006	9.260e-004	4.514e-006
4×36	2×2	48	144	0	5.614e-003	7.835e-004	3.562e-003	9.108e-004
4×36	4×4	160	576	0	2.172e-003	4.614e-004	1.805e-003	5.527e-004
4×36	8×8	576	2304	0	1.108e-004	2.373e-005	9.436e-005	2.631e-005
4×36	16×16	2176	9216	4	2.639e-005	5.266e-006	2.534e-005	5.538e-006
4×36	32×32	8448	36864	32	2.927e-006	6.799e-007	2.747e-006	6.870e-007
4×49	2×2	48	196	0	4.542e-003	6.005e-004	4.921e-003	8.337e-004
4×49	4×4	160	784	0	8.402e-004	1.636e-004	7.217e-004	1.935e-004
4×49	8×8	576	3136	0	1.552e-004	4.562e-005	1.430e-004	5.143e-005
4×49	16×16	2176	12544	4	9.719e-005	1.807e-005	8.479e-005	1.872e-005
4×49	32×32	8448	50176	32	5.750e-005	1.117e-007	3.547e-005	1.175e-007
4×49	64×64	33280	200704	305	3.293e-007	6.974e-008	3.112e-007	7.074e-008

Literaturverzeichnis

[Zu finiten Elementen]

[B] Klaus Böhmer. *Variational Crimes in Petrov-Galerkin Methods for Operator Equations and Bifurcation*. Bisher nicht veröffentlicht.

[BS] Susanne C. Brenner, L. Ridgway Scott. *The Mathematical Theory of Finite Element Methods*. Springer-Verlag New York, 1994.

[Zu den Kollokationsverfahren von Doedel]

[D] Eusebius Doedel. *On the construction of discretizations of elliptic partial differential equations*. J. Difference Equations and Applications, Vol. 3, 389-416, 1997.

[DS] Eusebius Doedel, Hamid Sharifi. *Collocation Methods for Continuation Problems in Nonlinear Elliptic PDEs*. Issue on Continuation Methods in Fluid Mechanics, D. Henry and A. Bergeon, eds., Notes on Numer. Fluid. Mech., Vol. 74, 105-118, Vieweg, 2000.

[Zur numerischen Quadratur]

[CR] R. Cools, P. Rabinowitz. *Monomial cubature rules since "Stroud": a compilation*. J. Comput. Appl. Math., 48:309-326, 1993.

[C] R. Cools. *Monomial cubature rules since "Stroud": a compilation – part 2*. J. Comput. Appl. Math., 112(1-2): 21–27, 1999.

[Q] Department of Computer Science, Katholieke Universiteit Leuven. *Encyclopedia of Cubature Formulas*. Internet-Datenbank, entstanden 1998, wird laufend aktualisiert. *http://www.cs.kuleuven.ac.be/ nines/research/ecf/ecf.html*

[Zur allgemeinen Theorie elliptischer Randwertprobleme]

[H] Wolfgang Hackbusch. *Theorie und Numerik elliptischer Differentialgleichungen*. B.G.Teubner-Verlag Stuttgart, 1986.

[Zu Grundlagen der Numerik]

[IK] E. Isaacson, H. B. Keller. *Analysis of Numerical Methods*. John Wiley New York, 1966.

[S] Hans Rudolf Schwarz. *Numerische Mathematik*. B.G.Teubner-Verlag Stuttgart, 4.Auflage 1997.

[Allgemeine Grundlagen]

[Ba] Heinz Bauer. *Maß- und Integrationstheorie*. Walter de Gruyter-Verlag Berlin, 2.Auflage 1992.

Index