

NUMERIK PARTIELLER DIFFERENTIALGLEICHUNGEN I

VORLESUNG VOM WS 2012/2013

MARIO OHLBERGER

Institut für Numerische und Angewandte Mathematik
Fachbereich Mathematik und Informatik
Westfälische Wilhelms-Universität Münster

Dieses Skript beruht auf meiner Vorlesung *Numerik Partieller Differentialgleichungen I* vom Wintersemester 2008 an der Westfälischen Wilhelms-Universität Münster. Zentraler Inhalt dieser Vorlesung sind Finite Elemente Verfahren zur Diskretisierung von elliptischen und parabolischen Differentialgleichungen. Die numerische Behandlung von hyperbolischen Erhaltungsgleichungen ist Gegenstand der Vorlesung *Numerik Partieller Differentialgleichungen II*.

Mario Ohlberger

Inhaltsverzeichnis

0 Einleitung	1
0.1 Modellierung mit partiellen Differentialgleichungen	2
0.1.1 Erhaltungsgleichungen	2
0.1.2 Variationsgleichungen	5
0.2 Grundideen numerischer Methoden	7
0.2.1 Finite Differenzen	7
0.2.2 Finite Volumen	8
0.2.3 Finite Elemente	8
0.3 Notation und theoretische Grundlagen	10
0.3.1 Differentialoperatoren und Hölderräume $C^{m,\alpha}$	10
0.3.2 L^p -Räume	12
0.3.3 Satz von Gauß und partielle Integration	14
1 Theorie für elliptische Randwertprobleme	17
1.1 Elliptische Randwertprobleme in Divergenzform	17
1.2 Grundlagen aus der Funktionalanalysis	19
1.2.1 Aussagen in Hilberträumen	19
1.2.2 Sobolevräume	23
1.2.3 Einbettungssätze und Poincarésche Ungleichung	28
1.3 Variationsformulierung elliptischer Randwertprobleme	31
1.4 Regularität für elliptische Randwertprobleme	35
2 Finite Elemente Verfahren	37
2.1 Ritz-Galerkin Verfahren und abstrakte Fehlerabschätzung	37
2.2 Finite Elemente Verfahren	41
2.2.1 Simpliziale Lagrange Elemente	41
2.2.2 Verallgemeinerung und weitere gebräuchliche Elemente	50
2.3 Approximationssätze und Fehlerabschätzungen	53
2.3.1 Interpolationabschätzungen	53
2.3.2 A priori Fehlerabschätzung für Finite Elemente Verfahren	60

2.3.3	L^2 -Fehlerabschätzung und das Lemma von Aubin-Nitsche	61
2.4	Randapproximation	65
2.5	A posteriori Fehlerabschätzungen und Adaptivität	80
2.5.1	Residuenfehlerschätzer	81
2.5.2	H^1 -Interpolation	85
2.5.3	Lokalisierte Abschätzungen für den Residuenfehlerschätzer	91
2.5.4	Adaptive Verfahren basierend auf Fehlerschätzern	99
2.6	Quadraturformeln	101
3	Parabolische Differentialgleichungen	113
3.1	Sobolevräume und Bochnerräume	116
3.2	Schwache Formulierung der Wärmeleitungsgleichung	119
3.3	Semidiskrete Finite Elemente Methode	124
3.4	Zeitschrittverfahren	132
3.5	Stabilität und Konvergenz	138
3.6	A posteriori Fehlerabschätzung	145
4	CG-Verfahren	159
4.1	Gradientenverfahren	159
4.2	CG-Verfahren	162
4.3	Kondition der Steifigkeitsmatrix	164

Abbildungsverzeichnis

2.1	Lineares Lagrange Element.	44
2.2	Kubisches Lagrange Element.	48
2.3	Kubisches Hermit Element.	51
2.4	Agyris Element.	52
2.5	Glatt berandetes Gebiet Ω mit Gebietsapproximation $(\Omega_h, \mathcal{T}_h)$.	66

Kapitel 0

Einleitung

Partielle Differentialgleichungen sind Gleichungen, die eine oder mehrere Funktionen mit ihren partiellen Ableitungen in Beziehung setzen. Sie sind von besonderer Bedeutung, da sich viele natürliche oder technische Prozesse mit Hilfe solcher Gleichungen beschreiben lassen. Beispiele und gleichzeitig Prototypen von partiellen Differentialgleichungen in einer Raumdimension sind

i) die **Poisson-Gleichung**

$$-\partial_{xx}u(x) = f(x), \quad x \in \Omega \subset \mathbb{R},$$

ii) die **Wärmeleitungsgleichung**

$$\partial_t u(x, t) - \partial_{xx}u(x, t) = f(x, t), \quad (x, t) \in \Omega_T \subset \mathbb{R} \times \mathbb{R}^+,$$

iii) und die **Wellengleichung**

$$\partial_{tt}u(x, t) = c\partial_{xx}u(x, t), \quad (x, t) \in \Omega_T \subset \mathbb{R} \times \mathbb{R}^+.$$

Dies sind Differentialgleichungen zweiter Ordnung, da maximal zweite partielle Ableitungen der gesuchten Funktion u auftreten. Die Gleichungen sind elliptisch, parabolisch, bzw. hyperbolisch und vertreten damit die wichtigsten Klassen von partiellen Differentialgleichungen zweiter Ordnung. In den Übungsaufgaben werden analytische Lösungsansätze für diese Prototypen diskutiert.

Da es oft schwierig oder gar unmöglich ist, Lösungen solcher partieller Differentialgleichungen analytisch in geschlossener Form anzugeben, sind wir auf numerische Verfahren angewiesen die Approximationen der exakten Lösung

liefern. In dieser Vorlesung werden wir uns mit einer Klasse von Approximationen detailliert auseinander setzen, der Finite Elemente Methode. Bevor wir uns jedoch mit numerischen Verfahren beschäftigen, wollen wir im folgenden sehen, wie sich beispielsweise physikalische Prozesse mit Hilfe partieller Differentialgleichungen modellieren lassen.

0.1 Modellierung mit partiellen Differentialgleichungen

Wir wollen in diesem Abschnitt einerseits Modelle für konkrete Problemstellungen ableiten, aber gleichzeitig allgemeine Prinzipien der Modellierung aufzeigen.

0.1.1 Erhaltungsgleichungen

Betrachten wir als erstes Beispiel die Ausbreitung eines Tintenkleckses in Wasser. Bezeichnen wir mit u die Konzentration der Tinte im Wasser, so sind wir also an der räumlichen und zeitlichen Entwicklung von $u : \Omega \times \mathbb{R}^+ \rightarrow [0, 1]$ interessiert, wobei $\Omega \subset \mathbb{R}^d$ ein Gebiet sei. Zur Modellierung dieser Entwicklung verwenden wir ein Grundprinzip der Kontinuumsmechanik, das Erhaltungsprinzip.

Definition 0.1 (Erhaltungsprinzip)

- a) **Physikalisches Prinzip:** Für eine extensive Zustandsgröße, wie z.B. Masse, Impuls, oder Energie, gilt: Die Änderung der Zustandsgröße in einem beliebigen Volumen V , kann nur durch den Transport der Größe über den Rand des Volumens V geschehen.
- b) **Mathematisches Äquivalent:** Ist $u(x, t)$ die Dichteverteilung einer extensiven Zustandsgröße, so gilt für ein beliebiges Teilvolumen $V \subset \Omega$:

$$\frac{d}{dt} \int_V u(x, t) dx = - \int_{\partial V} \mathbf{q}(x, t) \cdot \mathbf{n}(x, t) d\sigma(x). \quad (1)$$

Dabei ist \mathbf{n} die äußere Normale an den Rand von V und \mathbf{q} die Flussdichte der Zustandsgröße.

Um von diesem allgemeinen Prinzip zu einem mathematischen Modell zur Beschreibung der Ausbreitung des Tintenkleckses zu kommen, benötigen wir

eine physikalische Beziehung zwischen der Flussdichte \mathbf{q} und der Konzentration u . Nehmen wir an, dass das zugrundeliegende Wasser ruht, so ist es z.B. physikalisch sinnvoll anzunehmen, dass der Fluss proportional zum negativen Gradienten der Dichte ist, wir erhalten also folgendes Gesetz:

$$\mathbf{q}(x, t) = -D\nabla u(x, t).$$

Dabei bezeichnet $\nabla := (\partial_{x_1}, \dots, \partial_{x_d})$ den Gradienten bezüglich der Ortsvariablen x und der Proportionalitätsfaktor $D > 0$ wird Diffusionskoeffizient genannt. Setzen wir diese Beziehung in die allgemeine Erhaltungsgleichung ein, so erhalten wir

$$\frac{d}{dt} \int_V u(x, t) dx = \int_{\partial V} D\nabla u(x, t) \cdot \mathbf{n}(x, t) d\sigma(x).$$

Somit haben wir bereits ein mathematisches Modell zur Beschreibung der Ausbreitung des Tintenklecks erhalten. Mit Hilfe des Gaußschen Satzes können wir dieses Modell in eine partielle Differentialgleichung transformieren:

$$\begin{aligned} \frac{d}{dt} \int_V u(x, t) dx &= \int_{\partial V} D\nabla u(x, t) \cdot \mathbf{n}(x, t) d\sigma(x) \\ &\stackrel{\text{Satz von Gauß}}{=} \int_V \operatorname{div}(D\nabla u(x, t)) dx. \end{aligned}$$

und durch Vertauschen von Differentiation und Integration:

$$\int_V \partial_t u(x, t) dx = \int_V \operatorname{div}(D\nabla u(x, t)) dx.$$

Da diese integrale Beziehung für beliebige Teilgebiete $V \subset \Omega$ gilt, folgt schließlich

$$\partial_t u(x, t) = \operatorname{div}(D\nabla u(x, t)), \quad \forall (x, t) \in \Omega \times \mathbb{R}^+.$$

Speziell folgt für konstanten Diffusionskoeffizienten D in einer Raumdimension ($\Omega \subset \mathbb{R}$):

$$\partial_t u(x, t) - D \partial_{xx} u(x, t) = 0.$$

Wir sehen also, dass das mathematische Modell zur Beschreibung der Ausbreitung eines Tintenklecks in ruhenden Wasser die *Wärmeleitungsgleichung* ist, die in diesem Zusammenhang auch instationäre Diffusionsgleichung

genannt wird.

Im allgemeinen hat die instationäre Diffusionsgleichung unendlich viele Lösungen, um eine physikalisch relevante Lösung auszuwählen müssen wir Anfangs- und Randwerte vorschreiben.

Definition 0.2 (Anfangs-Randwertproblem für die instationäre Diffusionsgleichung)

Sei $\Omega \subset \mathbb{R}^d$ ein Gebiet (offen, zusammenhängend und beschränkt), dann heißt eine Funktion $u \in C^1([0, T]; C^2(\Omega) \cap C^0(\bar{\Omega}))$ eine klassische Lösung des Dirichletproblems für die instationäre Diffusionsgleichung (homogene Wärmeleitungsgleichung), falls gilt

$$\begin{aligned}\partial_t u(x, t) - \operatorname{div}(D \nabla u(x, t)) &= 0, & \forall (x, t) \in \Omega \times (0, T), \\ u(x, t) &= g(x, t), & \forall (x, t) \in \partial\Omega \times [0, T], \\ u(x, 0) &= u_0(x), & \forall x \in \Omega.\end{aligned}$$

Die Aufgabe der Angewandten Mathematik ist es nun folgende Fragen zu beantworten:

- Existiert eine Lösung des Anfangs-Randwertproblems und ist diese eindeutig?
- Welche Regularität hat die Lösung?
- Ist die Lösung eine beschränkte Funktion?
- Können wir die Lösung in geschlossener Form angeben?
- Was sind numerische Verfahren zur Approximation der Lösung?
- Konvergieren die numerischen Verfahren gegen die exakte Lösung? (Mit welcher Konvergenzrate?)
- Wie können wir die Ergebnisse einer numerischen Approximation visualisieren?
- Stimmt das mathematische Modell mit dem physikalischen Experiment überein? (Validierung)

Neben diesen allgemeinen Fragestellungen kann man z.B. auch an dem Langzeitverhalten der Konzentrationsverteilung u interessiert sein. Ist bei dem be-

trachteten Anfangs-Randwertproblem die Randkonzentration g unabhängig von der Zeit, d.h. $g(x, t) = \bar{g}(x) \forall t \in R^+$, so erhalten wir folgendes Theorem:

Theorem 0.3 (Langzeitverhalten/Stationäre Diffusionsgleichung)

Gilt $g(x, t) = \bar{g}(x) \forall t \in R^+$ in dem Anfangs-Randwertproblem für die instationäre Diffusionsgleichung, so konvergiert die Lösung u für große Zeiten gegen eine Funktion $\bar{u} \in C^2(\Omega) \cap C^0(\bar{\Omega})$, die nicht mehr von der Zeit abhängt und \bar{u} ist eine klassische Lösung des Dirichletproblems für die stationären Diffusionsgleichung, d.h. \bar{u} löst

$$\begin{aligned} -\operatorname{div}(D\nabla\bar{u}(x)) &= 0, & \forall x \in \Omega, \\ \bar{u}(x, t) &= \bar{g}(x), & \forall x \in \partial\Omega. \end{aligned}$$

Die so erhaltene stationäre Diffusionsgleichung ist eine mehrdimensionale Verallgemeinerung der bereits eingeführten Poisson-Gleichung.

Eine Stärke der Modellierung mit partiellen Differentialgleichungen ist, dass sich viele unterschiedliche physikalische Prozesse mit den gleichen mathematischen Modellen beschreiben lassen. Im nächsten Abschnitt werden wir dies anhand eines anderen Prinzips zur Herleitung von partiellen Differentialgleichungen verdeutlichen.

0.1.2 Variationsgleichungen

Neben dem Erhaltungsprinzip spielt in der Physik das Prinzip der *Energie minimierung* eine wichtige Rolle. Dieses Prinzip ermöglicht uns z.B. die Beschreibung des Verhaltens elastischer Körper. Dazu sei $u(x, t) \in \mathbb{R}^d$ die Auslenkung/der Verschiebevektor eines elastischen Körpers im Punkt x zum Zeitpunkt t und $\varepsilon(u) := \frac{1}{2}(\nabla u + \nabla u^\top)$ der Verzerrungstensor. Dann ist die potentielle Gesamtenergie eines belasteten, elastischen Körpers gegeben durch

$$E(u) = \frac{1}{2} \int_{\Omega} \sigma : \varepsilon(u) dx - \int_{\Omega} f u dx.$$

Dabei ist σ der symmetrische Spannungstensor und f eine äußere Volumenkraft. Das Skalarprodukt für Matrizen $A : B$ ist dabei definiert durch $A : B = \operatorname{tr}(A \top B) = \sum_{i,j=1}^n a_{ij} b_{ij}$.

Für idealisierte Materialien ist der Spannungstensor proportional zum Verzerrungstensor, d.h. es gilt das lineare Materialgesetz

$$\sigma(u) = A \varepsilon(u)$$

mit dem symmetrischen und positiv definiten Elastizitätstensor A . Zur Modellierung elastischer Körper mit partiellen Differentialgleichungen wenden wir auf die Energie $E(u)$ das folgende Minimierungsprinzip an.

Definition 0.4 (Energieminimierung/Variationsprinzip)

a) Physikalisches Prinzip *Ein physikalisches System strebt immer in den Zustand minimaler Energie.*

b) Mathematische Äquivalenz: *Sei $\bar{u}(x, t)$ eine Zustandsvariable und $E(\bar{u})$ die Energie eines Systems in Abhängigkeit von \bar{u} . Dann strebt \bar{u} gegen einen optimalen Zustand $u = u(x)$, der die Energie minimiert, d.h. falls E genügend glatt ist gilt*

$$\frac{d}{d\varepsilon} E(u + \varepsilon\varphi) \Big|_{\varepsilon=0} = 0$$

für beliebige zulässige Variationen φ .

Berechnen wir

$$\frac{d}{d\varepsilon} E(u + \varepsilon\varphi) \Big|_{\varepsilon=0} = 0$$

für die potentielle Energie elastischer Körper, so erhalten wir die Gleichung

$$\int_{\Omega} A\varepsilon(u) : \varepsilon(\varphi) dx = \int_{\Omega} f\varphi dx$$

für alle zulässigen Variationen φ . Nehmen wir an, dass u genügend oft differenzierbar ist, so folgt mit partieller Integration für Testfunktionen (Variationen) φ , die auf dem Rand von Ω verschwinden:

$$\int_{\Omega} \left(-\operatorname{div}(A\varepsilon(u)) - f \right) \varphi dx = 0.$$

Mit dem Hauptsatz der Variationsrechnung folgt somit die Differentialgleichung

$$-\operatorname{div}(A\varepsilon(u)) = f.$$

oder in ausführlicher Notation

$$-\sum_{i=1}^d \sum_{k,l=1}^d \partial_{x_i} A_{ijkl} \varepsilon(u)_{kl} = f_i, \quad \forall j = 1, \dots, d.$$

Im einfachsten eindimensionalen Fall, der Auslenkung eines dünnen Drahtes, reduziert sich diese Differentialgleichung zu

$$-\partial_x(A\partial_x u) = f,$$

und im einfachsten mehrdimensionalen Fall, der Auslenkung einer dünnen Membran, zu

$$-\operatorname{div}(A\nabla u) = f.$$

Wir erhalten also auch hier zur Beschreibung der Auslenkung elastischer Körper in Spezialfällen die Poisson-Gleichung.

0.2 Grundideen numerischer Methoden

In diesem Abschnitt wollen wir drei Klassen von numerischen Diskretisierungsverfahren vorstellen, die in der Praxis besonders relevant sind.

0.2.1 Finite Differenzen

Idee: Approximation von Differentialoperatoren durch Differenzenquotienten

Ersetzt man die Ableitungen in einer Differentialgleichungen durch geeignete Differenzenquotienten basierend auf einem regulären Gitter, so erhält man ein endlichdimensionales System von linearen oder nichtlinearen Gleichungen für die approximierenden Funktionswerte in den Gitterpunkten.

In einer Raumdimension kann man beispielsweise eine Ableitung durch folgende Differenzenquotienten approximieren:

$$\begin{aligned} u'(x) &\approx \partial^{+h} u(x) := \frac{u(x+h) - u(x)}{h}, & \text{Vorwärtsdifferenzenquotient,} \\ u'(x) &\approx \partial^{-h} u(x) := \frac{u(x) - u(x-h)}{h}, & \text{Rückwärtsdifferenzenquotient,} \\ u'(x) &\approx \partial^{c,h} u(x) := \frac{u(x+h) - u(x-h)}{2h}, & \text{Zentraler Differenzenquotient.} \end{aligned}$$

All diese Approximationen konvergieren für $h \rightarrow 0$ gegen $u'(x)$, falls u stetig differenzierbar ist. Baut man auf solchen Differenzenquotienten ein numerisches Verfahren auf, so führt jedoch nicht jede Wahl auf ein konvergentes Verfahren.

Betrachtet man zum Beispiel die lineare Transportgleichung

$$\partial_t u(x, t) + a \partial_x u(x, t) = 0,$$

so muss der Differenzenquotient zur Approximation von $\partial_x u$ in Abhängigkeit vom Vorzeichen des Koeffizienten a gewählt werden. Ist a positiv, so muss der Rückwärtsdifferenzenquotient gewählt werden, während für negatives a nur der Vorwärtsdifferenzenquotient zu einem konvergenten Verfahren führt. Die lineare Transportgleichung gehört zur Klasse der hyperbolischen Erhaltungsgleichungen. Diesen Typ von Differentialgleichungen und mögliche Diskretisierungsverfahren werden wir in der Vorlesung Numerik Partieller Differentialgleichungen II behandeln. In dieser Vorlesung werden wir Finite Differenzen Approximationen lediglich in Kapitel 3 zur Zeitdiskretisierung parabolischer Probleme verwenden.

Für eine ausführliche Diskussion solcher Verfahren zur Diskretisierung elliptischer Probleme verweisen wir beispielsweise auf [GR94].

0.2.2 Finite Volumen

Idee: Diskretes Erhaltungsprinzip

Zerlegt man ein Berechnungsgebiet $\Omega \subset \mathbb{R}^d$ in paarweise disjunkte Teilgebiete $T_i \subset \Omega$, so dass gilt $\bar{\Omega} = \bigcup_{i=1}^N \bar{T}_i$, so kann man von dem Erhaltungsprinzip 0.1 diskrete Verfahren ableiten. Hierzu bezeichnen wir mit $u_i(t)$ den Mittelwert von $u(\cdot, t)$ auf dem Teilgebiet T_i . Dann erhalten wir aus (1) mit dem Hauptsatz der Differential- und Integralrechnung

$$u_i(t^{n+1}) - u_i(t^n) = \frac{1}{|T_i|} \int_{t^n}^{t^{n+1}} \int_{\partial T_i} \mathbf{q}(u) \cdot \mathbf{n} \, d\sigma(x) dt.$$

Aus dieser Darstellung erhält man eine Finite Volumen Diskretisierung, wenn man den kontinuierlichen Fluss \mathbf{q} durch einen numerischen Fluss \mathbf{q}_h ersetzt, der nur von u_i und den Mittelwerten u_j auf benachbarten Teilgebieten T_j abhängt.

Finite Volumen Verfahren werden wir detaillierter in der Vorlesung Numerik Partieller Differentialgleichungen II im Kontext von Erhaltungsgleichungen behandeln. An dieser Stelle wollen wir auf folgende Literatur verweisen [Krö97, BO04].

0.2.3 Finite Elemente

Idee: Energieminimierung in endlich dimensionalen Teilräumen

Finite Elemente sind ein Spezialfall von so genannten Galerkin Verfahren. Die Idee der Galerkin Verfahren beruht auf dem Prinzip der Energieminimierung.

Sei also V ein Funktionenraum und $E : V \rightarrow \mathbb{R}$ ein Energiefunktional. In Abschnitt 1.2 hatten wir gesehen, dass die Lösung u vieler physikalischer Fragestellungen formal wie folgt definiert werden können:

$$u = \arg \min_{v \in V} E(v).$$

Aus diesem kontinuierlichen Minimierungsproblem erhält man ein diskretes Minimierungsproblem, wenn man den Funktionenraum V durch einen endlichdimensionalen Teilraum $V_h \subset V$ ersetzt. Somit lässt sich die Lösung $u_h \in V_h$ eines allgemeinen Galerkin Verfahrens wie folgt schreiben

$$u_h = \arg \min_{v_h \in V_h} E(v_h).$$

Analog zum kontinuierlichen Variationsprinzip, ist die diskrete Lösung u_h dann charakterisiert durch

$$\frac{d}{d\varepsilon} E(u_h + \varepsilon v_h) \Big|_{\varepsilon=0} = 0, \quad \text{für alle } v_h \in V_h,$$

falls E genügend regulär ist. Da V_h endlichdimensional ist können wir uns bei dieser Variation auf eine Basis von V_h einschränken. Sei also $\Phi := \{\varphi_i | i = 1, \dots, N\}$ eine Basis von V_h , d.h. $N = \dim(V_h)$ und $u_h = \sum_{i=1}^N u_i \varphi_i$ die zugehörige Basisdarstellung von u_h , so folgt

$$\frac{d}{d\varepsilon} E\left(\sum_{i=1}^N u_i \varphi_i + \varepsilon \varphi_j\right) \Big|_{\varepsilon=0} = 0, \quad \text{für alle } j = 1, \dots, N.$$

Dies ist ein System von N linearen oder nichtlinearen Gleichungen für die Unbekannten $u_i, i = 1, \dots, N$, das mit numerischen Methoden gelöst werden kann.

Finite Elemente Approximationen erhält man aus diesem allgemeinen Prinzip durch spezielle Wahl des endlichdimensionalen Teilraums V_h und zugehöriger Basis Φ mit der Eigenschaft, dass

$$\frac{d}{d\varepsilon} E\left(\varphi_i + \varepsilon \varphi_j\right) \Big|_{\varepsilon=0} \neq 0$$

nur für eine kleine Anzahl von Paaren (i, j) gilt.

In Kapitel 2 werden wir Finite Elemente Verfahren zur Approximation von elliptischen Differentialgleichungen genauer betrachten und numerisch analysieren. Auch in Kapitel 3 werden wir zur Ortsdiskretisierung parabolischer Differentialgleichungen Finite Elemente Methoden heranziehen. Als weiterführende Literatur wollen wir schon an dieser Stelle auf einige Lehrbücher verweisen [Bra97, BS08, Cia87, Sch91]. Hilfreich sind auch Vorlesungsskripte, die im Netz zu finden sind [Dzi, Ran].

0.3 Notation und theoretische Grundlagen

In diesem Abschnitt wollen wir Notationen festlegen und einige Grundlagen aus der Analysis wiederholen, auf die wir im Laufe der Vorlesung immer wieder Bezug nehmen werden. Alle Sätze werden ohne Beweise angegeben. Für eine detailliertere Darstellung verweisen wir auf die Literatur zur Analysis in mehreren Raumdimensionen.

0.3.1 Differentialoperatoren und Hölderräume $C^{m,\alpha}$

Zunächst führen wir einige Notation für Differentialoperatoren ein und definieren die Hölderräume $C^{m,\alpha}(\Omega)$ als Verallgemeinerung der Funktionenräume $C^m(\Omega)$, der m -mal stetig differenzierbaren Funktionen.

Definition 0.5 (Multiindex und partielle Ableitungen)

Sei $u : \mathbb{R}^d \rightarrow \mathbb{R}$ eine skalare Funktion mehrerer Veränderlicher. Dann heißt $D^\alpha u$ partielle Ableitung k -ter Ordnung von u , falls gilt $|\alpha| = k$ und

$$D^\alpha u := \frac{\partial^\alpha u}{\partial x^\alpha} := \left(\frac{\partial}{\partial x_1} \right)^{\alpha_1} \dots \left(\frac{\partial}{\partial x_d} \right)^{\alpha_d} u,$$

wobei $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}_0^d$ ein Multiindex ist mit der Schreibweise

$$|\alpha| := \sum_{i=1}^d \alpha_i, \quad x^\alpha := x_1^{\alpha_1} \dots x_d^{\alpha_d}.$$

Abkürzend werden wir im Folgenden für Funktionen $u(x, t), (x, t) \in \mathbb{R} \times \mathbb{R}$ die Notationen

$$\partial_t u(x, t) := \frac{\partial}{\partial t} u(x, t), \quad \partial_x u(x, t) := \frac{\partial}{\partial x} u(x, t).$$

verwenden. Ebenso schreiben wir für Funktionen $u(x), x \in \mathbb{R}^d$

$$\partial_i u(x) := \partial_{x_i} u(x) := \frac{\partial}{\partial x_i} u(x).$$

Analog kürzen wir k -te partielle Ableitungen in Richtung von x_i ab durch $\partial_i^k u$.

Definition 0.6 (Divergenz, Gradient, Rotation und Laplace Operator)

Seien $u : R^d \rightarrow R$ und $\mathbf{v} : R^d \rightarrow R^d$ genügend oft stetig differenzierbar. Dann definieren wir den Gradienten $\nabla u = \text{grad } u$ von u durch

$$\nabla u(x) := \text{grad } u(x) := \left(\partial_1 u(x), \dots, \partial_d u(x) \right)^\top.$$

Die Divergenz $\nabla \cdot \mathbf{v} = \text{div } \mathbf{v}$ des Vektorfeldes \mathbf{v} definieren wir als

$$\nabla \cdot \mathbf{v}(x) = \text{div } \mathbf{v}(x) := \sum_{i=1}^d \partial_i v_i(x).$$

Der Laplace Operator Δu von u ist definiert durch

$$\Delta u(x) := \nabla \cdot (\nabla u(x)) = \sum_{i=1}^d \partial_i^2 u(x).$$

Schließlich definieren wir für $d = 3$ die Rotation $\nabla \times \mathbf{v} = \text{rot } \mathbf{v}$ des Vektorfeldes \mathbf{v} durch

$$\nabla \times \mathbf{v} = \text{rot } \mathbf{v} := \left(\partial_2 v_3(x) - \partial_3 v_2(x), \partial_3 v_1(x) - \partial_1 v_3(x), \partial_1 v_2(x) - \partial_2 v_1(x) \right)^\top.$$

Definition 0.7 (Hölderräume)

Sei $C^m(\Omega, \mathbb{R}^n)$ für $\Omega \subset \mathbb{R}^d$ der Raum aller m -mal stetig differenzierbarer Funktionen, so schreiben wir für $n = 1$ auch $C^m(\Omega) := C^m(\Omega, \mathbb{R})$. Versehen mit der Supremumsnorm $\|\cdot\|_\infty$ ist $C^m(\Omega, \mathbb{R}^n)$ ein vollständiger Vektorraum. Wir definieren die Hölder-Konstante von $u \in C^0(\Omega)$ zu $\alpha \in [0, 1]$ durch

$$hö_1(u, \bar{\Omega}) := \sup \left\{ \frac{|u(x) - u(y)|}{|x - y|^\alpha} ; x, y \in \bar{\Omega}, x \neq y \right\}.$$

und die Hölderräume durch

$$C^{m,\alpha}(\bar{\Omega}) := \{u \in C^m(\Omega) | \, hö_1(\partial^s u, \bar{\Omega}) < \infty \text{ für } |s| = m\}.$$

$C^{m,\alpha}(\bar{\Omega})$ ist mit der Norm

$$\|u\|_{C^{m,\alpha}(\bar{\Omega})} := \sum_{|s| \leq m} \|\partial^s u\|_\infty + \sum_{|s|=m} hö_1(\partial^s u, \bar{\Omega})$$

ein Banachraum. Eine Funktion $C^{0,1}(\bar{\Omega})$ heißt Lipschitz-stetig mit Lipschitzkonstante $L := lip(u, \bar{\Omega}) := hö_1(u, \bar{\Omega})$.

0.3.2 L^p -Räume**Definition 0.8 (Banachraum und Dualraum)**

Sei V ein normierter Vektorraum. Ein normierter Raum $(V, \|\cdot\|)$ heißt vollständig, falls alle Cauchy-Folgen in V bzgl. $\|\cdot\|$ in V konvergieren. Ein vollständiger normierter Raum heißt **Banachraum**.

Ist V ein Banachraum, so ist durch

$$V' := \{\varphi : V \rightarrow \mathbb{R} | \, \varphi \text{ linear}, \|\varphi\|_{V'} := \sup_{v \in V, \|v\|=1} |\varphi(v)| < \infty\}$$

der Dualraum V' von V definiert. Der Dualraum V' eines Banachraums ist, ausgestattet mit der Norm $\|\cdot\|_{V'}$, wieder ein Banachraum.

Definition 0.9 (Operatoren)

Seien U, V normierte Vektorräume, $D \subset U$. Wir bezeichnen eine Abbildung $T : D \rightarrow V$ als **Operator**. Dabei gilt:

(i) T heißt **stetig** in $u \in D$, g.d.w.

$$\forall \varepsilon > 0 \exists \delta > 0 \forall v \in D : \|u - v\|_U < \delta \implies \|T(u) - T(v)\|_V < \varepsilon.$$

(ii) T heißt **stetig** in D , g.d.w.

T ist stetig für alle $u \in D$.

(iii) T heißt **Lipschitz-stetig**, g.d.w.

$$\text{es existiert ein } L > 0 \forall u, v \in D : \|T(u) - T(v)\|_V < L\|u - v\|_U.$$

Satz 0.10 (Banachscher Fixpunktsatz)

Seien X ein Banachraum, $D \subset X$ abgeschlossen, und $T : D \rightarrow D$ eine Kontraktion, d.h. T ist Lipschitz-stetig mit Lipschitzkonstante $L < 1$. Dann hat T genau einen Fixpunkt $u \in D$.

Definition 0.11 (Hilbertraum)

Ein Vektorraum X mit Skalarprodukt $(\cdot, \cdot)_X$ heißt **Prähilbertraum**. Ein vollständiger Prähilbertraum heißt **Hilbertraum**.

Definition 0.12 (L^p -Räume)

Mit $L^p(\Omega)$, $p \in [1, \infty)$ bezeichnen wir den Raum aller Lebesgue-messbarer Funktionen u , so dass u^p Lebesgue-integrierbar ist. Ausgestattet mit der Norm

$$\|u\|_p := \left(\int_{\Omega} |u|^p \right)^{1/p}$$

ist $L^p(\Omega)$ ein Banachraum.

Auf $L^2(\Omega)$ ist durch $\int_{\Omega} uv$ ein Skalarprodukt definiert. Dadurch wird $L^2(\Omega)$ zu einem Hilbertraum.

Mit $L^{\infty}(\Omega)$ bezeichnen wir den Raum aller Lebesgue-messbarer Funktionen u , deren Supremumsnorm $\|u\|_{\infty} := \text{ess sup } |u|$ beschränkt ist. Auch dieser Raum ist mit der zugehörigen Norm ein Banachraum.

Für $1 < p, q < \infty$ mit $\frac{1}{p} + \frac{1}{q} = 1$ gilt: $L^q(\Omega)$ ist isomorph zum Dualraum von $L^p(\Omega)$.

Satz 0.13 (Höldersche Ungleichung)

Sei $u \in L^p(\Omega)$ und $v \in L^q(\Omega)$ mit $\frac{1}{p} + \frac{1}{q} = 1$. Dann gilt die Abschätzung

$$\|uv\|_1 \leq \|u\|_p \|v\|_q.$$

Insbesondere ist also $uv \in L^1(\Omega)$.

Für $p = q = 2$ erhält man als Spezialfall die Cauchysche Ungleichung.

Satz 0.14 (Youngsche Ungleichung)

Seien $a, b > 0$, und gelte $\frac{1}{p} + \frac{1}{q} = 1$, dann gilt die Abschätzung

$$ab \leq \frac{1}{p}a^p + \frac{1}{q}b^q.$$

Insbesondere folgt für $p = q = 2$ mit für $\varepsilon > 0$

$$ab \leq \frac{\varepsilon}{2}a^2 + \frac{1}{\varepsilon 2}b^2.$$

Satz 0.15 (Fundamentalsatz der Variationsrechnung)

Sei $u \in L^1_{loc}(\Omega)$, $\Omega \subset \mathbb{R}^n$ offen. Dann sind äquivalent:

$$\int_{\Omega} uv = 0 \quad \forall v \in C_0^{\infty}(\Omega),$$

und

$$u = 0 \text{ fast überall in } \Omega.$$

(siehe [Alt92, 2.11])

0.3.3 Satz von Gauß und partielle Integration

Definition 0.16 (Lipschitz-Gebiet)

Eine offene Teilmenge $\Omega \subset \mathbb{R}^d$ heißt Lipschitz-Gebiet, g.d.w. der Rand von Ω sich lokal in einer jeweils geeigneten Richtung als Graph einer Lipschitz-stetigen Funktion schreiben lässt und Ω auf einer Seite des Graphen liegt.

Satz 0.17 (Satz von Gauß für Lipschitz Gebiet)

Sei $\Omega \subset \mathbb{R}^d$ ein Lipschitz-Gebiet und $\mathbf{v} \in C^0(\bar{\Omega}; \mathbb{R}^d) \cap C^1(\Omega; \mathbb{R}^d)$ ein Vektorfeld mit $\operatorname{div}(\mathbf{v}) \in L^1(\Omega)$. Dann gilt

$$\int_{\Omega} \operatorname{div}(\mathbf{v}) dx = \int_{\partial\Omega} \mathbf{v} \cdot \mathbf{n} d\sigma(x).$$

Dabei bezeichnet \mathbf{n} die äußere Normale an den Rand von Ω .

(siehe [Alt92, A 5.9] für eine Verallgemeinerung.)

Lemma 0.18 (Partielle Integration)

Seien $\Omega \subset \mathbb{R}^d$ ein Gebiet, so dass der Gaußsche Integralsatz gilt und $u \in C^1(\bar{\Omega})$ und $\mathbf{v} \in C^1(\bar{\Omega}; \mathbb{R}^d)$. Dann gilt

$$\int_{\Omega} \nabla u(x) \cdot \mathbf{v}(x) dx = - \int_{\Omega} u(x) \operatorname{div}(\mathbf{v})(x) dx + \int_{\partial\Omega} u(x) \mathbf{v}(x) \cdot \mathbf{n}(x) d\sigma(x).$$

(folgt aus dem Satz von Gauß.)

Kapitel 1

Theorie schwacher Lösungen für elliptische Randwertprobleme

1.1 Elliptische Randwertprobleme in Divergenzform

In diesem Abschnitt werden wir eine Klasse von elliptischen Randwertproblemen vorstellen, die im Kontext dieser Vorlesung behandelt werden. Es handelt sich dabei um skalare elliptische Differentialgleichungen zweiter Ordnung in Divergenzform. Der einfachste Spezialfall dieser Klasse ist die Poisson-Gleichung

Definition 1.1 (Poisson-Problem)

Sei $\Omega \subset \mathbb{R}^d$ ein Gebiet (offen, zusammenhängend und beschränkt) mit glattem Rand und sei $f \in C^0(\Omega)$ gegeben. Dann heißt $u \in C^2(\Omega) \cap C^0(\bar{\Omega})$ Lösung des homogenen Dirichlet-Problems für die Poisson-Gleichung, falls gilt

$$-\Delta u(x) = f(x) \quad \forall x \in \Omega, \quad (1.1)$$

$$u(x) = 0 \quad \forall x \in \partial\Omega. \quad (1.2)$$

Bemerkung 1.2 (Allgemeinere Randwerte)

Anstelle des homogenen Dirichlet-Problems kann man das Poisson-Problem auch mit anderen Randwerten betrachten. Die wichtigsten Klassen von Randwertproblemen erhält man, indem man die Randbedingung (1.2) durch eine der folgenden Bedingungen ersetzt:

- **Dirichlet-Randwerte:**

$$u(x) = g_D(x) \quad \forall x \in \partial\Omega,$$

für eine gegebenen Funktion $g_D \in C^0(\partial\Omega)$.

- **Neumann-Randwerte:**

$$\nabla u(x) \cdot n(x) = g_N(x) \quad \forall x \in \partial\Omega,$$

mit einer gegebenen Funktion $g_N \in C^0(\partial\Omega)$.

- **Robin-Randwerte:**

$$\nabla u(x) \cdot n(x) + \alpha(x)u(x) = g_R(x) \quad \forall x \in \partial\Omega,$$

mit gegebenen Funktionen $\alpha, g_R \in C^0(\partial\Omega)$.

Neben diesen Problemen kann man auch Probleme mit gemischten Randbedingungen betrachten, d.h. für eine disjunkte Zerlegungen des Randes $\partial\Omega = \Gamma_D \cup \Gamma_N \cup \Gamma_R$ geben wir auf Γ_D Dirichlet-Randwerte vor, auf Γ_N Neumann-Randwerte und auf Γ_R Robin-Randwerte.

In dieser Vorlesung werden wir uns der Einfachheit halber nur mit dem Poisson-Problem beschäftigen. Die gesamte Methodik und numerische Analysis lässt sich jedoch ohne größeren Aufwand auf die Klasse elliptischer Differentialgleichungen zweiter Ordnung in Divergenzform verallgemeinern.

Definition 1.3 (Allgemeines Randwertproblem in Divergenzform)

Seien $\Omega \subset \mathbb{R}^d$ ein Gebiet mit glattem Rand und folgende Datenfunktionen gegeben:

$$A = (a_{ij})_{i,j=1}^d \in C^1(\Omega; \mathbb{R}^{d \times d}), \quad a = (a_i)_{i=1}^d \in C^1(\Omega; \mathbb{R}^d),$$

$$b = (b_i)_{i=1}^d \in C^0(\Omega; \mathbb{R}^d), \quad c \in C^0(\Omega; \mathbb{R}), \quad f \in C^0(\Omega; \mathbb{R}).$$

Dann heißt $u \in C^2(\Omega) \cap C^0(\bar{\Omega})$ Lösung des homogenen Dirichlet-Problems für eine elliptische Differentialgleichungen zweiter Ordnung in Divergenzform, falls gilt

$$\begin{aligned} -\nabla \cdot (A(x)\nabla u(x)) - \nabla \cdot (a(x)u(x)) &+ b(x)\nabla u(x) \\ &+ c(x)u(x) = f(x) \quad \forall x \in \Omega, \end{aligned} \quad (1.3)$$

$$u(x) = 0 \quad \forall x \in \partial\Omega. \quad (1.4)$$

Neben dieser Klasse linearer elliptischer Probleme kann man auch nichtlineare Verallgemeinerungen betrachten, z.B. Probleme der Form

$$-\nabla \cdot A(x, u, \nabla u) + B(x, u, \nabla u) = 0.$$

Hierzu sei z.B. auf [GT77, Chapter 13] verwiesen.

In dieser Vorlesung werden wir den Begriff einer schwachen Lösung von elliptischen Differentialgleichungen einführen und in diesem Kontext die eindeutige Lösbarkeit des Poisson-Problems beweisen. Hierzu benötigen wir einige Grundlagen aus der Funktionalanalysis, die wir im folgenden Abschnitt vorstellen.

1.2 Grundlagen aus der Funktionalanalysis

Wie wir bereits in der Einleitung gesehen haben, lassen sich elliptische Differentialgleichungen aus dem Minimierungsprinzip ableiten. Dadurch erhalten wir eine Variationsformulierung die bei hinreichender Regularität der Lösung äquivalent zur Definition einer klassischen Lösung im Sinne von Definition 1.1 ist (siehe Übungsaufgabe).

Ist die Lösung des Variationsproblems nicht hinreichend regulär, so kann das Problem immer noch eine eindeutige Lösung haben, die dann aber nicht mehr eine klassischen Lösung darstellt. Um später die Existenz und Eindeutigkeit von Lösungen in der Variationsformulierung zeigen zu können, werden wir nun einige Grundlagen aus der Funktionalanalysis behandeln.

1.2.1 Aussagen in Hilberträumen

Satz 1.4 (Rieszscher Darstellungssatz)

Seien X ein reeller Hilbertraum mit Skalarprodukt $(\cdot, \cdot)_X$ und $f \in X'$ ein Element aus dem Dualraum von X . Dann gibt es genau ein $x_0 \in X$, so dass für alle $x \in X$ gilt:

$$f(x) = (x, x_0)_X. \quad (1.5)$$

Beweis: Die Idee des Beweises beruht auf dem Minimierungsprinzip. Zunächst definieren wir das Funktional $I : X \rightarrow \mathbb{R}$ durch

$$I(x) := \frac{1}{2} \|x\|_X^2 - f(x).$$

Die Behauptung ist dann gezeigt, falls wir die Existenz und Eindeutigkeit eines Minimums $x_0 \in X$ des Funktionals I zeigen, wie wir im letzten Schritt des Beweises sehen werden.

Um die Existenz und Eindeutigkeit eines Minimums $x_0 \in X$ von I zu zeigen, definieren wir

$$d := \inf_{x \in X} I(x).$$

1. Schritt: Wir zeigen $-\infty < d < \infty$.

$d < \infty$ ist klar, da $X \neq \emptyset$. Aus der Definition der Norm im Dualraum folgt

$$f(x) \leq \|f\|_{X'} \|x\|_X$$

und somit erhalten wir mit der Youngschen Ungleichung für ein $\varepsilon > 0$

$$I(x) \geq \frac{1}{2} \|x\|_X^2 - \|f\|_{x'} \|x\|_X \geq \left(\frac{1}{2} - \frac{\varepsilon}{2}\right) \|x\|_X^2 - \frac{1}{2\varepsilon} \|f\|_{x'}^2.$$

Wählen wir z.B. $\varepsilon = 1$ so folgt die Behauptung, da

$$I(x) \geq -\frac{1}{2\varepsilon} \|f\|_{x'}^2 > -\infty$$

für alle $x \in X$ gilt.

2. Schritt: Wir zeigen die Konvergenz einer Minimalfolge.

Nach der Definition des Infimums gibt es eine sogenannte Minimalfolge $(x_m)_{m \in \mathbb{N}}$, $x_m \in X$ mit der Eigenschaft

$$\lim_{m \rightarrow \infty} I(x_m) = d.$$

Da, X ein Banachraum ist, reicht es zu zeigen, dass die Minimalfolge eine Cauchyfolge ist. Wir haben mit der Parallelogrammidentität

$$\begin{aligned} \|x_m - x_n\|_X^2 &= 2\left(\|x_m\|_X^2 + \|x_n\|_X^2\right) - \|x_m + x_n\|_X^2 \\ &= 4I(x_m) + 4I(x_n) - 8I\left(\frac{x_m + x_n}{2}\right) \\ &\leq 4I(x_m) + 4I(x_n) - 8d \\ &\rightarrow 4d + 4d - 8d = 0 \quad \text{für } m, n \rightarrow \infty. \end{aligned}$$

Also ist $(x_m)_{m \in \mathbb{N}}$ konvergent und wir nennen den Grenzwert x_0 .

3. Schritt: Zeige: Für x_0 gilt $I(x_0) = \inf_{x \in X} I(x)$.

Da die Norm eines normierten Raumes stetig ist und alle Elemente eines Dualraums stetig sind, ist das Funktional I stetig. Die Behauptung folgt dann aus der Stetigkeit von I .

4. Schritt: Wir zeigen noch, dass aus der Minimierungseigenschaft Gleichung (1.5) folgt:

Für $x \in X$ und $\varepsilon \in \mathbb{R}$ gilt auch $x_0 + \varepsilon x \in X$ und somit folgt, dass die Funktion $\Phi(\varepsilon) := I(x_0 + \varepsilon x)$ in $\varepsilon = 0$ ein Minimum hat. Aus

$$\Phi(\varepsilon) = \frac{1}{2} \|x_0\|_X^2 - f(x_0) + \frac{1}{2} \varepsilon^2 \|x\|_X^2 - \varepsilon f(x) + \varepsilon (x, x_0)_X$$

folgt dann mit $\Phi'(\varepsilon)|_{\varepsilon=0} = 0$ die Behauptung (1.5).

Die Eindeutigkeit von x_0 ist klar, aufgrund der Linearität von (1.5). \square

Der Rieszsche Darstellungssatz liefert uns ein nützliches Werkzeug bezogen auf Skalarprodukte von Hilberträumen. Eine Verallgemeinerung davon für nicht notwendigerweise symmetrische Bilinearformen liefert der Satz von Lax-Milgram. Mit seiner Hilfe werden wir später in der Lage sein Existenz und Eindeutigkeit von schwachen Lösungen für elliptische Probleme in Divergenzform zu beweisen.

Satz 1.5 (Lax-Milgram)

Sei X ein reeller Hilbertraum und $B : X \times X \rightarrow \mathbb{R}$ eine stetige Bilinearform, d.h. es existiert eine Konstante c_1 , so dass für alle $x, y \in X$ gilt

$$|B(x, y)| \leq c_1 \|x\|_X \|y\|_X. \quad (1.6)$$

Weiter sei B koerativ, d.h. es existiert eine Konstante c_0 , so dass für alle $x \in X$ gilt

$$B(x, x) \geq c_0 \|x\|_X^2. \quad (1.7)$$

Dann gibt es zu jedem $f \in X'$ genau ein $u \in X$, so dass für alle $\varphi \in X$ gilt

$$B(u, \varphi) = f(\varphi). \quad (1.8)$$

Beweis: **1. Schritt:** Definition einer linearen Abbildung $T : X \rightarrow X$:

Für festes $x_0 \in X$ definiere

$$F(x) = B(x_0, x), \quad x \in X.$$

Dann ist F linear und stetig mit $\|F\|_{X'} \leq c_1 \|x_0\|_X$. Nach dem Rieszschen Darstellungssatz gibt es dann ein Element $Tx_0 \in X$, so dass gilt

$$B(x_0, x) = F(x) = (x, Tx_0)_X, \quad \forall x \in X.$$

Dadurch wird eine Abbildung $T : X \rightarrow X$ definiert, die linear ist, da für alle $x, x_1, x_2 \in X$ und $\lambda_1, \lambda_2 \in \mathbb{R}$ gilt:

$$\begin{aligned} (T(\lambda_1 x_1 + \lambda_2 x_2), x)_X &= B(\lambda_1 x_1 + \lambda_2 x_2, x) = \lambda_1 B(x_1, x) + \lambda_2 B(x_2, x) \\ &= \lambda_1 (T(x_1), x)_X + \lambda_2 (T(x_2), x)_X \\ &= (\lambda_1 T(x_1) + \lambda_2 T(x_2), x)_X. \end{aligned}$$

Da diese Gleichung für alle $x \in X$ gilt, erhalten wir

$$T(\lambda_1 x_1 + \lambda_2 x_2) = \lambda_1 T(x_1) + \lambda_2 T(x_2).$$

2. Schritt: T ist stetig:

Aus der Stetigkeit der Bilinearform folgt

$$\|Tx\|_X^2 = (Tx, Tx)_X = B(x, Tx) \leq c_1 \|x\|_X \|Tx\|_X$$

und demnach

$$\|Tx\|_X \leq c_1 \|x\|_X.$$

3. Schritt: T ist injektiv:

Mit der Cauchy-Schwarzschen Ungleichung folgt

$$c_0 \|x\|_X^2 \leq B(x, x) = (x, Tx)_X \leq \|x\|_X \|Tx\|_X,$$

und somit

$$c_0 \|x\|_X \leq \|Tx\|_X.$$

Also folgt aus $Tx = 0$, dass $x = 0$ ist und somit folgt die Injektivität aus der Linearität von T .

Wir zeigen: Der Bildbereich $R(T) := \{y \in X \mid \exists x \in X : Tx = y\}$ ist abgeschlossen:

Sei also $y_m \in R(T)$ eine konvergente Folge mit $\lim_{m \rightarrow \infty} y_m = y$, so wollen wir zeigen, dass $y \in R(T)$ ist, d.h. dass es ein $x \in X$ gibt mit $Tx = y$. Nach Annahme gibt es zu jedem y_m ein $x_m \in X$ mit $Tx_m = y_m$. Für die Folge $(x_m)_{m \in \mathbb{N}}$ gilt

$$\begin{aligned} c_0 \|x_m - x_l\|_X &\leq \|T(x_m - x_l)\|_X = \|Tx_m - Tx_l\|_X \\ &= \|y_m - y_l\|_X \rightarrow 0 \quad (m, l \rightarrow \infty). \end{aligned}$$

Also ist $(x_m)_{m \in \mathbb{N}}$ eine Cauchyfolge und aufgrund der Vollständigkeit von X existiert ein $x \in X$ mit $\lim_{m \rightarrow \infty} x_m = x$. Wegen der Stetigkeit von T folgt

hieraus schließlich $Tx = y$.

4. Schritt: T ist surjektiv:

Dazu nehmen wir zunächst an, dass $R(T) \subset X$ ein echter Teilraum wäre, d.h. $R(T) \neq X$ gilt. Dann existiert ein Element $x_0 \in R(T)^\perp \setminus \{0\}$, i.e. $(x_0, y)_X = 0$ für alle $y \in R(T)$ und es gilt

$$c_0 \|x_0\|_X^2 \leq B(x_0, x_0) = (x_0, Tx_0) \stackrel{Tx_0 \in R(T)}{=} 0.$$

Dies ist ein Widerspruch zu $x_0 \neq 0$ und somit folgt die Surjektivität.

5. Schritt: Beweis der Behauptung:

Sei $f \in X'$, dann existiert nach dem Rieszschen Darstellungssatz genau ein $x_0 \in X$, so dass für alle $\varphi \in X$ gilt

$$(\varphi, x_0)_X = f(\varphi).$$

Nach der Konstruktion von T existiert dann aber auch ein $u \in X$ mit $Tu = x_0$. Also folgt für alle $\varphi \in X$:

$$f(\varphi) = (\varphi, x_0)_X = (\varphi, Tu)_X = B(u, \varphi).$$

Somit folgt die Existenz einer Lösung von (1.8). Die Eindeutigkeit folgt direkt aus der Koerzivität der Bilinearform, denn sind u_1, u_2 zwei Lösungen von (1.8), so gilt

$$\begin{aligned} c_0 \|u_1 - u_2\|_X^2 &\leq B(u_1 - u_2, u_1 - u_2) = B(u_1, u_1 - u_2) - B(u_2, u_1 - u_2) \\ &= f(u_1 - u_2) - f(u_1 - u_2) = 0, \end{aligned}$$

was wiederum $u_1 = u_2$ impliziert. \square

1.2.2 Sobolevräume

In diesem Abschnitt wollen wir schwache Differenzierbarkeit von Funktionen definieren und Funktionenräume für schwach differenzierbar Funktionen definieren, die sogenannten Sobolevräume. Wir werden uns dabei im Wesentlichen auf die Angabe von Resultaten und Sätzen beschränken und verweisen für detaillierte Beweise auf die Vorlesung von Herrn Burger, oder auf die Literatur [Alt92, BS08].

Definition 1.6 ($L^1_{loc}(\Omega)$ und $C_0^\infty(\Omega)$)

Sei $\Omega \subset \mathbb{R}^d$ ein Gebiet. Dann bezeichnet $C_0^\infty(\Omega)$ alle Funktionen in $C^\infty(\Omega)$ mit kompaktem Träger und der Raum der lokal integrierbaren Funktionen $L^1_{loc}(\Omega)$ ist definiert durch

$$L^1_{loc}(\Omega) := \{u \in L^1(K) \quad \forall K \subset \Omega, K \text{ kompakt}\}.$$

Definition 1.7 (Schwache Ableitung)

Sei $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}_0^d$ ein Multiindex. Eine Funktion $u \in L^1_{loc}(\Omega)$ besitzt die schwache Ableitung $v_\alpha \in L^1_{loc}(\Omega)$, wenn für alle Testfunktionen $\varphi \in C_0^\infty(\Omega)$ gilt

$$\int_{\Omega} u D^\alpha \varphi = (-1)^{|\alpha|} \int_{\Omega} v_\alpha \varphi.$$

Wir schreiben dann auch $v_\alpha = D^\alpha u$ für die schwache Ableitung.

Beispiel 1.8 (Schwache Ableitung von $|x|$)

Sei $\Omega = (-1, 1)$ und $u(x) = |x|$. Dann ist $u'(x) = \text{sign}(x)$ die schwache Ableitung von u .

Beweis: Es gilt für beliebige $\varphi \in C_0^\infty(\Omega)$

$$\begin{aligned} \int_{-1}^1 u(x) \varphi'(x) dx &= \int_{-1}^0 u(x) \varphi'(x) dx + \int_0^1 u(x) \varphi'(x) dx \\ &= \int_{-1}^0 (-x) \varphi'(x) dx + \int_0^1 x \varphi'(x) dx \\ &= - \int_{-1}^0 (-1) \varphi(x) dx - [(-x) \varphi]_{-1}^0 - \int_0^1 1 \varphi(x) dx + [x \varphi]_0^1 \\ &= \int_{-1}^1 \text{sign}(x) \varphi(x) dx. \quad \square \end{aligned}$$

Im Gegensatz zu $|x|$ ist $\text{sign}(x)$ nicht schwach differenzierbar (Übungsaufgabe).

Definition 1.9 (Sobolevräume)

Seien $m \in \mathbb{N}_0$, $p \in [1, \infty]$ und $u \in L^1_{loc}(\Omega)$. Wir nehmen an, dass alle schwachen partielle Ableitungen $D^\alpha u$ existieren für $|\alpha| \leq m$. Dann definieren wir die Sobolevnormen $\|u\|_{m,p} = \|u\|_{H^{m,p}(\Omega)}$ durch

$$\|u\|_{H^{m,p}(\Omega)} := \left(\sum_{|\alpha| \leq m} \|D^\alpha u\|_{L^p(\Omega)}^p \right)^{1/p},$$

falls $1 \leq p < \infty$ und für $p = \infty$ als

$$\|u\|_{H^{m,\infty}(\Omega)} := \max_{|\alpha| \leq m} \|D^\alpha u\|_{L^\infty(\Omega)}.$$

Schließlich definieren wir die Sobolevräume $H^{m,p}(\Omega)$ durch

$$H^{m,p}(\Omega) := \{u \in L^1_{loc}(\Omega) \mid \|u\|_{H^{m,p}(\Omega)} < \infty\}.$$

Für $p = 2$ schreiben wir auch $H^m(\Omega) := H^{m,2}(\Omega)$. Außerdem definieren wir die Sobolev-Halbnormen durch

$$|u|_{H^{m,p}(\Omega)} := \left(\sum_{|\alpha|=m} \|D^\alpha u\|_{L^p(\Omega)}^p \right)^{1/p},$$

falls $1 \leq p < \infty$ und für $p = \infty$ als

$$|u|_{H^{m,\infty}(\Omega)} := \max_{|\alpha|=m} \|D^\alpha u\|_{L^\infty(\Omega)}.$$

Bemerkung 1.10

Anstelle von $H^{m,p}(\Omega)$ werden die Sobolevräume in der Literatur auch oft mit $W^{m,p}(\Omega)$ bezeichnet.

Beispiel 1.11

Seien $\Omega := B_{1/2}(0) \subset \mathbb{R}^2$ und $u(x) := \log |\log |x||$, $x \in \Omega$. Dann gilt: $u \in H^{1,2}(\Omega)$, aber $u \notin C^0(\Omega)$. D.h. Funktionen in H^1 sind in mehreren Raumdimensionen nicht notwendigerweise stetig (Beweis siehe Übungsaufgabe)!

Beispiel 1.12

Wir betrachten $\Omega := B_1(0) \subset \mathbb{R}^d$, $u(x) = |x|^s$, für $s \in \mathbb{R}$ und $x \in \Omega$. Dann ist $u \in H^{1,p}(\Omega)$, wenn gilt $s > 1 - \frac{d}{p}$.

Zum Beweis dieser Aussage muß zunächst gezeigt werden, dass es ein $p \geq 1$

gibt, so dass $u \in L^p(\Omega)$ ist. Wegen

$$|u(x)|^p \leq |x|^{sp}$$

ist $u \in L^p(\Omega)$, falls $sp > -d$. Für den weiteren Beweis verweisen wir auf die Übungsaufgaben.

Wir haben die Sobolevräume eingeführt, weil die klassischen Funktionenräume bezüglich den L^p -Normen nicht vollständig sind. Für die Sobolevräume gilt hingegen der für uns wichtige Satz:

Satz 1.13 (Vollständigkeit der Sobolevräume)

Sei $\Omega \subset \mathbb{R}^d$ ein Gebiet. Dann ist $H^{m,p}(\Omega)$, $1 \leq p \leq \infty$, $m \in \mathbb{N}_0$ ein Banachraum. $H^m(\Omega)$ ist ein Hilbertraum mit dem Skalarprodukt

$$(u, v)_{H^m(\Omega)} := \sum_{|\alpha| \leq m} (D^\alpha u, D^\alpha v)_{L^2(\Omega)}.$$

Sobolevräume sind also insbesondere vollständig. Dennoch erben Sie viele Eigenschaften von klassischen Funktionenräumen. Dass dies der Fall ist, liegt an der folgenden Approximationseigenschaft von stetig differenzierbaren Funktionen.

Satz 1.14

Sei $\Omega \subset \mathbb{R}^d$ ein Gebiet. Dann ist $C^\infty(\Omega) \cap H^{m,p}(\Omega)$ dicht in $H^{m,p}(\Omega)$, d.h. $H^{m,p}(\Omega)$ -Funktionen lassen sich durch beliebig oft differenzierbare Funktionen approximieren.

(Beweis siehe Übungsaufgaben)

Bemerkung 1.15 (Rechenregeln für schwache Ableitungen)

Aufgrund der Tatsache, dass klassisch differenzierbare Funktionen dicht in $H^{m,p}(\Omega)$ liegen, sieht man leicht ein, dass sich bekannte Rechenregeln für klassische Ableitungen, wie etwa die Produktregel oder die Kettenregel auf schwache Ableitungen übertragen.

Ferner sei bemerkt, dass sich auch der Gaußsche Integralsatz und die daraus resultierenden Formeln zur partiellen Integration direkt auf entsprechend schwach differenzierbare Funktionen in $H^{m,p}(\Omega)$ übertragen.

Da wir uns mit Randwertproblemen befassen wollen, ist es notwendig zu klären in welchem Sinne wir bei Sobolevräumen von Randwerten reden können, da die Funktionen zunächst nur bis auf Nullmengen definiert sind und der Rand eines Gebietes eine Nullmenge darstellt, auf der man L^p -Funktionen beliebig abändern kann. In der folgenden Definition klären wir zunächst, was wir unter 0-Randwerten in schwachem Sinne verstehen wollen.

Definition 1.16 (Schwache Nullrandwerte)

Für $1 \leq p < \infty$ und $m \in \mathbb{N}$ definieren wir die Sobolevräume mit Nullrandwerten $H_0^{m,p}(\Omega)$ durch

$$H_0^{m,p}(\Omega) := \overline{C_0^m(\Omega)}^{\|\cdot\|_{H^{m,p}(\Omega)}}.$$

Zur Definition von schwachen Nullrandwerten machen wir uns also die Approximation durch klassisch differenzierbaren Funktionen mit Nullrandwerten zu nutze. Wichtig ist dabei allerdings, dass die Vollständigkeit nicht verloren geht. Dies drückt sich in folgenden Satz aus.

Satz 1.17

Für $1 \leq p < \infty$ und $m \in \mathbb{N}$ ist $H_0^{m,p}(\Omega)$ ein abgeschlossener Teilraum von $H^{m,p}(\Omega)$. Insbesondere ist also $H_0^{m,p}(\Omega)$ ein Banachraum und die Norm $\|\cdot\|_{H^{m,p}(\Omega)}$ überträgt sich auf $H_0^{m,p}(\Omega)$.

Dass aus der Definition von $H_0^{m,p}(\Omega)$ tatsächlich folgt, dass solche Funktionen Randwerte besitzen, drückt der folgende Satz aus.

Satz 1.18 (Spursatz)

Seien $\Omega \subset \mathbb{R}^d$ ein Lipschitz-Gebiet und $1 \leq p < \infty$. Dann gibt es einen linearen Spur-Operator $B : H^{1,p}(\Omega) \rightarrow L^p(\partial\Omega)$, so dass für $u \in H^{1,p}(\Omega) \cap C^0(\bar{\Omega})$ gilt

$$Bu = u|_{\partial\Omega}.$$

Insbesondere gilt für $u \in H_0^{1,p}(\Omega)$

$$Bu = 0.$$

Zuletzt wollen wir noch eine Notation für die Dualräume von Sobolevräumen einführen.

Definition 1.19 (Dualräume $H^{-m,p'}(\Omega)$)

Für $1 \leq p, p' \leq \infty$ mit $\frac{1}{p} + \frac{1}{p'} = 1$ bezeichnen wir mit

$$H^{-m,p'}(\Omega) := \left(H_0^{m,p}(\Omega) \right)'$$

den Dualraum von $H_0^{m,p}(\Omega)$. Wir schreiben $H^{-m}(\Omega)$ für $H^{-m,2}(\Omega)$.

Wir führen hier die Dualräume ein, damit es uns möglich sein wird Differentialgleichungen zu betrachten, deren rechte Seite ein Funktional und keine Funktion ist.

1.2.3 Einbettungssätze und Poincarésche Ungleichung

Ein wichtiges Hilfsmittel zum Verständnis und Umgang mit Sobolevräumen sind Sobolevsche Einbettungssätze. Solche Sätze sagen z.B. aus, dass $H^{m,p}(\Omega)$ -Funktionen in $H^{m-1,p^*}(\Omega)$ liegen, falls p, p^* geeignet gewählt werden. Außerdem kann gezeigt werden, dass für hinreichend große m, p $H^{m,p}(\Omega)$ -Funktionen sogar klassisch differenzierbar sind. Auch bei den Sobolevschen Einbettungssätzen werden wir keine Beweise angeben und verweisen z.B. auf [Alt92, Sätze 8.7 und 8.8].

Satz 1.20 (1. Sobolevscher Einbettungssatz)

Seien $\Omega \subset \mathbb{R}^d$ ein Gebiet (offen und beschränkt) und $m_1 \geq m_2$, $m_1, m_2 \in \mathbb{N}_0$, $p_1, p_2 \in [1, \infty)$. Dann gilt:

$$H_0^{m_1,p_1}(\Omega) \hookrightarrow H_0^{m_2,p_2}(\Omega), \text{ falls } m_1 - \frac{d}{p_1} > m_2 - \frac{d}{p_2}.$$

Insbesondere ist

$$H_0^{m,p}(\Omega) \hookrightarrow L^{p^*}(\Omega), \text{ mit } p^* < \frac{dp}{d - mp}, \text{ falls } mp < d$$

ist. Ist Ω ein Lipschitz-Gebiet, so gelten die obigen Aussagen auch für die Sobolevräume ohne Nullrandwerte.

Satz 1.21 (2. Sobolevscher Einbettungssatz)

Seien $\Omega \subset \mathbb{R}^d$ ein Gebiet und $m_1 \geq m_2$, $m_1, m_2 \in \mathbb{N}_0$, $p \in [1, \infty)$. Dann gilt für $0 \leq \alpha \leq 1$

$$H_0^{m_1, p}(\Omega) \hookrightarrow C^{m_2, \alpha}(\bar{\Omega}), \text{ falls } m_1 - \frac{d}{p} > m_2 + \alpha.$$

Ist Ω ein Lipschitz-Gebiet, so gelten die obigen Aussagen auch für die Sobolevräume ohne Nullrandwerte.

Beispiel 1.22

In einer Raumdimension ($d = 1$) sind Funktionen in $H^1(\Omega)$ stetig, denn es ist $1 - \frac{1}{2} = \frac{1}{2} > 0$. Für $d > 1$ ist dieser Schluß nicht möglich, da in diesem Fall $1 - \frac{d}{2} \leq 0$ gilt.

Zum Abschluß dieses Abschnitts wollen wir nun die sogenannte Poincarésche Ungleichung angeben und beweisen. Eine nicht unwichtige Folgerung aus dieser Ungleichung ist es, dass die Halbnorm $|\cdot|_{H^1(\Omega)}$ auf $H_0^1(\Omega)$ bereits eine Norm ist.

Satz 1.23 (Poincarésche Ungleichung)

Sei $\Omega \subset \mathbb{R}^d$ ein Gebiet mit Durchmesser $D := \text{diam}(\Omega)$ und $p \in [1, \infty)$. Dann gibt es eine Konstante $c_p \leq 2D$, so dass für alle $v \in H_0^{1,p}(\Omega)$ gilt

$$\|v\|_{L^p(\Omega)} \leq c_p \|\nabla v\|_{L^p(\Omega)}. \quad (1.9)$$

Beweis: Es reicht die Abschätzung (1.9) für $v \in C_0^1(\Omega)$ zu zeigen. Da aufgrund von Satz 1.14 $C_0^1(\Omega)$ dicht in $H_0^{1,p}(\Omega)$ liegt, folgt die Aussage dann auch für Funktionen in $H_0^{1,p}(\Omega)$, was wir wie folgt einsehen. Sei $(v_m)_{m \in \mathbb{N}}$ eine Folge in $C_0^1(\Omega)$ mit $v := \lim_{m \rightarrow \infty} v_m \in H_0^{1,p}(\Omega)$ und gelte die Ungleichung (1.9) für Funktionen in $C_0^1(\Omega)$. Dann folgt

$$\begin{aligned} \|v\|_{L^p(\Omega)} &\leq \|v_m\|_{L^p(\Omega)} + \|v_m - v\|_{L^p(\Omega)} \\ &\leq c_p \|\nabla v_m\|_{L^p(\Omega)} + \|v_m - v\|_{L^p(\Omega)} \\ &\leq c_p \|\nabla v\|_{L^p(\Omega)} + c_p \|\nabla(v_m - v)\|_{L^p(\Omega)} + \|v_m - v\|_{L^p(\Omega)} \\ &\rightarrow c_p \|\nabla v\|_{L^p(\Omega)} \quad \text{für } m \rightarrow \infty. \end{aligned}$$

Sei nun also $v \in C_0^1(\Omega)$ und definiere eine Fortsetzung \tilde{v} von v auf \mathbb{R}^d durch

$$\tilde{v}(x) := \begin{cases} v(x), & x \in \Omega \\ 0, & \text{sonst} \end{cases}.$$

Sei nun $C > 0$ so gewählt, dass gilt $\Omega \subset [-C, C] \times \mathbb{R}^{d-1}$, dann folgt wegen $\tilde{v} \in C_0^1(\Omega; \mathbb{R}^d)$ aus dem Hauptsatz der Differential- und Integralrechnung

$$\tilde{v}(x) = \int_{-C}^x \partial_{x_1} \tilde{v}(s, x_2, \dots, x_d) ds.$$

Damit folgt weiter für p, p' mit $\frac{1}{p} + \frac{1}{p'} = 1$:

$$\begin{aligned} |\tilde{v}(x)|^p &\leq \left| \int_{-C}^x \partial_{x_1} \tilde{v}(s, x_2, \dots, x_d) ds \right|^p \\ &\leq \left(\int_{-C}^x |\partial_{x_1} \tilde{v}(s, x_2, \dots, x_d)| ds \right)^p \\ &\leq \left(\int_{-C}^C |\partial_{x_1} \tilde{v}(s, x_2, \dots, x_d)|^p ds \right)^{p/p} \left(\int_{-C}^C 1^{p'} ds \right)^{p/p'} \\ &\leq (2C)^{p/p'} \int_{-C}^C |\partial_{x_1} \tilde{v}(s, x_2, \dots, x_d)|^p ds. \end{aligned}$$

Integration dieser Gleichung über x_1 liefert dann

$$\int_{-C}^C |\tilde{v}(x)|^p dx_1 \leq (2C)^{p/p'+1} \int_{-C}^C |\partial_{x_1} \tilde{v}(s, x_2, \dots, x_d)|^p ds.$$

Integration über die Richtungen x_2, \dots, x_d ergibt schließlich

$$\int_{\mathbb{R}^d} |\tilde{v}(x)|^p dx \leq (2C)^{p/p'+1} \int_{\mathbb{R}^d} |\partial_{x_1} \tilde{v}(x)|^p dx$$

und somit folgt aufgrund der Definition von \tilde{v} mit $p/p' + 1 = p$

$$\left(\int_{\Omega} |v(x)|^p dx \right)^{1/p} \leq 2C \left(\int_{\Omega} |\partial_{x_1} v(x)|^p dx \right)^{1/p}.$$

Die Behauptung folgt durch analoges vorgehen für die partiellen Ableitungen in die Richtungen x_2 bis x_d . \square

Der Beweis der Poincarésche Ungleichung zeigt die Abhängigkeit der Konstanten c_p von Ω . Weitaus schwieriger ist es für bestimmte Geometrieklassen von Gebieten Ω optimale Konstanten c_p zu bestimmen. Man kann zeigen, dass die optimale Poincaré-Konstante der Kehrwert des kleinsten Eigenwertes des Laplaceoperators $-\Delta$ mit homogenen Dirichletrandwerten ist.

Wir haben nun alle funktionalanalytischen Grundlagen vorbereitet, um uns wieder der Analysis von elliptischen Randwertproblemen zuzuwenden.

1.3 Variationsformulierung elliptischer Randwertprobleme

Wir wollen nun von schwachen Ableitungen und Sobolevräumen gebrauch machen, um die Definition von klassischen Lösungen elliptischer Differentialgleichungen zu verallgemeinern. Dazu betrachten wir zunächst das Randwertproblem für die Poisson-Gleichung (1.1), (1.2).

Definition 1.24 (Schwache Formulierung des Poisson-Problems)

Seien Ω ein Lipschitz-Gebiet und $f \in L^2(\Omega)$ gegeben. Eine Funktion $u \in H_0^1(\Omega)$ heißt schwache Lösung des Randwertproblems für die Poisson-Gleichung (1.1), (1.2), falls für alle Testfunktionen $\varphi \in H_0^1(\Omega)$ gilt

$$\int_{\Omega} \nabla u(x) \cdot \nabla \varphi(x) dx = \int_{\Omega} f(x) \varphi(x) dx.$$

Satz 1.25 (Variationsprinzip)

Seien Ω ein Lipschitz-Gebiet und $f \in L^2(\Omega)$ gegeben. Dann sind äquivalent:

- i) $u \in H_0^1(\Omega)$ ist schwache Lösung des Randwertproblems für die Poisson-Gleichung.
- ii) Es gilt

$$u = \arg \min_{v \in H_0^1(\Omega)} I(v)$$

wobei das Energiefunktional $I : X \rightarrow \mathbb{R}$ gegeben ist durch

$$I(v) := \frac{1}{2} |v|_{H^1(\Omega)}^2 - (f, v).$$

Dabei bezeichnet (\cdot, \cdot) das L^2 -Skalarprodukt.

Beweis: (Siehe Übungsaufgabe)

Satz 1.26 (Existenz und Eindeutigkeit von Lösungen)

Seien Ω ein Lipschitz-Gebiet und $f \in L^2(\Omega)$. Dann existiert genau eine schwache Lösung $u \in H_0^1(\Omega)$ des Randwertproblems für die Poisson-Gleichung.

Beweis: Der Beweis folgt aus dem Rieszschen Darstellungssatz, denn durch

$$(u, \varphi)_{H_0^1(\Omega)} := \int_{\Omega} \nabla u(x) \cdot \nabla \varphi(x) dx$$

wird aufgrund der Poincarésche Ungleichung ein Skalarprodukt auf dem Hilbertraum $H_0^1(\Omega)$ definiert. Außerdem wird für $f \in L^2(\Omega)$ durch

$$F(\varphi) := \int_{\Omega} f(x) \varphi(x) dx$$

ein Funktional $F \in H_0^{-1}(\Omega)$ definiert, da gilt:

$$H_0^1(\Omega) \subset L^2(\Omega) = (L^2(\Omega))' \subset H_0^{-1}(\Omega).$$

□

Analog zur schwachen Formulierung des Poisson Problems wollen wir nun auch die schwache Lösung eines allgemeinen linearen elliptischen Problems in Divergenzform definieren.

Definition 1.27 (Schwache Formulierung von elliptischen Differentialgleichungen in Divergenzform)

Seien Ω ein Lipschitz-Gebiet, $X := H_0^1(\Omega)$ und $f \in X' = H_0^{-1}(\Omega)$. Seien weiter folgende Datenfunktionen gegeben

$$A = (a_{ij})_{i,j=1}^d \in L^\infty(\Omega; R^{d \times d}), \quad a = (a_i)_{i=1}^d \in L^\infty(\Omega; R^d), \\ b = (b_i)_{i=1}^d \in L^\infty(\Omega; R^d), \quad c \in L^\infty(\Omega; R),$$

wobei die sogenannte L-Bedingung erfüllt sei, d.h. es gibt eine Konstante $\alpha > 0$, so dass für alle $\xi \in \mathbb{R}^d$ und alle $\chi \in \mathbb{R}$ gilt

$$(A(x)\xi, \xi) + (a(x)\chi, \xi) + (b(x), \xi\chi) + c(x)\chi^2 \geq \alpha \|\xi\|^2 \quad \forall x \in \mathbb{R}^d. \quad (1.10)$$

Mit Hilfe der Datenfunktionen definieren wir eine Bilinearform $B : X \times X \rightarrow \mathbb{R}$ durch

$$B(u, v) := \int_{\Omega} A(x)\nabla u(x) \cdot \nabla v(x) + a(x)u(x) \cdot \nabla v(x) \\ + b(x) \cdot \nabla u(x)v(x) + c(x)u(x)v(x) dx.$$

Dann heißt $u \in H_0^1(\Omega)$ schwache Lösung des Randwertproblems in Divergenzform (1.3), (1.4), falls für alle Testfunktionen $\varphi \in H_0^1(\Omega)$ gilt

$$B(u, \varphi) = f(\varphi). \quad (1.11)$$

Satz 1.28

Unter den Voraussetzungen aus Definition 1.27 gibt es genau eine schwache Lösung $u \in H_0^1(\Omega)$ des Randwertproblems in Divergenzform (1.11).

Beweis: Zum Beweis dieses Satzes müssen wir zeigen, dass die Voraussetzungen des Satz von Lax-Milgram erfüllt sind. Wir müssen also zeigen, dass die Bilinearform B stetig und koerziv ist.

Stetigkeit von B :

Mit der Cauchy-Schwarzschen Ungleichung folgt

$$\begin{aligned}
 B(u, v) &= \int_{\Omega} A(x) \nabla u(x) \cdot \nabla v(x) + a(x) u(x) \cdot \nabla v(x) \\
 &\quad + b(x) \cdot \nabla u(x) v(x) + c(x) u(x) v(x) dx \\
 &\leq \|A\|_{\infty} \|\nabla u\| \|\nabla v\| + \|a\|_{\infty} \|u\| \|\nabla v\| \\
 &\quad + \|b\|_{\infty} \|\nabla u\| \|v\| + \|c\|_{\infty} \|u\| \|v\| \\
 &\leq c_1 \|u\|_{H^1(\Omega)} \|v\|_{H^1(\Omega)},
 \end{aligned}$$

wobei hier stets $\|\cdot\|$ die L^2 -Norm bezeichne und für die Stetigkeitskonstante c_1 gilt

$$c_1 = \|A\|_{\infty} + \|a\|_{\infty} + \|b\|_{\infty} + \|c\|_{\infty}.$$

Koerzivität von B :

Aufgrund der L-Bedingungen (1.10) folgt:

$$\begin{aligned}
 B(u, u) &= \int_{\Omega} A(x) |\nabla u(x)|^2 + a(x) u(x) \cdot \nabla u(x) \\
 &\quad + b(x) \cdot \nabla u(x) u(x) + c(x) |u(x)|^2 dx \\
 &\geq \alpha \|\nabla u\|^2.
 \end{aligned}$$

Mit der Poincaré Ungleichung folgt weiter

$$B(u, u) \geq \alpha \|\nabla u\|^2 \geq \frac{\alpha}{1 + c_p^2} \|u\|_{H^1(\Omega)}^2,$$

so dass die Koerzivität mit $c_0 = \frac{\alpha}{1 + c_p^2}$ gezeigt ist. Mit dem Satz von Lax-Milgram folgt nun die Behauptung.

Bemerkung 1.29 (Reduktion auf Nullrandwerte)

Zur Betrachtung von allgemeinen Dirichletrandwerten, kann man wie folgt vorgehen.

Seien $g_D \in H^1(\Omega)$ und $f \in L^2(\Omega)$. Dann ist $u \in H^1(\Omega)$ schwache Lösung von

$$-\Delta u = f \text{ in } \Omega, \quad u = g_D \text{ auf } \partial\Omega,$$

wenn gilt $\tilde{u} := u - g_D \in H_0^1(\Omega)$ und für alle $\varphi \in H_0^1(\Omega)$

$$(\nabla u, \nabla \varphi) = (f, \varphi). \tag{1.12}$$

Dabei ist zu bemerken, dass mit der Definition von \tilde{u} (1.12) äquivalent ist zu

$$(\nabla \tilde{u}, \nabla \varphi) = (f, \varphi) - (\nabla g_D, \nabla \varphi).$$

Die Existenz und Eindeutigkeit einer Lösung folgt nun durch die Einschicht, dass durch $\tilde{f}(\varphi) := (f, \varphi) - (\nabla g_D, \nabla \varphi)$ ein Element $\tilde{f} \in H_0^{-1}(\Omega)$ definiert wird.

1.4 Regularität für elliptische Randwertprobleme

Wie man in einer Raumdimension unmittelbar einsieht, haben elliptische Differentialgleichungen einen sogenannten regularisierenden Effekt. Betrachtet man z.B. in einer Raumdimension die Differentialgleichung

$$u''(x) = f(x)$$

mit einer stetigen Funktion f , so folgt mit Hilfe des Hauptsatzes der Differential- und Integralrechnung, dass bereits $u \in C^2$ gelten muß. In mehreren Raumdimensionen gilt diese Aussage jedoch nicht mehr, wenn man u'' durch Δu ersetzt. Vielmehr kann man unter geeigneten Voraussetzungen zeigen, dass aus $f \in C^{0,\alpha}(\Omega)$, $\alpha > 0$ folgt, dass $u \in C^{2,\alpha}(\Omega)$ ist. Da wir uns im Kontext von Finite Elemente Verfahren mit schwachen Lösungen beschäftigen, wollen wir an dieser Stelle ein Regularitätsresultat in Sobolevräumen angeben.

Satz 1.30 (Satz von Friedrich)

Seien Ω ein Gebiet mit glatten Rand (es gelte mindestens $\partial\Omega \in C^2$), oder ein konvexe Lipschitz-Gebiet und $f \in L^2(\Omega)$. Dann gilt für die schwache Lösungen $u \in H_0^1(\Omega)$ des Poisson Problems (1.1), (1.2)

$$u \in H^2(\Omega)$$

und es gibt eine Konstante $C > 0$, so dass folgende Abschätzung gilt

$$\|u\|_{H^2(\Omega)} \leq C\|f\|.$$

(Einen Beweis für glatte Gebiete findet man z.B. in [Alt92].)

Bemerkung 1.31

Betrachtet man nicht konvexe Lipschitz-Gebiete, so geht i.A. der regularisierende Effekt von elliptischen Differentialgleichungen verloren. Ein Beispiel hierfür wird in den Übungsaufgaben diskutiert.

Kapitel 2

Finite Elemente Verfahren für elliptische Differentialgleichungen

In diesem Kapitel werden wir uns mit der Approximation von elliptischen Differentialgleichungen zweiter Ordnung mit Hilfe der Finiten Elemente Methode (FEM) beschäftigen. Da die Finite Elemente Verfahren ein Spezialfall der allgemeineren Ritz-Galerkin Verfahren darstellen, wollen wir uns im folgenden Abschnitt zunächst Ritz-Galerkin Verfahren einführen und eine abstrakte Fehlerabschätzung beweisen.

2.1 Ritz-Galerkin Verfahren und abstrakte Fehlerabschätzung

Wie durch Kapitel 2 motiviert, wollen wir uns in diesem Abschnitt mit der Approximation von Variationsproblemen beschäftigen, die die Voraussetzungen des Satzes von Lax-Milgram erfüllen.

Definition 2.1 (Variationsproblem)

Seien X ein reeller Hilbertraum, $B : X \times X \rightarrow \mathbb{R}$ eine stetige und koerzive Bilinearform mit Stetigkeitsmodul c_1 und Koerzivitätskonstante c_0 und $f \in X'$. Dann bezeichnen wir mit $u \in X$ die eindeutige Lösung des Variationsproblems

$$B(u, \varphi) = f(\varphi), \quad \forall \varphi \in X. \quad (2.1)$$

Definition 2.2 (Ritz-Galerkin Verfahren)

Seien X, B wie in Definition 2.1 und $X_h \subset X$ ein Teilraum. Dann ist die Ritz-Galerkin Approximation $u_h \in X_h$ definiert durch

$$B(u_h, \varphi_h) = f(\varphi_h), \quad \forall \varphi_h \in X_h.$$

Bemerkung 2.3

Die Existenz und Eindeutigkeit von u und u_h folgen unmittelbar aus dem Satz von Lax-Milgram 1.5, da der Teilraum X_h wieder ein Hilbertraum mit dem aus X geerbten Skalarprodukt ist.

Satz 2.4 (Abstrakte Fehlerabschätzung / Lemma von Cea)

Seien X, X_h, B, u und u_h wie in den Definition 2.1, 2.2 definiert. Dann gilt die abstrakte Fehlerabschätzung

$$\|u - u_h\|_X \leq \frac{c_1}{c_0} \inf_{v_h \in X_h} \|u - v_h\|_X.$$

Außerdem gilt die Galerkin-Orthogonalität

$$B(u - u_h, v_h) = 0 \quad \forall v_h \in X_h.$$

Beweis: Wir zeigen zunächst die Galerkin-Orthogonalität. Dazu sei $v_h \in X_h$ und es folgt, mit $X_h \subset X$

$$B(u - u_h, v_h) = B(u, v_h) - B(u_h, v_h) = f(v_h) - f(v_h) = 0.$$

Mit der Stetigkeit und Koerzivität von B folgt nun weiter

$$\begin{aligned} c_0 \|u - u_h\|_X^2 &\leq B(u - u_h, u - u_h) = B(u - u_h, u - v_h) \\ &\leq c_1 \|u - u_h\|_X \|u - v_h\|_X. \end{aligned}$$

Nach Division mit $\|u - u_h\|_X$ folgt die Behauptung, indem wir auf beiden Seiten zum Infimum übergehen. \square

Bemerkung 2.5

Die abstrakte Fehlerabschätzung zeigt, dass der Fehler zwischen Ritz-Galerkin Approximation und exakter Lösung abgeschätzt werden kann durch die Bestapproximation in dem Teilraum X_h . Die weitere numerische Analysis beruht

somit allein auf der Approximationstheorie.

Beispiel 2.6 (Mögliche Wahl von Teilräumen)

Betrachten wir konkret ein elliptisches Problem in Divergenzform, d.h. $X = H_0^1(\Omega)$ auf einem Gebiet $\Omega \subset \mathbb{R}^d$, so sind neben den Finite Elemente Verfahren, die wir im nächsten Abschnitt betrachten wollen, vor allem folgende Wahlen von Teilräumen gebräuchlich:

- Polynomräume $X_h := \mathbb{P}^{k(h)}(\Omega) \cap \{v_h \in C^0(\bar{\Omega}) \mid v_h = 0 \text{ auf } \partial\Omega\}$, wobei $\mathbb{P}^{k(h)}(\Omega)$ der Raum der Polynome mit Grad $\leq k(h)$ über Ω sind. Die zugehörigen Verfahren nennt man Spektralverfahren.
- $X_h := \text{span}\{u_i \in X \mid Lu_i = \lambda_i u_i, i = 1, \dots, k(h)\}$, wobei u_i die i-te Eigenfunktion des zugrundeliegenden Differentialoperators L ist.
- $X_h := \text{span}\{u_i \in X \mid \Delta u_i = \lambda_i u_i, i = 1, \dots, k(h)\}$, wobei u_i die i-te Eigenfunktion des Laplace-Operators Δ ist.

Folgerung 2.7 (Matrix-Vektor Form von Ritz-Galerkin Verfahren)

Seien X, X_h, B, u und u_h wie in den Definition 2.1, 2.2 definiert und sei zudem X_h endlichdimensional mit Dimension $N(h) := \dim(X_h)$. Ist dann $\Phi := \{\varphi_1, \dots, \varphi_{N(h)}\}$ eine Basis von X_h so folgt mit der Darstellung $u_h = \sum_{i=1}^{N(h)} U_i \varphi_i$ aus der Definition von u_h :

$$B\left(\sum_{i=1}^{N(h)} U_i \varphi_i, \varphi_j\right) = f(\varphi_j), \quad \forall j = 1, \dots, N(h).$$

Durch Ausnützen der Linearität von B im 1. Argument folgt weiter

$$\sum_{i=1}^{N(h)} B(\varphi_i, \varphi_j) U_i = f(\varphi_j), \quad \forall j = 1, \dots, N(h).$$

Definieren wir also die Matrix $\mathbf{S} \in \mathbb{R}^{N \times N}$ durch

$$\mathbf{S}_{ji} := B(\varphi_i, \varphi_j), \quad \forall i, j = 1, \dots, N(h).$$

und die Vektoren $\mathbf{u}, \mathbf{f} \in \mathbb{R}^N$ durch

$$\begin{aligned} \mathbf{u}_i &:= U_i, \quad \forall i = 1, \dots, N(h), \\ \mathbf{f}_i &:= f(\varphi_i), \quad \forall i = 1, \dots, N(h), \end{aligned}$$

so ist u_h genau dann Lösung des Ritz-Galerkin Verfahrens, wenn \mathbf{u} das folgende lineare Gleichungssystem löst

$$\mathbf{S} \mathbf{u} = \mathbf{f}.$$

2.2 Finite Elemente Verfahren

Finite Elemente Verfahren sind Spezialfälle von Ritz-Galerkin Verfahren für eine bestimmte Klasse von Teilräumen $X_h \subset X$. Die Konstruktion von X_h im Fall von Finite Elemente Verfahren beruht auf einer Zerlegung des Gebietes Ω in nichtüberlappende Teilgebiete, die selbst wiederum einfache Geometrische Objekte sind. Die einfachste Klasse von Finiten Elementen sind Lagrange-Elemente auf simplizialen Gittern. Diese wollen wir im folgenden Unterabschnitt einführen.

2.2.1 Simpliziale Lagrange Elemente

Ziel dieses Abschnitts ist die Konstruktion geeigneter endlichdimensionaler Teilräume X_h , die auf einer simplizialen Zerlegung/Triangulierung des Gebietes Ω beruhen. In zwei Raumdimensionen besteht das Rechengitter aus Dreiecken, in drei Raumdimensionen aus Tetraedern. Eingeschränkt auf einen Simplex wird eine Funktion aus X_h dann ein Polynom mit $\text{Grad} \leq k, k \in \mathbb{N}$ sein.

Definition 2.8 (Simplex)

Seien $s \in \{1, \dots, d\}$ und $a_0, \dots, a_s \in \mathbb{R}^d$ Punkte, so dass $(a_j - a_0)_{j=1, \dots, s}$ linear unabhängig sind. Dann heißt

$$T := \{x \in \mathbb{R}^d \mid x = \sum_{i=0}^s \lambda_i a_i, 0 \leq \lambda_i, \sum_{i=0}^s \lambda_i = 1\}$$

nicht degeneriertes s -dimensionales Simplex im \mathbb{R}^d .

Die Punkte a_0, \dots, a_s heißen Ecken des Simplex. Ist $r \in \{1, \dots, s\}$ und $\tilde{a}_0, \dots, \tilde{a}_r \in \{a_0, \dots, a_s\}$, so heißt

$$\tilde{T} := \{x \in \mathbb{R}^d \mid x = \sum_{i=0}^r \lambda_i \tilde{a}_i, 0 \leq \lambda_i, \sum_{i=0}^r \lambda_i = 1\}$$

r -dimensionales Seitensimplex von T . Die eindimensionalen Seitensimplexe heißen Kanten, die nulldimensionalen Ecken und die $s - 1$ dimensionale Flächen.

Wir bezeichnen mit T_0 den Simplex zu den Punkten $a_0 = 0, a_i = e_i, i = 1, \dots, d$. T_0 heißt d -dimensionaler Einheitssimplex.

Der Durchmesser von T ist gegeben durch

$$h(T) := \text{diam}(T) = \max_{i,j=1}^s |a_i - a_j|.$$

Mit $\rho(T) := 2 \sup\{R \mid B_R(x_0) \subset T\}$ bezeichnen wir den Innenkugeldurchmesser von T und mit $\sigma(T) := \frac{h(T)}{\rho(T)}$ den Quotienten aus h und ρ .

Bemerkung 2.9

Im \mathbb{R}^2 ist ein zweidimensionales Simplex das Dreieck mit den Ecken a_0, a_1, a_2 , im \mathbb{R}^3 ist ein dreidimensionales Simplex der Tetraeder mit den Ecken a_0, a_1, a_2, a_3 , sechs eindimensionalen und vier zweidimensionalen Seitensimplexen. Ein s -dimensionales Simplex besitzt $\binom{s+1}{r+1}$ r -dimensionale Seitensimplexe.

Die Größen $\lambda_0, \dots, \lambda_s \in [0, 1]$ sind Koordinaten, die dem Simplex besonders gut angepaßt sind.

Definition 2.10 (Baryzentrische Koordinaten)

Die baryzentrischen Koordinaten $\lambda_0, \dots, \lambda_s \in [0, 1]$ eines Punktes $x \in T$ des s -dimensionalen Simplex T sind die Lösung des linearen Gleichungssystems

$$x = \sum_{i=0}^s \lambda_i a_i, \quad \sum_{i=0}^s \lambda_i = 1.$$

Der Schwerpunkt x_s von T ist definiert durch

$$x_s := \frac{1}{s+1} \sum_{i=0}^s \tilde{a}_i$$

und hat die baryzentrischen Koordinaten $\lambda_i = \frac{1}{s+1}$.

Für die Eckpunkte a_k von T sind die baryzentrischen Koordinaten gegeben durch $\lambda_k = 1, \lambda_i = 0, i \neq k$.

Lemma 2.11 (Referenzabbildung)

Sei T_0 das Einheitssimplex im \mathbb{R}^s . Dann ist jedes s -dimensionale Simplex T affin äquivalent zu T_0 . Die eindeutige affine Abbildung $F : T_0 \rightarrow T, F(x) = Ax + b, A \in \mathbb{R}^{s \times s}, b \in \mathbb{R}^s, \det A \neq 0$ mit $F(e_j) = a_j, j = 0, \dots, s$ heißt Referenzabbildung. F ist invertierbar und es gelten die Abschätzungen

$$\begin{aligned} \|\nabla F\| &= \|A\| \leq \frac{h(T)}{\rho(T_0)}, \\ \|\nabla F^{-1}\| &= \|A^{-1}\| \leq \frac{h(T_0)}{\rho(T)}, \\ c\rho(T)^s &\leq |\det(\nabla F)| = |\det A| \leq Ch(T)^s. \end{aligned}$$

Beweis: Zunächst sehen wir ein, dass $\{e_j | j = 1, \dots, s\}$ und $\{a_j - a_0 | j = 1, \dots, s\}$ zwei Basen des \mathbb{R}^s sind. Definieren wir dann die Matrix A als Basistransformation

$$Ae_j = a_j - a_0, \quad j = 1, \dots, s,$$

so ist die Referenzabbildung F gegeben durch $F(x) = Ax + a_0$.

Für die weiteren Aussagen verweisen wir auf die Übungsaufgaben.

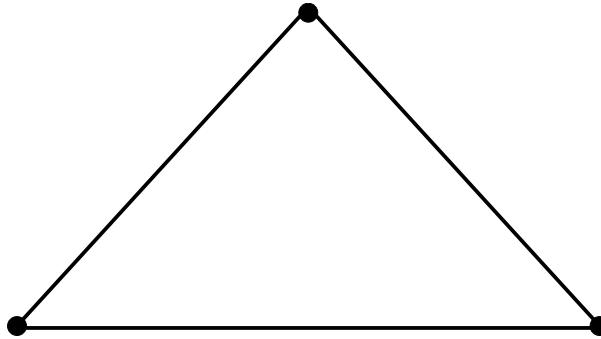


Abbildung 2.1: Lineares Lagrange Element.

Definition 2.12 (Zulässige Triangulierung)

Sei $\Omega \subset \mathbb{R}^d$ ein Gebiet und

$$\mathcal{T}_h := \{T_j \mid j = 1, \dots, m, T_j \text{ ist } d\text{-dimensionaler Simplex im } \mathbb{R}^d\}.$$

\mathcal{T}_h heißt zulässige Triangulierung der Feinheit h und Güte ρ von Ω falls gilt

- $\bar{\Omega} = \bigcup_{j=1}^m T_j$, $\partial\Omega = \bigcup_{j=1}^m \tilde{T}_j$, wobei die \tilde{T}_j Flächen der Simplexe T_j sind.
- Für je zwei $T_1, T_2 \in \mathcal{T}_h$ mit $S := T_1 \cap T_2$ gilt: $S = \emptyset$ oder S ist $(d - k)$ -dimensionales Seitensimplex von T_1 und T_2 für ein $k \in \{0, \dots, d\}$.
- $h := \max_{j=1, \dots, m} h(T_j)$, $\rho := \min_{j=1, \dots, m} \rho(T_j)$.

Zur Definition von Finite Elemente Räumen basierend auf einer Triangulierung \mathcal{T}_h müssen wir nun lediglich lokale Funktionenräume auf den Simplexen $T \in \mathcal{T}_h$ angeben und festlegen, wie solche lokalen Funktionen global zusammengesetzt werden. Ein Tripel bestehend aus einem geometrischen Objekt T , einer lokalen Basis Φ und lokalen Freiheitsgraden N wollen wir im folgenden Element nennen.

Definition 2.13 (Lineares simpliziales Lagrange Element)

Sei $T \subset \mathbb{R}^d$ ein d -dimensionaler Simplex. Sei $\mathcal{N} := \{a_i \mid i = 0, \dots, d\}$ die Menge der Ecken von T . Dann ist durch die Angabe von Werten in den Punkten $a_i \in \mathcal{N}$ eindeutig eine lineare Funktion $p \in \mathbb{P}^1(T)$ definiert. Durch $\Phi := \{\varphi_i \mid \varphi_i(a_i) = \delta_{ik}, i, k = 0, \dots, d\}$ ist eine nodale Basis von $\mathbb{P}^1(T)$ gegeben. Wir nennen das Tripel (T, Φ, \mathcal{N}) lineares simpliziales Lagrange Element. Die Basisfunktionen $\varphi_i \in \Phi$ werden Formfunktionen oder im Englischen shape functions genannt und \mathcal{N} ist die Menge der nodale Variablen.

Bemerkung 2.14

Zur Wohldefiniertheit des linearen simplizialen Lagrange Elements müssen wir einsehen, dass die Basis Φ von $\mathbb{P}^1(T)$ durch die Bedingungen $\varphi_i(a_k) = \delta_{ik}$, $i, k = 0, \dots, d$ eindeutig bestimmt ist. Eine Funktion $u_h \in \mathbb{P}^1(T)$ hat die Form $u_h(x) = \alpha_0 + \sum_{i=1}^d \alpha_i x_i$. Somit ist $\varphi_i(a_k) = \delta_{ik}$, $k = 0, \dots, d$ ein lineares Gleichungssystem mit $d+1$ Gleichungen und $d+1$ Unbekannten. Die eindeutige Lösbarkeit folgt also, wenn wir eine Lösung angeben können. Dies kann auf einfache Weise mit Hilfe der baryzentrischen Koordinaten erreicht werden. Seien dazu $\lambda := \{\lambda_0, \dots, \lambda_1\}$ die baryzentrischen Koordinaten eines Punktes $x \in T$. Dann setzen wir

$$\varphi_i(x) = \varphi_i(x(\lambda)) = \sum_{j=0}^d \delta_{ij} \lambda_j.$$

Die Behauptung folgt dann aus der Eigenschaft, dass für $\lambda = e_{k+1} \in \mathbb{R}^{d+1}$ gilt $a_k = x(e_{k+1})$:

$$\varphi_i(a_k) = \varphi_i(x(e_{k+1})) = \sum_{j=0}^d \delta_{ij} (e_{k+1})_{j+1} = \sum_{j=0}^d \delta_{ij} \delta_{jk} = \delta_{ik}.$$

Beispiel 2.15 (Lineares Lagrange Element für $d = 2$)

Sei $d = 2$. Wir betrachten das Einheitsdreieck T_0 mit Eckpunkten $(0, 0)$, $(1, 0)$ und $(0, 1)$. Die Formfunktionen sind dann gegeben durch

$$\begin{aligned} \tilde{\varphi}_0(x, y) &= 1 - x - y, \\ \tilde{\varphi}_1(x, y) &= x, \\ \tilde{\varphi}_2(x, y) &= y. \end{aligned}$$

Sind $p(\tilde{N}_0), p(\tilde{N}_1), p(\tilde{N}_2)$ Funktionswerte einer linearen Funktion $\tilde{p} \in \mathbb{P}^1(T_0)$ in den Eckpunkten $\tilde{N}_0 := (0, 0)$, $\tilde{N}_1 := (1, 0)$ und $\tilde{N}_2 := (0, 1)$, so ist p gegeben durch

$$\tilde{p}(x, y) = \sum_{i=0}^2 \tilde{p}(N_i) \tilde{\varphi}_i(x, y).$$

Für ein beliebiges Dreieck $T \subset \mathbb{R}^2$ erhält man das lineare Lagrange Element mit Hilfe der Referenzabbildung $F : T_0 \rightarrow T$ (vgl. Lemma 2.11), wie wir wie

folgt einsehen. Wir definieren die Formfunktionen $\varphi_i, i = 0, 1, 2$ durch

$$\varphi_i(x, y) := \tilde{\varphi}_i(F^{-1}(x, y))$$

Da F eine affine Abbildung ist, folgt $\varphi \in \mathbb{P}^1(T_0)$. Außerdem gilt für die Eckpunkte a_0, a_1, a_2 von T nach der Definition der Referenzabbildung

$$\varphi_i(a_j) = \tilde{\varphi}_i(F^{-1}(a_j)) = \tilde{\varphi}_i(\tilde{N}_j) = \delta_{ij}, \quad i, j = 0, 1, 2.$$

Sind also $p(a_0), p(a_1), p(a_2)$ Funktionswerte einer linearen Funktion $p \in \mathbb{P}^1(T)$ in den Eckpunkten a_0, a_1 und a_2 , so ist p gegeben durch

$$p(x, y) = \sum_{i=0}^2 p(a_i) \varphi_i(x, y).$$

Bemerkung 2.16

Das Beispiel 2.14 zeigt, dass es ausreicht ein Finites Element auf einer Referenzgeometrie zu definieren. Durch die Referenzabbildung erhält man dann die entsprechende Klasse von Elementen auf beliebigen Geometrien im Raum.

Ein Finites Element legt lediglich einen lokalen Funktionenraum auf einem Simplex T fest. Um zu einem Unterraum von $H_0^1(\Omega)$ zu gelangen, müssen wir zusätzlich festlegen auf welche Weise die lokalen Funktionen global zusammengesetzt werden. Dies geschieht in der folgenden Definition.

Definition 2.17 (Linearer Finite Elemente Raum S_h^1)

i) Sei $\Omega \subset \mathbb{R}^d$ und \mathcal{T}_h eine zulässige Triangulierung von Ω . Wir definieren den Raum der linearen Finite Elemente auf simplizialen Gittern S_h^1 durch

$$S_h^1 := \{v_h \in C^0(\Omega) \mid v_h|_T \in \mathbb{P}^1(T), T \in \mathcal{T}_h\}.$$

ii) Sind $\bar{a}_j, j = 1, \dots, N$ die Ecken der Triangulierung \mathcal{T}_h , so ist eine Funktion $v_h \in S_h^1$ durch die Vorgabe von Funktionswerten in den Ecken $v_h(\bar{a}_j)$ eindeutig definiert. Insbesondere gilt $\dim(S_h^1) = N$.

Eine Basis von S_h^1 ist durch die Funktionen

$$\bar{\varphi}_i \in S_h^1, \bar{\varphi}_i(\bar{a}_j) = \delta_{ij}, \quad i, j = 1, \dots, N$$

gegeben. Diese Basis heißt Knotenbasis.

iii) Ist $(T_0, \tilde{\Phi}, \tilde{N})$ das lineare Lagrange Element auf dem Einheitssimplex T_0 und $v_h \in S_h^1$ gegeben durch

$$v_h(x) = \sum_{i=1}^N v_h(\bar{a}_i) \bar{\varphi}_i(x)$$

so gilt für beliebige Simplexe $T \in \mathcal{T}_h$ mit Ecken a_0, \dots, a_d

$$v_h|_T(x) = \sum_{i=1}^d v_h(a_i) \tilde{\varphi}_i(F^{-1}(x)),$$

wobei $F : T_0 \rightarrow T$ die Referenzabbildung und $\tilde{\varphi}_i \in \tilde{\Phi}$ die Formfunktionen des Referenzelements T_0 sind.

Bemerkung 2.18

Um zu zeigen, dass die obige Definition Sinn macht, stellen wir zunächst fest, dass $v_h|_T \in \mathbb{P}^1(T)$ durch die Vorgabe von Werten in den Eckpunkten der Triangulierung wohldefiniert ist. Zu zeigen ist dann noch die globale Stetigkeit von v_h . Dies folgt jedoch direkt daraus, dass für zwei Simplexe $T_1, T_2 \in \mathcal{T}_h$ mit einer gemeinsamen Fläche $S = T_1 \cap T_2$ gilt $v_h|_S \in \mathbb{P}^1(S)$. Analog zum ersten Teil folgt, dass $v_h|_S$ durch die Vorgabe von Werten in den Eckpunkten der Triangulierung eindeutig bestimmt ist, was wiederum $v_h \in C^0(\Omega)$ impliziert.

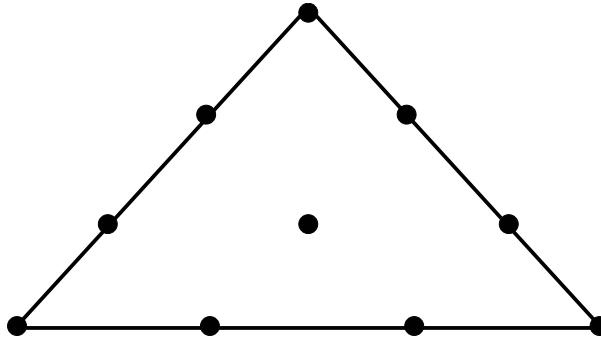


Abbildung 2.2: Kubisches Lagrange Element.

Definition 2.19 (Lineares Finite Elemente Verfahren)

Sei $\Omega \subset \mathbb{R}^d$ und \mathcal{T}_h eine zulässige Triangulierung von Ω . Seien $X := H_0^1(\Omega)$ und $X_h := S_{h,0}^1 := S_h^1 \cap \{v \in C^0(\bar{\Omega}) | v = 0 \text{ auf } \partial\Omega\}$. Weiter seien eine stetige und koerzive Bilinearform $B : X \times X \rightarrow \mathbb{R}$ und ein $f \in X'$ gegeben.

Dann ist $X_h \subset X$ ein Teilraum und $u_h \in X_h$ heißt Lösung des linearen Finite Elemente Verfahrens für das Variationsproblem (2.1), falls gilt

$$B(u_h, v_h) = f(v_h), \quad \forall v_h \in X_h.$$

Nach der Diskussion linearer Lagrange Elemente, wollen wir nun allgemeine Lagrange Elemente k -ter Ordnung definieren. Dabei werden wir auf Beweise der Wohldefiniertheit verzichten und verweisen hierfür auf die Literatur [Bra97, BS08, Cia87, Sch91].

Definition 2.20 (Allgemeines Lagrange Element)

Sei $T \subset \mathbb{R}^d$ ein d -dimensionaler Simplex. Wir definieren auf T das sogenannte Lagrange Gitter k -ter Ordnung $G_k(T)$ durch

$$G_k(T) := \{x = \sum_{j=0}^d \lambda_j a_j | \lambda_j \in \{\frac{m}{k} | m = 0, \dots, k\}, 0 \leq \lambda_j, \sum_{j=0}^d \lambda_j = 1\}.$$

Die Anzahl der Punkte in $G_k(T)$ ist $g_k := \binom{d+k}{k}$. Durch die Angabe von Werten $p(N_i)$ in den Punkten $N_i \in \mathcal{N} := G_k(T)$, $i = 1, \dots, g_k$, ist eindeutig eine Funktion $p \in \mathbb{P}^k(T)$ definiert. Durch $\Phi := \{\varphi_i | \varphi_i(N_k) = \delta_{ik}, i, k = 0, \dots, g_k\}$ ist eine nodale Basis des Raumes $\mathbb{P}^k(T)$ gegeben. Wir nennen das Tripel (T, Φ, \mathcal{N}) simpliziales Lagrange Element von Grad k .

Definition 2.21 (Lagrange Finite Elemente Raum S_h^k)

i) Sei $\Omega \subset \mathbb{R}^d$ und \mathcal{T}_h eine zulässige Triangulierung von Ω . Wir definieren den Raum der Lagrange Finite Elemente auf simplizialen Gittern S_h^k durch

$$S_h^k := \{v_h \in C^0(\Omega) \mid v_h|_T \in \mathbb{P}^k(T), T \in \mathcal{T}_h\}.$$

ii) Wir definieren das Lagrange-Gitter k -ter Ordnung $G_k(\mathcal{T}_h)$ durch

$$G_k(\mathcal{T}_h) := \{\bar{x}_j \mid j = 1, \dots, N\} := \bigcup_{T \in \mathcal{T}_h} G_k(T).$$

Sind $\bar{x}_j, j = 1, \dots, N$ die Punkte des Lagrange-Gitters der Triangulierung \mathcal{T}_h , so ist eine Funktion $v_h \in S_h^k$ durch die Vorgabe von Funktionswerten $v_h(\bar{x}_j)$ eindeutig definiert. Eine Basis von S_h^k ist durch folgende Funktionen gegeben

$$\bar{\varphi}_i \in S_h^k, \bar{\varphi}_i(\bar{x}_j) = \delta_{ij}, \quad i, j = 1, \dots, N$$

iii) Ist $(T_0, \tilde{\Phi}, \tilde{N})$ das Lagrange Element k -ter Ordnung auf dem Einheitssimplex T_0 und $v_h \in S_h^k$ gegeben durch

$$v_h(x) = \sum_{i=1}^N v_h(\bar{x}_i) \bar{\varphi}_i(x),$$

so gilt für beliebige Simplexe $T \in \mathcal{T}_h$ mit Punkten $x_i \in G_k(T)$

$$v_h|_T(x) = \sum_{i=1}^{g_k} v_h(x_i) \tilde{\varphi}_i(F^{-1}(x)),$$

wobei $F : T_0 \rightarrow T$ die Referenzabbildung und $\tilde{\varphi}_i \in \tilde{\Phi}$ die Formfunktionen des Lagrange-Referenzelements k -ter Ordnung sind.

2.2.2 Verallgemeinerung und weitere gebräuchliche Elemente

Im letzten Abschnitt haben wir Lagrange Elemente auf simplizialen Gittern betrachtet. Im Folgenden wollen wir zunächst die Definition eines Finiten Elements verallgemeinern und schließlich weitere gebräuchliche Elemente einführen.

Definition 2.22 (Finites Element)

Seien

- (i) $T \subset \mathbb{R}^d$ eine beschränkte, abgeschlossene Teilmenge mit nicht leerem Inneren und Lipschitz-Rand, die Element-Geometrie,
- (ii) $\Phi := \{\varphi_1, \dots, \varphi_k\}$ linear unabhängige Funktionen auf T , die Formfunktionen und $\mathcal{P} := \text{span } \Phi$ der zugehörige diskrete Funktionenraum mit $\dim(\mathcal{P}) = k$,
- (iii) $\mathcal{N} := \{N_1, \dots, N_k\}$ einen Basis von \mathcal{P}' , die Menge der lokalen Freiheitsgrade.

Dann heißt das Tripel (T, Φ, \mathcal{N}) Finites Element.

Die lokalen Freiheitsgrade \mathcal{N} heißen nodale Variablen und die Formfunktionen Φ nodale Basis, falls für alle $i, j = 1, \dots, k$ gilt

$$N_i(\varphi_j) = \delta_{ij}.$$

Bemerkung 2.23

- (i) Durch die Wahl von \mathcal{N} als Basis des Dualraums von \mathcal{P} verallgemeineren wir die Definition eines Finiten Elements aus Abschnitt 2.2.1.

Für die Menge von Eckpunkten $\{a_i | i = 1, \dots, d\}$ aus der Definition des linearen Lagrange Elements, definieren wir beispielsweise $N_i \in \mathcal{P}'$ durch

$$N_i(p) := p(a_i), \quad \forall p \in \mathcal{P}.$$

Dann folgt für alle $\varphi_j \in \Phi$

$$N_i(\varphi_j) = \varphi_j(a_i) = \delta_{ij}.$$

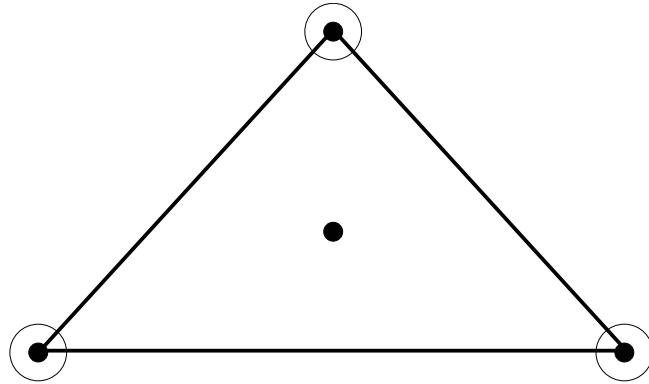


Abbildung 2.3: Kubisches Hermit Element.

- (ii) Anstelle von Simplizes können auch andere geometrische Elemente verwendet werden. Auf Rechtecken ($d = 2$) und Würfeln ($d = 3$) verwendet man beispielsweise anstelle von linearen Formfunktionen, bi-, b.z.w. tri-lineare Funktionen. Auf den Einheitsquadern in \mathbb{R}^d führt man allgemein folgende Funktionenräume ein

$$Q^1([0, 1]^d) := \bigotimes_{i=1}^d \mathbb{P}^1([0, 1]).$$

Für $d = 2$ haben Polynome $p \in Q^1([0, 1]^2)$ also die Form

$$p(x, y) = a + bx + cy + dxy.$$

Die resultierenden Lagrange Finite Elemente Räume werden mit Q_h^k bezeichnet.

Bislang haben wir nur H^1 -Elemente betrachtet, da sich beispielsweise simpliciale Lagrange Elemente global nicht zu Funktionen in $H^m(\Omega)$ mit $m > 1$ zusammensetzen lassen. Moechte man H^m -Elemente für $m > 1$ definieren, so kommen als Freiheitsgrade nicht nur die Funktionswerte sondern auch die Werte der Ableitungen einer Funktion in Betracht. Das erste Element, dass in diese Richtung geht, aber noch kein H^2 -Element ist, ist das kubische Hermit Element.

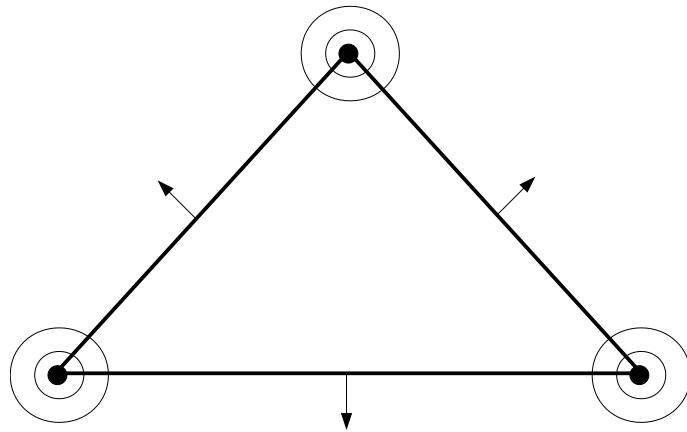


Abbildung 2.4: Agyris Element.

Definition 2.24 (Kubisches Hermit Element auf Dreiecken)

Sei $T \subset \mathbb{R}^2$ ein Dreieck. Sei $\{a_i | i = 0, \dots, 2\}$ die Menge der Ecken von T und x_s der Schwerpunkt von T . Dann ist durch die Angabe von Werten $p(a_i), \nabla p(a_i), i = 0, \dots, d$ und $p(x_s)$ eindeutig ein Polynom $p \in \mathbb{P}^3(T)$ definiert.

Die Menge der lokalen Freiheitsgrade \mathcal{N} ist gegeben durch

$$N_i(p) = \begin{cases} p(a_i), & i = 0, \dots, 2, \\ \partial_j p(a_i), & i = 3j, \dots, 3j + 2, \\ p(x_s), & i = 9. \end{cases}$$

Durch $\Phi := \{\varphi_j | N_i(\varphi_j) = \delta_{ij}, i, j = 0, \dots, 2\}$ ist eine nodale Basis von $\mathbb{P}^3(T)$ eindeutig definiert. Das Tripel (T, Φ, \mathcal{N}) heißt kubisches Hermit Element.

Als nächstes Element geben wir als ein echtes H^2 -Element das Agyris Element an.

Definition 2.25 (Agyris Element auf Dreiecken)

Sei $T \subset \mathbb{R}^2$ ein Dreieck. Seien $\{a_i | i = 0, \dots, 2\}$ die Menge der Ecken von T , $\{b_i | i = 0, \dots, 2\}$ die Mittelpunkte der Kanten von T und x_s der Schwerpunkt von T . Dann ist durch die Angabe von Werten $p(a_i), \nabla p(a_i), \nabla^2 p(a_i)$ und $\nabla p(b_i) \cdot n_i, i = 0, \dots, d$ eindeutig ein Polynom $p \in \mathbb{P}^5(T)$ definiert. Dabei ist n_i die äußere Normale an die i -te Kante von T .

2.3 Approximationssätze und Fehlerabschätzungen

In Abschnitt 2.1 haben wir gesehen, dass für Ritz-Galerkin Verfahren und somit auch für Finite Elemente Verfahren die abstrakte Fehlerabschätzung

$$\|u - u_h\|_X \leq \frac{c_1}{c_0} \inf_{v_h \in X_h} \|u - v_h\|_X.$$

gilt, wobei X, X_h, B, u und u_h wie in den Definition 2.1, 2.2 definiert sind. Ist $X_h = S_{h,0}^k$ der Lagrange Finite Elemente Raum, so können wir für Funktionen $u \in H^m(\Omega)$, falls m hinreichend groß ist, einen Interpolationsoperator $I_h : H^m(\Omega) \rightarrow S_{h,0}^k$ definieren. Aus der abstrakte Fehlerabschätzung erhalten wir dann

$$\|u - u_h\|_X \leq \frac{c_1}{c_0} \|u - I_h(u)\|_X.$$

Um also zu einer Fehlerabschätzung für Lagrange Finite Elemente Verfahren zu kommen, benötigen wir eine Abschätzung für den Interpolationsfehler $\|u - I_h(u)\|_X$.

2.3.1 Interpolationabschätzungen

Sei X ein Sobolevraum, $X_h \subset X$ ein Finite Elemente Raum und $I_h \in L(X, X_h)$ ein Interpolationsoperator. Ist $u \in X$, so sind wir an Fehlerabschätzungen der Form

$$\|u - I_h(u)\|_X \leq Ch^\alpha$$

mit einer möglichst kleinen Konstante $C > 0$ und einem möglichst großen Exponenten $\alpha > 0$ interessiert.

Da Finite Elemente Räume X_h lokal durch Funktionenräume auf den Elementgeometrien festgelegt sind, werden wir zur Herleitung von Interpolationabschätzungen zunächst lokale Approximationseigenschaften von Interpolationsoperatoren untersuchen. Dies geschieht mit Hilfe eines Skalierungsarguments (Transformation auf ein Referenzelement) und einer entsprechenden Abschätzung auf dem Referenzelement. Um dabei optimale Potenzen der Gitterweite h zu erhalten, dürfen nur höchste Ableitungen in der Norm vorkommen. Dies erreicht man durch sukzessives Anwenden der Poincaré Ungleichung für $H^{1,p}$ -Funktionen mit Mittelwert Null.

Satz 2.26 (Poincaré Ungleichung mit Mittelwert Null)

Seien $\Omega \subset \mathbb{R}^d$ ein konvexes Lipschitzgebiet, $u \in H^{m,p}(\Omega)$ und gelte

$$\int_{\Omega} D^{\alpha} u = 0, \quad \forall |\alpha| = 0, \dots, m-1.$$

Dann existiert eine Konstante $C > 0$, die nur von Ω, m und p abhängt, so dass gilt

$$\|u\|_{H^{m,p}(\Omega)} \leq C|u|_{H^{m,p}(\Omega)}.$$

(ohne Beweis)

Lemma 2.27

Zu $u \in H^{k+1,p}(\Omega)$ gibt es genau ein Polynom $p \in \mathbb{P}^k(\Omega)$, so dass gilt

$$\int_{\Omega} D^{\alpha} (u - p) = 0, \quad \forall |\alpha| = 0, \dots, k.$$

Beweis: Ein Polynom $p \in \mathbb{P}^k(\Omega)$ hat die Form

$$p(x) = \sum_{|\beta|=0}^k c_{\beta} x^{\beta}.$$

Wie müssen also zeigen, dass das lineare Gleichungssystem

$$\sum_{|\beta|=0}^k c_{\beta} \int_{\Omega} D^{\alpha} x^{\beta} dx = \int_{\Omega} D^{\alpha} u(x) dx$$

für die Unbekannten $c^{\beta}, |\beta| = 0, \dots, k$ eindeutig lösbar ist. Da die Anzahl der Gleichungen gleich der Anzahl der Unbekannten ist, reicht es aus die Eindeutigkeit der Lösung zu zeigen. Sei also

$$\sum_{|\beta|=0}^k c_{\beta} \int_{\Omega} D^{\alpha} x^{\beta} dx = 0, \quad \forall |\alpha| = 0, \dots, k$$

so folgt

$$\int_{\Omega} D^{\alpha} p = 0, \quad \forall |\alpha| = 0, \dots, k$$

und dies ist nur dann erfüllt, wenn $p = 0$ gilt. \square .

Satz 2.26 und Lemma 2.27 zusammen führen zu folgendem Satz.

Satz 2.28

Seien $\Omega \subset \mathbb{R}^d$ ein konvexes Lipschitzgebiet und $k \in \mathbb{N}$. Dann gibt es eine Konstante $C > 0$, so dass für alle $u \in H^{k+1,p}(\Omega)/\mathbb{P}^k(\Omega)$ gilt

$$\|u\|_{H^{k+1,p}(\Omega)/\mathbb{P}^k(\Omega)} \leq C|u|_{H^{k+1,p}(\Omega)},$$

dabei verwenden wir die Definition

$$\|u\|_{H^{k+1,p}(\Omega)/\mathbb{P}^k(\Omega)} := \inf_{p \in \mathbb{P}^k} \|u - p\|_{H^{k+1,p}(\Omega)}.$$

Wir haben nun die Vorbereitungen getroffen, um einen abstrakten Satz zur Interpolationabschätzungen beweisen zu können. Wir werden diesen Satz später zur lokalen Abschätzung auf dem Referenzelement T_0 verwenden.

Satz 2.29 (Variante des Lemmas von Bramble-Hilbert)

Seien $\Omega \subset \mathbb{R}^d$ ein konvexes Lipschitzgebiet und $k, m \in \mathbb{N}_0$, $p, q \geq 1$, so dass

$$E \in L(H^{k+1,p}(\Omega), H^{m,q}(\Omega))$$

eine Einbettung ist, die auf $\mathbb{P}_k(\Omega)$ die Identität ist. Sei weiter

$$I \in L(H^{k+1,p}(\Omega), H^{m,q}(\Omega))$$

ein Interpolationsoperator, der $\mathbb{P}_k(\Omega)$ invariant lässt, d.h. es gilt $Is = s, \forall s \in \mathbb{P}_k(\Omega)$. Dann gibt es eine Konstante $C > 0$, die nur von $\|E\|, \|I\|, \Omega, k, m, p$ und q abhängt, so dass für alle $u \in H^{k+1,p}(\Omega)$ gilt

$$\|E(u) - I(u)\|_{H^{m,q}(\Omega)} \leq C|u|_{H^{k+1,p}(\Omega)}.$$

Beweis: Für $s \in \mathbb{P}_k(\Omega)$ gilt

$$\begin{aligned} \|E(u) - I(u)\|_{H^{m,q}(\Omega)} &= \|E(u) - I(s) + I(s) - I(u)\|_{H^{m,q}(\Omega)} \\ &= \|E(u - s) - I(u - s)\|_{H^{m,q}(\Omega)} \\ &\leq (\|E\| + \|I\|)\|u - s\|_{H^{k+1,p}(\Omega)} \\ &\leq (\|E\| + \|I\|)\|u\|_{H^{k+1,p}(\Omega)/\mathbb{P}_k(\Omega)}. \end{aligned}$$

Mit Satz 2.28 folgt dann die Behauptung.

Im folgenden Satz wollen wir als Vorbereitung von Interpolationsabschätzungen auf beliebigen Elementgeometrien, das Skalierungsverhalten vom $H^{m,p}$ -Seminormen untersuchen.

Satz 2.30 (Transformationsabschätzungen)

Seien $T_0, T \subset \Omega$ offen, beschränkt und affin äquivalent, d.h. es gebe eine Abbildung $F : T_0 \rightarrow T$, $x = F(y) = Ay + b$ mit $T = F(T_0)$. Seien $m \in \mathbb{N}_0$ und $p \in [1, \infty]$. Dann gelten für $u \in H^{m,p}(T)$ und $v \in H^{m,p}(T_0)$ definiert durch $v(y) := u(F(y))$ die Abschätzungen:

$$\begin{aligned} |v|_{H^{m,p}(T_0)} &\leq c_1 \|A\|^m |\det A|^{-1/p} |u|_{H^{m,p}(T)}, \\ |u|_{H^{m,p}(T)} &\leq c_2 \|A^{-1}\|^m |\det A|^{1/p} |v|_{H^{m,p}(T_0)}. \end{aligned}$$

Beweis: Da C^m dicht in $H^{m,p}$ ist, reicht es aus die Behauptung für $u \in C^m(T) \cap H^{m,p}(T)$ und $v \in C^m(T_0) \cap H^{m,p}(T_0)$ zu zeigen.

Es gilt nach der Kettenregel mit $x = F(y)$

$$\partial_{y_j} v(y) = \partial_{y_j} u(F(y)) = \sum_{i=1}^d \partial_{x_i} u(F(y)) \partial_{y_j} F_i(y) = \sum_{i=1}^d \partial_{x_i} u(x) A_{ij}.$$

Somit folgt

$$|\partial_{y_j} v(y)| \leq \|A^\top \nabla_x u(F(y))\| \leq \|A\| \|\nabla_x u(x)\|.$$

Analog erhält man

$$|\partial_{y_i, y_j} v(y)| \leq \|A\|^2 \|\nabla_x^2 u(x)\|$$

und induktiv beweist man, dass für Multiindizes α mit $|\alpha| = m$ gilt

$$|D_y^\alpha v(y)| \leq \|A\|^m \left(\sum_{|\beta|=m} (D_x^\beta u(x))^2 \right)^{1/2} \leq c(d, m) \|A\|^m \sum_{|\beta|=m} |D_x^\beta u(x)|.$$

Somit folgt für die L^p -Norm durch Integration und Verwendung der Dreiecksungleichung

$$\|D^\alpha v\|_{L^p(T_0)} \leq c(d, m) \|A\|^m \sum_{|\beta|=m} \|(D_x^\beta u) \circ F\|_{L^p(T_0)}.$$

Weiter folgt mit Substitution

$$\begin{aligned} \|(D^\beta u) \circ F\|_{L^p(T_0)} &= \left(\int_{T_0} |D^\beta u(F(y))|^p dy \right)^{1/p} \\ &= \left(\int_T |D^\beta u(x)|^p |\det A^{-1}| dx \right)^{1/p} \\ &= |\det A|^{-1/p} \|D^\beta u\|_{L^p(T)}. \end{aligned}$$

Also erhalten wir insgesamt

$$\begin{aligned} |v|_{H^{m,p}(T_0)} &= \left(\sum_{|\alpha|=m} \|D^\alpha v\|_{L^p(T_0)}^p \right)^{1/p} \\ &\leq c(d, m) \|A\|^m |\det A|^{-1/p} \left(\sum_{|\beta|=m} \|D^\beta u\|_{L^p(T)}^p \right)^{1/p} \\ &= c(d, m) \|A\|^m |\det A|^{-1/p} |u|_{H^{m,p}(T)}. \end{aligned}$$

Die zweite Abschätzung folgt analog, indem man die Rollen von u und v vertauscht und F durch F^{-1} ersetzt. \square

Ist T_0 der Einheitssimplex und F die Referenzabbildung auf einen Simplex T , so erhalten wir mit den Abschätzungen aus Lemma 2.11, folgende Transformationsabschätzungen.

Korollar 2.31 (Transformation von Seminormen auf Simplizes)

Seien T ein d -dimensionaler Simplex, T_0 der zugehörige Einheitssimplex und $F : T_0 \rightarrow T$, $F(y) = Ay + b$ die Referenzabbildung (vgl. Lemma 2.11). Ist dann $u \in H^{m,p}(T)$ und $v \in H^{m,p}(T_0)$ definiert durch $v(y) := u(F(y))$, so gelten die Abschätzungen

$$\begin{aligned} |v|_{H^{m,p}(T_0)} &\leq c_1(d, m, p) \left(\frac{h(T)}{\rho(T_0)} \right)^m \rho(T)^{-d/p} |u|_{H^{m,p}(T)}, \\ |u|_{H^{m,p}(T)} &\leq c_2(d, m, p) \left(\frac{h(T_0)}{\rho(T)} \right)^m h(T)^{d/p} |v|_{H^{m,p}(T_0)}. \end{aligned}$$

Beweis: Die Behauptung folgt aus Satz 2.30 unter Verwendung der Abschätzungen

$$\|A\| \leq \frac{h(T)}{\rho(T_0)}, \|A^{-1}\| \leq \frac{h(T_0)}{\rho(T)}, c\rho(T)^s \leq |\det(\nabla F)| = |\det A| \leq Ch(T)^s$$

aus Lemma 2.11. \square .

Mit der Transformationsabschätzung 2.31 und dem Bramble-Hilbert Lemma (Satz 2.29) sind wir nun in der Lage eine lokale Interpolationsfehlerabschätzung auf einem Simplex T zu beweisen.

Satz 2.32 (Lokale Interpolationsfehlerabschätzung)

Seien T ein d -dimensionaler Simplex, T_0 der zugehörige Einheitssimplex und $F : T_0 \rightarrow T$, $F(y) = Ay + b$ die Referenzabbildung (vgl. Lemma 2.11). Seien weiter $k, m \in \mathbb{N}_0$, $p, q \geq 1$, so dass

$$E_0 \in L(H^{k+1,p}(T_0), H^{m,q}(T_0))$$

eine Einbettung ist, die auf $\mathbb{P}_k(T_0)$ die Identität ist. Sei weiter

$$I_0 \in L(H^{k+1,p}(T_0), H^{m,q}(T_0))$$

ein Interpolationsoperator, der $\mathbb{P}_k(T_0)$ invariant läßt, d.h. es gilt $I_0 s_0 = s_0, \forall s_0 \in \mathbb{P}_k(T_0)$. Dann ist durch

$$Iu := I_0(u \circ F) \circ F^{-1}$$

ein Interpolationsoperator $I \in L(H^{k+1,p}(T), H^{m,q}(T))$ definiert und es gilt die Fehlerabschätzung

$$|u - Iu|_{H^{m,q}(T)} \leq c\sigma(T)^{m+\frac{d}{p}} h(T)^{k+1-m+d(\frac{1}{q}-\frac{1}{p})} |u|_{H^{k+1,p}(T)}$$

für alle $u \in H^{k+1,p}(T)$ mit $c = c(k, m, p, q, T_0, \|I_0\|)$.

Beweis: Die Wohldefiniertheit von I ist klar aufgrund der Invertierbarkeit der Referenzabbildung. Sei also $u \in H^{k+1,p}(T)$. Dann ist $Eu := E_0(u \circ F) \circ F^{-1} \in H^{m,q}(T)$. Durch Anwendung der zweiten Transformationsabschätzung aus Corollar 2.31 erhalten wir also

$$|Eu - Iu|_{H^{m,q}(T)} \leq c \left(\frac{h(T_0)}{\rho(T)} \right)^m h(T)^{d/q} |(Eu - Iu) \circ F|_{H^{m,q}(T_0)}. \quad (2.2)$$

Mit der Definition von E und I folgt weiter

$$|(Eu - Iu) \circ F|_{H^{m,q}(T_0)} = |E_0(u \circ F) - I_0(u \circ F)|_{H^{m,q}(T_0)}.$$

Mit dem Bramble-Hilbert Lemma (Satz 2.29) folgt hieraus

$$|(Eu - Iu) \circ F|_{H^{m,q}(T_0)} \leq c |u \circ F|_{H^{k+1,p}(T_0)}.$$

Insgesamt folgt also aus Gleichung (2.2)

$$|Eu - Iu|_{H^{m,q}(T)} \leq c \left(\frac{h(T_0)}{\rho(T)} \right)^m h(T)^{d/q} |u \circ F|_{H^{k+1,p}(T_0)}.$$

Mit der erste Transformationsabschätzung aus Corollar 2.31 erhalten wir weiter

$$\begin{aligned}
 |Eu - Iu|_{H^{m,q}(T)} &\leq c \left(\frac{h(T_0)}{\rho(T)} \right)^m h(T)^{d/q} \left(\frac{h(T)}{\rho(T_0)} \right)^{k+1} \rho(T)^{-d/p} |u|_{H^{k+1,p}(T)} \\
 &= c \frac{h(T)^{k+1+d/q}}{\rho(T)^{m+d/p}} |u|_{H^{k+1,p}(T)} \\
 &= c \sigma(T)^{m+d(1/p)} h(T)^{k+1-m+d(1/q-1/p)} |u|_{H^{k+1,p}(T)}.
 \end{aligned}$$

Dabei ist $\sigma(T) = \frac{h(T)}{\rho(T)}$.

Bemerkung 2.33

Verwendet man im Beweis von Satz 2.32, die Identität $|\det A| = \frac{|T|}{|T_0|} = |\det(A^{-1})|^{-1}$, so lässt sich die Interpolationsfehlerabschätzung verbessern zu

$$|u - Iu|_{H^{m,q}(T)} \leq c \sigma(T)^{m-d \min\{0, (\frac{1}{q} - \frac{d}{p})\}} h(T)^{k+1-m+d(\frac{1}{q} - \frac{1}{p})} |u|_{H^{k+1,p}(T)}.$$

Die lokale Interpolationsfehlerabschätzung verwenden wir nun, um eine globale Abschätzung des Interpolationsfehlers auf polygonal berandeten Gebieten herzuleiten.

Satz 2.34 (Globale Interpolationsfehlerabschätzung)

Seien $\Omega \subset \mathbb{R}^d$ ein Lipschitzgebiet und \mathcal{T}_h ein zulässige Triangulierung auf Ω . Weiter sei S_h^k der Lagrange Finite Elemente Raum aus Definition 2.21. Gelte nun $p > \frac{d}{2}$, so ist durch

$$I_h u \in S_h^k, \quad I_h u(x) = u(x) \forall x \in G_k(\mathcal{T}_h)$$

ein Interpolationsoperator $I_h \in L(H^{2,p}(\Omega), S_h^k)$ definiert, der alle Elemente $v_h \in S_h^k$ invariant lässt. Für $m \in \{0, 1\}$ und $s \in \mathbb{N}_0$ mit $0 \leq s \leq k$ gilt die Interpolationsfehlerabschätzung

$$|u - I_h u|_{H^{2,p}(T)} \leq c \left(\sum_{T \in \mathcal{T}_h} \sigma(T)^{mp} h(T)^{(s+1-m)p} |u|_{H^{s+1,p}(T)}^p \right)^{1/p}$$

für alle $u \in H^{s+1,p}(\Omega)$. Ist $\sigma(T) \leq \sigma, \forall T \in \mathcal{T}_h$, so folgt

$$|u - I_h u|_{H^{2,p}(T)} \leq c \sigma^m h^{s+1-m} |u|_{H^{s+1,p}(\Omega)}.$$

Beweis: Aufgrund der Voraussetzung an p ist die Einbettung von $H^{2,p}(\Omega)$ nach $C^0(\bar{\Omega})$ gegeben und daher der Interpolationsoperator wohldefiniert. Sei also $u \in H^{s+1,p}(\Omega)$, so gilt

$$|u - I_h u|_{H^{m,p}(\Omega)}^p = \sum_{T \in \mathcal{T}_h} |u - I_h u|_{H^{m,p}(T)}^p.$$

Mit der lokalen Abschätzung aus Bemerkung 2.33, folgt weiter

$$|u - I_h u|_{H^{m,p}(\Omega)}^p \leq c \sum_{T \in \mathcal{T}_h} \sigma(T)^{mp} h(T)^{(k+1-m)p} |u|_{H^{s+1,p}(T)}^p.$$

Gilt $\sigma(T) \leq \sigma, \forall T \in \mathcal{T}_h$, so folgt mit $h(T) \leq h, \forall T \in \mathcal{T}_h$ der zweite Teil der Behauptung. \square

2.3.2 A priori Fehlerabschätzung für Finite Elemente Verfahren

Mit dem Lemma von Cea und der Interpolationsfehlerabschätzung aus dem vorherigen Abschnitt haben wir nun alle Werkzeuge zur Hand, um eine A priori Fehlerabschätzung für Lagrange Finite Elemente Verfahren auf simplizialen Gittern beweisen zu können.

Satz 2.35 (A priori Fehlerabschätzung)

Seien $\Omega \subset \mathbb{R}^d, d \leq 3$ ein Lipschitzgebiet und \mathcal{T}_h eine zulässige Triangulierung von Ω , so dass die Bedingung $\sigma(T) \leq \sigma, \forall T \in \mathcal{T}_h$ erfüllt ist. Sei $f \in L^2(\Omega)$ und $u \in H_0^1(\Omega)$ eine Lösung des homogenen Dirichletproblems für die Poissonsleichung

$$\int_{\Omega} \nabla u \cdot \nabla v = \int_{\Omega} f v, \quad \forall v \in H_0^{1,2}(\Omega)$$

und sei für $k \in \mathbb{N}$, $u_h \in S_{h,0}^k$ die Lösung des Finite Elemente Verfahrens

$$\int_{\Omega} \nabla u_h \cdot \nabla v_h = \int_{\Omega} f v_h, \quad \forall v_h \in S_{h,0}^k.$$

Liegt nun $u \in H^{s+1}(\Omega)$ für ein $s \in \{1, \dots, k\}$, so gibt es eine Konstante $c > 0$, die nur von d, s, k, σ und Ω abhängt, so dass gilt

$$\|u - u_h\|_{H^1(\Omega)} \leq ch^s |u|_{H^{s+1}(\Omega)}.$$

Beweis: In Abschnitt 2.1 haben wir gesehen, dass für Ritz-Galerkin Verfahren und somit auch für Finite Elemente Verfahren die abstrakte Fehlerabschätzung

$$\|u - u_h\|_X \leq \frac{c_1}{c_0} \inf_{v_h \in X_h} \|u - v_h\|_X.$$

gilt, wobei X, X_h, B, u und u_h wie in den Definition 2.1, 2.2 definiert sind. Insbesondere gilt diese Abschätzung also auch in unserem Fall mit $X = H_0^{1,2}(\Omega)$ und $X_h = S_{h,0}^k$. Sei I_h die Lagrange Interpolation aus Satz 2.34, so folgt weiter

$$\|u - u_h\|_{H_0^1(\Omega)} \leq \frac{c_1}{c_0} \|u - I_h u\|_{H_0^1(\Omega)}.$$

Für $p = 2, m = 1, d \leq 3$ sind dann alle Voraussetzung des Satzes 2.34 erfüllt und wir erhalten

$$\|u - u_h\|_{H_0^1(\Omega)} \leq \frac{c_1}{c_0} c_3 \sigma^m h^{s+1-1} |u|_{H^{s+1}(\Omega)}.$$

Mit $c := \frac{c_1}{c_0} c_3 \sigma^m$ folgt dann die Behauptung. \square

Bemerkung 2.36

- (i) Bei der a priori Fehlerabschätzung mußten wir voraussetzen, dass die Lösung u der Poissons-Gleichung mindestens in $H^2(\Omega)$ liegt. Nach dem Satz von Friedrich 1.30 ist dies für $f \in L^2(\Omega)$ der Fall, falls Ω entweder ein Gebiet mit glatten Rand (es gelte mindestens $\partial\Omega \in C^2$) ist, oder falls Ω ein konkaves Lipschitz-Gebiet ist.
- (ii) Da die Voraussetzungen des Lemmas von Cea (Satz 2.4) auch für allgemeine elliptische Probleme in Divergenzform (vgl. Definition 1.27) erfüllt sind, gilt die a priori Fehlerabschätzung aus Satz 2.35 auch, wenn wir die Poissons-Gleichung durch ein allgemeines elliptisches Problem in Divergenzform ersetzen.

2.3.3 L^2 -Fehlerabschätzung und das Lemma von Aubin-Nitsche

Seien u und u_h wie in Satz 2.35 definiert. Da $\|u - u_h\|_{L^2(\Omega)} \leq \|u - u_h\|_{H^1(\Omega)}$ ist, folgt aus Satz 2.35 auch eine Abschätzung für den L^2 -Fehler der Form

$$\|u - u_h\|_{L^2(\Omega)} \leq ch^s |u|_{H^{s+1}(\Omega)}.$$

Andererseits wissen wir jedoch aufgrund der Interpolationsfehlerabschätzung aus Satz 2.34, dass für den Interpolationsfehler in der L^2 -Norm gilt

$$\|u - I_h u\|_{L^2(\Omega)} \leq ch^{s+1} |u|_{H^{s+1}(\Omega)}.$$

Es stellt sich also die Frage, ob die Fehlerabschätzung bzgl. der L^2 -Norm optimal ist und unter welchen Voraussetzungen wir die Fehlerabschätzung verbessern können. Eine Antwort darauf liefert das abstrakte Lemma von Aubin-Nitsche.

Satz 2.37 (Lemma von Aubin-Nitsche)

Seien X, H Hilberträume, $X_h \subset X$ ein endlichdimensionaler Teilraum und gelte die Einbettung

$$E : X \hookrightarrow H.$$

Seien $B : X \times X \rightarrow \mathbb{R}$ eine stetige (c_1) und koerzive (c_0) Bilinearform und $f \in X'$ gegeben. Sind dann $u \in X$ und $u_h \in X_h$ Lösungen der Variationsprobleme

$$\begin{aligned} B(u, \varphi) &= f(\varphi), & \forall \varphi \in X, \\ B(u_h, \varphi_h) &= f(\varphi_h), & \forall \varphi_h \in X_h, \end{aligned}$$

so gilt

$$\|E(u - u_h)\|_H \leq c_1 \|u - u_h\|_X \sup_{r \in H \setminus \{0\}} \frac{\inf_{v_h \in X_h} \|v_r - v_h\|_X}{\|r\|_H}.$$

Dabei ist $v_r \in X$ zu gegebenem $r \in H$ die Lösung des dualen (adjungierten) Problems

$$B^*(v_r, \varphi) = (r, E(\varphi))_H, \quad \forall \varphi \in X,$$

wobei $B^*(v_r, \varphi) := B(\varphi, v_r)$ die zu B duale Bilinearform ist.

Beweis: Sei $r \in H$ gegeben, so definieren wir \tilde{f} durch

$$\tilde{f}(\varphi) := (r, E(\varphi))_H, \quad \forall \varphi \in X.$$

Da

$$|\tilde{f}(\varphi)| \leq \|r\|_H \|E(\varphi)\|_H \leq \|E\| \|r\|_H \|\varphi\|_X,$$

folgt $\tilde{f} \in X'$ mit $\|\tilde{f}\|_{X'} \leq \|E\| \|r\|_H$. Da aus der Stetigkeit und Koerzivität von B auch die Stetigkeit und Koerzivität von B^* folgt, existiert dann nach dem Satz von Lax-Milgram 1.5 genau ein $v_r \in X$ mit

$$B^*(v_r, \varphi) = (r, E(\varphi))_H, \quad \forall \varphi \in X.$$

Wählen wir nun in diesem dualen Problem als Testfunktion $\varphi = u - u_h \in X$, so erhalten wir unter Verwendung der Galerkin-Orthogonalität

$$\begin{aligned} (r, E(u - u_h))_H &= B^*(v_r, u - u_h) = B(u - u_h, v_r) \\ &= B(u - u_h, v_r - v_h), \quad \forall v_h \in X_h. \end{aligned}$$

Also folgt mit der Stetigkeit der Bilinearform

$$(r, E(u - u_h))_H = B(u - u_h, v_r - v_h) \leq c_1 \|u - u_h\|_X \|v_r - v_h\|_X.$$

Da dies für alle $v_h \in X_h$ gilt, folgt weiter

$$(r, E(u - u_h))_H \leq c_1 \|u - u_h\|_X \inf_{v_h \in X_h} \|v_r - v_h\|_X.$$

Da für alle $l \in H$ gilt

$$\|l\|_H = \sup_{r \in H \setminus \{0\}} \frac{|(r, l)_H|}{\|r\|_H}$$

folgt hieraus die Behauptung

$$\begin{aligned} \|E(u - u_h)\|_H &= \sup_{r \in H \setminus \{0\}} \frac{|(r, E(u - u_h))_H|}{\|r\|_H} \\ &\leq c_1 \|u - u_h\|_X \sup_{r \in H \setminus \{0\}} \frac{\inf_{v_h \in X_h} \|v_r - v_h\|_X}{\|r\|_H}. \quad \square \end{aligned}$$

Korollar 2.38 (L^2 -Fehlerabschätzung)

Seien die Voraussetzungen aus Satz 2.35 erfüllt. Zusätzlich existiere ein Konstante $c_* > 0$, so dass für alle $r \in L^2(\Omega)$ ein schwache Lösung $v_r \in H_0^1(\Omega)$ des dualen Poissonproblems

$$\int_{\Omega} \nabla \varphi \cdot \nabla v_r = \int_{\Omega} r \varphi, \quad \forall \varphi \in H_0^{1,2}(\Omega)$$

existiert, die H^2 -regulär ist ($v_r \in H^2(\Omega)$) und die folgende uniforme a priori Abschätzung erfüllt

$$\|v_r\|_{H^2(\Omega)} \leq c_* \|r\|_{L^2(\Omega)}.$$

Dann gibt es eine Konstante $C > 0$, die nicht von h und u abhängt, so dass gilt

$$\|u - u_h\|_{L^2(\Omega)} \leq Ch^{s+1} |u|_{H^{s+1}(\Omega)},$$

falls $u \in H^{s+1}(\Omega)$ für ein $s \in \{1, \dots, k\}$ ist.

Beweis: Wir wenden das Lemma von Aubin-Nitsche 2.37 an, wobei $X = H_0^1(\Omega)$, $H = L^2(\Omega)$ und $X_h = S_{h,0}^k \subset H_0^1(\Omega)$ seien und die Bilinearform B durch die schwache Formulierung der Poissons-Gleichung gegeben sei. Wir erhalten

$$\|E(u - u_h)\|_{L^2(\Omega)} \leq c_1 \|u - u_h\|_{H^1(\Omega)} \sup_{r \in L^2(\Omega) \setminus \{0\}} \frac{\inf_{v_h \in X_h} \|v_r - v_h\|_{H^1(\Omega)}}{\|r\|_L^2(\Omega)}.$$

Schätzen wir den Approximationsfehler

$$\inf_{v_h \in X_h} \|v_r - v_h\|_{H^1(\Omega)}$$

durch den Interpolationsfehler ab, so erhalten wir mit der Regularitätsabschätzung für v_r

$$\begin{aligned} \inf_{v_h \in X_h} \|v_r - v_h\|_{H^1(\Omega)} &\leq \|v_r - I_h(v_r)\|_{H^1(\Omega)} \\ &\leq c_2 h |v_r|_{H^2(\Omega)} \leq ch \|v_r\|_{H^2(\Omega)} \\ &\leq c_2 c_* h \|r\|_{L^2(\Omega)}. \end{aligned}$$

Zusammen mit der Fehlerabschätzung aus Satz 2.35 folgt also

$$\|E(u - u_h)\|_{L^2(\Omega)} \leq c_1 c_3 h^s |u|_{H^{s+1}(\Omega)} c_2 c_* h = Ch^{s+1} |u|_{H^{s+1}(\Omega)}.$$

mit $C = c_1 c_2 c_3 c_*$. □

2.4 Randapproximation

In den vorherigen Abschnitten hatten wir stets angenommen, dass \mathcal{T}_h eine zulässige Triangulierung eines Gebietes $\Omega \subset \mathbb{R}^d$ ist. Diese Annahme impliziert, dass Ω polygonal berandet sein muss. Dies ist auf der einen Seite eine große Einschränkung für viele Anwendungen und steht zum anderen im Widerspruch zu der Forderung, dass wir optimale a priori Fehlerabschätzungen für Finite Elemente Verfahren nur dann erhalten, wenn die exakte Lösung u in $H^{k+1}(\Omega)$ für $k \geq 1$ liegt. Solche Regularität ist i.A. nur zu erwarten, wenn Ω konvex ist, oder glatten Rand hat.

In diesem Abschnitt wollen wir daher die Approximation von Gebieten Ω mit glattem Rand durch polygonal berandete Gebiete Ω_h mit einbeziehen.

Als Modellproblem betrachten wir wieder das homogene Dirichletproblem für die Poisongleichung (vgl. auch Definition 1.24)

Definition 2.39 (Zulässige Gebietsapproximation)

Sei $\Omega \subset \mathbb{R}^d$ ein Gebiet mit glattem Rand. Dann heißt $(\Omega_h, \mathcal{T}_h)$ zulässige Gebietsapproximation der Feinheit h , falls \mathcal{T}_h zulässige Triangulierung von Ω_h mit Feinheit h ist und für alle Ecken x der Simplexe in \mathcal{T}_h gilt:

$$x \in \partial\Omega_h \implies x \in \partial\Omega.$$

Insbesondere ist Ω_h also polygonal berandet, alle Randflächen von Ω_h sind Flächen der Triangulierung \mathcal{T}_h und die Ecken der Triangulierung, die auf dem Rand von Ω_h liegen, liegen auch auf dem Rand von Ω .

Die neu hinzukommenden Schwierigkeiten in der Analyse von Finite Elemente Verfahren unter Berücksichtigung von Gebietsapproximationen wird klar, wenn wir uns die Definition solcher Verfahren im Detail anschauen.

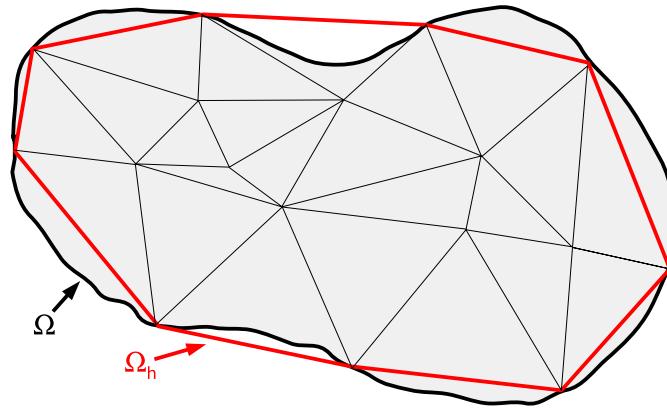


Abbildung 2.5: Glatt berandetes Gebiet Ω mit Gebietsapproximation $(\Omega_h, \mathcal{T}_h)$.

Definition 2.40 (Finite Elemente Verfahren mit Randapproximation)

Seien $\Omega \subset \mathbb{R}^d$ ein Gebiet mit glattem Rand und $(\Omega_h, \mathcal{T}_h)$ zulässige Gebietsapproximation der Feinheit h . Sei $X := H_0^1(\Omega)$ und $X_h := S_{h,0}^k \subset H_0^1(\Omega_h)$ der Lagrange Finite Elemente Raum mit Nullrandwerten auf \mathcal{T}_h . Seien weiter $f \in L^2(\Omega)$, $\bar{f} \in L^2(\Omega \cup \Omega_h)$ mit $\bar{f}|_\Omega = f$ und $B : X \times X \rightarrow \mathbb{R}$, $B_h : H_0^1(\Omega_h) \times H_0^1(\Omega_h) \rightarrow \mathbb{R}$ definiert durch

$$\begin{aligned} B(u, v) &:= \int_{\Omega} \nabla u \cdot \nabla v \, dx, & \forall u, v \in H_0^1(\Omega), \\ B_h(u, v) &:= \int_{\Omega_h} \nabla u \cdot \nabla v \, dx, & \forall u, v \in H_0^1(\Omega_h), \end{aligned}$$

so ist die Lösung $u \in X$ des homogenen Dirichletproblems für die Poisongleichung gegeben durch

$$B(u, \varphi) = (f, \varphi)_{L^2(\Omega)}, \quad \forall \varphi \in X.$$

Die Lösung des Finite Elemente Verfahrens mit Randapproximation $u_h \in X_h$ für dieses Problem ist dann gegeben durch

$$B_h(u_h, \varphi_h) = (\bar{f}, \varphi_h)_{L^2(\Omega_h)}, \quad \forall \varphi_h \in X_h.$$

Wie wir in der Definition sehen, ist das resultierende Finite Elemente Verfahren nicht konform, d.h. es gilt nicht $X_h \subset X$! Außerdem verwenden wir im Finite Elemente Verfahren eine Bilinearform B_h , die von der zulässigen Gebietsapproximation abhängt und die rechte Seite f muss durch eine Er-

weiterung \bar{f} ersetzt werden.

Zur Vorbereitung einer numerischen Analyse des Finite Elemente Verfahrens wollen wir uns zunächst mit grundlegenden Eigenschaften des Randes eines glatt berandeten Gebietes vertraut machen.

Lemma 2.41 (Distanzfunktion zum Rand)

Sei $\Omega \subset \mathbb{R}^d$ ein Gebiet mit glattem Rand. Dann ist die Distanzfunktion

$$\text{dist}(x, \partial\Omega) := \inf_{y \in \partial\Omega} |x - y|, \quad (x \in \mathbb{R}^d)$$

global Lipschitz-stetig mit Lipschitzkonstante 1, d.h. $\text{dist}(\cdot, \partial\Omega) \in C^{0,1}(\mathbb{R}^d)$.

Beweis: Der Rand von Ω ist abgeschlossen und beschränkt. Also ist die Distanzfunktion dist wohldefiniert. Seien $x, y \in \mathbb{R}^d$ und sei $z \in \partial\Omega$, so dass gilt $\text{dist}(x, \partial\Omega) = |x - z|$. Dann folgt $\text{dist}(y, \partial\Omega) \leq |y - z| \leq |x - y| + |x - z| = |x - y| + \text{dist}(x, \partial\Omega)$ und analog $\text{dist}(x, \partial\Omega) \leq |x - y| + \text{dist}(y, \partial\Omega)$. Daher gilt $|\text{dist}(x, \partial\Omega) - \text{dist}(y, \partial\Omega)| \leq |x - y|$. \square

Definition 2.42 (Orientierte Distanzfunktion)

Für ein Gebiet $\Omega \subset \mathbb{R}^d$ definieren wir die orientierte Distanzfunktion durch

$$d(x) := \begin{cases} \text{dist}(x, \partial\Omega) & , \text{falls } x \in \mathbb{R}^d \setminus \Omega \\ -\text{dist}(x, \partial\Omega) & , \text{falls } x \in \bar{\Omega}. \end{cases}$$

Wir sagen, dass $\partial\Omega \in C^{k,\alpha}$ ist für $k \in \mathbb{N}, 0 \leq \alpha \leq 1$, wenn es ein $\delta \geq 0$ gibt, so dass für eine Umgebung $U_\delta := \{x \in \mathbb{R}^d \mid \text{dist}(x, \partial\Omega) < \delta\}$ des Randes gilt

$$d \in C^{k,\alpha}(U_\delta) \quad \text{und} \quad \nabla d \neq 0 \text{ in } U_\delta.$$

Die orientierte Distanzfunktion beschreibt vollständig die Geometrie des Randes von Ω . Dies wird in dem folgenden Lemma ersichtlich.

Lemma 2.43 (Geometrische Eigenschaften)

Sei $\Omega \subset \mathbb{R}^d$ und gelte $\partial\Omega \in C^{k,\alpha}$ mit $k \geq 1$. Dann ist

$$n(x) = \nabla d(x) \quad \forall x \in U_\delta$$

und n ist die äußere Normale an $\partial\Omega$.

Ist $k \geq 2$, so genügt die Matrix $H \in \mathbb{R}^{d \times d}$ definiert durch

$$H_{ij} := \partial_{x_i, x_j} d, \quad i, j = 1, \dots, d,$$

in U_δ der Gleichung

$$Hn = 0.$$

Die Eigenwerte ungleich Null von H in $x \in \partial\Omega$ sind die Hauptkrümmungen $\kappa_1(x), \dots, \kappa_{d-1}(x)$ von $\partial\Omega$ im Punkt x .

Gilt für δ die folgende Abschätzung

$$\delta \leq \min_{x \in \partial\Omega} \delta(x), \quad \delta(x) := \min \left\{ \frac{1}{|\kappa_i(x)|} \mid i = 1, \dots, d-1 \right\},$$

so existiert zu jedem $x \in U_\delta$ genau ein $a(x) \in \partial\Omega$ mit $\text{dist}(x, \partial\Omega) = |x - a(x)|$ und es gilt die Darstellung

$$x = a(x) + d(x)n(x).$$

Beweis: Wir beweisen nur einen Teil der Behauptung.

Sei $x \in U_\delta$. Die Existenz von $a(x)$ mit $\text{dist}(x, \partial\Omega) = |x - a(x)|$ ist dann klar, da $\partial\Omega$ abgeschlossen ist.

Ansatz: $x = a(x) + d(x)v(x)$ mit $|v(x)| = 1$.

Zu zeigen: $v(x) = n(a(x))$.

Sei dazu $\varphi : [-1, 1] \rightarrow \partial\Omega$ differenzierbar mit $\varphi(0) = a(x)$. Definiere weiter $\psi(t) := |x - \varphi(t)|^2$, so gilt $\psi(t) \geq |x - \varphi(0)|^2 \quad \forall t \in [-1, 1]$ und somit

$$0 = \frac{d}{dt} \psi(t) \Big|_{t=0} = 2(x - \varphi(0), \varphi'(0)) = 2(x - a(x), \varphi'(0)) = 2(d(x)v(x), \varphi'(0)).$$

Aufgrund der Definition von φ ist φ' ein Tangentialvektor an $\partial\Omega$ in $a(x)$. Da φ beliebig gewählt werden kann folgt also $v(x) \perp$ Tangentialraum in $a(x)$. Da $|v(x)| = 1$ gilt, folgt also $v(x) = n(a(x))$.

Noch zu zeigen: $\nabla d(x) = n(x)$

Es ist $a(x + nh) = a(x)$ für $0 \leq h \leq \delta$, also folgt

$$hn = x + hn - x = (d(x + hn) - d(x))n(x).$$

Also gilt

$$1 = \frac{d(x + hn) - d(x)}{h} \rightarrow \nabla d(x)n(x) \quad (h \rightarrow 0).$$

Da die Lipschitzkonstante von d gleich 1 ist, folgt $|\nabla d(x)| = 1$ und somit insgesamt $\nabla d(x) = n(x)$.

Schließlich folgt aus $1 = |\nabla d(x)|^2$ durch ableiten nach x_j

$$0 = \sum_{i=1}^d 2\partial_{x_i, x_j} d(x) \partial_{x_i} d(x) = 2(H(x) \nabla d(x))_j. \quad \square$$

Das Lemma sagt insbesondere aus, dass die Breite δ des U_δ Streifens von der Krümmung des Randes von Ω abhängt.

Beispiel 2.44

Für $\Omega = B_R(x_0) \subset \mathbb{R}^d$ ist

$$\text{dist}(x, \partial\Omega) = ||x - x_0| - R| \quad \text{und} \quad d(x) = |x - x_0| - R.$$

Damit folgt

$$\nabla d(x) = \frac{x - x_0}{|x - x_0|} = n(x)$$

und

$$H_{ij}(x) = \frac{1}{R}(\delta_{ij} - n_i(x)n_j(x)), \quad (x \in \partial\Omega).$$

Die Eigenwerte von $H(x)$ sind $\kappa_0(x) = 0$ und $\kappa_i(x) = \frac{1}{R}$ für $i = 1, \dots, d-1$. Daher kann $\delta = R$ gewählt werden.

Im nächsten Satz wollen wir uns nun mit der Approximation von Gebieten mit glattem Rand durch Flächen von Simplizes beschäftigen.

Lemma 2.45

Sei $\Omega \subset \mathbb{R}^d$ und gelte $\partial\Omega \in C^2$. Ist dann T ein $d-1$ dimensionaler Seiten-simplex mit Ecken auf $\partial\Omega$ und $T \subset U_\delta$, so folgt

$$|d(x)| \leq ch(T)^2 \quad \forall x \in T.$$

Beweis: Ohne Einschränkung können wir annehmen, dass $T \in \mathbb{R}^{d-1}$ gilt (sonst Translation und Drehung). Dann folgt aus der lokalen Interpolationsfehlerabschätzung mit $I(d) = 0$:

$$||d||_{L^\infty(T)} = ||d - I(d)||_{L^\infty(T)} \leq ch(T)^2 |d|_{H^{2,\infty}(T)}.$$

$|d|_{H^{2,\infty}(T)}$ ist die maximale Krümmung von $\partial\Omega$, also insbesondere beschränkt.

\square

Um globale Eigenschaften für eine zulässige Gebietsapproximation $(\Omega_h, \mathcal{T}_h)$ zeigen zu können, wollen wir im Folgenden stets annehmen, dass $\partial\Omega_h \subset U_\delta$ gilt. Dazu definieren wir:

Definition 2.46 (Reguläre Gebietsapproximation)

Sei $\Omega \subset \mathbb{R}^d$ und $\partial\Omega \in C^2$. Dann heißt eine zulässige Gebietsapproximation $(\Omega_h, \mathcal{T}_h)$ regulär, falls gilt

$$\partial\Omega_h \subset U_\delta,$$

wobei U_δ eine hinreichend kleine Umgebung von $\partial\Omega$ ist (vgl. Lemma 2.43).

Außerdem habe jedes Simplex $T \in \mathcal{T}_h$ mindestens einen Eckpunkt in Ω und es gebe eine Konstante $\sigma > 0$ mit $\sigma(T) \leq \sigma, \forall T \in \mathcal{T}_h$.

Die lokale Eigenschaft aus Lemma 2.45 hat global zur Folge, dass wir folgende Eigenschaft für reguläre Gebietsapproximationen erhalten

$$|\Omega \setminus \Omega_h \cup \Omega_h \setminus \Omega| \leq ch^2.$$

Im folgenden Lemma werden wir nun wesentliche Abschätzungen zur Untersuchung der Randapproximation beweisen.

Lemma 2.47

Sei $\Omega \subset \mathbb{R}^d$, $\partial\Omega \in C^2$ und $(\Omega_h, \mathcal{T}_h)$ eine reguläre Gebietsapproximation. Dann gibt es eine Konstante $c > 0$, die nicht von h abhängt, so dass für alle $\varphi \in H^1(\Omega_h)$ gilt

$$\|\varphi\|_{L^2(\Omega_h \setminus \Omega)} \leq ch \|\varphi\|_{H^1(\Omega_h)}.$$

Gilt zusätzlich $\varphi \in H_0^1(\Omega_h)$, so folgt

$$\begin{aligned} \|\varphi\|_{L^2(\Omega_h \setminus \Omega)} &\leq ch^2 \|\varphi\|_{H^1(\Omega_h \setminus \Omega)}, \\ \|\varphi\|_{L^2(\partial\Omega \cap \Omega_h)} &\leq ch \|\varphi\|_{H^1(\Omega_h \setminus \Omega)}. \end{aligned}$$

Analog folgt in der L^1 -Norm

$$\|\varphi\|_{L^1(\Omega_h \setminus \Omega)} \leq ch^2 \|\varphi\|_{H^{1,1}(\Omega_h)}.$$

Beweis: Wir beweisen zunächst die letzte L^1 -Abschätzung des Lemmas.

Sei $\tilde{T} \subset \partial\Omega_h \subset U_\delta$ ein Seitenfläche der Triangulierung \mathcal{T}_h . Nach Lemma 2.43 gilt dann für alle $x \in \tilde{T}$ die eindeutige Darstellung

$$x = a(x) + d(x)n(x)$$

mit $a(x) \in \partial\Omega$ und $n(x) = n(a(x))$. Die Parallelflächen von $\partial\Omega$ sind durch die Isoflächen

$$d(x) = c, \quad -\delta < c < \delta$$

gegeben. Ohne Einschränkung sei $\tilde{T} \in \mathbb{R}^{d-1}$. Wir setzen

$$d(y)^+ := \max\{d(y), 0\},$$

$$D_h := D_h(\tilde{T}) := \{x \in \mathbb{R}^d \mid x = a(y) + sn(y), y \in \tilde{T} \setminus \partial\tilde{T}, s \in (0, d(y)^+)\}.$$

Dann ist D_h offen und $D_h \subset \mathbb{R}^d \setminus \Omega$. Sei $D_h \neq \emptyset$, so gilt nach Transformationsregel mit $x(y, s) = a(y) + sn(y)$, $\bar{\varphi}(y, s) = \varphi(x(y, s))$

$$\int_{D_h} \varphi(x) dx = \int_{\tilde{T}} \int_0^{d(y)^+} \bar{\varphi}(y, s) \left| \det \frac{\partial x}{\partial(y, s)} \right| ds dy.$$

Wie man leicht einsieht, gibt es Konstanten c_1, c_2 , die nur von der Geometrie von $\partial\Omega$ abhängen, so dass gilt

$$c_1 \leq \left| \det \frac{\partial x}{\partial(y, s)} \right| \leq c_2.$$

Sei nun $y \in \tilde{T}$ und $s_1, s_2 \in (0, d(y)^+)$, so gilt

$$\bar{\varphi}(y, s_1) - \bar{\varphi}(y, s_2) = \int_{s_2}^{s_1} \partial_s \bar{\varphi}(y, s) ds$$

und somit

$$|\bar{\varphi}(y, s_1)| \leq |\bar{\varphi}(y, s_2)| + \int_{-\delta}^{d(y)^+} |\partial_s \bar{\varphi}(y, s)| ds.$$

Integration von $-\delta$ bis $d(y)^+$ über s_2 ergibt dann

$$\delta |\bar{\varphi}(y, s_1)| \leq \int_{-\delta}^{d(y)^+} |\bar{\varphi}(y, s_2)| ds_2 + (\delta + d(y)^+) \int_{-\delta}^{d(y)^+} |\partial_s \bar{\varphi}(y, s)| ds.$$

Schließlich liefert Integration von 0 bis $d(y)^+$ über s_1

$$\begin{aligned} \delta \int_0^{d(y)^+} |\bar{\varphi}(y, s_1)| ds_1 &\leq d(y)^+ \int_{-\delta}^{d(y)^+} |\bar{\varphi}(y, s_2)| ds_2 \\ &\quad + d(y)^+ (\delta + d(y)^+) \int_{-\delta}^{d(y)^+} |\partial_s \bar{\varphi}(y, s)| ds. \end{aligned}$$

Nach Lemma 2.45 gilt

$$d(y)^+ \leq ch(\tilde{T})^2, \quad \forall y \in \tilde{T},$$

also erhalten wir nach Integration über $y \in \tilde{T}$ die Abschätzung

$$\begin{aligned} \delta \int_{\tilde{T}} \int_0^{d(y)^+} |\bar{\varphi}(y, s_1)| ds_1 dy &\leq ch(\tilde{T})^2 \int_{\tilde{T}} \int_{-\delta}^{d(y)^+} |\bar{\varphi}(y, s_2)| ds_2 dy \\ &\quad + ch(\tilde{T})^2 (\delta + ch(\tilde{T})^2) \int_{\tilde{T}} \int_{-\delta}^{d(y)^+} |\partial_s \bar{\varphi}(y, s)| ds dy. \end{aligned}$$

Insgesamt folgt also mit $U_\delta(\tilde{T}) := \{x = a(y) + sn(y) \mid y \in \tilde{T}, -\delta < s < d(y)^+\}$

$$\|\varphi\|_{L^1(D_h(\tilde{T}))} \leq ch(\tilde{T})^2 \left(\frac{1}{\delta} \|\varphi\|_{L^1(U_\delta(\tilde{T}))} + \left(1 + \frac{h^2}{\delta}\right) \|\nabla \varphi\|_{L^1(U_\delta(\tilde{T}))} \right).$$

Summieren wir diese Abschätzung auf über alle $\tilde{T} \in \partial\Omega_h$, so erhalten wir die letzte Abschätzung des Lemmas

$$\|\varphi\|_{L^1(\Omega \setminus \Omega_h)} \leq ch^2 \|\varphi\|_{H^{1,1}(\Omega_h)}.$$

Die erste Abschätzung folgt nun direkt

$$\begin{aligned} \|\varphi\|_{L^2(\Omega \setminus \Omega_h)}^2 &= \|\varphi^2\|_{L^1(\Omega \setminus \Omega_h)} \leq ch^2 \|\varphi^2\|_{H^{1,1}(\Omega_h)} \\ &\leq ch^2 (\|\varphi\|_{L^2(\Omega_h)}^2 + \|\varphi\|_{L^2(\Omega_h)} \|\nabla \varphi\|_{L^2(\Omega_h)}) \\ &\leq ch^2 \|\varphi\|_{H^{1,2}(\Omega_h)}^2. \end{aligned}$$

Gilt nun zusätzlich $\varphi|_{\partial\Omega} = 0$, so folgt $\bar{\varphi}(y, d(y)^+) = 0$ und folglich erhält man

$$|\bar{\varphi}(y, s_1)| \leq \int_0^{d(y)^+} |\partial_s \bar{\varphi}(y, s)| ds.$$

Integration über s_1 und y ergibt dann analog zur Vorgehensweise oben

$$\|\varphi\|_{L^1(D_h(\tilde{T}))} \leq ch(\tilde{T})^2 \|\nabla \varphi\|_{L^1(D_h(\tilde{T}))}.$$

Angewendet auf φ^2 folgt nun

$$\|\varphi\|_{L^2(D_h(\tilde{T}))}^2 \leq ch(\tilde{T})^2 (\|\varphi\|_{L^2(D_h(\tilde{T}))} \|\nabla \varphi\|_{L^2(D_h(\tilde{T}))})$$

und somit

$$\|\varphi\|_{L^2(D_h(\tilde{T}))} \leq ch(\tilde{T})^2 \|\nabla \varphi\|_{L^2(D_h(\tilde{T}))}.$$

Summiert man nun das Quadrat dieser Ungleichung über alle $\tilde{T} \in \partial\Omega_h$ auf, so erhält man schließlich

$$\|\varphi\|_{L^2(\Omega_h \setminus \Omega)} \leq ch^2 \|\nabla \varphi\|_{L^2(\Omega_h \setminus \Omega)}.$$

Bleibt noch die Abschätzung von $\|\varphi\|_{L^2(\partial\Omega \cap \Omega_h)}$ zu beweisen. Hierzu verwendet man $s_1 = 0$ und $s_2 = d(y)^+$ und erhält mit $\bar{\varphi}(y, d(y)^+) = 0$

$$|\bar{\varphi}(y, 0)| \leq \int_0^{d(y)^+} |\partial_s \bar{\varphi}(y, s)| ds.$$

Die Abschätzung folgt dann analog zu der Vorgehensweise bei den anderen Abschätzungen. \square

Mit dem Lemma 2.47 haben wir nun alles zur Hand um eine erste Fehlerabschätzung beweisen zu können.

Satz 2.48 (Abstrakte Fehlerabschätzung)

Es gelten die Bezeichnungen und Voraussetzungen aus Definition 2.40. Zusätzlich sei $(\Omega_h, \mathcal{T}_h)$ reguläre Approximation von Ω und für die exakte Lösung gelte $u \in H^2(\Omega)$. Sei \bar{u} eine Fortsetzung von u mit den Eigenschaften

$$\bar{u} \in H^1(\Omega \cup \Omega_h), \quad \bar{u} \in H^2(\Omega_h \setminus \Omega), \quad \bar{u}|_{\Omega} = u.$$

Dann gilt für die Finite Elemente Approximation $u_h \in X_h$ aus Definition 2.40 die Fehlerabschätzung

$$\begin{aligned} \|\bar{u} - u_h\|_{H^1(\Omega_h)} &\leq c \inf_{v_h \in X_h} \|\bar{u} - v_h\|_{H^1(\Omega_h)} \\ &\quad + ch \|\nabla \bar{u} \cdot n\|_{L^2(\partial\Omega \cap \Omega_h)} + ch^2 \|\bar{f} + \Delta \bar{u}\|_{L^2(\Omega_h \setminus \Omega)}. \end{aligned}$$

Dabei bezeichnet $[\nabla \bar{u} \cdot n] = (\nabla u|_{\Omega \cap \Omega_h} - \nabla \bar{u}|_{\Omega_h \setminus \Omega}) \cdot n$ den Sprung der Normalableitung auf $\partial\Omega \cap \Omega_h$.

Beweis: Wir beweisen zunächst eine Abschätzung, die die Galerkin-Orthogonalität in diesem Fall ersetzt. Sei dazu $\varphi_h \in X_h$. Dann gilt

$$B_h(\bar{u} - u_h, \varphi_h) = B_h(\bar{u}, \varphi_h) - (\bar{f}, \varphi_h)_{L^2(\Omega_h)}.$$

Wegen $\Omega_h = (\Omega \cap \Omega_h) \cup (\Omega_h \setminus \Omega)$ und $-\Delta u = f$ f.ü. auf Ω folgt dann

$$\begin{aligned}
 B_h(\bar{u} - u_h, \varphi_h) &= B_h(\bar{u}, \varphi_h) - (f, \varphi_h)_{L^2(\Omega \cap \Omega_h)} - (\bar{f}, \varphi_h)_{L^2(\Omega_h \setminus \Omega)} \\
 &= B_h(\bar{u}, \varphi_h) + (\Delta u, \varphi_h)_{L^2(\Omega \cap \Omega_h)} - (\bar{f}, \varphi_h)_{L^2(\Omega_h \setminus \Omega)} \\
 &\quad - (\Delta \bar{u}, \varphi_h)_{L^2(\Omega_h \setminus \Omega)} + (\Delta \bar{u}, \varphi_h)_{L^2(\Omega_h \setminus \Omega)} \\
 &= B_h(\bar{u}, \varphi_h) - (\bar{f} + \Delta \bar{u}, \varphi_h)_{L^2(\Omega_h \setminus \Omega)} \\
 &\quad + (\Delta u, \varphi_h)_{L^2(\Omega \cap \Omega_h)} + (\Delta \bar{u}, \varphi_h)_{L^2(\Omega_h \setminus \Omega)} \\
 &= B_h(\bar{u}, \varphi_h) - (\bar{f} + \Delta \bar{u}, \varphi_h)_{L^2(\Omega_h \setminus \Omega)} \\
 &\quad - (\nabla u, \nabla \varphi_h)_{L^2(\Omega \cap \Omega_h)} + (\nabla u \cdot n, \varphi_h)_{L^2(\partial(\Omega \cap \Omega_h))} \\
 &\quad - (\nabla \bar{u}, \nabla \varphi_h)_{L^2(\Omega_h \setminus \Omega)} + (\nabla \bar{u} \cdot n, \varphi_h)_{L^2(\partial(\Omega_h \setminus \Omega))}
 \end{aligned}$$

Da φ_h auf $\partial\Omega_h$ verschwindet, können die Randintegrale über $\partial(\Omega_h \setminus \Omega)$ und $\partial(\Omega \cap \Omega_h)$ durch entsprechende Integrale über $\partial\Omega \cap \Omega_h$ ersetzt werden, wobei zu beachten ist, dass die Normalen in entgegengesetzte Richtungen zeigen. Zusammen mit $(\Omega \cap \Omega_h) \cup (\Omega_h \setminus \Omega) = \Omega_h$ folgt dann

$$\begin{aligned}
 B_h(\bar{u} - u_h, \varphi_h) &= -(\bar{f} + \Delta \bar{u}, \varphi_h)_{L^2(\Omega_h \setminus \Omega)} + ((\nabla u|_{\Omega \cap \Omega_h} - \nabla \bar{u}|_{\Omega_h \setminus \Omega}) \cdot n, \varphi_h)_{L^2(\partial\Omega \cap \Omega_h)} \\
 &=: -(\bar{f} + \Delta \bar{u}, \varphi_h)_{L^2(\Omega_h \setminus \Omega)} + ([\nabla \bar{u} \cdot n], \varphi_h)_{L^2(\partial\Omega \cap \Omega_h)}
 \end{aligned}$$

wobei n jetzt die äußere Normale an $\Omega \cap \Omega_h$ ist. Mit Cauchy-Schwarz und Lemma 2.47 erhalten wir nun

$$\begin{aligned}
 B_h(\bar{u} - u_h, \varphi_h) &\leq \|[\nabla \bar{u} \cdot n]\|_{L^2(\partial\Omega \cap \Omega_h)} \|\varphi_h\|_{L^2(\partial\Omega \cap \Omega_h)} \\
 &\quad + \|\bar{f} + \Delta \bar{u}\|_{L^2(\Omega_h \setminus \Omega)} \|\varphi_h\|_{L^2(\Omega_h \setminus \Omega)} \\
 &\leq c \|[\nabla \bar{u} \cdot n]\|_{L^2(\partial\Omega \cap \Omega_h)} h \|\varphi_h\|_{H^1(\Omega_h)} \\
 &\quad + c \|\bar{f} + \Delta \bar{u}\|_{L^2(\Omega_h \setminus \Omega)} h^2 \|\varphi_h\|_{H^1(\Omega_h)} \\
 &= c \left(h \|[\nabla \bar{u} \cdot n]\|_{L^2(\partial\Omega \cap \Omega_h)} + h^2 \|\bar{f} + \Delta \bar{u}\|_{L^2(\Omega_h \setminus \Omega)} \right) \|\varphi_h\|_{H^1(\Omega_h)}.
 \end{aligned}$$

Diese Ungleichung ersetzt die Galerkin-Orthogonalität und so können wir jetzt analog zur Herleitung im Lemma von Cea vorgehen, indem wir die Koerzivität und Stetigkeit von B_h ausnutzen.

$$\begin{aligned}
 c_0 \|u_h - \varphi_h\|_{H^1(\Omega_h)}^2 &\leq B_h(u_h - \varphi_h, u_h - \varphi_h) \\
 &\leq B_h(\bar{u} - \varphi_h, u_h - \varphi_h) - B_h(\bar{u} - u_h, u_h - \varphi_h) \\
 &\leq \|\bar{u} - \varphi_h\|_{H^1(\Omega_h)} \|u_h - \varphi_h\|_{H^1(\Omega_h)} \\
 &\quad + c \left(h \|[\nabla \bar{u} \cdot n]\|_{L^2(\partial\Omega \cap \Omega_h)} + h^2 \|\bar{f} + \Delta \bar{u}\|_{L^2(\Omega_h \setminus \Omega)} \right) \|u_h - \varphi_h\|_{H^1(\Omega_h)}
 \end{aligned}$$

Also folgt

$$\begin{aligned} \|u_h - \varphi_h\|_{H^1(\Omega_h)} &\leq c\|\bar{u} - \varphi_h\|_{H^1(\Omega_h)} \\ &\quad + c\left(h\|[\nabla \bar{u} \cdot n]\|_{L^2(\partial\Omega \cap \Omega_h)} + h^2\|\bar{f} + \Delta \bar{u}\|_{L^2(\Omega_h \setminus \Omega)}\right). \end{aligned}$$

Mit der Dreiecksungleichung gilt weiter

$$\|\bar{u} - u_h\|_{H^1(\Omega_h)} \leq c\|\bar{u} - \varphi_h\|_{H^1(\Omega_h)} + \|u_h - \varphi_h\|_{H^1(\Omega_h)}$$

und somit erhalten wir insgesamt

$$\begin{aligned} \|\bar{u} - u_h\|_{H^1(\Omega_h)} &\leq c\|\bar{u} - \varphi_h\|_{H^1(\Omega_h)} \\ &\quad + c\left(h\|[\nabla \bar{u} \cdot n]\|_{L^2(\partial\Omega \cap \Omega_h)} + h^2\|\bar{f} + \Delta \bar{u}\|_{L^2(\Omega_h \setminus \Omega)}\right). \end{aligned}$$

Da dies für alle $\varphi_h \in X_h$ gilt, folgt hieraus die Behauptung. \square

Bemerkung 2.49

Die abstrakte Abschätzung in Satz 2.48 ist ein wichtiger Schritt in Richtung einer a priori Fehlerabschätzung. Dabei bleiben nun vor allem zwei Punkte offen.

- (i) Wie kommen wir von der Abschätzung für $\|\bar{u} - u_h\|_{H^1(\Omega_h)}$ zu einer Abschätzung für $\|u - u_h\|_{H^1(\Omega)}$?
- (ii) Wie müssen die Fortsetzungen \bar{f} und \bar{u} gewählt werden, um zu einer optimalen Fehlerabschätzung zu gelangen.

Beispiel 2.50 (Fortsetzung von f und u durch Null)

Setzen wir $\bar{u} = \bar{f} = 0$ auf $\Omega_h \setminus \Omega$, so reduziert sich die abstrakte Fehlerabschätzung aus Satz 2.48 auf

$$\|\bar{u} - u_h\|_{H^1(\Omega_h)} \leq c \inf_{v_h \in X_h} \|\bar{u} - v_h\|_{H^1(\Omega_h)} + ch\|[\nabla \bar{u} \cdot n]\|_{L^2(\partial\Omega \cap \Omega_h)}.$$

Dies sieht zunächst brauchbar aus, allerdings ist \bar{u} dann nicht in H^2 , so dass wir zur weiteren Abschätzung des Approximationsfehler nicht die optimalen Interpolationsabschätzungen verwenden könnten.

Es liegt also nahe, dass wir zur weiteren Untersuchung eine Fortsetzung \bar{u} von u benötigen, die auf $\Omega_h \cup \Omega$ in H^2 liegt. Eine solche Fortsetzung liefert das folgende Lemma.

Lemma 2.51 (Lokale H^2 -Fortsetzung)

Sei $\Omega \subset \mathbb{R}^d$ ein Gebiet mit $\partial\Omega \in C^3$. Weiter sei $\Omega_\delta := \Omega \cup U_\delta$. Dann ist durch

$$\bar{u}(x) := \begin{cases} u(x) & , \text{falls } x \in \Omega, \\ -u(x - 2d(x)n(x)) & , \text{falls } x \in \Omega_\delta \setminus \Omega \end{cases}$$

eine Fortsetzung $\bar{u} \in H^2(\Omega_\delta)$ von $u \in H^2(\Omega) \cap H_0^1(\Omega)$ definiert und es gilt

$$\|\bar{u}\|_{H^2(\Omega_\delta)} \leq c\|u\|_{H^2(\Omega)}.$$

Beweis: Aufgrund der Definition ist klar, dass $\bar{u} \in L^2(\Omega_\delta)$ ist und, dass gilt $\|\bar{u}\|_{L^2(\Omega_\delta)} \leq c\|u\|_{L^2(\Omega)}$ gilt.

Wir zeigen nun: $\bar{u} \in H^1(\Omega_\delta)$. Sei dazu $\varphi \in C_0^\infty(\Omega_\delta)$, so gilt

$$\begin{aligned} \int_{\Omega_\delta} \partial_i \varphi \bar{u} &= \int_{\Omega} \partial_i \varphi \bar{u} + \int_{\Omega_\delta \setminus \Omega} \partial_i \varphi \bar{u} \\ &= \int_{\partial\Omega} \varphi u|_{\Omega} n_i - \int_{\Omega} \varphi \partial_i u + \int_{\partial(\Omega_\delta \setminus \Omega)} \varphi \bar{u} n_i - \int_{\Omega_\delta \setminus \Omega} \varphi \partial_i \bar{u} \\ &= \int_{\partial\Omega} \varphi u|_{\Omega} n_i - \int_{\partial\Omega} \varphi \bar{u}|_{\Omega_h \setminus \Omega} n_i + \int_{\Omega_\delta} \varphi \partial_i \bar{u}. \end{aligned}$$

Da $d = 0$ auf $\partial\Omega$ ist, folgt für $x \in \partial\Omega$

$$u(x) = -u(x) = -u(x - 2d(x)n(x)) = \bar{u}(x).$$

Also ist \bar{u} schwach differenzierbar und man sieht leicht ein, dass $\partial_i \bar{u} \in L^2(\Omega_\delta)$ ist und dass es eine Konstante c gibt mit

$$\|\nabla \bar{u}\|_{L^2(\Omega_\delta)} \leq c\|\nabla u\|_L^2(\Omega).$$

Untersuchen wir also nun, ob $\bar{u} \in H^2(\Omega_\delta)$ ist. Für $\varphi \in C_0^\infty(\Omega_\delta)$ gilt

$$\begin{aligned} \int_{\Omega_\delta} \partial_j \varphi \partial_i \bar{u} &= \int_{\Omega} \partial_j \varphi \partial_i u + \int_{\Omega_\delta \setminus \Omega} \partial_j \varphi \partial_i \bar{u} \\ &= \int_{\partial\Omega} \varphi \partial_i u|_{\Omega} n_j - \int_{\Omega} \varphi \partial_j \partial_i u + \int_{\partial(\Omega_\delta \setminus \Omega)} \varphi \partial_i \bar{u}|_{\Omega_\delta \setminus \Omega} n_j - \int_{\Omega_\delta \setminus \Omega} \varphi \partial_j \partial_i \bar{u} \\ &= \int_{\Omega_\delta} \varphi \partial_j \partial_i \bar{u} + \int_{\partial\Omega} \varphi n_j (\partial_i u|_{\Omega} - \partial_i \bar{u}|_{\Omega \setminus \Omega}). \end{aligned}$$

Wir müssen also zeigen, dass das Integral über den Rand verschwindet. Betrachten wir dazu \bar{u} für $x \in \Omega_\delta \setminus \Omega$, so gilt

$$\bar{u}(x) = -u(x - 2d(x)n(x))$$

und somit

$$\partial_i \bar{u} = - \sum_{k=1}^d \partial_k u(x - 2d(x)n(x))(\delta_{ik} - 2d(x)H_{ik}(x) - 2n_i(x)n_k(x)).$$

Werten wir dies aus für $x \in \partial\Omega$, so erhalten wir mit $d(x) = 0$

$$\partial_i \bar{u} = - \sum_{k=1}^d \partial_k u(x)(\delta_{ik} - 2n_i(x)n_k(x)).$$

Um nachzuweisen, dass $\bar{u} \in H^2(\Omega_\delta)$ ist, müssen wir also zeigen, dass für $x \in \partial\Omega$ gilt

$$\partial_i u = - \sum_{k=1}^d \partial_k u(x)(\delta_{ik} - 2n_i(x)n_k(x)).$$

Dazu stellen wir zunächst fest, dass aufgrund der Nullrandwerte von u die Ableitungen in alle Tangentialrichtungen des Randes Null sind, d.h. es gilt auf $\partial\Omega$

$$\partial_k u = (\nabla u \cdot n)n_k.$$

Hiermit folgt dann

$$\begin{aligned} - \sum_{k=1}^d \partial_k u(x)(\delta_{ik} - 2n_i(x)n_k(x)) &= -\partial_i u + 2(\nabla u \cdot n)n_i \sum_{k=1}^d n_k^2 \\ &= -\partial_i u + 2\partial_i u = \partial_i u. \end{aligned}$$

Damit haben wir die zweifache schwache Differenzierbarkeit gezeigt und die Abschätzung

$$\|\bar{u}\|_{H^2(\Omega)} \leq c\|u\|_{H^2(\Omega)}$$

folgt durch nachrechnen. □

Mit der lokalen H^2 -Fortsetzung haben wir nun alle Vorbereitungen getroffen, um eine a priori Fehlerabschätzung mit Randapproximation zu beweisen.

Satz 2.52 (A priori Fehlerabschätzung mit Randapproximation)

Sei $\Omega \subset \mathbb{R}^d, d \leq 3$ ein Gebiet mit $\partial\Omega \in C^3$. Sei $f \in L^2(\Omega)$ und sei $u \in H_0^1(\Omega) \cap H^2(\Omega)$ die schwache Lösung von

$$-\Delta u = f \quad \text{f. ü. in } \Omega, \quad u = 0 \quad \text{auf } \partial\Omega.$$

Seien weiter $(\Omega_h, \mathcal{T}_h)$ eine reguläre Gebietsapproximation von Ω und $\bar{f} \in L^2(\Omega \cup \Omega_h)$ eine Fortsetzung von f mit $\|f\|_{L^2(\Omega \cup \Omega_h)} \leq c\|f\|_{L^2(\Omega)}$. Ist dann $u_h \in S_{h,0}^k \subset H_0^1(\Omega_h)$ die Finite Elemente Lösung aus Definition 2.40, so gilt die a priori Fehlerabschätzung

$$\|u - u_h\|_{H^1(\Omega)} \leq ch\|u\|_{H^2(\Omega)}.$$

Dabei sei u_h durch Null außerhalb von Ω_h fortgesetzt.

Beweis: Sei \bar{u} die lokale H^2 Fortsetzung aus Lemma 2.51. Dann liefert die abstrakte Fehlerabschätzung 2.48

$$\|\bar{u} - u_h\|_{H^1(\Omega_h)} \leq c \inf_{v_h \in X_h} \|\bar{u} - v_h\|_{H^1(\Omega_h)} + ch^2 \|\bar{f} + \Delta \bar{u}\|_{L^2(\Omega_h \setminus \Omega)}.$$

Mit der Interpolationsabschätzung und den Abschätzungen für die Fortsetzungen \bar{f} und \bar{u} folgt dann

$$\|\bar{u} - u_h\|_{H^1(\Omega_h)} \leq ch\|u\|_{H^2(\Omega)}.$$

Also folgt wegen $u_h = 0$ außerhalb von Ω_h

$$\begin{aligned} \|\bar{u} - u_h\|_{H^1(\Omega)} &\leq \|\bar{u} - u_h\|_{H^1(\Omega \cup \Omega_h)} \\ &\leq \|\bar{u} - u_h\|_{H^1(\Omega_h)} + \|\bar{u} - u_h\|_{H^1(\Omega \setminus \Omega_h)} \\ &\leq \|\bar{u} - u_h\|_{H^1(\Omega_h)} + \|u\|_{H^1(\Omega \setminus \Omega_h)} \\ &\leq ch\|u\|_{H^2(\Omega)} + c\|u\|_{H^1(\Omega \setminus \Omega_h)}. \end{aligned}$$

Analog zu Lemma 2.47 zeigt man für $\Omega \setminus \Omega_h$

$$\|u\|_{H^1(\Omega \setminus \Omega_h)} \leq ch\|u\|_{H^2(\Omega)}.$$

Damit ist die Abschätzung gezeigt. \square

Die Fehlerabschätzung zeigt, dass für stückweise lineare Finite Elemente Verfahren die optimale Konvergenzordnung h erhalten bleibt. Leider ist dies nicht mehr für Lagrange Elemente höherer Ordnung der Fall. Mit etwas

zusätzlichem Aufwand kann man für quadratische Elemente zeigen, dass die Abschätzung mit $h^{3/2}$ gilt. Um auch für Lagrange Elemente höherer Ordnung optimale Konvergenzraten zu erhalten, muss die Randapproximation verbessert werden. Dies liefern beispielsweise Isoparametrische Elemente, die eine polynomiale Randapproximation erlauben. Auf eine genauere Darstellung und Analyse dieser Methodik verweisen wir auf die Literatur (Ciarlet. The Finite element Method for elliptic problems. North-Holland, 1978. Abschnitt 4.3).

2.5 A posteriori Fehlerabschätzungen und Adaptivität

Bislang haben wir nur a priori Fehlerabschätzungen bewiesen. Solche Abschätzungen geben uns asymptotische Informationen über das Konvergenzverhalten der Verfahren. Zudem liefern sie nur globale Informationen über den Fehler und keine Information, wo der Fehler besonders groß oder besonders klein ist. In vielen Anwendungen ist man jedoch daran interessiert für konkrete Gitter eine Abschätzung des globalen Fehlers oder sogar der lokalen Verteilung des Fehlers zu erhalten. Liegt der geschätzte Fehler dann im Toleranzbereich, so wird die Approximation akzeptiert, falls nicht, muss versucht werden die Finite Elemente Lösung durch globale oder lokale Gitterverfeinerung zu verbessern.

Stellen wir uns also vor, wir hätten eine Fehlerabschätzung der Form

$$\|u - u_h\| \leq \eta_h(u_h),$$

wobei $\eta_h(u_h)$ berechnet werden kann, sobald die Approximation u_h bekannt ist. Eine solche Fehlerabschätzung heißt a posteriori Fehlerabschätzung. Sie gibt eine rigorose berechenbare obere Schranke des Fehlers an. Jedoch müssen wir feststellen, dass eine solche Abschätzung alleine noch nicht sehr hilfreich ist. Z.B. könnte $\eta_h(u_h) = \text{const}$ eine obere Schranke sein, aber keine Informationen über den tatsächlichen Fehler liefern. Wir führen daher folgende Begriffe ein:

Definition 2.53 (Zuverlässige und effiziente Fehlerschätzer)

Ein Fehlerschätzer $\eta_h(u_h)$ heißt zuverlässig für den Fehler $\|u - u_h\|$, falls gilt

$$\|u - u_h\| \leq \eta_h(u_h).$$

Er heißt effizient mit Konstante $c > 0$, falls gilt

$$c \eta_h(u_h) \leq \|u - u_h\|.$$

Im folgenden werden wir einen Fehlerschätzer herleiten, der zuverlässig und effizient ist. Darüberhinaus werden wir zeigen, dass der Fehlerschätzer lokale

Informationen über den Fehler enthält.

2.5.1 Residuenfehlerschätzer

Auch in diesem Abschnitt wollen wir uns auf die Analyse des homogenen Dirichletproblems für die Poissons-Gleichung beschränken. Sei also stets $\Omega \subset \mathbb{R}^d$ ein Gebiet, $f \in H^{-1}(\Omega)$ und $u \in X := H_0^1(\Omega)$ schwache Lösung von

$$-\Delta u = f \quad \text{in } \Omega,$$

d.h. mit $B(u, v) := \int_{\Omega} \nabla u \cdot \nabla v$ gilt

$$B(u, \varphi) = f(\varphi), \quad \forall \varphi \in X.$$

Weiter sei \mathcal{T}_h eine zulässige Triangulierung von Ω und $u_h \in X_h := S_{h,0}^k$ die Lagrange Finite Elemente Lösung, d.h.

$$B(u_h, \varphi_h) = f(\varphi_h) \quad \forall \varphi_h \in X_h.$$

Ist eine Gleichung $Au - f = 0$ gegeben, so bezeichnet man mit $R(v) := Av - f$ das Residuum von v . Das Residuum ist also ein Maß dafür, wie gut v die Gleichung erfüllt. Betrachten wir nun die Differentialgleichung $-\Delta u = f$, so sind wir an dem Residuum “ $\Delta u_h + f$ ” interessiert, jedoch ist zunächst unklar in welchem Sinne wir ‘ $\Delta u_h + f$ ’ verstehen sollen.

Lemma 2.54 (Der Laplace Operator)

Sein $\Omega \subset \mathbb{R}^d$ ein Gebiet mit Lipschitzrand. Dann ist durch

$$-\Delta u(\varphi) := B(u, \varphi) := \int_{\Omega} \nabla u \cdot \nabla \varphi \quad \forall \varphi \in H_0^1(\Omega)$$

ein Isomorphismus $-\Delta : H_0^1(\Omega) \rightarrow H^{-1}(\Omega)$ gegeben und es gilt für alle $u \in H_0^1(\Omega)$

$$\|-\Delta u\|_{H^{-1}(\Omega)} \leq \|\nabla u\|_{L^2(\Omega)}.$$

Beweis: Sei $u \in H_0^1(\Omega)$ gegeben, dann gilt für alle $\varphi \in H_0^1(\Omega)$

$$|-\Delta u(\varphi)| \leq \|\nabla u\|_{L^2(\Omega)} \|\nabla \varphi\|_{L^2(\Omega)} \leq \|\nabla u\|_{H^1(\Omega)} \|\nabla \varphi\|_{H^1(\Omega)}.$$

Also folgt

$$\|-\Delta u\|_{H^{-1}(\Omega)} := \sup_{\varphi \in H^1(\Omega) \setminus \{0\}} \frac{|-\Delta u(\varphi)|}{\|\varphi\|_{H^1(\Omega)}} \leq \|\nabla u\|_{H^1(\Omega)}.$$

Da $-\Delta u$ linear ist folgt $-\Delta u \in H^{-1}(\Omega)$. Auch $-\Delta$ ist linear, denn für $\varphi \in H_0^1(\Omega)$ gilt

$$\begin{aligned}-\Delta(u_1 + u_2)(\varphi) &= B(u_1 + u_2, \varphi) = B(u_1, \varphi) + B(u_2, \varphi) \\ &= -\Delta(u_1)(\varphi) + (-\Delta(u_2)(\varphi)).\end{aligned}$$

Insgesamt folgt also

$$-\Delta \in L(H_0^1(\Omega), H_{-1}(\Omega)), \quad \|-\Delta\|_{L(H_0^1(\Omega), H_{-1}(\Omega))} \leq 1.$$

Injectivität und Surjektivität folgen aus der eindeutigen Lösbarkeit des homogenen Dirichletproblems. Dabei geht die Poincaréungleichung ein, denn für $u \in H_0^1(\Omega)$ gilt

$$-\Delta u(u) = \|\nabla u\|_{L^2(\Omega)}^2 \geq \frac{1}{1 + c_p^2} \|u\|_{H^1(\Omega)}^2.$$

Zusammen mit der oberen Abschätzung

$$-\Delta u(u) \leq \|-\Delta u\|_{H^{-1}(\Omega)} \|u\|_{H^1(\Omega)} \leq \|-\Delta\|_{L(H_0^1(\Omega), H_{-1}(\Omega))} \|u\|_{H^1(\Omega)}^2$$

folgt also

$$\frac{1}{1 + c_p^2} \leq \|-\Delta\|_{L(H_0^1(\Omega), H_{-1}(\Omega))} \leq 1.$$

□

Da $X_h \subset X$ ist, hat $-\Delta u_h$ jetzt also einen Sinn. Analog ist dann auch $-\Delta_h : X_h \rightarrow X'_h$ wohldefiniert durch

$$-\Delta_h u_h(\varphi_h) := B(u_h, \varphi_h).$$

Da $H^{-1}(\Omega) = X' \subset X'_h$ ist, können wir die Gleichung

$$-\Delta_h u_h = f \quad \text{in } X_h$$

für alle $f \in X'$ lösen. Außerdem gilt

$$-\Delta_h = -\Delta|_{X_h}.$$

Dabei ist klar, dass $-\Delta u_h$ nicht gleich $-\Delta_h u_h$ ist. In diesem abstrakten Setting führen wir nun das Residuum ein und erhalten damit einen zuverlässigen und effizienten Fehlerschätzer.

Lemma 2.55 (Residuum und Residuumsfehlerschätzer)

Seien die Voraussetzungen von Lemma 2.54 erfüllt und sei $f \in H^{-1}(\Omega)$. Dann ist das Residuum $R : H_0^1(\Omega) \rightarrow H^{-1}(\Omega)$ der Poisongleichung definiert durch

$$R(v) := -\Delta v - f, \quad \forall v \in H_0^1(\Omega).$$

Ist $u \in H_0^1(\Omega)$ die schwache Lösung der Poisongleichung und $u_h \in S_{h,0}^k$ die zugehörige Finite Elemente Approximation, so gilt

$$\sqrt{\frac{1}{1 + c_p^2}} \|\nabla(u - u_h)\|_{L^2(\Omega)} \leq \|R(u_h)\|_{H^{-1}(\Omega)} \leq \|\nabla(u - u_h)\|_{L^2(\Omega)}.$$

Dabei ist c_p die Poincarékonstante zu $H_0^1(\Omega)$.

Beweis: Es ist

$$\begin{aligned} \|R(u_h)\|_{H^{-1}(\Omega)} &= \|-(f + \Delta u_h)\|_{H^{-1}(\Omega)} = \|-\Delta u + \Delta u_h\|_{H^{-1}(\Omega)} \\ &= \|\Delta(u_h - u)\|_{H^{-1}(\Omega)} \leq \|\nabla(u - u_h)\|_{L^2(\Omega)}. \end{aligned}$$

Andererseits ist wegen $u - u_h \in H_0^1(\Omega)$

$$R(u_h)(u - u_h) = \|\nabla(u - u_h)\|_{L^2(\Omega)}^2.$$

Also folgt

$$\begin{aligned} \|R(u_h)\|_{H^{-1}(\Omega)} &\geq \frac{|R(u_h)(u - u_h)|}{\|u - u_h\|_{H^1(\Omega)}} \\ &= \|\nabla(u - u_h)\|_{L^2(\Omega)} \frac{\|\nabla(u - u_h)\|_{L^2(\Omega)}}{\|u - u_h\|_{H^1(\Omega)}} \\ &\geq \|\nabla(u - u_h)\|_{L^2(\Omega)} \sqrt{\frac{1}{1 + c_p^2}}. \end{aligned}$$

□

Lemma 2.55 zeigt, dass das Residuum $R(u_h)$ in der H^{-1} -Norm gemessen, zumindest global, ein zuverlässiger und effizienter Fehlerschätzer für den Fehler in der H^1 -Norm ist. Dabei bleiben zwei Fragen jedoch offen

- 1) Kann der Fehlerschätzer lokalisiert werden?
- 2) Wie lässt sich die H^{-1} -Norm des Residuums numerisch berechnen?

Lemma 2.56 (Lokalisierung des Residuums)

Seien die Voraussetzungen von Lemma 2.55 erfüllt und zusätzlich $f \in L^2(\Omega)$. Dann gilt für alle $\varphi \in H_0^1(\Omega)$ und alle $\varphi_h \in S_{h,0}^k$

$$\begin{aligned} R(u_h)(\varphi) &= \sum_{T \in \mathcal{T}_h} \left(- (f + \Delta u_h, \varphi - \varphi_h)_{L^2(T)} \right. \\ &\quad \left. + \frac{1}{2} \sum_{S \subset \partial T \setminus \partial \Omega} ([\nabla u_h \cdot n], \varphi - \varphi_h)_{L^2(S)} \right). \end{aligned}$$

Dabei Bezeichnet $[\nabla u_h \cdot n]$ den Sprung der Normalenableitung von u_h über die Seitenfläche.

Beweis: Mit der Definition des Residuums folgt

$$R(u_h)(\varphi) = - \int_{\Omega} f \varphi + \int_{\Omega} \nabla u_h \cdot \nabla \varphi = \sum_{T \in \mathcal{T}_h} \int_T \left(- f \varphi + \nabla u_h \cdot \nabla \varphi \right).$$

Da $u_h|_T$ ein Polynom k -ten Grades ist können wir auf T lokal partiell integrieren und erhalten

$$\int_T \nabla u_h \cdot \nabla \varphi = \int_{\partial T} \nabla u_h \cdot n \varphi + \int_T -\Delta u_h \varphi.$$

Da φ Nullrandwerte hat folgt also insgesamt

$$R(u_h)(\varphi) = \sum_{T \in \mathcal{T}_h} \left(- (f + \Delta u_h, \varphi)_{L^2(T)} + \sum_{S \subset \partial T \setminus \partial \Omega} (\nabla u_h \cdot n, \varphi)_{L^2(S)} \right).$$

Verwenden wir nun, dass Seitenflächen S in der Summation doppelt vorkommen, so erhalten wir durch Umsortierung der Summation

$$R(u_h)(\varphi) = \sum_{T \in \mathcal{T}_h} \left(- (f + \Delta u_h, \varphi)_{L^2(T)} + \sum_{S \subset \partial T \setminus \partial \Omega} \frac{1}{2} ([\nabla u_h \cdot n], \varphi)_{L^2(S)} \right).$$

Aufgrund der Galerkinorthogonalität gilt weiter $R(u_h)(\varphi_h) = 0$, also folgt

$$R(u_h)(\varphi) = R(u_h)(\varphi) - R(u_h)(\varphi_h) = R(u_h)(\varphi - \varphi_h).$$

Zusammen mit der lokalen Darstellung von $R(u_h)(\varphi)$ folgt hieraus die Behauptung. \square

Um eine lokalisierte Abschätzung von $\|R(u_h)\|_{H^{-1}(\Omega)}$ zu erhalten, könnten wir mit Hilfe der lokalen Darstellung nun wie folgt vorgehen

$$\begin{aligned}
 |R(u_h)(\varphi)| &\leq \sum_{T \in \mathcal{T}_h} \left((|f + \Delta u_h|, |\varphi - \varphi_h|)_{L^2(T)} \right. \\
 &\quad \left. + \frac{1}{2} \sum_{S \subset \partial T \setminus \partial \Omega} (|[\nabla u_h \cdot n]|, |\varphi - \varphi_h|)_{L^2(S)} \right) \\
 &\leq \sum_{T \in \mathcal{T}_h} \|f + \Delta u_h\|_{L^2(T)} \|\varphi - \varphi_h\|_{L^2(T)} \\
 &\quad + \sum_{T \in \mathcal{T}_h} \frac{1}{2} \sum_{S \subset \partial T \setminus \partial \Omega} \|[\nabla u_h \cdot n]\|_{L^2(S)} \|\varphi - \varphi_h\|_{L^2(S)}.
 \end{aligned}$$

Da diese Abschätzung für alle $\varphi_h \in S_{h,0}^k$ gilt, können wir zum Infimum übergehen und anschließend zur lokalen Abschätzung des Approximationsfehlers

$$\inf_{\varphi_h \in S_{h,0}^k} \|\varphi - \varphi_h\|_{L^2(T)}$$

Interpolationsabschätzungen verwenden. Dabei tritt jedoch das Problem auf, dass φ hier nur in $H_0^1(\Omega)$ liegt und wir daher keine Lagrangeinterpolation verwenden können. Daher werden wir im folgenden Unterabschnitt eine H^1 -Interpolation einführen, für die die gewohnten Interpolationsabschätzungen erhalten bleiben.

2.5.2 H^1 -Interpolation

In Abschnitt 2.3.1 haben wir Interpolationsabschätzungen für die Lagrangeinterpolation hergeleitet. Da die Lagrangeinterpolation auf einer Punktauswertung beruht, können wir eine solche Interpolation und die dazugehörigen Interpolationsfehlerabschätzungen allerdings nicht auch H^1 -Funktionen in mehreren Raumdimensionen anwenden.

Wir wollen daher im Folgenden einen Interpolationsoperator definieren, der ohne Punktauswertungen auskommt und für den die Interpolationsabschätzungen aus Abschnitt 2.3.1 erhalten bleiben. Die Idee ist dabei, die Auswertung einer Funktion in einem Gitterpunkt durch den Mittelwert der Funktion in einer Umgebung des Gitterpunktes zu ersetzen. Dazu formulieren wir folgendes Lemma

Lemma 2.57

Seien $\Omega \subset \mathbb{R}^d$ ein Gebiet, \mathcal{T}_h eine zulässige Triangulierung von Ω und $X_h := S_h^k$ der Raum der linearen Lagrange Elemente. Ist x ein Knoten des Lagrange-gitters $G(\mathcal{T}_h)$ und $\varphi \subset X_h$ die zugehörige Basisfunktion, so definieren wir den Träger von φ durch

$$M := \text{supp}(\varphi)$$

und den Durchmesser $h(M)$ von M durch

$$h(M) := \sup_{x,y \in M} |x - y|.$$

Sei nun

$$X_h(M) := \{v_h|_M \mid v_h \in X_h\}.$$

Dann gibt es eine Abbildung $P_M : L^2(M) \rightarrow X_h(M)$, so dass für alle $v \in L^2(M)$ gilt

$$\|v - P_M(v)\|_{L^2(M)} = \inf_{\varphi_h \in X_h(M)} \|v - \varphi_h\|_{L^2(M)}.$$

Insbesondere ist $\|P_M\| \leq 1$.

Beweis: P_M ist die L^2 -Projektion vom $L^2(M)$ auf den Teilraum $X_h(M)$. Existenz und Eindeutigkeit folgen aus dem Projektionssatz, angewendet auf den Hilbertraum $L^2(M)$.

Korollar 2.58

Mit den Voraussetzungen und Notationen aus Lemma 2.57 gilt für all $v \in H^1(M)$

$$\begin{aligned} \|v - P_M(v)\|_{L^2(M)} &\leq ch(M) \|\nabla v\|_{L^2(M)} \\ |v - P_M(v)|_{H^1(M)} &\leq c \|\nabla v\|_{L^2(M)}. \end{aligned}$$

Beweis: Da konstante Funktionen in $X_h(M)$ liegen gilt

$$\|v - P_M(v)\|_{L^2(M)} \leq \|v - \frac{1}{|M|} \int_M v\|_{L^2(M)}$$

Da die Mittelwertbildung den Raum $\mathbb{P}^0(M)$ invariant lässt, folgt analog zu den lokalen Interpolationsabschätzungen aus Abschnitt 2.3.1 mit einem Skalierungsargument und dem Lemma von Bramble-Hilbert

$$\begin{aligned} \|v - P_M(v)\|_{L^2(M)} &\leq \|v - \frac{1}{|M|} \int_M v\|_{L^2(M)} \\ &\leq ch(M) \|\nabla v\|_{L^2(M)}. \end{aligned}$$

Die zweite Ungleichung sehen wir wie folgt mit einem Skalierungsargument (inverse Ungleichung) ein

$$\begin{aligned} \|\nabla(v - P_M(v))\|_{L^2(M)} &\leq \|\nabla v\|_{L^2(M)} + \|\nabla P_M(v)\|_{L^2(M)} \\ &\leq \|\nabla v\|_{L^2(M)} + \|\nabla(P_M(v) - \frac{1}{|M|} \int_M v)\|_{L^2(M)} \\ &\leq \|\nabla v\|_{L^2(M)} + \|\nabla P_M(v - \frac{1}{|M|} \int_M v)\|_{L^2(M)} \\ &\leq \|\nabla v\|_{L^2(M)} + \frac{c}{h(M)} \|P_M(v - \frac{1}{|M|} \int_M v)\|_{L^2(M)} \\ &\leq \|\nabla v\|_{L^2(M)} + \frac{c}{h(M)} \|v - \frac{1}{|M|} \int_M v\|_{L^2(M)} \\ &\leq \|\nabla v\|_{L^2(M)} + c \|\nabla v\|_{L^2(M)}. \end{aligned}$$

□

Mit Hilfe der lokalen Projektionen P_M wollen wir nun eine Interpolation auf $H^1(\Omega)$ definieren.

Satz 2.59 (H^1 -Interpolation/Clément-Interpolation)

Seien die Voraussetzungen und Notationen aus Lemma 2.57 gegeben und gelte $\sigma(T) \leq \sigma$ für alle $T \in \mathcal{T}_h$. Dann gibt es einen Interpolationsoperator

$$I_h \in L(H^1(\Omega), X_h),$$

so dass für alle $v \in H^1(\Omega)$ gilt

$$\begin{aligned} \|v - I_h v\|_{L^2(\Omega)} &\leq ch \|\nabla v\|_{L^2(\Omega)}, \\ \|\nabla(v - I_h v)\|_{L^2(\Omega)} &\leq c \|\nabla v\|_{L^2(\Omega)}. \end{aligned}$$

Beweis: Sei $N := \dim(X_h)$, $\bar{a}_i \in G(\mathcal{T}_h)$, $\bar{\varphi}_i \in X_h$ die nodale Basisfunktion zu

\bar{a}_j und \bar{M}_i der Träger von $\bar{\varphi}_i$. Dann definieren wir I_h durch

$$(I_h v)(x) := \sum_{i=1}^N (P_{M_i} v)(\bar{a}_i) \bar{\varphi}_i(x).$$

Zu $T \in \mathcal{T}_h$ definieren eine Umgebung $U(T)$ durch

$$U(T) := \bigcup_{T' \cap T \neq \emptyset} T'.$$

Seien $a_j \in G(T), j = 1, \dots, m$ die Lagrange knoten auf dem Simplex T , und φ_j die zugehörigen Formfunktionen der Lagrange Elemente. Dann folgt wegen $\sum_{j=1}^m \varphi_j = 1$ und $|\varphi_j| \leq 1$

$$\begin{aligned} \|v - I_h v\|_{L^2(T)} &= \left\| v - \sum_{j=1}^m (P_{M_j} v)(a_j) \varphi_j(x) \right\|_{L^2(T)} \\ &= \left\| \sum_{j=1}^m (v - (P_{M_j} v)(a_j)) \varphi_j \right\|_{L^2(T)} \\ &\leq \sum_{j=1}^m \|v - (P_{M_j} v)(a_j)\|_{L^2(T)} \\ &\leq \sum_{j=1}^m \left(\|v - P_{M_j} v\|_{L^2(T)} + \|P_{M_j} v - (P_{M_j} v)(a_j)\|_{L^2(T)} \right) \\ &\leq \sum_{j=1}^m \left(\|v - P_{M_j} v\|_{L^2(M_j)} + \|P_{M_j} v - (P_{M_j} v)(a_j)\|_{L^2(T)} \right) \\ &\leq \sum_{j=1}^m c_h(M_j) \|\nabla v\|_{L^2(M_j)} + \sum_{j=1}^m \|P_{M_j} v - (P_{M_j} v)(a_j)\|_{L^2(T)}. \end{aligned}$$

Da $P_{M_j} v \in X_h$, gilt auf T die Darstellung

$$(P_{M_j} v)(x) = \sum_{k=1}^m (P_{M_j} v)(a_k) \varphi_k(x)$$

und wir erhalten die Abschätzung

$$\begin{aligned}
 \|P_{M_j}v - (P_{M_j}v)(a_j)\|_{L^2(T)} &\leq \sum_{k=1}^m \|(P_{M_j}v)(a_k) - (P_{M_j}v)(a_j)\|_{L^2(T)} \\
 &= \sum_{k=1}^m |T|^{1/2} |(P_{M_j}v)(a_k) - (P_{M_j}v)(a_j)| \\
 &\leq \sum_{k=1}^m |T|^{1/2} h(M_j) \|\nabla P_{M_j}v\|_{L^\infty(T)} \\
 &\leq c \sum_{k=1}^m h(M_j) \|\nabla P_{M_j}v\|_{L^2(T)}.
 \end{aligned}$$

Insgesamt folgt also

$$\begin{aligned}
 \|v - I_h v\|_{L^2(T)} &\leq \sum_{j=1}^m c h(M_j) \|\nabla v\|_{L^2(M_j)} + c \sum_{k=1}^m h(M_j) \|\nabla P_{M_j}v\|_{L^2(T)} \\
 &\leq c h \|\nabla v\|_{L^2(U(T))}.
 \end{aligned}$$

Quadrieren und aufsummieren über alle $T \in \mathcal{T}_h$ ergibt die erste Abschätzung. Die zweite Abschätzung erhält man analog zur Vorgehensweise für die zweite Abschätzung in Korollar 2.58. \square

Wollen wir die Interpolation aus Satz 2.59 auf Funktionen in $H_0^1(\Omega)$ anwenden, so ist nicht klar, ob $I_h v$ Nullrandwerte hat, also wieder in $H_0^1(\Omega)$ liegt. Um dies zu erreichen, definieren wir für $v \in H_0^1(\Omega)$ eine modifizierte Interpolierende durch

$$(I_{h,0}v)(x) := \sum_{i=1}^{N_0} (P_{M_i}v)(\bar{a}_i) \bar{\varphi}_i,$$

wobei $\bar{a}_i \in G(\mathcal{T}_h) \cap \Omega$, $i = 1, \dots, N_0$ nur die inneren Punkte des Lagrangegeitters bezeichnen. Dann ist klar, dass $I_{h,0}v|_{\partial\Omega} = 0$ ist. Allerdings müssen dann die lokalen Abschätzungen aus dem Beweis von Satz 2.59 für Randsimplizes neu bewiesen werden. Dabei verwenden wir den folgenden Spursatz, der in den Übungen bewiesen wird.

Satz 2.60 (Spursatz)

Sei T ein d -dimensionaler Simplex im \mathbb{R}^d mit $(d-1)$ -dimensionalem Seiten-simplex S . Dann gibt es eine Konstante $c > 0$, so dass für alle $v \in H^1(T)$

gilt

$$\|v\|_{L^2(S)} \leq c \left(h(T)^{-1/2} \|v\|_{L^2(T)} + h(T)^{1/2} \|\nabla v\|_{L^2(T)} \right).$$

Beweis: Siehe Übungsaufgabe.

Durch Anwendung dieses Spursatzes erhalten wir nun auch eine Interpolationsfehlerabschätzung für $I_{h,0}$.

Korollar 2.61 (H_0^1 -Interpolation)

Seien die Voraussetzungen aus Satz 2.59 erfüllt und sei $X_h := S_{h,0}^k \subset H_0^1(\Omega)$. Dann existiert ein Interpolationsoperator

$$I_{h,0} \in L(H_0^1(\Omega), X_h),$$

so für alle $v \in H_0^1(\Omega)$ gilt

$$\begin{aligned} \|v - I_{h,0}v\|_{L^2(\Omega)} &\leq ch \|\nabla v\|_{L^2(\Omega)}, \\ \|\nabla(v - I_{h,0}v)\|_{L^2(\Omega)} &\leq c \|\nabla v\|_{L^2(\Omega)}. \end{aligned}$$

Beweis: Zum Beweis dieses Satzes kann analog zum Beweis von Satz 2.59 vorgegangen werden. Dabei müssen lediglich die lokalen Abschätzungen für Randsimplizes neu hergeleitet werden. Sei also $T \in \mathcal{T}_h$ ein Simplex mit $T \cap \partial\Omega \neq \emptyset$. Dann gilt

$$\begin{aligned} \|v - I_{h,0}v\|_{L^2(T)} &= \|v - \sum_{a_j \in G(T) \setminus \partial\Omega} (P_{M_j}v)(a_j) \varphi_j(x)\|_{L^2(T)} \\ &\leq \|v - \sum_{a_j \in G(T)} (P_{M_j}v)(a_j) \varphi_j(x)\|_{L^2(T)} \\ &\quad + \left\| \sum_{a_j \in G(T) \cap \partial\Omega} (P_{M_j}v)(a_j) \varphi_j(x) \right\|_{L^2(T)} \\ &\leq \|v - \sum_{a_j \in G(T)} (P_{M_j}v)(a_j) \varphi_j(x)\|_{L^2(T)} \\ &\quad + \sum_{a_j \in G(T) \cap \partial\Omega} |T|^{1/2} |(P_{M_j}v)(a_j)| \end{aligned}$$

Den ersten Summanden können wir jetzt analog zum Beweis von Satz 2.59 abschätzen. Für den zweiten Summanden benötigen wir eine Abschätzung

für

$$|(P_{M_j}v)(a_j)|, \quad a_j \in G(T) \cap \partial\Omega.$$

Sei dazu $S \subset \partial\Omega$ eine Seitenfläche von T mit $a_j \in S \subset T \subset M_j$, dann gilt

$$\begin{aligned} |(P_{M_j}v)(a_j)| &\leq \|P_{M_j}v\|_{L^\infty(S)} \\ &\leq ch(T)^{\frac{d-1}{2}} \|P_{M_j}v\|_{L^2(S)} \quad (\text{inverse Ungleichung}) \\ &= ch(T)^{\frac{d-1}{2}} \|v - P_{M_j}v\|_{L^2(S)} \quad (\text{da } v = 0 \text{ auf dem Rand}) \\ &\stackrel{2.59}{\leq} ch(T)^{\frac{d-1}{2}} \left(h(T)^{-1/2} \|v - P_{M_j}v\|_{L^2(T)} + h(T)^{1/2} \|\nabla(v - P_{M_j}v)\|_{L^2(T)} \right) \\ &\leq ch(T)^{\frac{d-2}{2}} \left(\|v - P_{M_j}v\|_{L^2(M_j)} + h(T) \|\nabla(v - P_{M_j}v)\|_{L^2(M_j)} \right) \\ &\stackrel{2.58}{\leq} ch(T)^{\frac{d-2}{2}} \left(h(M_j) + h(T) \right) \|\nabla v\|_{L^2(M_j)}. \end{aligned}$$

Setzen wir diese Abschätzung in die lokale Abschätzung für $\|v - I_{h,0}v\|_{L^2(T)}$ ein, so erhalten wir analog zur Vorgehensweise im Beweis von Satz 2.59 die Behauptung. \square

2.5.3 Lokalisierte Abschätzungen für den Residuenfehlerschätzer

Wir wollen nun an unsere Diskussion nach Lemma 2.56 anknüpfen, um zu einer lokalisierten Abschätzung des Residuums zu gelangen. Mit Hilfe der H_0^1 -Interpolation sind wir in der Lage folgenden Satz zu beweisen.

Satz 2.62 (Lokalisierte obere Abschätzung des Residuums)

Seien die Voraussetzungen und Notationen aus Lemma 2.56 gegeben. Dann gilt

$$\sqrt{\frac{1}{1 + c_p^2}} \|\nabla(u - u_h)\|_{L^2(\Omega)} \leq \|R(u_h)\|_{H^{-1}(\Omega)} \leq c \eta_h(u_h).$$

Dabei ist

$$\eta_h(u_h) := \left(\sum_{T \in \mathcal{T}_h} \eta_T(u_h)^2 \right)^{1/2}$$

mit den lokalen Fehlerindikatoren $\eta_T(u_h)$ definiert durch

$$\eta_T(u_h)^2 = h(T)^2 \|f + \Delta u_h\|_{L^2(T)}^2 + \frac{1}{2} h(T) \sum_{S \subset \partial T \setminus \partial \Omega} \left\| [\nabla u \cdot n] \right\|_{L^2(S)}^2.$$

Beweis: Die erste Abschätzung ist bereits in Lemma 2.55 enthalten. Zu zeigen bleibt also

$$\|R(u_h)\|_{H^{-1}(\Omega)} \leq c \eta_h(u_h).$$

Dazu starten wir mit der Abschätzung, die wir in Anschluss an Lemma 2.56 hergeleitet haben

$$\begin{aligned} |R(u_h)(\varphi)| &\leq \sum_{T \in \mathcal{T}_h} \|f + \Delta u_h\|_{L^2(T)} \|\varphi - \varphi_h\|_{L^2(T)} \\ &\quad + \sum_{T \in \mathcal{T}_h} \frac{1}{2} \sum_{S \subset \partial T \setminus \partial \Omega} \|[\nabla u_h \cdot n]\|_{L^2(S)} \|\varphi - \varphi_h\|_{L^2(S)}. \end{aligned}$$

Wählen wir nun $\varphi_h = I_{h,0}\varphi$, so folgt mit dem Spursatz 2.60

$$\|\varphi - \varphi_h\|_{L^2(S)} \leq c \left(h(T)^{-1/2} \|\varphi - I_{h,0}\varphi\|_{L^2(T)} + h(T)^{1/2} \|\nabla(\varphi - I_{h,0}\varphi)\|_{L^2(T)} \right)$$

und mit den lokalen Abschätzungen aus dem Beweis von Satz 2.61

$$\begin{aligned} \|\varphi - I_{h,0}\varphi\|_{L^2(T)} &\leq c h(T) \|\nabla \varphi\|_{L^2(U(T))}, \\ \|\nabla(\varphi - I_{h,0}\varphi)\|_{L^2(T)} &\leq c \|\nabla \varphi\|_{L^2(U(T))}. \end{aligned}$$

Setzen wir dies in die Abschätzung von $|R(u_h)(\varphi)|$ ein, so erhalten wir

$$\begin{aligned} |R(u_h)(\varphi)| &\leq c \sum_{T \in \mathcal{T}_h} \left(h(T) \|f + \Delta u_h\|_{L^2(T)} \|\nabla \varphi\|_{L^2(U(T))} \right. \\ &\quad \left. + \frac{1}{2} \sum_{S \subset \partial T \setminus \partial \Omega} \|[\nabla u_h \cdot n]\|_{L^2(S)} (h(T)^{-1/2} h(T) + h(T)^{1/2}) \|\nabla \varphi\|_{L^2(U(T))} \right) \\ &\leq c \left(\sum_{T \in \mathcal{T}_h} c h(T) \|f + \Delta u_h\|_{L^2(T)} \right. \\ &\quad \left. + \sum_{T \in \mathcal{T}_h} \frac{1}{2} \sum_{S \subset \partial T \setminus \partial \Omega} h(T)^{1/2} \|[\nabla u_h \cdot n]\|_{L^2(S)} \right) \|\varphi\|_{H^1(\Omega)}. \end{aligned}$$

Division durch $\|\varphi\|_{H^1(\Omega)}$ und Supremumsbildung über $\varphi \in H^1(\Omega) \setminus \{0\}$ ergibt dann die Behauptung. \square

In Satz 2.62 haben wir bewiesen, dass der Fehlerschätzer $c\eta_h(u_h)$ zuverlässig ist. Im Gegensatz zu $\|R(u_h)\|_{H_{-1}(\Omega)}$ ist $c\eta_h(u_h)$ auch berechenbar, wenn man berücksichtigt, dass man die Konstante c auch konkret ausrechnen kann. Da wir bisher nur gezeigt haben, dass $c\eta_h(u_h)$ eine obere Schranke für den Fehler ist, steht aber noch aus die Effizienz von $\eta_h(u_h)$ zu zeigen.

Im Folgenden werden wir sogar noch mehr zeigen, und zwar, dass sich sogar die lokalen Fehlerindikatoren $\eta_T(u_h)$ durch den Fehler auf einer Umgebung von T abschätzen lassen. Konkret zeigen wir

$$\eta_T(u_h) \leq \|\nabla(u - u_h)\|_{L^2(U(T))} + \text{"Datenapproximationsfehler".}$$

D.h. der Fehlerindikator $\eta_T(u_h)$ ist klein, wenn der lokale Fehler in einer Umgebung von T klein ist.

Um eine solche lokale Aussage zu erhalten, müssen wir die globalen Normen in geeigneter Weise lokalisieren. Dies erreichen wir durch die Multiplikation mit sogenannten "Bubble"-Funktionen, die wir im Folgenden einführen und untersuchen.

Lemma 2.63 (Element-Bubbles)

Seien T ein d -dimensionaler Simplex im \mathbb{R}^d und $\lambda_i(x), i = 0, \dots, d$ die Baryzentrischen Koordinaten von $x \in T$. Wir definieren die Element-Bubble-Funktion $\psi_T : \mathbb{R}^d \rightarrow \mathbb{R}$ durch

$$\psi_T(x) := \begin{cases} c_0 \prod_{i=0}^d \lambda_i(x), & \text{falls } x \in T, \\ 0, & \text{sonst,} \end{cases}$$

wobei $c_0 := (d+1)^{d+1}$. Dann erfüllt ψ_T folgende Eigenschaften

- (i) $\text{supp } \psi_T \subset T, \quad \psi_T|_T \in \mathbb{P}^{d+1}(T), \quad 0 \leq \psi_T \leq 1, \quad \max_{x \in T} \psi_T(x) = 1,$
- (ii) $\int_T \psi_T(x) dx = c_d |T|$, mit $c_d := \frac{d!(d+1)^{d+1}}{(2d+1)!},$
- (iii) $\|\nabla \psi_T\|_{L^2(T)} \leq \frac{c}{h(T)} \|\psi_T\|_{L^2(T)}.$

Beweis: Aufgrund der Definition von ψ_T gilt $\psi_T|_{\partial T} = 0$, somit folgt $\text{supp } \psi_T \subset T$. Weiterhin sind $\psi_T|_T \in \mathbb{P}^{d+1}(T)$ und $0 \leq \psi_T$ klar aufgrund der Definition. Wegen der Symmetrie wird das Maximum von ψ_T im Schwerpunkt von T angenommen und es ist aufgrund der Definition von c_0 :

$$\psi_T(x_s) = c_0 \left(\frac{1}{d+1} \right)^{d+1} = 1.$$

Die zweite Eigenschaft sieht man wie folgt ein.

$$\int_T \psi_T = (d+1)^{d+1} \int_T \lambda_0(x) \cdots \lambda_d(x) dx = (d+1)^{d+1} \frac{d!}{(2d+1)!} |T|.$$

Die dritte Eigenschaft folgt aus einem Skalierungsargument. □

Lemma 2.64 (Kanten-Bubbles)

Seien T, T' d -dimensionale Simplexe im \mathbb{R}^d mit gemeinsamer Seitenfläche $S = T \cap T'$ und seien $\lambda_i(x), i = 0, \dots, d$, b.z.w. $\lambda'_i(x), i = 0, \dots, d$ die Baryzentrischen Koordinaten von $x \in T$, b.z.w. $x \in T'$. Dabei nehmen wir an, dass die Eckpunkte der Dreiecke so durchnummieriert sind, dass a_0 , b.z.w. a'_0 der Fläche S gegenüber liegen, was $\lambda_0(x) = \lambda'_0(x) = 0 \forall x \in S$ impliziert. Wir definieren die Kanten-Bubble-Funktion $\psi_S : \mathbb{R}^d \rightarrow \mathbb{R}$ durch

$$\psi_S(x) := c_0 \begin{cases} \prod_{i=1}^d \lambda_i(x), & \text{falls } x \in T, \\ \prod_{i=1}^d \lambda'_i(x), & \text{falls } x \in T', \\ 0, & \text{sonst,} \end{cases}$$

wobei $c_0 := d^d$. Dann erfüllt ψ_S folgende Eigenschaften

- (i) $\text{supp } \psi_S \subset T \cup T'$, $\psi_S|_T \in \mathbb{P}^d(T)$, $\psi_S|_{T'} \in \mathbb{P}^d(T')$,
- $0 \leq \psi_S \leq 1$, $\max_{x \in T \cup T'} \psi_S(x) = 1$,
- (ii) $\int_S \psi_S(x) dx = \tilde{c}_d |S|$, mit $\tilde{c}_d := \frac{d!(d)^d}{(2d)!}$,
- (iii) $\|\nabla \psi_S\|_{L^2(T \cup T')} \leq \frac{c}{h(S)} \|\psi_S\|_{L^2(T \cup T')}$.

Beweis: Der Beweis folgt analog zum Beweis von Satz 2.63. □

Wir haben nun alle Vorbereitungen getroffen, um die lokale Effizienz des Fehlerschätzers zeigen zu können. Der Einfachheit halber wollen wir dies aber nur für lineare Finite Elemente Approximationen u_h ausführen.

Lemma 2.65 (Effizienz des Fehlerschätzers)

Seien die Voraussetzungen und Notationen aus Lemma 2.56 gegeben und speziell $u_h \in S_{h,0}^1$ lineare Finite Elemente Approximation. Wir definieren für $T \in \mathcal{T}_h$ und $f \in L^2(\Omega)$, den Mittelwert von f auf T durch

$$f_T := \frac{1}{|T|} \int_T f(x) dx.$$

Dann gilt die lokale Abschätzung

$$\eta_T(u_h) \leq c \|\nabla(u - u_h)\|_{L^2(\omega(T))} + \sum_{T' \in \omega(T)} h(T') \|f - f_{T'}\|_{L^2(T')},$$

dabei ist die Umgebung $\omega(T)$ definiert durch $\omega(T) := \bigcup_{T' \in \mathcal{T}_h | T \cap T' = S} T'$, wobei S eine Seitenfläche ist.

Beweis: Der Beweis gliedert sich in drei Teile. Im ersten Teil werden wir eine Abschätzung für $h(T)\|f + \Delta u_h\|_{L^2(T)}$ herleiten, im zweiten Teil eine Abschätzung für den Sprungterm $h(T)^{1/2}\|[\nabla u_h \cdot n]\|_{L^2(S)}$. Im dritten Teil werden wir dann die beiden Abschätzungen zu einer Abschätzung für $\eta_T(u_h)$ zusammenfassen.

1. Schritt: Abschätzung für $h(T)\|f + \Delta u_h\|_{L^2(T)}$.

Da wir u_h als stückweise linear angenommen haben, gilt

$$\|f + \Delta u_h\|_{L^2(T)} = \|f\|_{L^2(T)} \leq \|f - f_T\|_{L^2(T)} + \|f_T\|_{L^2(T)}.$$

Mit der Bubble-Funktion ψ_T und Lemma 2.63 folgt

$$c_d \|f_T\|_{L^2(T)}^2 = c_d |f_T|^2 |T| = |f_T|^2 \int_T \psi_T = \int_T (f_T)^2 \psi_T.$$

Mit $v_T := f_T \psi_T$ folgt dann

$$c_d \|f_T\|_{L^2(T)}^2 = \int_T f_T v_T = \int_T f v_T + \int_T (f_T - f) v_T.$$

Da $u_h|T \in \mathbb{P}^1$ und da v_T Nullrandwerte auf ∂T hat, gilt weiter

$$0 = \int_T \Delta u_h v_T = \int_T \nabla u_h \cdot \nabla v_T = \int_\Omega \nabla u_h \cdot \nabla v_T.$$

Außerdem folgt aus $\int_\Omega \nabla u \cdot \nabla \varphi = \int_\Omega f \varphi$ mit $\varphi = v_T$

$$\int_T \nabla u \cdot \nabla v_T = \int_\Omega \nabla u \cdot \nabla v_T = \int_\Omega f v_T = \int_T f v_T.$$

Setzen wir diese beiden Beziehungen in die Abschätzung für $c_d \|f_T\|_{L^2(T)}^2$ ein, so erhalten wir

$$\begin{aligned} c_d \|f_T\|_{L^2(T)}^2 &= \int_T \nabla u \cdot \nabla v_T - \int_T \nabla u_h \cdot \nabla v_T + \int_T (f_T - f) v_T \\ &= \int_T \nabla(u - u_h) \cdot \nabla v_T + \int_T (f_T - f) v_T \\ &\leq \|\nabla(u - u_h)\|_{L^2(T)} \|\nabla v_T\|_{L^2(T)} + \|f - f_T\|_{L^2(T)} \|v_T\|_{L^2(T)}. \end{aligned}$$

Mit Lemma 2.63 folgt weiter

$$\begin{aligned} \|\nabla v_T\|_{L^2(T)} &= |f_T| \|\nabla \psi_T\|_{L^2(T)} \leq \frac{c}{h(T)} |f_T| \|\psi_T\|_{L^2(T)} \\ &\leq \frac{c}{h(T)} |f_T| |T|^{1/2} = \frac{c}{h(T)} \|f_T\|_{L^2(T)}. \end{aligned}$$

und

$$\|v_T\|_{L^2(T)} \leq |f_T| \|\psi_T\|_{L^2(T)} \leq |f_T| |T|^{1/2} = \|f_T\|_{L^2(T)}.$$

Damit erhalten wir insgesamt

$$c_d \|f_T\|_{L^2(T)}^2 \leq \left(\frac{c}{h(T)} \|\nabla(u - u_h)\|_{L^2(T)} + \|f - f_T\|_{L^2(T)} \right) \|f_T\|_{L^2(T)}.$$

Also folgt

$$\|f_T\|_{L^2(T)} \leq c \left(\frac{c}{h(T)} \|\nabla(u - u_h)\|_{L^2(T)} + \|f - f_T\|_{L^2(T)} \right),$$

und somit nach Multiplikation mit $h(T)$

$$h(T) \|f_T\|_{L^2(T)} \leq c \left(\|\nabla(u - u_h)\|_{L^2(T)} + h(T) \|f - f_T\|_{L^2(T)} \right).$$

2. Schritt: Abschätzung für $h(T)^{1/2} \|[\nabla u_h \cdot n]\|_{L^2(S)}$.

Wir gehen hier analog zum ersten Schritt vor und setzen

$$v_S := [\nabla u_h \cdot n] \psi_S.$$

Dann gilt mit Lemma 2.64 und da $[\nabla u_h \cdot n]$ konstant auf S ist

$$\begin{aligned} \|[\nabla u_h \cdot n]\|_{L^2(S)}^2 &= [\nabla u_h \cdot n]^2 |S| = \frac{1}{\tilde{c}_d} [\nabla u_h \cdot n]^2 \int_S \psi_S \\ &= \frac{1}{\tilde{c}_d} \int_S [\nabla u_h \cdot n] v_S. \end{aligned}$$

Mit $\omega(S) := T \cup T'$ gilt andereseits

$$\int_{\omega(S)} f v_S = \int_{\Omega} f v_S = \int_{\Omega} \nabla u \cdot \nabla v_S = \int_{\omega(S)} \nabla u \cdot \nabla v_S$$

und

$$\begin{aligned}
\int_{\Omega} \nabla u_h \cdot \nabla v_S &= \int_{\omega(S)} \nabla u_h \cdot \nabla v_S \\
&= \int_T \nabla u_h \cdot \nabla v_S + \int_{T'} \nabla u_h \cdot \nabla v_S \\
&= \int_{\partial T} \nabla u_h \cdot n v_S + \int_{\partial T'} \nabla u_h \cdot n v_S \\
&= \int_S [\nabla u_h \cdot n] v_S.
\end{aligned}$$

Zusammen erhalten wir also

$$\begin{aligned}
\tilde{c}_d \|[\nabla u_h \cdot n]\|_{L^2(S)}^2 &= \int_{\omega(S)} \nabla u_h \cdot \nabla v_S \\
&= \int_{\omega(S)} \nabla(u_h - u) \cdot \nabla v_S + \int_{\omega(S)} f v_S \\
&\leq \|\nabla(u - u_h)\|_{L^2(\omega(S))} \|\nabla v_S\|_{L^2(\omega(S))} + \|f\|_{L^2(\omega(S))} \|v_S\|_{L^2(\omega(S))}.
\end{aligned}$$

Mit Lemma 2.64 gilt weiter

$$\begin{aligned}
\|v_S\|_{L^2(\omega(S))} &= \|[\nabla u_h \cdot n]\| \|v_S\|_{L^2(\omega(S))} \\
&\leq \|[\nabla u_h \cdot n]\| |\omega(S)|^{1/2} \\
&= \|[\nabla u_h \cdot n]\|_{L^2(S)} |S|^{-1/2} |\omega(S)|^{1/2} \\
&\leq ch(T)^{1/2} \|[\nabla u_h \cdot n]\|_{L^2(S)}
\end{aligned}$$

und

$$\begin{aligned}
\|\nabla v_S\|_{L^2(\omega(S))} &\leq \frac{c}{h(T)} \|v_S\|_{L^2(\omega(S))} \\
&\leq ch(T)^{-1/2} \|[\nabla u_h \cdot n]\|_{L^2(S)}.
\end{aligned}$$

Dabei haben wir verwendet, dass aufgrund der Regularität von \mathcal{T}_h ($\sigma(T) \leq \sigma$) gilt

$$\begin{aligned}
|\omega(S)| &= |T \cup T'| \leq c|T|, \\
|T| &\leq c|S| h(T).
\end{aligned}$$

Damit folgt insgesamt für die Kantensprünge

$$\begin{aligned}
h(T)^{1/2} \|[\nabla u_h \cdot n]\|_{L^2(S)} &\leq c \|\nabla(u - u_h)\|_{L^2(\omega(S))} \\
&\quad + ch(T) (\|f\|_{L^2(T)} + \|f\|_{L^2(T')}).
\end{aligned}$$

3. Schritt: Abschätzung für $\eta_T(u_h)$.

Mit den Abschätzungen aus den Schritten 1 und 2 erhalten wir nun

$$\begin{aligned}
 \eta_T(u_h) &\leq ch(T) \|f\|_{L^2(\Omega)}^2 + c \frac{1}{2} h(T)^{1/2} \sum_{S \subset \partial T \setminus \partial \Omega} \left\| \left[\nabla u \cdot n \right] \right\|_{L^2(S)} \\
 &\leq c \left(\|\nabla(u - u_h)\|_{L^2(T)} + h(T) \|f - f_T\|_{L^2(T)} \right) \\
 &\quad + c \sum_{T' \cap T = S} \left(\|\nabla(u - u_h)\|_{L^2(\omega(S))} + h(T) (\|f\|_{L^2(T)} + \|f\|_{L^2(T')}) \right) \\
 &\leq c \left(\|\nabla(u - u_h)\|_{L^2(T)} + h(T) \|f - f_T\|_{L^2(T)} \right) \\
 &\quad + c \sum_{T' \cap T = S} \|\nabla(u - u_h)\|_{L^2(T')} + c \sum_{T' \cap T = S} h(T') \|f - f_{T'}\|_{L^2(T')}.
 \end{aligned}$$

Mit der Definition von $\omega(T)$ folgt hieraus die Behauptung. \square

2.5.4 Adaptive Verfahren basierend auf Fehlerschätzern

Wir setzen nun voraus, dass ein lokalisierter Fehlerschätzer der Form

$$\|u - u_h\|^2 \leq \eta_h(u_h)^2 := \sum_{T \in \mathcal{T}_h} \eta_T(u_h)^2$$

gegeben ist. Darauf aufbauend wollen wir eine Adaptionstrategie angeben, die zu einem adaptiven Verfahren führt, so dass der Fehler $\|u - u_h\|$ durch eine vorgegebene Toleranz TOL beschränkt ist.

Definition 2.66 (Gleichverteilungsstrategie)

Sei $\Omega \subset \mathbb{R}^d$ ein Gebiet und \mathcal{T}_h eine zulässige Makrotriangulierung von Ω . Weiter sei eine Toleranz $TOL > 0$ und ein Parameter $\Theta \in (0, 1)$ gegeben. Dann definieren wir folgenden Algorithmus.

EQUALDISTRIBUTION($\mathcal{T}_h, TOL, \Theta$)

```

1  repeat
2       $\mathcal{T}_h^0 := \mathcal{T}_h$ 
3       $[u_h, \eta_h] := \text{FEM}(\mathcal{T}_h)$ 
4      if  $\eta_h > TOL$ 
5          then
6              for  $T \in \mathcal{T}_h$ 
7                  do (berechne Markierungen)
8                      if  $\eta_T^2 > \frac{TOL^2}{\#(\mathcal{T}_h)}$ 
9                          then  $M_h := \text{MARKREFINE}(T)$ 
10                     else if  $\eta_T^2 < \Theta \frac{TOL^2}{\#(\mathcal{T}_h)}$ 
11                         then  $M_h := \text{MARKCOARSEN}(T)$ 
12                      $\mathcal{T}_h := \text{ADAPT}(M_h, \mathcal{T}_h)$ 
13                 until  $\eta_h \leq TOL$ 
14     return  $(\mathcal{T}_h, u_h, \eta_h)$ 

```

Dabei bezeichnet $\#(\mathcal{T}_h)$ die Anzahl der Simplizes in \mathcal{T}_h und M_h ist eine Indikatorfunktion (z.B. $M_h|_T = 1$, falls T verfeinert werden soll, $M_h|_T = -1$, falls T vergröbert werden soll und Null sonst).

Der Gleichverteilungsalgorithmus liefert (falls er konvergiert) ein adaptives Gitter \mathcal{T}_h mit einer Finite Elemente Lösung u_h und Fehlerschätzer η_h , so dass gilt $\eta_h \leq TOL$. Dabei werden die Simplizes derart verfeinert oder vergröbert, dass die Fehler $\eta_T, T \in \mathcal{T}_h$ alle von der gleichen Größenordnung sind. Insbesondere folgt aus

$$\eta_T^2 \leq \frac{TOL^2}{\#(\mathcal{T}_h)}.$$

durch Summation über alle Simplizes

$$\eta_h^2 := \sum_{T \in \mathcal{T}_h} \eta_T^2 \leq \sum_{T \in \mathcal{T}_h} \frac{TOL^2}{\#(\mathcal{T}_h)} = \#(\mathcal{T}_h) \frac{TOL^2}{\#(\mathcal{T}_h)} = TOL^2.$$

2.6 Quadraturformeln

Wie wir in Abschnitt 2.1 gesehen haben, führen Finite Elemente Verfahren auf große lineare Gleichungssysteme. Beim Aufstellen dieser Gleichungssysteme, müssen Integrale über das Gebiet Ω berechnet werden, z.B. für die rechte Seite

$$\int_{\Omega} f(x) \varphi_j(x) dx,$$

oder für die Steifigkeitsmatrix

$$\int_{\Omega} A(x) \nabla \varphi_i(x) \cdot \nabla \varphi_j(x) dx.$$

Dabei sind φ_i, φ_j Basen des Finite Elemente Raums, während f und A beliebige Funktionen sein können. Deshalb ist eine exakte Integration hier oft nicht möglich. Statt dessen werden die Integrale durch zusammengesetzte Quadraturformeln ersetzt, d.h. zunächst zerlegen wir das Integral über Ω

$$\int_{\Omega} v(x) dx = \sum_{T \in \mathcal{T}_h} \int_T v(x) dx.$$

Anschließend verwenden wir die Transformationsformel, um die Integrale über T durch Integrale über dem Referenelement T_0 auszudrücken

$$\int_T v(x) dx = \int_{T_0} v(F(y)) |\det \nabla F(y)| dy.$$

Schließlich wenden wir zur Integration auf dem Referenzelement eine Quadraturformel an, d.h.

$$\int_{T_0} \tilde{v}(y) dy \approx \sum_{i=1}^Q \omega_i \tilde{v}(q_i).$$

Verwenden wir solche Approximationen für die Integration, so führt dies dazu, dass wir die exakte Bilinearform B und rechte Seite f durch Approximationen B_h , b.z.w. f_h ersetzen. Das Finite Elemente Verfahren mit Quadratur kann dann geschrieben werden als: Finde $u_h \in X_h$ mit

$$B_h(u_h, \varphi_h) = f_h(\varphi_h) \quad \forall \varphi_h \in X_h.$$

In diesem Abschnitt wollen wir uns mit Quadraturformeln für Simplizes beschäftigen und insbesondere die Fragen stellen, welche Güte die Quadraturformeln haben müssen, damit die asymptotische Konvergenzordnung der Finite Elemente Verfahren nicht zerstört wird. Abstrakt lässt sich diese Frage relativ leicht beantworten, wie der folgende Satz zeigt.

Satz 2.67 (Abstrakte Fehlerabschätzung mit Quadratur)

Sei X ein Hilbertraum und $B : X \times X \rightarrow \mathbb{R}$ eine Bilinearform, die koerziv (c_0) und stetig (c_1) sei. Weiter sei $X_h \subset X$ ein endlichdimensionaler Teilraum. Auf X_h sei eine Bilinearform $B_h : X_h \times X_h \rightarrow \mathbb{R}$ definiert, die ebenfalls koerziv mit Konstante \bar{c}_0 sei. Außerdem gebe es ein r , so dass für alle $v_h \in X_h$ gilt

$$|B_h(u_h, v_h) - B(u_h, v_h)| \leq c_2 h^r \|v_h\|_X,$$

Wobei $u_h \in X_h$ die Lösung des Ritz-Galerkin Verfahrens sei, d.h.

$$B(u_h, \varphi_h) = f(\varphi_h), \quad \forall \varphi_h \in X_h.$$

Dabei sei $f \in X'$. Weiter sei $f_h \in X'_h$ und erfülle

$$|f(v_h) - f_h(v_h)| \leq c_3 h^r \|v_h\|_X, \quad \forall v_h \in X_h.$$

Dann gibt es genau ein $\tilde{u}_h \in X_h$ mit

$$B_h(u_h, \varphi_h) = f_h(\varphi_h), \quad \forall \varphi_h \in X_h$$

und es gilt die Fehlerabschätzung

$$\|u_h - \tilde{u}_h\|_X \leq \frac{c_2 + c_3}{\bar{c}_0} h^r.$$

Beweis: Mit den Notationen aus dem Satz gilt

$$\begin{aligned} \bar{c}_0 \|u_h - \tilde{u}_h\|_X^2 &\leq B_h(u_h - \tilde{u}_h, u_h - \tilde{u}_h) \\ &= B_h(u_h, u_h - \tilde{u}_h) - B_h(\tilde{u}_h, u_h - \tilde{u}_h) \\ &= B_h(u_h, u_h - \tilde{u}_h) - B(u_h, u_h - \tilde{u}_h) \\ &\quad + B(u_h, u_h - \tilde{u}_h) - B_h(\tilde{u}_h, u_h - \tilde{u}_h) \\ &= B_h(u_h, u_h - \tilde{u}_h) - B(u_h, u_h - \tilde{u}_h) \\ &\quad + f(u_h - \tilde{u}_h) - f_h(u_h - \tilde{u}_h) \\ &\leq |B_h(u_h, u_h - \tilde{u}_h) - B(u_h, u_h - \tilde{u}_h)| \\ &\quad + |f(u_h - \tilde{u}_h) - f_h(u_h - \tilde{u}_h)| \\ &\leq c_2 h^r \|u_h - \tilde{u}_h\|_X + c_3 h^r \|u_h - \tilde{u}_h\|_X. \end{aligned}$$

Also folgt die Behauptung durch Division mit $\bar{c}_0 \|u_h - \tilde{u}_h\|_X$. \square

Durch den abstrakten Satz ist klar, worauf wir bei der Approximation der Integrale zu achten haben. Zum einen muss die resultierende Bilinearform B_h gleichmässig koerziv sein, zum anderen muß die Approximationsordnung, so gewählt werden, dass $r \geq s$ ist, falls s die Konvergenzordnung des Finite Elemente Verfahrens ist.

Als nächstes wollen wir definieren, was wir unter einer Quadraturformel auf dem Referenzelement verstehen.

Definition 2.68 (Quadraturformel)

Sei $T_0 \subset \mathbb{R}^d$ ein Gebiet. Eine Quadraturformel Q auf T_0 für Funktionen $f \in L^1(\Omega)$ mit Gewichten $\omega_l \in \mathbb{R}, l = 1, \dots, L$ und Quadraturpunkten $q_l \in T_0, l = 1, \dots, L$ ist definiert durch

$$Q(f) := \sum_{l=1}^L \omega_l \tilde{f}(q_l).$$

Dabei ist $\tilde{f}(q_l)$ definiert durch

$$\tilde{f}(q_l) := \lim_{\varepsilon \rightarrow 0} \frac{1}{|B_\varepsilon(q_l)|} \int_{B_\varepsilon(q_l)} f(y) dy.$$

Ist $P(T_0) \subset L^1(\Omega)$ ein Funktionenraum, so heißt Q exakt auf $P(T_0)$, wenn für alle $p \in P(T_0)$ gilt

$$Q(p) = \int_{T_0} p(y) dy.$$

Im folgenden werden wir der Einfachheit halber immer $f(x)$ schreiben und verstehen darunter $\tilde{f}(x)$ im Sinne von Definition 2.68.

Wir wollen nun Quadraturformeln auf Simplizes definieren, die exakt auf Polynomräumen sind. Sei dazu $P(T)$ ein Polynomraum mit $\dim(P(T)) = s$. Ein allgemeiner Ansatz sind die Interpolationsquadraturen. Dazu sei eine Menge von Stützstellen $x_i \in T, i = 1, \dots, s$ gegeben und $L_i^s \in \mathbb{P}(T)$ seien die zugehörigen Lagrange-Polynome, definiert durch $L_i^s(x_j) = \delta_{ij}, i, j = 1, \dots, s$.

$1, \dots, s$. Dann ist die zugehörige Interpolationsquadratur definiert durch

$$Q(f) := \sum_{i=1}^s \omega_i f(x_i),$$

mit den Gewichten

$$\omega_i := \int_T L_i^s(x) dx.$$

Aus numerischer Sicht sollten die Quadraturpunkte x_i dabei so gewählt werden, dass die Quadraturformel exakt auf Polynomen von möglichst großem Polynomgrad sind, wobei jedoch gleichzeitig aus Stabilitätsgründen darauf geachtet werden muss, dass die Gewichte nicht negativ werden.

Für Dreiecke in \mathbb{R}^2 wollen wir konkrete Quadraturformeln angeben. Dabei ist folgendes Lemma hilfreich.

Lemma 2.69

Seien $T \subset \mathbb{R}^d$ ein d -dimensionaler Simplex und $\alpha \in \mathbb{N}_0^{d+1}$ ein Multiindex, so gilt für baryzentrische Koordinaten $\lambda = (\lambda_0, \dots, \lambda_d)$ die Integrationsformel

$$\int_T \lambda(x)^\alpha dx = \frac{\alpha! d!}{(|\alpha| + d)!} |T|.$$

(Beweis siehe Übungsaufgaben)

Satz 2.70 (Quadraturen für Simplizes)

Sei $T \subset \mathbb{R}^d$ ein d -dimensionaler Simplex mit Eckpunkten $a_i, i = 0, \dots, d$ und Schwerpunkt x_s . Dann gilt:

(i) Die Quadraturformel

$$Q_1(f) := |T| f(x_s), \quad x_s \text{ Schwerpunkt von } T,$$

ist exakt auf $\mathbb{P}^1(T)$.

(ii) Die Quadraturformel

$$Q_2(f) := \frac{1}{3} |T| \sum_{0 \leq i < j \leq d} f(a_{ij}), \quad a_{ij} := \frac{1}{2}(a_i + a_j),$$

ist für $d = 2$ exakt auf $\mathbb{P}^2(T)$.

(iii) Die Quadraturformel

$$Q_3(f) := \frac{1}{60}|T| \left(3 \sum_{i=0}^d f(a_i) + 8 \sum_{0 \leq i < j \leq d} f(a_{ij}) + 27f(x_s) \right),$$

ist für $d = 2$ exakt auf $\mathbb{P}^3(T)$.

Beweis: Wir beweisen die Aussagen i) und ii).

Zu i) Jedes $p \in \mathbb{P}^1(T)$ hat die Darstellung

$$p(x) = \sum_{i=0}^d p(a_i) \lambda_i(x), \quad x \in T.$$

Also folgt mit Lemma 2.69 und der Linearität von p

$$\begin{aligned} \int_T p(x) dx - |T|p(x_s) &= \sum_{i=0}^d p(a_i) \int_T \lambda_i(x) dx - |T|p(x_s) \\ &= \sum_{i=0}^d p(a_i) \frac{d!}{(1+d)!} |T| - |T|p(x_s) \\ &= |T| \left(\frac{1}{d+1} \sum_{i=0}^d p(a_i) - p(x_s) \right) = 0. \end{aligned}$$

Zu ii) Zunächst kann man zeigen, dass $p \in \mathbb{P}^2$ mit Hilfe der baryzentrischen Koordinaten in der Lagrangebasis wie folgt dargestellt werden kann

$$p(x) = \sum_{j=0}^d p(a_j) \lambda_j(x) (2\lambda_j(x) - 1) + 4 \sum_{j=0}^d \sum_{i=0}^{j-1} p(a_{ij}) \lambda_i(x) \lambda_j(x).$$

Also folgt mit Lemma 2.69

$$\begin{aligned}
\int_T p(x) dx &= \sum_{j=0}^d p(a_j) \left(\int_T 2\lambda_j(x)^2 dx - \int_T \lambda_j(x) dx \right) \\
&\quad + 4 \sum_{j=0}^d \sum_{i=0}^{j-1} p(a_{ij}) \int_T \lambda_i(x) \lambda_j(x) dx \\
&= \sum_{j=0}^d p(a_j) \left(2|T| \frac{2!d!}{(d+2)!} - |T| \frac{1!d!}{(d+1)!} \right) + 4 \sum_{j=0}^d \sum_{i=0}^{j-1} p(a_{ij}) |T| \frac{1!d!}{(2+d)!} \\
&= |T| \left(\frac{2-d}{(d+1)(d+2)} \sum_{j=0}^d p(a_j) + \frac{4}{(d+1)(d+2)} \sum_{j=0}^d \sum_{i=0}^{j-1} p(a_{ij}) \right).
\end{aligned}$$

Setzt man $d = 2$, so verschwindet die Summe über die Eckpunkte und man erhält $Q_2(p)$. \square

Nachdem wir Quadraturformeln eingeführt haben, wollen wir untersuchen, wie gut die resultierenden approximierenden Bilinearformen und rechten Seiten approximiert werden (vgl. Satz 2.67). Zuvor wollen wir allerdings noch die Koerzivität der approximierenden Bilinearform untersuchen. Dazu haben wir folgendes Lemma.

Lemma 2.71 (Koerzivität von B_h)

Sei $A \in C(\mathbb{R}^d; \mathbb{R}^{d \times d})$ uniform elliptisch, d.h. es gibt eine Konstante $c_0 > 0$, so dass

$$\xi^\top A(x) \xi \geq c_0 |\xi|^2, \quad \forall \xi, x \in \mathbb{R}^d.$$

Weiter seien $\Omega \subset \mathbb{R}^d$ und \mathcal{T}_h eine zulässige Triangulierung von Ω mit $\sigma(T) \leq \sigma$ für alle $T \in \mathcal{T}_h$. Die Bilinearform $B : H^1(\Omega) \times H^1(\Omega) \rightarrow \mathbb{R}$ sei definiert durch

$$B(u, v) := \int_{\Omega} A(x) \nabla u(x) \cdot \nabla v(x) dx.$$

Ist Q eine Quadraturformel auf dem Referenzelement T_0 , so definieren wir $B_h : S_h^k \times S_h^k \rightarrow \mathbb{R}$ durch

$$B_h(u_h, v_h) := \sum_{T \in \mathcal{T}_h} |\det \nabla F_T| Q((A \nabla u_h \cdot \nabla v_h) \circ F_T).$$

Dabei ist $F_T : T^0 \rightarrow T$ die Referenzabbildung für eine Element $T \in \mathcal{T}_h$. Hat die Quadraturformel Q nur positive Gewichte $\omega_l > 0$ und ist sie exakt auf \mathbb{P}^{2k-2} , so existiert eine Konstante $c > 0$, so dass für alle $v_h \in S_h^k$ gilt

$$B_h(v_h, v_h) \geq \frac{c_0}{c} |v_h|_{H^1(\Omega)}.$$

(Beweis siehe Übungsaufgabe)

Wenden wir uns nun der Approximation von Integralen durch Quadraturformeln zu.

Lemma 2.72 (Approximationseigenschaften von Quadraturen)

Sei $k \in \mathbb{N}$ so gewählt, dass die Einbettung $H^{k+1,p}(T) \hookrightarrow L^\infty(T)$ gilt. Sei Q_0 eine Quadraturformel auf dem Referenzelement T_0 , die exakt auf $\mathbb{P}^m(T_0)$ mit $m \geq k$ ist. Dann ist durch

$$Q(v) := |\det \nabla F_T| Q_0(v \circ F)$$

eine Quadratur auf T gegeben, die exakt auf $\mathbb{P}^m(T)$ ist.

Definiert man das Fehlerfunktional $G : H^{k+1,p}(T) \rightarrow \mathbb{R}$ durch

$$G(v) := \int_T v(x) dx - Q(v),$$

so gilt für alle $w \in H^{k+1,p}(T)$

$$|G(v)| \leq ch(T)^{d(1-1/p)+k+1} |v|_{H^{k+1,p}(T)}.$$

Beweis: Definiert man G_0 durch $G_0(w) := \int_{T_0} w - Q_0(w)$, so gilt für $v \in H^{k+1,p}(T)$:

$$|G(v)| = |\det \nabla F_T| |G_0(v \circ F)|.$$

Zur Abschätzung von G_0 gehen wir analog zum Bramble Hilbert Lemma 2.29 vor. Sei dazu $p \in \mathbb{P}^m(T_0)$, so gilt mit Satz 2.28

$$\begin{aligned} |G_0(w)| &= |G_0(w) + G_0(p)| \leq |G_0|_{(H^{k+1,p}(T_0))'} |w + p|_{H^{k+1,p}(T_0)} \\ &\leq c |G_0|_{(H^{k+1,p}(T_0))'} |w|_{H^{k+1,p}(T_0)}. \end{aligned}$$

Setzen wir diese Abschätzung für $w = v \circ F$ ein, so folgt durch Rücktransformation

$$\begin{aligned}|G(v)| &\leq c|\det \nabla F_T| \|v \circ F\|_{H^{k+1,p}(T_0)} \\ &\leq c|\det \nabla F_T|^{1-1/p} |\nabla F|^{k+1} \|v\|_{H^{k+1,p}(T)}.\end{aligned}$$

Hieraus folgt die Behauptung mit Lemma 2.11. \square

Lemma 2.72 liefert uns zwar Fehlerabschätzungen, jedoch können wir diese mit Blick auf das abstrakte Resultat in Satz 2.67 nicht verwenden, da wir dort Abschätzungen der Form $ch^r \|v\|_{H^1(\Omega)}$ benötigen. Um dies zu erreichen, müssen wir ausnutzen, dass die Bilinearform und die rechten Seite Integrale über Produkte von Funktionen sind, bei denen bei der Approximation mindestens ein Faktor ein Polynom ist, d.h. es gilt z.B. $v = fv_h$. Da wir bei diesen Abschätzungen wieder Gebrauch vom Bramble-Hilbert Lemma machen werden, wollen wir dieses hier noch einmal in einer anderen Variante formulieren.

Satz 2.73 (Bramble-Hilbert Lemma II)

Sei T ein Lipschitzgebiet, $k \in \mathbb{N}$ und $p \in [1, \infty]$. Sei $G \in (H^{k,p}(T))'$ und gelte $G(p) = 0$ für alle $p \in \mathbb{P}^k(T)$. Dann existiert eine Konstante $c > 0$, so dass für alle $v \in H^{k,p}(T)$ gilt

$$|G(v)| \leq c\|G\| \|v\|_{H^{k,p}(T)}.$$

(ohne Beweis)

Satz 2.74 (Approximation der Bilinearform)

Seien $\Omega \subset \mathbb{R}^d$ ein Gebiet, \mathcal{T}_h eine zulässige Triangulierung von Ω mit $\sigma(T) \leq \sigma$, $k \in \mathbb{N}$ und $A \in L^\infty(\Omega; \mathbb{R}^{d \times d})$, $A|_T \in H^{k,\infty}(T; \mathbb{R}^{d \times d})$ uniform elliptisch, d.h. es gibt eine Konstante $c_0 > 0$, so dass

$$\xi^\top A(x) \xi \geq c_0 |\xi|^2, \quad \forall \xi, x \in \mathbb{R}^d.$$

Die Bilinearform $B : H^1(\Omega) \times H^1(\Omega) \rightarrow \mathbb{R}$ sei definiert durch

$$B(u, v) := \int_{\Omega} A(x) \nabla u(x) \cdot \nabla v(x) dx.$$

Ist Q eine Quadraturformel auf dem Referenzelement T_0 , so definieren wir $B_h : S_h^k \times S_h^k \rightarrow \mathbb{R}$ durch

$$B_h(u_h, v_h) := \sum_{T \in \mathcal{T}_h} |\det \nabla F_T| Q((A \nabla u_h \cdot \nabla v_h) \circ F_T).$$

Dabei ist $F_T : T^0 \rightarrow T$ die Referenzabbildung für eine Element $T \in \mathcal{T}_h$.

Wir definieren für $f \in H^{-1}(\Omega)$ die Lösungen $u \in H_0^1(\Omega)$ und $u_h \in S_{h,0}^k$ durch

$$\begin{aligned} B(u, \varphi) &= f(\varphi), \quad \forall \varphi \in H_0^1(\Omega), \\ B(u_h, \varphi_h) &= f(\varphi_h), \quad \forall \varphi_h \in S_{h,0}^k. \end{aligned}$$

Ist dann Q exakt auf $\mathbb{P}^{2k-1}(T_0)$, gilt $u \in H^{k+1}(\Omega)$ und ist $k > \frac{d}{2} - 1$, so gilt für alle $v_h \in S_{h,0}^k$

$$|B(u_h, v_h) - B_h(u_h, v_h)| \leq ch^k \|v_h\|_{H^1(\Omega)} \|u\|_{H^{k+1}(\Omega)}.$$

Beweis: Seien $p, q \in \mathbb{P}^{k-1}(T_0)$ und seien G, G_0 wie in Lemma 2.72 definiert. Dann gilt für Funktionen $a, r \in H^{k,\infty}(T_0)$

$$\begin{aligned} |G_0(rq)| &\leq c|rq|_{L^\infty(T_0)} \leq c|r|_{L^\infty(T_0)} \|q\|_{L^2(T_0)} \\ &\leq c\|r\|_{H^{k,\infty}(T_0)} \|q\|_{L^2(T_0)}. \end{aligned}$$

Ist $q \in \mathbb{P}^{k-1}(T_0)$ fest gewählt, so erfüllt $G_0(\cdot q)$ also die Voraussetzungen des Satzes 2.73 und es folgt

$$|G_0(rq)| \leq c|r|_{H^{k,\infty}(T_0)} \|q\|_{L^2(T_0)}.$$

Setzt man $r = ap$, so folgt weiter

$$\begin{aligned} |G_0(apq)| &\leq c|ap|_{H^{k,\infty}(T_0)} \|q\|_{L^2(T_0)} \\ &\leq c \sum_{s=1}^k |a|_{H^{k-s,\infty}(T_0)} |p|_{H^{s,\infty}(T_0)} \|q\|_{L^2(T_0)} \\ &= c \sum_{s=1}^{k-1} |a|_{H^{k-s,\infty}(T_0)} |p|_{H^{s,\infty}(T_0)} \|q\|_{L^2(T_0)} \\ &\leq c \sum_{s=1}^{k-1} |a|_{H^{k-s,\infty}(T_0)} |p|_{H^{s,2}(T_0)} \|q\|_{L^2(T_0)} \end{aligned}$$

Dabei haben wir die Produktregel ausgenutzt und verwendet, dass die k -ten Ableitungswen von p Null sind. Mit einem Skalierungsargument folgt dann für $p, q \in \mathbb{P}^{k-1}(T)$ und $a \in H^{k,\infty}(T)$

$$\begin{aligned} |G(apq)| &\leq ch(T)^k \sum_{s=1}^{k-1} |a|_{H^{k-s,\infty}(T)} |p|_{H^{s,2}(T)} \|q\|_{L^2(T)} \\ &\leq ch(T)^k \|a\|_{H^{k,\infty}(T)} \|p\|_{H^{k-1,2}(T)} \|q\|_{L^2(T)}. \end{aligned}$$

Damit ergibt sich für die Approximation der Bilinearform

$$\begin{aligned} |B(u_h, v_h) - B_h(u_h, v_h)| &\leq \sum_{i,j=1}^d \sum_{T \in \mathcal{T}_h} |G(a_{ij} \partial_i u_h \partial_j v_h)| \\ &\leq c \sum_{i,j=1}^d \sum_{T \in \mathcal{T}_h} h(T)^k \|a_{ij}\|_{H^{k,\infty}(T)} \|\partial_i u_h\|_{H^{k-1,2}(T)} \|\partial_j v_h\|_{L^2(T)} \\ &\leq c \max_{i,j=1}^d \max_{T \in \mathcal{T}_h} \|a_{ij}\|_{H^{k,\infty}(T)} \sum_{T \in \mathcal{T}_h} h(T)^k \|\nabla u_h\|_{H^{k-1,2}(T)} \|\nabla v_h\|_{L^2(T)} \\ &\leq ch^k \left(\sum_{T \in \mathcal{T}_h} \|u_h\|_{H^k(T)}^2 \right)^{1/2} |v_h|_{H^1(\Omega)}. \end{aligned}$$

Um den Beweis des Satzes abzuschließen, müssen wir noch zeigen, dass gilt

$$\left(\sum_{T \in \mathcal{T}_h} \|u_h\|_{H^k(T)}^2 \right)^{1/2} \leq c \|u\|_{H^{k+1}(\Omega)}.$$

Dies sehen wir wie folgt ein. Es ist

$$\begin{aligned} \left(\sum_{T \in \mathcal{T}_h} \|u_h\|_{H^k(T)}^2 \right)^{1/2} &\leq \left(\sum_{T \in \mathcal{T}_h} \|u_h - I_h u\|_{H^k(T)}^2 \right)^{1/2} \\ &\quad + \left(\sum_{T \in \mathcal{T}_h} \|I_h u - u\|_{H^k(T)}^2 \right)^{1/2} + \left(\sum_{T \in \mathcal{T}_h} \|u\|_{H^k(T)}^2 \right)^{1/2}. \end{aligned}$$

Mit inverser Ungleichung erhalten wir für den ersten Term der rechten Seite

$$\left(\sum_{T \in \mathcal{T}_h} \|u_h - I_h u\|_{H^k(T)}^2 \right)^{1/2} \leq ch^{1-k} |u_h - I_h u|_{H^1(\Omega)}.$$

Somit folgt mit der Interpolationsabschätzung und dem Konvergenzresultat

$$\begin{aligned}
 \left(\sum_{T \in \mathcal{T}_h} \|u_h\|_{H^k(T)}^2 \right)^{1/2} &\leq ch^{1-k} \|u_h - I_h u\|_{H^1(\Omega)} \\
 &\quad + \|I_h u - u\|_{H^k(\Omega)} + \|u\|_{H^k(\Omega)} \\
 &\leq ch^{1-k} (\|u_h - u\|_{H^1(\Omega)} + \|u - I_h u\|_{H^1(\Omega)}) \\
 &\quad + \|I_h u - u\|_{H^k(\Omega)} + \|u\|_{H^k(\Omega)} \\
 &\leq ch^{1-k} (ch^k \|u\|_{H^{k+1}(\Omega)} + ch^k \|u\|_{H^{k+1}(\Omega)}) \\
 &\quad + ch \|u\|_{H^{k+1}(\Omega)} + \|u\|_{H^k(\Omega)} \\
 &\leq c(h \|u\|_{H^{k+1}(\Omega)} + \|u\|_{H^k(\Omega)}) \\
 &\leq c \|u\|_{H^{k+1}(\Omega)}.
 \end{aligned}$$

Damit ist der Satz bewiesen. □

Satz 2.75 (Approximation der rechten Seite)

Seien $\Omega \subset \mathbb{R}^d$ ein Gebiet, \mathcal{T}_h eine zulässige Triangulierung von Ω mit $\sigma(T) \leq \sigma$, $k \in \mathbb{N}$ mit $k > \frac{d}{2}$. Weiter sei $f \in L^2(\Omega)$ gegeben mit $f|_T \in H^k(T)$. Ist Q eine Quadraturformel auf dem Referenzelement T_0 , so definieren wir $f_h \in (S_{h,0}^k)'$ durch

$$f_h(v_h) := \sum_{T \in \mathcal{T}_h} |\det \nabla F_T| Q(f \circ F_T).$$

Dabei ist $F_T : T^0 \rightarrow T$ die Referenzabbildung für eine Element $T \in \mathcal{T}_h$. Ist dann Q exakt auf $\mathbb{P}^{2k-2}(T_0)$, so gilt für alle $v_h \in S_{h,0}^k$

$$|\int_{\Omega} f v_h - f_h(v_h)| \leq ch^k \left(\sum_{T \in \mathcal{T}_h} \|f\|_{H^k(T)}^2 \right)^{1/2} \|v_h\|_{H^1(\Omega)}.$$

Beweis: Der Beweis geht analog zum Beweis von Satz 2.74, weshalb wir hier auf eine detaillierte Darstellung verzichten.

Mit den bewiesenen Approximationsabschätzungen können wir nun das Hauptresultat dieses Abschnitts beweisen.

Satz 2.76 (Fehlerabschätzung mit Quadratur)

Mit den Voraussetzungen und Notationen der Sätze 2.71, 2.74 und 2.75 gibt es genau eine diskrete Lösung $\tilde{u}_h \in S_{h,0}^k$ von

$$B_h(\tilde{u}_h, v_h) = f_h(v_h), \quad \forall v_h \in S_{h,0}^k$$

und es gilt die Fehlerabschätzung

$$\|u - \tilde{u}_h\|_{H^1(\Omega)} \leq ch^k \left(\|u\|_{H^{k+1}(\Omega)} + \left(\sum_{T \in \mathcal{T}_h} \|f\|_{H^k(T)}^2 \right)^{1/2} \right).$$

Beweis: Mit der Dreiecksungleichung gilt mit den Definitionen aus Satz 2.67

$$\|u - \tilde{u}_h\|_{H^1(\Omega)} \leq \|u - u_h\|_{H^1(\Omega)} + \|u_h - \tilde{u}_h\|_{H^1(\Omega)}.$$

Das Resultat folgt direkt aus den Sätzen 2.35 und 2.67 unter Zuhilfenahme der Resultate aus den Sätzen 2.71, 2.74 und 2.75. \square

Kapitel 3

Finite Elemente Verfahren für parabolische Differentialgleichungen

In diesem Kapitel wollen wir uns mit der Finite Elemente Diskretisierung der zeitabhängigen Wärmeleitungsgleichung

$$\partial_t u(x, t) - \operatorname{div}(D \nabla u(x, t)) = f, \quad \forall (x, t) \in \Omega \times (0, T)$$

beschäftigen. Wie wir in der Einleitung (Kap. 0) gesehen haben, ist die Wärmeleitungsgleichung der Prototyp einer parabolischen Differentialgleichung zweiter Ordnung und modelliert beispielsweise die zeitabhängige Temperaturverteilung. In diesem Fall ist u die Temperatur, D der Wärmeleitfähigkeitstensor und f eine äußere Wärmequelle.

Im folgenden Abschnitt wollen wir uns zunächst mit der einfachsten Form dieser Gleichung beschäftigen, d.h. wir setzen $D = Id$.

Definition 3.1 (Anfangsrandwertproblem für die Wärmeleitungsgleichung)

Sei $\Omega \subset \mathbb{R}^d$ ein Gebiet und $\Omega_t := \Omega \times (0, \infty)$ das zugehörige Orts-Zeit-Gebiet. Eine Funktion $u : \Omega_t \rightarrow \mathbb{R}$ heißt klassische Lösung der Wärmeleitungsgleichung mit homogenen Dirichletrandwerten falls gilt

(i) **(Regularität)** $u \in C^0((0, \infty) \times \bar{\Omega}) \cap C^2((0, \infty) \times \Omega)$,

(ii) **(Differentialgleichung)**

$$\partial_t u(x, t) - \Delta u(x, t) = f(x, t), \quad \forall (x, t) \in \Omega_t,$$

(iii) **(Randwerte)** $u(x, t) = 0, \quad \forall (x, t) \in \partial\Omega \times (0, \infty)$,

(iv) **(Anfangswerte)** $u(x, 0) = u_0(x), \quad \forall x \in \Omega$.

Zur Lösung des Anfangsrandwertproblems gibt es unterschiedliche Ansätze. Ein Möglichkeit ist der Variationsansatz. Dazu nehmen wir an, dass es Funktionen $w(t)$ und $v(x)$ mit $v = 0$ auf $\partial\Omega$ gibt mit

$$u(x, t) = w(t)v(x).$$

Aus der Differentialgleichung folgt dann durch einsetzen

$$\partial_t w(t)v(x) - w(t)\Delta v(x) = 0,$$

oder

$$\frac{\partial_t w(t)}{w(t)} = \frac{\Delta v(x)}{v(x)}.$$

Da dies für alle t und alle x gelten soll, muß also ein $\lambda \in \mathbb{R}$ existieren mit

$$\frac{\partial_t w(t)}{w(t)} = -\lambda, \quad \frac{-\Delta v(x)}{v(x)} = \lambda.$$

Aus der Gleichung für w folgt durch Integration

$$w(t) = w(0)e^{-\lambda t}.$$

Die Gleichung für v hingegen führt auf das Eigenwertproblem für den Laplace-Operator

$$-\Delta v(x) = \lambda v(x), \quad \forall x \in \Omega, \quad v(x) = 0, \quad \forall x \in \partial\Omega.$$

Dieser Ansatz zeigt, dass man Lösungen der Wärmeleitungsgleichung mit Hilfe des Eigenwertproblems für den Laplace-Operator konstruieren kann. Man kann insbesondere zeigen, dass es für beschränkte Gebiete Ω eine Folge von Eigenwerten $0 < \lambda_1 \leq \lambda_2 \leq \dots$ mit zugehörigen Eigenfunktionen u_1, u_2, \dots von $-\Delta$ mit Nullrandwerten gibt. Die Eigenvektoren bilden ein vollständiges Orthonormalsystem in $L^2(\Omega)$. Mithilfe dieses Orthonormalsystems kommt man zu folgendem Satz.

Satz 3.2 (Klassische Lösung der Wärmeleitungsgleichung)

Sei $\Omega \subset \mathbb{R}^d$, mit $d \leq 3$ ein Gebiet mit glattem Rand und sei $u_0 \in C_0^2(\Omega)$. Dann ist durch

$$u(x, t) := \sum_{j=1}^{\infty} e^{-\lambda_j t} a_{0,j} u_j(x), \quad a_{0,j} := (u_0, u_j)_{L^2(\Omega)}$$

eine klassische Lösung des Anfangsrandwertproblems 3.1 gegeben.
(ohne Beweis)

In dieser Vorlesung wollen wir klassische Lösungen nicht weiter verfolgen und uns stattdessen mit schwachen Lösungen der Wärmeleitungsgleichung beschäftigen. Um die Existenz von schwachen Lösungen zu zeigen, werden wir uns einer Methode bedienen, die auf der Energieminimierung beruht. Dazu halten wir zunächst folgendes Resultat fest.

Lemma 3.3 (A priori Abschätzung)

Jede klassische Lösung u im Sinne von Definition 3.1 erfüllt für alle $t \in [0, T]$ die Gleichung

$$\frac{1}{2} \|u(\cdot, t)\|_{L^2(\Omega)}^2 + \int_0^t \|\nabla u(\cdot, s)\|_{L^2(\Omega)}^2 ds = \frac{1}{2} \|u_0(\cdot)\|_{L^2(\Omega)}^2. \quad (3.1)$$

Beweis: Wir starten mit der Wärmeleitungsgleichung aus Definition 3.1 multiplizieren mit u und integrieren über Ω und $(0, t)$. Wir erhalten mit partieller Integration

$$\begin{aligned} 0 &= \int_0^t \int_{\Omega} (\partial_t u - \Delta u) u = \int_0^t \int_{\Omega} \left(\partial_t \frac{1}{2} u^2 + |\nabla u|^2 \right) \\ &= \int_0^t \frac{d}{dt} \frac{1}{2} \int_{\Omega} |u(x, t)|^2 dx dt + \int_0^t \|\nabla u(\cdot, s)\|_{L^2(\Omega)}^2 ds \\ &= \frac{1}{2} \|u(\cdot, t)\|_{L^2(\Omega)}^2 - \frac{1}{2} \|u_0(\cdot)\|_{L^2(\Omega)}^2 + \int_0^t \|\nabla u(\cdot, s)\|_{L^2(\Omega)}^2 ds \end{aligned}$$

Damit ist die Behauptung bewiesen. \square

Auf dieser a priori Abschätzung beruhen viele Methoden, mit denen die Existenz von schwachen Lösungen gezeigt werden kann. Insbesondere liefert Gleichung (3.1) eine Abschätzung für die Norm von u im Funktionenraum $L^2(0, T; H_0^1(\Omega))$, den wir weiter unten noch definieren werden. Die Vorgehensweise sieht dabei immer wie folgt aus.

- (1) Konstruiere für $k \in \mathbb{N}$ approximative Lösungen u^k .
- (2) Zeige die a priori Abschätzung (3.1), d.h.

$$\|u^k\|_{L^2(0, T; H_0^1(\Omega))} \leq C,$$

mit einer Konstanten $C > 0$, die nicht von k abhängt.

(3) Finde einen Grenzwert u für eine Teilfolge von $(u^k)_{k \in \mathbb{N}}$.

(4) Zeige, dass u eine schwache Lösung ist.

Im folgenden wollen wir dieses Prozedere für die sogenannte Rothe-Methode durchführen, dabei wird die Differentialgleichung lediglich in der Zeit diskretisiert. Eine andere Möglichkeit wäre eine reine Ortsdiskretisierung.

Um das Existenzresultat vorzubereiten wollen wir jedoch zunächst schwache Funktionenräume für Funktionen einführen, die von der Zeit und dem Ort abhängen und einige Aussagen aus der Funktionalanalysis angeben.

3.1 Sobolevräume und Bochnerräume

Definition 3.4 (Die Räume $L^2(0, T; H_0^1(\Omega))$ und $L^2(0, T; H^{-1}(\Omega))$)

Wir definieren Sobolevräume für Funktionen $u : \Omega \times (0, T) \rightarrow \mathbb{R}$ wie folgt

$$\begin{aligned} L^2(0, T; L^2(\Omega)) &:= L^2((0, T) \times \Omega), \\ L^2(0, T; H^1(\Omega)) &:= \{u \in L^2((0, T) \times \Omega) \mid \nabla u \in L^2((0, T) \times \Omega)\}. \end{aligned}$$

Als natürliche Norm auf $L^2(0, T; H^1(\Omega))$ definieren wir

$$\|u\|_{L^2(0, T; H^1(\Omega))} := \left(\int_0^T \|u(\cdot, t)\|_{H^1(\Omega)}^2 dt \right)^{1/2}.$$

Beide Räume sind Hilberträume mit den entsprechenden kanonischen Skalarprodukten. Wir definieren weiter für Funktionen mit Nullrandwerten

$$\begin{aligned} L^2(0, T; H_0^1(\Omega)) &:= \{u \in L^2(0, T; H^1(\Omega)) \mid u(\cdot, t) \in H_0^1(\Omega) \text{ für f.a. } t \in (0, T)\}, \\ L^2(0, T; H^{-1}(\Omega)) &:= (L^2(0, T; H_0^1(\Omega)))'. \end{aligned}$$

Bemerkung 3.5 (Bochnerintegral und Bochnerräume)

Ist X ein Banachraum so kann man verallgemeinerte Funktionen $u : [0, T] \rightarrow X$ betrachten, und für solche Abbildungen den Begriff des Bochnerintegrals einführen. Allgemein definiert man dann die Bochnerräume

$$L^p(0, T; X)$$

als die Menge aller stark meßbaren Funktionen $u : [0, T] \rightarrow X$ mit

$$\|u\|_{L^p(0, T; X)} := \left(\int_0^T \|u(t)\|_X^p dt \right)^{1/p} < \infty.$$

für $1 \leq p < \infty$ und

$$\|u\|_{L^\infty(0, T; X)} := \text{ess} \sup_{0 \leq t \leq T} \|u(t)\|_X < \infty.$$

Für mehr Details siehe Übungsaufgaben und [Eva98, Abschnitte E.5 und 5.9.2].

Im folgenden werden wir die Begriffe des Bochnerintegrals und der Bochnerräume ohne detaillierte Einführung verwenden.

Definition 3.6 (Der Raum $C(0, T; X)$)

Der Raum

$$C(0, T; X)$$

besteht aus allen stetigen Funktionen $u : [0, T] \rightarrow X$ mit

$$\|u\|_{C(0, T; X)} := \max_{0 \leq t \leq T} \|u(t)\|_X < \infty.$$

Definition 3.7 (Schwache Zeitableitung)

Sei $u \in L^1(0, T; X)$. Wir sagen $v \in L^1(0, T; X)$ ist die schwache (Zeit-)Ableitung von u , falls gilt

$$\int_0^T \varphi'(t)u(t)dt = - \int_0^T \varphi(t)v(t)dt, \quad \forall \varphi \in C_0^\infty(0, T).$$

Wir schreiben u' für die schwache Ableitung v von u .

Für die schwache Formulierung der Wärmeleitungsgleichung benötigen wir noch folgendes Resultat.

Satz 3.8 (Einbettung)

Sei $u \in L^2(0, T; H_0^1(\Omega))$ mit schwacher Ableitung $u' \in L^2(0, T; H^{-1}(\Omega))$. Dann gilt

(i) $u \in C(0, T; L^2(\Omega))$ (nach eventueller Abänderung auf einer Nullmenge),

(ii) Die Abbildung $t \rightarrow \|u(t)\|_{L^2(\Omega)}^2$ ist absolut stetig und es gilt

$$\frac{d}{dt} \|u(t)\|_{L^2(\Omega)}^2 = 2\langle u'(t), u(t) \rangle, \quad \text{für f.a. } t \in [0, T].$$

(iii) Es gilt die Abschätzung

$$\|u\|_{C(0, T; L^2(\Omega))} \leq C(\|u\|_{L^2(0, T; H_0^1(\Omega))} + \|u'\|_{L^2(0, T; H^{-1}(\Omega))}).$$

Dabei hängt die Konstante C nur vom Endzeitpunkt T ab.

(ohne Beweis)

Für Schritt (3) im angestrebten Existenzbeweis benötigen wir noch folgenden Satz aus der Funktionalanalysis.

Satz 3.9 (Schwache Folgenkompaktheit)

Sei X ein reflexiver Banachraum – d.h. X ist isomorph zu seinem Bidualraum X'' – und $(x_k)_{k \in \mathbb{N}}$ eine Folge in X mit $\|x_k\| \leq C$ unabhängig von k . Dann existiert eine Teilfolge mit einem schwachen Grenzwert $x \in X$. D.h. es existiert ein $x \in X$ und eine Teilfolge $(k_j)_{j \in \mathbb{N}}$, so dass für alle $\varphi \in X'$ gilt

$$\varphi(x_{k_j}) \rightarrow \varphi(x) \quad \text{für } j \rightarrow \infty.$$

(Beweis siehe [Alt92, 5.7 Satz].)

Bemerkung 3.10

Die zeitabhängigen Funktionenräume aus Definition 3.4 sind alle Hilberträume und damit insbesondere reflexiv. Also haben in diesen Räumen beschränkte Folgen schwach konvergente Teilfolgen.

3.2 Schwache Formulierung der Wärmeleitungsgleichung

Wir haben nun alle Definitionen zur Hand, um den Begriff der schwachen Lösung der Wärmeleitungsgleichung einzuführen.

Definition 3.11 (Schwache Lösung)

Sei $\Omega \subset \mathbb{R}^d$ ein Lipschitzgebiet. Eine Funktion $u : \Omega \times [0, T] \rightarrow \mathbb{R}$ heißt schwache Lösung der Wärmeleitungsgleichung 3.1 für Anfangswerte $u_0 \in L^2(\Omega)$ und rechte Seite $f \in L^2(0, T; H^{-1}(\Omega))$, falls gilt

(1) **(Regularität)** $u \in L^2(0, T; H_0^1(\Omega))$ und $u' \in L^2(0, T; H^{-1}(\Omega))$,

(2) **(Schwache Form der Differentialgleichung mit Randbedingung)**

$$\langle u'(t), \varphi \rangle + (\nabla u(t), \nabla \varphi)_{L^2(\Omega)} = \langle f(t), \varphi \rangle, \quad \forall \varphi \in H_0^1(\Omega), \text{ f.a. } t \in (0, T),$$

(3) **(Anfangswerte)** $u(\cdot, 0) = u_0$.

Bemerkung 3.12

Die Definition der Anfangswerte macht Sinn, da nach Satz 3.6 gilt, dass u stetig in der Zeit als Abbildung nach $L^2(\Omega)$ ist.

Um die Existenz von schwachen Lösungen der Wärmeleitungsgleichung zu zeigen, werden wir nun der Rothe-Methode folgend eine Zeitdiskretisierung der Wärmeleitungsgleichung einführen.

Definition 3.13 (Zeitdiskretisierung)

Es gelten die Voraussetzungen und Notationen aus Definition 3.11. Für das Zeitintervall $[0, T]$ definieren wir eine reguläre Zerlegung $I_N := \{t^0, \dots, t^N\}$ der Feinheit $\Delta t := T/N$, durch $t^n := n\Delta t$. Weiter seien die Teilintervalle I_n definiert durch $I_n := [t^n, t^{n+1})$ für $n = 0, \dots, N-1$.

Eine zeitdiskrete Approximation $u^n \in H_0^1(\Omega)$ der Lösung $u(t^n)$ ist dann definiert durch die Vorschrift

$$\begin{aligned} u^0 &= u_0, \\ \left(\frac{u^{n+1} - u^n}{\Delta t}, \varphi \right)_{L^2(\Omega)} + (\nabla u^{n+1}, \nabla \varphi)_{L^2(\Omega)} &= 0, \quad \forall \varphi \in H_0^1(\Omega). \end{aligned} \quad (3.2)$$

Zu den Funktionswerten $(u^n)_{n=0, \dots, N}$ definieren wir eine in der Zeit stückweise konstante Funktion $\bar{u}^N : [0, T] \rightarrow H_0^1(\Omega)$ durch

$$\bar{u}^N|_{I_n} := u^{n+1}, \quad \text{für } n = 0, \dots, N-1.$$

und eine stetige und stückweise lineare Funktion $\tilde{u}^N : [0, T] \rightarrow L^2(\Omega)$ durch

$$\tilde{u}^N(t)|_{I_n} := \frac{t^{n+1} - t}{\Delta t} u^n + \frac{t - t^n}{\Delta t} u^{n+1}, \quad \text{für } t \in I_n \text{ und } n = 0, \dots, N-1.$$

Das folgende Lemma gibt uns eine uniforme a priori Abschätzung für die Approximationen.

Lemma 3.14 (A priori Abschätzung)

Für $u_0 \in H_0^1(\Omega)$ sind die Approximationen \bar{u}^N und \tilde{u}^N aus Definition 3.13 beschränkt in $L^2(0, T; H_0^1(\Omega))$. Die schwache Zeitableitung $\partial_t \tilde{u}^N$ existiert in $L^2(0, T; H^{-1}(\Omega))$ und es gilt

$$\partial_t \tilde{u}^N(\cdot, t) = \Delta \bar{u}^N(\cdot, t) \quad \text{in } H^{-1}(\Omega) \text{ für f.a. } t \in (0, T). \quad (3.3)$$

Insbesondere ist also $\partial_t \tilde{u}^N$ beschränkt in $L^2(0, T; H^{-1}(\Omega))$.

Beweis: Wir testen in Gleichung (3.2) mit $\varphi = u^{n+1}$ und erhalten

$$\begin{aligned} 0 &= \frac{1}{\Delta t} (\|u^{n+1}\|_{L^2(\Omega)}^2 - (u^n, u^{n+1})_{L^2(\Omega)}) + \|\nabla u^{n+1}\|_{L^2(\Omega)}^2 \\ &\geq \frac{1}{\Delta t} (\|u^{n+1}\|_{L^2(\Omega)}^2 - \frac{1}{2} (\|u^{n+1}\|_{L^2(\Omega)}^2 + \|u^n\|_{L^2(\Omega)}^2)) + \|\nabla u^{n+1}\|_{L^2(\Omega)}^2. \end{aligned}$$

Dabei haben wir die Cauchy-Schwarzsche Ungleichung und die Youngsche Ungleichung verwendet, d.h. $(u^n, u^{n+1}) \leq \frac{1}{2} (\|u^n\|^2 + \|u^{n+1}\|^2)$. Multiplikation

mit Δt und Summation über n ergibt

$$\begin{aligned} 0 &\geq \sum_{n=0}^{N-1} \Delta t \|\nabla u^{n+1}\|_{L^2(\Omega)}^2 + \sum_{n=0}^{N-1} \frac{1}{2} (\|u^{n+1}\|_{L^2(\Omega)}^2 - \|u^n\|_{L^2(\Omega)}^2) \\ &= \sum_{n=0}^{N-1} \Delta t \|\nabla u^{n+1}\|_{L^2(\Omega)}^2 + \frac{1}{2} \|u^N\|_{L^2(\Omega)}^2 - \frac{1}{2} \|u_0\|_{L^2(\Omega)}^2. \end{aligned}$$

Also erhalten wir als diskretes Analogon zu (3.1) die Abschätzung

$$\frac{1}{2} \|\bar{u}^N(t^N)\|_{L^2(\Omega)}^2 + \int_0^T \|\nabla \bar{u}^N\|_{L^2(\Omega)}^2 \leq \frac{1}{2} \|u_0\|_{L^2(\Omega)}^2.$$

Somit ist der erste Teil der Behauptung bewiesen. Die schwache Differenzierbarkeit von \tilde{u}^N und Gleichung (3.3) folgen direkt aus der Definition von \tilde{u}^N . Die Beschränktheit von $\partial_t \tilde{u}^N$ in $L^2(0, T; H^{-1}(\Omega))$ folgt dann aus (3.3) aufgrund der Tatsache, dass \bar{u}^N uniform in $L^2(0, T; H_0^1(\Omega))$ beschränkt ist. \square

Aufgrund von Lemma 3.14 und der schwachen Folgenkomaktheit nach Satz 3.9 gilt für die Folge von Approximationen $(\bar{u}^N)_{N \in \mathbb{N}}, (\tilde{u}^N)_{N \in \mathbb{N}}$, dass jeweils Teilfolgen existieren (wieder $(\bar{u}^N)_{N \in \mathbb{N}}$ und $(\tilde{u}^N)_{N \in \mathbb{N}}$ genannt, die schwach in $L^2(0, T; H_0^1(\Omega))$ gegen Grenzwerte \bar{u} und \tilde{u} konvergieren. Wir zeigen zunächst, dass gilt $\bar{u} = \tilde{u}$.

Lemma 3.15

Seien \bar{u} und \tilde{u} Grenzwerte von Teilfolgen der Approximationen $(\bar{u}^N)_{N \in \mathbb{N}}, (\tilde{u}^N)_{N \in \mathbb{N}}$ aus Definition 3.13. Dann gilt

$$\bar{u} = \tilde{u}.$$

Beweis: Sei $\varphi \in C_0^\infty((0, T) \times \Omega)$, so gilt nach Taylorentwicklung

$$\begin{aligned} \int_0^T (\bar{u}^N(t), \varphi(t))_{L^2(\Omega)} &= \sum_{n=1}^N \Delta t (u^n, \varphi(t^n))_{L^2(\Omega)} + \mathcal{O}(\Delta t), \\ \int_0^T (\tilde{u}^N(t), \varphi(t))_{L^2(\Omega)} &= \sum_{n=1}^N \Delta t \int_0^1 ((1-\mu)u^{n-1} + \mu u^n, \varphi(t^{n-1} + \mu \Delta t))_{L^2(\Omega)} \\ &= \sum_{n=1}^N \Delta t \left(\frac{1}{2} (u^{n-1}, \varphi(t^{n-1}))_{L^2(\Omega)} + \frac{1}{2} (u^n, \varphi(t^n))_{L^2(\Omega)} \right) + \mathcal{O}(\Delta t). \end{aligned}$$

Im Grenzübergang für $N \rightarrow \infty$ gilt also $\langle \bar{u}, \varphi \rangle = \langle \tilde{u}, \varphi \rangle$ für alle $\varphi \in C_0^\infty((0, T) \times \Omega)$ und somit folgt mit dem Hauptsatz der Variationsrechnung $\bar{u} = \tilde{u}$. \square

Wir können nun zeigen, dass der Grenzwert $\bar{u} = \tilde{u}$ die schwache Form der Differentialgleichung löst.

Lemma 3.16

Sei $u := \bar{u} = \tilde{u} \in L^2(0, T; H_0^1(\Omega))$, wobei \bar{u} und \tilde{u} Grenzwerte von Teilstreichen der Approximationen $(\bar{u}^N)_{N \in \mathbb{N}}, (\tilde{u}^N)_{N \in \mathbb{N}}$ aus Definition 3.13 seien. Dann ist $\partial_t u \in L^2(0, T; H^{-1}(\Omega))$ wohldefiniert und u löst die schwache Form der Wärmeleitungsgleichung

$$\langle u'(t), \varphi \rangle + (\nabla u(t), \nabla \varphi)_{L^2(\Omega)} = 0, \quad \forall \varphi \in H_0^1(\Omega).$$

Beweis: Sei $\varphi \in C_0^\infty((0, T) \times \Omega)$, so folgt aus Lemma 3.14

$$\langle \partial_t \tilde{u}^N, \varphi \rangle = \langle \Delta \tilde{u}^N, \varphi \rangle.$$

Nun gilt einerseits

$$\langle \partial_t \tilde{u}^N, \varphi \rangle = -\langle \tilde{u}^N, \partial_t \varphi \rangle \rightarrow -\langle u, \partial_t \varphi \rangle = \langle \partial_t u, \varphi \rangle$$

und andererseits

$$\langle \Delta \tilde{u}^N, \varphi \rangle = \langle \tilde{u}^N, \Delta \varphi \rangle \rightarrow \langle u, \Delta \varphi \rangle = \langle \Delta u, \varphi \rangle.$$

Also gilt

$$\langle \partial_t u, \varphi \rangle = \langle \Delta u, \varphi \rangle.$$

Weiter folgt, da $u \in L^2(0, T; H_0^1(\Omega))$ ist, dass Δu ein lineares stetiges Funktional auf $L^2(0, T; H_0^1(\Omega))$ ist, also gilt $\Delta u \in L^2(0, T; H^{-1}(\Omega))$ und wegen der Gleichung auch $\partial_t u \in L^2(0, T; H^{-1}(\Omega))$. \square

Zur Existenz einer schwachen Lösung der Wärmeleitungsgleichung fehlt nun lediglich noch die Annahme der Anfangswerte und ein Argument, dass die Aussage aus Lemma 3.14 auf Anfangswerte $u_0 \in L^2(\Omega)$ verallgemeinert werden kann.

Wir fassen das Ergebnis in folgendem Existenzsatz zusammen.

**Satz 3.17 (Existenz einer schwachen Lösung der Wärmeleitungs-
gleichung)**

Sei $\Omega \subset \mathbb{R}^d$ ein Lipschitzgebiet, $u_0 \in L^2(\Omega)$ und $T > 0$. Dann existiert eine schwache Lösung u der Wärmeleitungs-
gleichung (vgl. Definition 3.11).

Beweis: **1. Schritt:** Approximation der Anfangswerte.

In den Lemmata 3.14 bis 3.16 haben wir die Existenz einer Lösung für Daten $u_0 \in H_0^1(\Omega)$ gezeigt. Ist $u_0 \in L^2(\Omega)$, so erhalten wir mit der Dichtheit von $H_0^1(\Omega)$ in $L^2(\Omega)$ ein äquivalentes Resultat, indem wir u_0 durch eine Folge von Funktionen $u_0^N \in H_0^1(\Omega)$ approximieren mit $u_0^N \rightarrow u_0$ in $L^2(\Omega)$.

2. Schritt: Annahme der Anfangswerte.

Für $\varphi \in C_0^\infty([0, T) \times \Omega)$ gilt die Gleichung

$$\int_0^T \langle \partial_t \tilde{u}^N(\cdot, t), \varphi(\cdot, t) \rangle dt + \int_0^T \langle \tilde{u}^N(\cdot, t), \partial_t \varphi(\cdot, t) \rangle dt = -\langle u_0^N, \varphi(\cdot, 0) \rangle.$$

Da die Funktionen $\partial_t \tilde{u}^N$ uniform in $L^2(0, T; H^{-1}(\Omega))$ beschränkt sind, exi-
stiert eine Teilfolge, die schwach in $L^2(0, T; H^{-1}(\Omega))$ konvergiert. Dadurch
können wir in allen Summanden zum Grenzwert übergehen und erhalten,
dass u die Anfangswerte in schwachem Sinne annimmt, d.h. es gilt

$$\int_0^T \langle \partial_t u(\cdot, t), \varphi(\cdot, t) \rangle dt + \int_0^T \langle u(\cdot, t), \partial_t \varphi(\cdot, t) \rangle dt = -\langle u_0, \varphi(\cdot, 0) \rangle.$$

Aufgrund von Satz 3.8 folgt, dass $u \in C^0(0, T; L^2(\Omega))$ ist. Daher folgt $u(\cdot, 0) = u_0$. \square

Bemerkung 3.18 (Eindeutigkeit und Verallgemeinerungen)

(1) Die Eindeutigkeit einer schwachen Lösung folgt direkt aus der Energie-
gleichung (3.1) für u , die aufgrund der absoluten Stetigkeit der Norm (vgl.
Satz 3.8 (ii)) auch für schwache Lösungen der Wärmeleitungs-
gleichung gilt.

(2) Der hier vorgestellte Existenzbeweis lässt sich verallgemeinern auf rechte
Seiten $f \neq 0$, Dirichletrandwerte $g_D \neq 0$ und auf allgemeine elliptische Diffe-
rentialoperatoren anstelle des Laplace-Operators.

3.3 Semidiskrete Finite Elemente Methode

Wir wollen in diesem Abschnitt zunächst nur die Ortsdiskretisierung von parabolischen Differentialgleichungen mit Hilfe von Finite Elemente Verfahren untersuchen. Dazu betrachten wir als Prototyp eines parabolischen Anfangswertproblems die inhomogene Wärmeleitungsgleichung.

Definition 3.19 (Inhomogene Wärmeleitungsgleichung)

Sei $\Omega \subset \mathbb{R}^d$ ein Lipschitzgebiet. Eine Funktion $u : \Omega \times [0, T]$ heißt schwache Lösung der Wärmeleitungsgleichung 3.1 für Anfangswerte $u_0 \in L^2(\Omega)$ und rechte Seite $f \in L^2(0, T; H^{-1}(\Omega))$, falls gilt

(1) **(Regularität)** $u \in L^2(0, T; H_0^1(\Omega))$ und $u' \in L^2(0, T; H^{-1}(\Omega))$,

(2) **(Schwache Form der Differentialgleichung mit Randbedingung)**

$$\langle u'(t), \varphi \rangle + (\nabla u(t), \nabla \varphi)_{L^2(\Omega)} = \langle f(t), \varphi \rangle, \quad \forall \varphi \in H_0^1(\Omega), \text{f.a. } t \in (0, T),$$

(3) **(Anfangswerte)** $u(\cdot, 0) = u_0$.

Die Finite Elemente Semi-Diskretisierung wird nun analog zu Kapitel 2 definiert und führt auf ein System gewöhnlicher Anfangswertprobleme für die Koeffizientenfunktionen $u_i(t)$ der Basisdarstellung $u_h(x, t) := \sum_{i=1}^N u_i(t) \varphi_i(x)$.

Definition 3.20 (Finite Elemente Semi-Diskretisierung)

Seien $\Omega \subset \mathbb{R}^d$ ein Gebiet und \mathcal{T}_h ein zulässige Triangulierung von Ω . Sei $X_h \subset H_0^1(\Omega)$ ein endlichdimensionaler Finite Elemente Raum über \mathcal{T}_h mit Dimension N und Basisfunktionen $\varphi_1, \dots, \varphi_N$. Die Finite Elemente Semi-Diskretisierung $u_h \in C^1(0, T; X_h)$ der inhomogenen Wärmeleitungsgleichung aus Definition 3.10 ist dann für Anfangswerte $u_0 \in H_0^1(\Omega)$ definiert durch

$$(\partial_t u_h(\cdot, t), \varphi_h)_{L^2(\Omega)} + (\nabla u_h(\cdot, t), \nabla \varphi_h)_{L^2(\Omega)} = \langle f(\cdot, t), \varphi_h \rangle, \quad \forall \varphi_h \in X_h, \text{f.a. } t \in (0, T),$$

$$u_h(\cdot, 0) = I_h(u_0).$$

Dabei ist $I_h : H_0^1(\Omega) \rightarrow X_h$ eine H^1 -Interpolation.

Bemerkung 3.21 (Anfangswertproblem für die Koeffizientenfunktionen)

Mit der Basisdarstellung

$$u_h(x, t) = \sum_{i=1}^N u_i(t) \varphi_i(x)$$

erhält man für die zeitabhängigen Koeffizientenfunktionen $u_i(t)$ aus der Definition von u_h

$$\sum_{i=1}^N \frac{d}{dt} u_i(t) (\varphi_i, \varphi_j)_{L^2(\Omega)} + \sum_{i=1}^N u_i(t) (\nabla \varphi_i, \nabla \varphi_j)_{L^2(\Omega)} = \langle f(\cdot, t), \varphi_j \rangle, j = 1, \dots, N.$$

Definieren wir die Massmatrix $M \in \mathbb{R}^{N \times N}$ durch

$$M_{ji} := (\varphi_i, \varphi_j)_{L^2(\Omega)}$$

und die Steifigkeitsmatrix $S \in \mathbb{R}^{N \times N}$ durch

$$S_{ji} := (\nabla \varphi_i, \nabla \varphi_j)_{L^2(\Omega)},$$

so folgt für den Koeffizientenvektor $U(t) := (u_1(t), \dots, u_N(t))^\top$ mit der rechten Seite $F(t) := (f_0(t), \dots, f_N(t))^\top$, $f_i(t) := \langle f(\cdot, t), \varphi_i \rangle, i = 0, \dots, N$

$$MU'(t) + SU(t) = F.$$

Dies ist ein System von linearen gewöhnlichen Differentialgleichungen für den Koeffizientenvektor U . Zusammen mit den Anfangsdaten, erhält man also ein gewöhnliches Anfangswertproblem.

Wir wollen nun die Konvergenz der Finite Elemente Semi-Diskretisierung untersuchen. Da die Fehlerabschätzungen auf der Idee der Energieabschätzung beruhen, wollen wir dazu zunächst einige a priori Abschätzungen für die inhomogene Wärmeleitungsgleichung beweisen. Dabei nehmen wir an, dass die Lösung der Wärmeleitungsgleichung genügend glatt ist.

Lemma 3.22 (A priori Abschätzungen)

Sei $u \in L^2(0, T; H_0^1(\Omega))$ die schwache Lösung der inhomogenen Wärmeleitungsgleichung aus Definition 3.10. Sind f, u_0 und u genügend regulär, so

gelten für $t \in [0, T]$ die folgenden a priori Abschätzungen

$$\|u(\cdot, t)\| \leq \int_0^t \|f(\cdot, s)\| ds + \|u_0\|, \quad (3.4)$$

$$\|u(\cdot, t)\| \leq \int_0^t e^{\lambda_1(s-t)} \|f(\cdot, s)\| ds + e^{-\lambda_1 t} \|u_0\|, \quad (3.5)$$

$$\int_0^t \|\partial_t u(\cdot, s)\|^2 ds + \|\nabla u(\cdot, t)\|^2 \leq \int_0^t \|f(\cdot, s)\|^2 ds + \|\nabla u_0\|^2, \quad (3.6)$$

$$\|\partial_t u(\cdot, t)\| \leq \int_0^t e^{\lambda_1(s-t)} \|\partial_t f(\cdot, s)\| ds + e^{-\lambda_1 t} (\|u_0\|_{H^2(\Omega)} + \|f(\cdot, 0)\|). \quad (3.7)$$

Dabei bezeichnet $\|\cdot\|$ die $L^2(\Omega)$ -Norm und λ_1 ist der kleinste Eigenwert von $-\Delta$, so dass die folgende optimale Poincaré-Ungleichung gilt

$$\lambda_1 \|u\|^2 \leq \|\nabla u\|^2.$$

Beweis: Abschätzung (3.1):

Wir setzen $\varphi = u(\cdot, t)$ in der schwachen Form der Wärmeleitungsgleichung ein und erhalten, falls u genügend regulär ist

$$(\partial_t u(t), u(t)) + \|\nabla u(t)\|^2 = (f(t), u(t)).$$

Dabei bezeichne (\cdot, \cdot) das Skalarprodukt in $L^2(\Omega)$. Hieraus folgt

$$\frac{1}{2} \frac{d}{dt} \|u(t)\|^2 + \|\nabla u(t)\|^2 \leq \|f(t)\| \|u(t)\|.$$

Hieraus folgt mit $\frac{1}{2} \frac{d}{dt} \|u(t)\|^2 = \|u(t)\| \frac{d}{dt} \|u(t)\|$

$$\frac{d}{dt} \|u(t)\| \leq \|f(t)\|.$$

Durch Integration erhalten wir hieraus die Abschätzung (3.1), wobei wir dabei den Term $\|\nabla u(t)\|^2$ nicht ausgenutzt haben.

Abschätzung (3.2):

Verwenden wir die Poincaré-Ungleichung

$$\lambda_1 \|u\|^2 \leq \|\nabla u\|^2,$$

so können wir die Abschätzung (3.1) verbessern. Zunächst erhalten wir

$$\frac{1}{2} \frac{d}{dt} \|u(t)\|^2 + \lambda_1 \|u(t)\|^2 \leq \|f(t)\| \|u(t)\|,$$

beziehungsweise

$$\frac{d}{dt}||u(t)|| + \lambda_1||u(t)|| \leq ||f(t)||.$$

Dies ist äquivalent zu

$$\frac{d}{dt}(e^{\lambda_1 t}||u(t)||) \leq e^{\lambda_1 t}||f(t)||,$$

da gilt $\frac{d}{dt}(e^{\lambda_1 t}||u(t)||) = e^{\lambda_1 t}(\lambda_1||u(t)|| + \frac{d}{dt}||u(t)||)$. Integration von 0 bis t liefert dann

$$e^{\lambda_1 t}||u(t)|| - e^0||u(0)|| \leq \int_0^t e^{\lambda_1 s}||f(s)||ds.$$

Somit haben wir (3.2) bewiesen.

Abschätzung (3.3):

Ist u genügend regulär, so können wir als Testfunktion φ auch $\partial_t u$ wählen. Dies führt zu

$$||\partial_t u(t)||^2 + \frac{1}{2} \frac{d}{dt} ||\nabla u(t)||^2 = (f(t), \partial_t u(t)).$$

Mit Cauchy-Schwarz und der Youngschen Ungleichung erhalten wir hieraus

$$||\partial_t u(t)||^2 + \frac{1}{2} \frac{d}{dt} ||\nabla u(t)||^2 \leq \frac{1}{2} ||f(t)|| + \frac{1}{2} ||\partial_t u(t)||^2.$$

Somit erhalten wir die Abschätzung

$$||\partial_t u(t)||^2 + \frac{d}{dt} ||\nabla u(t)||^2 \leq ||f(t)||$$

und nach Integration über t

$$\int_0^t ||\partial_t u(s)||^2 ds + ||\nabla u(t)||^2 \leq \int_0^t ||f(s)|| ds + ||\nabla u_0||^2.$$

Dies beweist (3.3).

Abschätzung (3.4):

Differenzieren wir die schwache Formulierung zunächst nach t , so erhalten wir für $\partial_t u(t)$ die schwache Form

$$(\partial_{tt} u(t), \varphi) + (\nabla \partial_t u(t), \nabla \varphi) = (\partial_t f(t), \varphi).$$

Wählt man nun $\varphi = \partial_t u$ und geht analog zu (3.2) vor, so erhält man die Abschätzung

$$\|\partial_t u(t)\| \leq \int_0^t e^{\lambda_1(s-t)} \|\partial_t f(s)\| ds + e^{-\lambda_1 t} \|\partial_t u(0)\|.$$

Bei genügender Regularität gilt

$$\partial_t u(0) = \Delta u_0 + f(0),$$

also erhalten wir insgesamt Abschätzung (3.4). \square

Wir verwenden die Ideen der Energieabschätzungen, um den Approximationsfehler, der durch die Finite Elemente Semi-Diskretisierung entsteht, abzuschätzen.

Dazu werden wir wie folgt vorgehen:

- (1) Herleitung einer schwachen Differentialgleichung für den Fehler $u - u_h$
- (2) Definition der Galerkin-Projektion $P_h u(t) \in X_h$ von u
- (4) Herleitung einer schwachen Differentialgleichung für den Fehler $P_h u - u_h$
- (4) Wahl einer geeigneten Testfunktion: $\varphi_h := (P_h u - u_h)(t)$
- (5) Energieabschätzung für den Fehler $P_h u - u_h$
- (6) Verwenden der Approximationseigenschaften der Galerkin-Projektion

Satz 3.23 (A priori Fehlerabschätzung)

Sei u eine genügend glatte Lösung der inhomogenen Wärmeleitungsgleichung aus Definition 3.7 und $u_h \in C^1(0, T; X_h)$, $X_h := S_{h,0}^k$ die Finite Elemente Semi-Diskretisierung aus Definition 3.10. Zusätzlich existiere ein Konstante $c_* > 0$, so dass für alle $g \in L^2(\Omega)$ eine schwache Lösung $v_g \in H_0^1(\Omega)$ des dualen Poissonproblems

$$\int_{\Omega} \nabla \varphi \cdot \nabla v_g = \int_{\Omega} g \varphi, \quad \forall \varphi \in H_0^{1,2}(\Omega)$$

existiert, die H^2 -regulär ist ($v_g \in H^2(\Omega)$) und die folgende uniforme a priori Abschätzung erfüllt

$$\|v_g\|_{H^2(\Omega)} \leq c_* \|g\|_{L^2(\Omega)}.$$

Dann gilt für $t > 0$

$$\begin{aligned} \|u(t) - u_h(t)\|_{L^2(\Omega)} &\leq e^{-\lambda_1 t} \|u_0 - u_h(0)\|_{L^2(\Omega)} \\ &+ ch^{k+1} \left(\|u(t)\|_{H^{k+1}(\Omega)} + e^{-\lambda_1 t} \|u_0\|_{H^{k+1}(\Omega)} + \int_0^t e^{\lambda_1(s-t)} \|\partial_t u(s)\| ds \right). \end{aligned}$$

Beweis: Subtrahieren wir die schwache Form der Finite Elemente Semi-Diskretisierung von der schwachen Form der Differentialgleichung, so erhalten wir bei genügenden Regularität der Lösung u

$$(\partial_t(u - u_h)(t), \varphi_h) + (\nabla(u - u_h)(t), \nabla\varphi_h) = 0, \quad \text{für alle } \varphi_h \in S_{h,0}^k.$$

Wir definieren die Galerkin-Projektion (oder auch Ritz-Projektion) $P_h u(t) \in X_h$ von $u(t)$ durch

$$(\nabla P_h u(t), \nabla\varphi_h) = (\nabla u(t), \nabla\varphi_h), \quad \forall \varphi_h \in X_h, t \in [0, T].$$

Als Fehlergleichung für $P_h u - u_h$ erhalten wir dann aus der Fehlergleichung für $u - u_h$ die Gleichung

$$\begin{aligned} (\partial_t(P_h u - u_h)(t), \varphi_h) + (\nabla(P_h u - u_h)(t), \nabla\varphi_h) &= (\partial_t(P_h u - u)(t), \varphi_h), \\ &\quad \text{für alle } \varphi_h \in S_{h,0}^k. \end{aligned}$$

Wählen wir als Testfunktion $\varphi = (P_h u - u_h)(t)$, so folgt

$$\begin{aligned} (\partial_t(P_h u - u_h)(t), (P_h u - u_h)(t)) + \|\nabla(P_h u - u_h)(t)\|^2 \\ = (\partial_t(P_h u - u)(t), (P_h u - u_h)(t)). \end{aligned}$$

Analog zur a priori Abschätzung (3.1) erhalten wir hieraus

$$\|(P_h u - u_h)(t)\| \leq e^{-\lambda_1 t} \|(P_h u - u_h)(0)\| + \int_0^t e^{\lambda_1(s-t)} \|\partial_t(P_h u - u)(s)\| ds.$$

Mit der Dreiecksungleichung folgt dann für den Fehler $u - u_h$

$$\begin{aligned} \|(u - u_h)(t)\| &\leq \|(u - P_h u)(t)\| + \|(P_h u - u_h)(t)\| \\ &\leq \|(u - P_h u)(t)\| + e^{-\lambda_1 t} \|(P_h u - u_h)(0)\| \\ &\quad + \int_0^t e^{\lambda_1(s-t)} \|\partial_t(P_h u - u)(s)\| ds. \end{aligned}$$

Um zu der Fehlerabschätzung des Satzes zu kommen, müssen wir nun noch die Fehler $(u - P_h u)(t)$ und $\partial_t(P_h u - u)(t)$ in der $L^2(\Omega)$ -Norm abschätzen.

Eine optimale Fehlerabschätzung hierfür liefert uns die L^2 -Fehlerabschätzung aus Korollar 2.38 (mit Aubin-Nitsche Trick)

$$\begin{aligned} \|(u - P_h u)(t)\| &\leq ch^{k+1} |u(t)|_{H^{k+1}(\Omega)}, \\ \|(\partial_t u - P_h(\partial_t u))(t)\| &\leq ch^{k+1} |\partial_t u(t)|_{H^{k+1}(\Omega)}. \end{aligned}$$

Zusammen mit der Abschätzung

$$\begin{aligned} \|(P_h u - u_h)(0)\| &\leq \|(P_h u - u)(0)\| + \|(u - u_h)(0)\| \\ &\leq ch^{k+1} |u(0)|_{H^{k+1}(\Omega)} + \|(u - u_h)(0)\| \end{aligned}$$

folgt dann die Behauptung. \square

Bemerkung 3.24

Im Beweis haben wir verwendet, dass man die Zeitableitung mit der Galerkin-Projektion vertauschen darf, d.h. es gilt $\partial_t(P_h u) = P_h(\partial_t u)$. Dies sieht man wie folgt ein: Nach Definition von P_h gilt

$$\begin{aligned} (\nabla P_h u(t), \nabla \varphi_h) &= (\nabla u(t), \nabla \varphi_h), \quad \forall \varphi_h \in X_h, \\ (\nabla P_h(\partial_t u)(t), \nabla \varphi_h) &= (\nabla \partial_t u(t), \nabla \varphi_h), \quad \forall \varphi_h \in X_h. \end{aligned}$$

Differenziert man die erste Gleichung nach t , so folgt

$$(\nabla \partial_t(P_h u)(t), \nabla \varphi_h) = (\nabla \partial_t u(t), \nabla \varphi_h), \quad \forall \varphi_h \in X_h.$$

Wir erhalten also

$$(\nabla P_h(\partial_t u)(t), \nabla \varphi_h) = (\nabla \partial_t(P_h u)(t), \nabla \varphi_h), \quad \forall \varphi_h \in X_h$$

und somit gilt $\partial_t(P_h u) = P_h(\partial_t u)$.

Nachdem wir eine Fehlerabschätzung in der $L^\infty(0, T; L^2(\Omega))$ -Norm gezeigt haben, wollen wir im nächsten Satz eine Abschätzung für den Gradienten beweisen.

Satz 3.25 (Fehlerabschätzung im Gradienten)

Unter den Voraussetzungen von Satz 3.13 gilt für $t > 0$ die Fehlerabschätzung

$$\begin{aligned} \|\nabla(u - u_h)(t)\|_{L^2(\Omega)} &\leq c \|\nabla(u - u_h)(0)\|_{L^2(\Omega)} \\ &\quad + ch^k \left(\|u_0\|_{H^{k+1}(\Omega)} + \|u(t)\|_{H^{k+1}(\Omega)} + \left(\int_0^t \|\partial_t u(s)\|_{H^k(\Omega)}^2 ds \right)^{1/2} \right). \end{aligned}$$

Beweis: Analog zum Beweis der L^2 -Abschätzung erhalten wir durch Subtraktion der schwachen Formulierungen von u und u_h

$$(\partial_t(u - u_h)(t), \varphi_h) + (\nabla(u - u_h)(t), \nabla\varphi_h) = 0, \quad \text{für alle } \varphi_h \in S_{h,0}^k.$$

Wieder erhalten wir mit der Galerkin-Projektion $P_h u(t) \in X_h$, definiert durch

$$(\nabla P_h u(t), \nabla\varphi_h) = (\nabla u(t), \nabla\varphi_h), \quad \forall \varphi_h \in S_{h,0}^k, t \in [0, T]$$

die Fehlerequation für $P_h u - u_h$

$$(\partial_t(P_h u - u_h)(t), \varphi_h) + (\nabla(P_h u - u_h)(t), \nabla\varphi_h) = (\partial_t(P_h u - u)(t), \varphi_h),$$

für alle $\varphi_h \in S_{h,0}^k$.

Der a priori Abschätzung (3.2) folgend wählen wir hier die Zeitableitung des Fehlers als Testfunktion, d.h. $\varphi = \partial_t(P_h u - u_h)(t)$. Wir erhalten

$$\|\partial_t(P_h u - u_h)(t)\|^2 + \frac{1}{2} \frac{d}{dt} \|\nabla(P_h u - u_h)(t)\|^2 = (\partial_t(P_h u - u)(t), (P_h u - u_h)(t)).$$

Jetzt können wir analog zur Herleitung der a priori Abschätzung (3.2) vorgehen. Wir erhalten

$$\begin{aligned} \frac{d}{dt} \|\nabla(P_h u - u_h)(t)\|^2 &\leq \|\partial_t(P_h u - u)(t)\|^2 \\ &\leq ch^{2k} \|\partial_t u(t)\|_{H^k(\Omega)}^2. \end{aligned}$$

Durch Integration in der Zeit von 0 bis t erhalten wir weiter

$$\begin{aligned} \|\nabla(P_h u - u_h)(t)\|^2 &\leq \|\nabla(P_h u - u_h)(0)\|^2 + ch^{2k} \int_0^t \|\partial_t u(s)\|_{H^k(\Omega)}^2 ds \\ &\leq (\|\nabla(P_h u_0 - u_0)\| + \|\nabla(u_0 - u_h(0))\|)^2 + ch^{2k} \int_0^t \|\partial_t u(s)\|_{H^k(\Omega)}^2 ds \\ &\leq c \|\nabla(u - u_h)(0)\|^2 + ch^{2k} \|u_0\|_{H^{k+1}(\Omega)}^2 + ch^{2k} \int_0^t \|\partial_t u(s)\|_{H^k(\Omega)}^2 ds. \end{aligned}$$

Mit der Dreiecksungleichung erhalten wir schließlich

$$\|\nabla(u - u_h)(t)\| \leq \|\nabla(u - P_h u)(t)\| + \|\nabla(P_h u - u_h)(t)\|$$

und somit folgt die Behauptung mit der Fehlerabschätzung für die Galerkin-Projektion

$$\|\nabla(u - P_h u)(t)\| \leq ch^k \|u(t)\|_{H^{k+1}(\Omega)}.$$

□

3.4 Zeitschrittverfahren

Die Finite-Elemente Semidiskretisierung aus Abschnitt 3.3 führt auf ein Anfangswertproblem für ein Differentialgleichungssystem der Form

$$\frac{d}{dt}U(t) = BU(t) + \tilde{F}(t), \quad (3.8)$$

Dabei ist $U(t)$ der Koeffizientenvektor der Semi-Diskretisierung u_h , $B := -M^{-1}S$ und $\tilde{F} := M^{-1}F(t)$ (vgl. Abschnitt 3.3).

Nun könnte man denken, dass man zur Zeitdiskretisierung ein beliebiges Verfahren zur Lösung von Anfangswertproblemen einsetzen kann. Problematisch dabei ist jedoch, dass (3.5) ein steifes Anfangswertproblem darstellt, da die Systemmatrix B von negativen h -Potenzen abhängt. In klassische Fehlerabschätzungen für Anfangswertprobleme geht die Lipschitz-Konstante, also $\|B\|$ ein. Überträgt man also solche Techniken auf diesen Fall, so erhält man Fehlerabschätzungen mit positiven Potenzen in der Zeitschrittweite Δt und eventuell negativen Potenzen in h .

Im Fall einfacher Diskretisierungen bezüglich der Zeit kann man den Gesamtfehler auch direkt abschätzen, ohne die Aufspaltung in räumlichen und zeitlichen Diskretisierungsfehler vorzunehmen. Wir wollen uns dies zunächst für die implizite Euler-Diskretisierung ansehen.

Definition 3.26 (Implizite Euler Diskretisierung)

Es gelten die Notationen und Bezeichnungen aus Abschnitt 3.3. Sei $u_h : [0, T] \times \Omega \rightarrow \mathbb{R}$ die Finite Elemente Semi-Diskretisierung aus Definition 3.10. Für das Zeitintervall $[0, T]$ definieren wir eine reguläre Zerlegung $I_M := \{t^0, \dots, t^M\}$ der Feinheit $\Delta t := T/M$, durch $t^m := m\Delta t$. Weiter seien die Teilintervalle I_m definiert durch $I_m := [t^m, t^m + 1)$ für $m = 0, \dots, M-1$. Die implizite Euler Diskretisierung $u_h^m \in S_{h,0}^k$ der Lösung $u_h(\cdot, t^m)$, $m = 0, \dots, M$, ist dann definiert durch die Vorschrift

$$\begin{aligned} u_h^0 &= u_h(0), \\ \left(\frac{u_h^{m+1} - u_h^m}{\Delta t}, \varphi_h \right)_{L^2(\Omega)} + (\nabla u_h^{m+1}, \nabla \varphi_h)_{L^2(\Omega)} &= \langle f(\cdot, t^{m+1}), \varphi_h \rangle, \quad (3.9) \\ \forall \varphi_h \in S_{h,0}^k. \end{aligned}$$

Bemerkung 3.27 (Iterationsvorschrift für die Koeffizientenvektoren)

Seien $M \in \mathbb{R}^{N \times N}$ die Massenmatrix, $S \in \mathbb{R}^{N \times N}$ die Steifigkeitsmatrix und $F^m := F(t^m) = (\langle f(\cdot, t^m), \varphi_i \rangle)_{i=1,\dots,N}$ der rechte Seite Vektor aus Bemerkung 3.11. Dann gilt für die Koeffizientenvektoren U^m von u_h^m die Iterationsvorschrift

$$(M + \Delta t S)U^{m+1} = MU^m + \Delta t F^m.$$

In jedem Zeitschritt muss also ein lineares Gleichungssystem gelöst werden mit der Matrix $A = M + \Delta t S$. Die Matrix A ist symmetrisch und positiv definit. Insbesondere ist die implizite Euler Diskretisierung $U^m, m = 0, \dots, M$ wohldefiniert.

Wir wollen nun den Beweis der L^2 -Fehlerabschätzung aus Satz 3.13 im voll diskreten Fall imitieren, um zu einer Abschätzung des Gesamtfehlers der impliziten Euler Diskretisierung zu gelangen.

Satz 3.28 (A priori Fehlerabschätzung)

Es gelten die Voraussetzungen aus Satz 3.13. Weiter sei $u_h^m \in S_{h,0}^k, m = 0, \dots, M$ die impliziten Euler Diskretisierung aus Definition 3.16. Dann gilt die Fehlerabschätzung

$$\begin{aligned} \|u(t^m) - u_h^m\|_{L^2(\Omega)} &\leq \|u_0 - u_h^0\|_{L^2(\Omega)} + \Delta t \int_0^{t^m} \|\partial_{tt} u(s)\|_{L^2(\Omega)} ds \\ &\quad + ch^{k+1} \left(\|u_0\|_{H^{k+1}(\Omega)} + \int_0^{t^m} \|\partial_t u(s)\|_{H^{k+1}(\Omega)} ds \right). \end{aligned}$$

Beweis: Wir immitieren den Beweis des semi-diskreten Falls. Mit der Galerkin-Projektion $P_h u$ (vgl. Beweis von Satz 3.13) gilt

$$u(t^m) - u_h^m = (u(t^m) - P_h u(t^m)) + (P_h u(t^m) - u_h^m).$$

Der erste Summand lässt sich analog zum Beweis von Satz 3.13 abschätzen durch

$$\|u(t^m) - P_h u(t^m)\| \leq ch^{k+1} \|u(t^m)\|_{H^{k+1}(\Omega)}.$$

Wegen

$$u(t^m) = u(0) + \int_0^{t^m} \partial_t u(s) ds$$

lässt sich dies weiter abschätzen zu

$$\|u(t^m) - P_h u(t^m)\| \leq ch^{k+1} (\|u_0\|_{H^{k+1}(\Omega)} + \int_0^{t^m} \|\partial_t u(s)\|_{H^{k+1}(\Omega)} ds).$$

Der zweite Summand $(P_h u(t^m) - u_h^m)$ ist schwieriger zu behandeln. Wieder haben wir das Ziel durch Subtraktion der schwachen Formulierungen für u und u_h^m zu einer Fehlerequation für $(P_h u(t^m) - u_h^m)$ zu gelangen. Es gilt für u_h^m

$$\left(\frac{u_h^m - u_h^{m-1}}{\Delta t}, \varphi_h \right) + (\nabla u_h^m, \nabla \varphi_h) = \langle f(\cdot, t^m), \varphi_h \rangle$$

und für u erhalten wir zur Zeit $t = t^m$

$$\left(\partial_t u(t^m), \varphi_h \right) + (\nabla u(t^m), \nabla \varphi_h) = \langle f(\cdot, t^m), \varphi_h \rangle.$$

Mit der Definition der Galerkin-Projektion folgt aus der Gleichung für $u(t^m)$

$$\left(\partial_t u(t^m), \varphi_h \right) + (\nabla P_h u(t^m), \nabla \varphi_h) = \langle f(\cdot, t^m), \varphi_h \rangle.$$

Definieren den Fehler $e_h^m := (P_h u(t^m) - u_h^m)$, so erhalten wir zusammen mit der Gleichung für u_h^m

$$\begin{aligned} & \left(\frac{e_h^m - e_h^{m-1}}{\Delta t}, \varphi_h \right) + (\nabla e_h^m, \nabla \varphi_h) \\ &= \left(\frac{P_h u(t^m) - P_h u(t^{m-1})}{\Delta t}, \varphi_h \right) - \left(\partial_t u(t^m), \varphi_h \right) \\ &= \left(P_h \left(\frac{u(t^m) - u(t^{m-1})}{\Delta t} \right) - \partial_t u(t^m), \varphi_h \right) \\ &= \left((P_h - Id) \left(\frac{u(t^m) - u(t^{m-1})}{\Delta t} \right), \varphi_h \right) \\ & \quad + \left(\left(\frac{u(t^m) - u(t^{m-1})}{\Delta t} \right) - \partial_t u(t^m), \varphi_h \right). \end{aligned}$$

Wir wählen nun als Testfunktion $\varphi_h := e_h^m$, so folgt

$$\begin{aligned} & \left(\frac{e_h^m - e_h^{m-1}}{\Delta t}, e_h^m \right) + \|\nabla e_h^m\|^2 \\ & \leq \left(\|(P_h - Id) \left(\frac{u(t^m) - u(t^{m-1})}{\Delta t} \right)\| + \left\| \left(\frac{u(t^m) - u(t^{m-1})}{\Delta t} \right) - \partial_t u(t^m) \right\| \right) \|e_h^m\| \end{aligned}$$

Mit $\varepsilon_m := \varepsilon_{1,m} + \varepsilon_{2,m}$, $\varepsilon_{1,m} := \|(P_h - Id) \left(\frac{u(t^m) - u(t^{m-1})}{\Delta t} \right)\|$ und $\varepsilon_{2,m} := \left\| \left(\frac{u(t^m) - u(t^{m-1})}{\Delta t} \right) - \partial_t u(t^m) \right\|$ folgt also nach Multiplikation mit Δt

$$\|e_h^m\|^2 - (e_h^{m-1}, e_h^m) + \Delta t \|\nabla e_h^m\|^2 \leq \Delta t \varepsilon_m \|e_h^m\|.$$

Dies ergibt weiter

$$\|e_h^m\| \leq \|e_h^{m-1}\| + \Delta t \varepsilon_m.$$

Summation über m ergibt dann

$$\|e_h^m\| \leq \|e_h^0\| + \Delta t \sum_{j=1}^m \varepsilon_j.$$

Analog zum Beweis von Satz 3.13 erhalten wir zunächst für den Anfangsfehler

$$\begin{aligned} \|e_h^0\| &\leq \|u_h^0 - u_0\| + \|u_0 - P_h u_0\| \\ &\leq \|u_h^0 - u_0\| + ch^{k+1} \|u_0\|_{H^{k+1}(\Omega)}. \end{aligned}$$

Bleibt also noch die Abschätzung der Terme ε_m . Betrachten wir zunächst $\varepsilon_{1,m}$. Es gilt

$$\begin{aligned} \varepsilon_{1,m} &= \|(P_h - Id) \frac{1}{\Delta t} \int_{t^{m-1}}^{t^m} \partial_t u(s) ds\| \\ &= \frac{1}{\Delta t} \int_{t^{m-1}}^{t^m} \|(P_h - Id) \partial_t u(s)\| ds \\ &\leq c \frac{h^{k+1}}{\Delta t} \int_{t^{m-1}}^{t^m} \|\partial_t u(s)\|_{H^{k+1}(\Omega)} ds. \end{aligned}$$

Aufsummieren und multiplizieren mit Δt ergibt

$$\Delta t \sum_{j=1}^m \varepsilon_{1,j} \leq ch^{k+1} \int_0^{t^m} \|\partial_t u(s)\|_{H^{k+1}(\Omega)} ds.$$

Wenden wir uns nun dem Term $\varepsilon_{2,m}$ zu. Es ist

$$\begin{aligned} \left(\frac{u(t^m) - u(t^{m-1})}{\Delta t} \right) - \partial_t u(t^m) &= \frac{1}{\Delta t} \int_{t^{m-1}}^{t^m} \partial_t u(s) ds - \partial_t u(t^m) \\ &= \frac{1}{\Delta t} \int_{t^{m-1}}^{t^m} \left(\partial_t u(s) - \partial_t u(t^m) \right) ds \\ &= \frac{1}{\Delta t} \int_{t^{m-1}}^{t^m} \int_s^{t^m} \partial_{tt} u(r) dr ds. \end{aligned}$$

Aufsummieren und multiplizieren mit Δt ergibt für $\varepsilon_{2,m}$:

$$\Delta t \sum_{j=1}^m \varepsilon_{2,j} \leq \sum_{j=1}^m \int_{t^{j-1}}^{t^j} \int_s^{t^j} \|\partial_{tt} u(r)\| dr ds \leq \Delta t \int_0^{t^m} \|\partial_{tt} u(s)\| ds.$$

Fügen wir alle Abschätzungen zusammen, so haben wir die Behauptung bewiesen. \square

Bemerkung 3.29

Für die Zeitdiskretisierung haben wir bewußt die implizite Euler-Diskretisierung gewählt, da implizite Diskretisierungen besondere Stabilitätseigenschaften besitzen (A-Stabilität). Insbesondere zeigt die Fehlerabschätzung in Satz 3.18, dass die implizite Euler-Diskretisierung konvergiert, ohne dass wir eine Bedingung an die Zeitschrittweite fordern müssen.

Im folgenden wollen wir auch explizite Diskretisierungen betrachten. Wir definieren dazu als Verallgemeinerung der implizite Euler-Diskretisierung die Klasse der θ -Verfahren.

Definition 3.30 (θ -Verfahren)

Es gelten die Notationen und Bezeichnungen aus Definition 3.16. Für $\theta \in [0, 1]$ definieren wir das θ -Verfahren durch die Iterationsvorschrift

$$\begin{aligned} u_h^0 &= u_h(0), \\ \left(\frac{u_h^{m+1} - u_h^m}{\Delta t}, \varphi_h \right) &+ \theta(\nabla u_h^{m+1}, \nabla \varphi_h) + (1 - \theta)(\nabla u_h^m, \nabla \varphi_h) \quad (3.10) \\ &= \langle \theta f(\cdot, t^{m+1}) + (1 - \theta) f(\cdot, t^m), \varphi_h \rangle, \quad \forall \varphi_h \in S_{h,0}^k. \end{aligned}$$

Dabei bezeichnet (\cdot, \cdot) das Skalarprodukt in $L^2(\Omega)$.

Bemerkung 3.31

Mit den Bezeichnungen aus Bemerkung 3.17 erhalten wir als lineares Gleichungssystem für den Koeffizientenvektor U^{m+1} von u_h^{m+1}

$$(M + \Delta t \theta S)U^{m+1} = (M - \Delta t(1 - \theta)S)U^m + \Delta t(\theta F^{m+1} + (1 - \theta)F^m).$$

Die Matrix $A := M + \Delta t \theta S$ ist symmetrisch und positiv definit, so dass die Existenz einer eindeutigen Lösung des Gleichungssystems sichergestellt ist.

Als Spezialfälle erhalten wir

- für $\theta = 1$ das implizite Euler Verfahren,
- für $\theta = 0$ das explizite Euler Verfahren,

- für $\theta = 1/2$ das Crank-Nicolson-Verfahren.

Das Crank-Nicolson-Verfahren ist ein Verfahren 2. Ordnung, während alle anderen Verfahren von 1. Ordnung in der Zeit sind.

Auch für das θ -Verfahren kann man versuchen analog zur Vorgehensweise für das implizite Eulerverfahren direkt a priori Fehlerabschätzungen zu zeigen. Statt diesen Weg zu gehen, wollen wir analog zur Vorgehensweise bei Anfangswertproblemen, zunächst die Stabilität der Verfahren untersuchen. Auch hier wird dann aus Konsistenz und Stabilität die Konvergenz folgen.

3.5 Stabilität und Konvergenz

Definition 3.32 (Stabilität)

Es gelten die Bezeichnungen aus Abschnitt 3.4. Ein numerisches Zeitschritt-Verfahren zur Diskretisierung der Wärmeleitungsgleichung heißt stabil (bzgl. der $L^\infty(0, T; L^2(\Omega))$ -Norm), falls es eine Konstante C gibt, so dass für die Approximationen $u_h^m \in X_h$ von $u(\cdot, t^m)$ gilt

$$\|u_h^m\|_{L^2(\Omega)} \leq C, \quad \text{uniform in } h \text{ und } m.$$

Entsprechend heißt das Verfahren stabil bezüglich einer anderen Norm $\|\cdot\|_X$, falls eine uniforme Abschätzung der Approximationen in dieser Norm gilt.

Satz 3.33 (Stabilität des θ -Verfahrens)

Seien $u_h^m \in S_{h,0}^k$, $m = 1, \dots, M$ die Approximationen des θ -Verfahrens aus Definition 3.20, wobei das zugrundeliegende Gitter \mathcal{T}_h regulär sei, d.h. $\sigma(T) \leq \sigma$, $\forall T \in \mathcal{T}_h$. Gelte weiter $f \in L^\infty(0, T; L^2(\Omega))$. Ist $\theta \in [0, \frac{1}{2}]$, so gelte zusätzlich die Bedingung

$$(1 - \theta)\Delta t \leq \frac{h_{\min}^2}{c},$$

wobei $h_{\min} := \min_{T \in \mathcal{T}_h} h(T)$ und c die optimale Konstante der folgenden inversen Ungleichung sei

$$\|\nabla v_h\|_{L^2(\Omega)} \leq \frac{c}{h_{\min}} \|v_h\|_{L^2(\Omega)}, \quad \forall v_h \in S_{h,0}^k.$$

Dann ist das θ -Verfahren stabil und es gilt die Abschätzung

$$\|u_h^m\|_{L^2(\Omega)} \leq \|u_h^0\|_{L^2(\Omega)} + C \max_{t \in [0, T]} \|f(\cdot, t)\|_{L^2(\Omega)}.$$

mit einer Konstanten $C := \sqrt{\frac{c^2 T}{2}}$.

Beweis: Wir wählen zunächst $\varphi_h = \theta u_h^{m+1} + (1 - \theta)u_h^m$ in der Definition des θ -Verfahrens und erhalten

$$T_1 + T_2 = T_3$$

mit

$$\begin{aligned} T_1 &:= \left(\frac{u_h^{m+1} - u_h^m}{\Delta t}, \theta u_h^{m+1} + (1 - \theta) u_h^m \right), \\ T_2 &:= \|\nabla(\theta u_h^{m+1} + (1 - \theta) u_h^m)\|^2, \\ T_3 &:= (\theta f(\cdot, t^{m+1}) + (1 - \theta) f(\cdot, t^m), \theta u_h^{m+1} + (1 - \theta) u_h^m). \end{aligned}$$

Betrachten wir zunächst den Term T_1 :

$$\begin{aligned} T_1 &= \frac{1}{\Delta t} \left(\theta \|u_h^{m+1} - u_h^m\|^2 + (u_h^{m+1} - u_h^m, u_h^m) \right) \\ &= \frac{1}{\Delta t} \left(\theta \|u_h^{m+1} - u_h^m\|^2 - \frac{1}{2} \|u_h^{m+1} - u_h^m\|^2 + \frac{1}{2} \|u_h^{m+1}\|^2 - \frac{1}{2} \|u_h^m\|^2 \right) \\ &= \frac{(\theta - \frac{1}{2})}{\Delta t} \|u_h^{m+1} - u_h^m\|^2 + \frac{1}{2\Delta t} \left(\|u_h^{m+1}\|^2 - \|u_h^m\|^2 \right). \end{aligned}$$

Von T_2 wollen wir lediglich die Positivität nutzen. Betrachten wir also noch T_3 :

$$\begin{aligned} T_3 &\leq \|\theta f(\cdot, t^{m+1}) + (1 - \theta) f(\cdot, t^m)\| \|\theta u_h^{m+1} + (1 - \theta) u_h^m\| \\ &\leq \|\theta f(\cdot, t^{m+1}) + (1 - \theta) f(\cdot, t^m)\| c_p \|\nabla(\theta u_h^{m+1} + (1 - \theta) u_h^m)\| \\ &\leq \frac{c_p^2}{4\varepsilon} \|\theta f(\cdot, t^{m+1}) + (1 - \theta) f(\cdot, t^m)\|^2 + \varepsilon \|\nabla(\theta u_h^{m+1} + (1 - \theta) u_h^m)\|^2. \end{aligned}$$

Dabei haben wir im letzten Schritt die Youngsche Ungleichung mit einem $\varepsilon > 0$ verwendet, das noch frei gewählt werden kann. Zusammen erhalten wir also nach Multiplikation mit $2\Delta t$

$$\begin{aligned} \|u_h^{m+1}\|^2 - \|u_h^m\|^2 + 2(\theta - \frac{1}{2}) \|u_h^{m+1} - u_h^m\|^2 + 2\Delta t(1 - \varepsilon) \|\nabla(\theta u_h^{m+1} + (1 - \theta) u_h^m)\|^2 \\ \leq \frac{c_p^2 \Delta t}{2\varepsilon} \|\theta f(\cdot, t^{m+1}) + (1 - \theta) f(\cdot, t^m)\|^2. \end{aligned}$$

Der Koeffizient $\theta - \frac{1}{2}$ ist nur dann größer oder gleich Null, wenn gilt $\theta \geq \frac{1}{2}$.
Betrachten wir daher zunächst diesen Fall, so erhalten wir mit $\varepsilon = 1$

$$\|u_h^{m+1}\|^2 - \|u_h^m\|^2 \leq \frac{c_p^2 \Delta t}{2} \|\theta f(\cdot, t^{m+1}) + (1 - \theta) f(\cdot, t^m)\|^2.$$

Summation über m ergibt also

$$\begin{aligned} \|u_h^m\|^2 &\leq \|u_h^0\|^2 + \frac{c_p^2}{2} \sum_{j=0}^{m-1} \Delta t \|\theta f(\cdot, t^{m+1}) + (1 - \theta) f(\cdot, t^m)\|^2 \\ &\leq \|u_h^0\|^2 + \frac{c_p^2 T}{2} \max_{t \in [0, T]} \|f(\cdot, t)\|^2. \end{aligned}$$

Somit haben wir für $\theta \geq \frac{1}{2}$ die Behauptung bewiesen mit einer Konstanten $C := \sqrt{\frac{c_p^2 T}{2}}$.

Für den Fall $\theta < \frac{1}{2}$ kommen wir mit der bisherigen Vorgehensweise nicht weiter. Wir gehen hier erneut von der Definition des θ -Verfahrens aus und testen diesmal mit $\varphi = u_h^{m+1}$. Wir erhalten nach Multiplikation mit Δt

$$(u_h^{m+1} - u_h^m, u_h^m) + \Delta t \theta \|\nabla u_h^{m+1}\|^2 + \Delta t (1 - \theta) (\nabla u_h^m, \nabla u_h^{m+1}) \\ = \Delta t (\theta f(\cdot, t^{m+1}) + (1 - \theta) f(\cdot, t^m), u_h^{m+1}).$$

mit $(a - b)a = \frac{1}{2}(a^2 + (a - b)^2 - b^2)$ und $ab = -\frac{1}{2}((a - b)^2 - a^2 - b^2)$ folgt hieraus

$$\frac{1}{2} \left(\|u_h^{m+1}\|^2 + \|u_h^{m+1} - u_h^m\|^2 - \|u_h^m\|^2 \right) + \Delta t \theta \|\nabla u_h^{m+1}\|^2 \\ - \frac{\Delta t (1 - \theta)}{2} \left(\|\nabla (u_h^{m+1} - u_h^m)\|^2 - \|\nabla u_h^{m+1}\|^2 - \|\nabla u_h^m\|^2 \right) \\ \leq \Delta t \|\theta f(\cdot, t^{m+1}) + (1 - \theta) f(\cdot, t^m)\| \|u_h^{m+1}\|.$$

Weiter erhalten wir nach Multiplikation mit 2

$$(\|u_h^{m+1}\|^2 - \|u_h^m\|^2) + \|u_h^{m+1} - u_h^m\|^2 + \Delta t (1 + \theta) \|\nabla u_h^{m+1}\|^2 \\ + \Delta t (1 - \theta) \|\nabla u_h^m\|^2 - \Delta t (1 - \theta) \|\nabla (u_h^{m+1} - u_h^m)\|^2 \\ \leq \frac{\Delta t c_p^2}{2\varepsilon} \|\theta f(\cdot, t^{m+1}) + (1 - \theta) f(\cdot, t^m)\|^2 + \Delta t \varepsilon \|\nabla u_h^{m+1}\|^2.$$

Mit der inversen Ungleichung erhalten wir weiter

$$(\|u_h^{m+1}\|^2 - \|u_h^m\|^2) + \|u_h^{m+1} - u_h^m\|^2 + \Delta t (1 + \theta - \varepsilon) \|\nabla u_h^{m+1}\|^2 \\ + \Delta t (1 - \theta) \|\nabla u_h^m\|^2 - \frac{\Delta t (1 - \theta) c^2}{h_{\min}^2} \|(u_h^{m+1} - u_h^m)\|^2 \\ \leq \frac{\Delta t c_p^2}{2\varepsilon} \|\theta f(\cdot, t^{m+1}) + (1 - \theta) f(\cdot, t^m)\|^2.$$

Mit $\varepsilon = 1 + \theta$ folgt also

$$(\|u_h^{m+1}\|^2 - \|u_h^m\|^2) + \Delta t (1 - \theta) \|\nabla u_h^m\|^2 + \left(1 - \frac{\Delta t (1 - \theta) c^2}{h_{\min}^2} \right) \|(u_h^{m+1} - u_h^m)\|^2 \\ \leq \frac{\Delta t c_p^2}{2(1 + \theta)} \|\theta f(\cdot, t^{m+1}) + (1 - \theta) f(\cdot, t^m)\|^2.$$

Mit den Annahmen an θ und Δt in diesem Fall sind der zweite und dritte Summand auf der linken Seite positiv und wir erhalten analog zum vorherigen

Fall durch aufsummieren über m

$$\begin{aligned} \|u_h^m\|^2 &\leq \|u_h^0\|^2 + \frac{c_p^2}{2(1+\theta)} \sum_{j=0}^{m-1} \Delta t \|\theta f(\cdot, t^{m+1}) + (1-\theta)f(\cdot, t^m)\|^2 \\ &\leq \|u_h^0\|^2 + \frac{c_p^2 T}{2(1+\theta)} \max_{t \in [0, T]} \|f(\cdot, t)\|^2. \end{aligned}$$

Somit haben wir auch für $\theta < \frac{1}{2}$ die Behauptung bewiesen. \square

Der Stabilitätssatz 3.32 zeigt, dass das θ -Verfahren für $\theta \in [\frac{1}{2}, 1]$ unbedingt stabil ist, während für $\theta \in [0, \frac{1}{2})$ eine Bedingung an die Zeitschrittweite gestellt werden muß. Nach der Stabilität wollen wir nun im nächsten Satz die Konvergenz des θ -Verfahrens untersuchen.

Satz 3.34 (A priori Fehlerabschätzung für das θ -Verfahren)

Seien u eine genügend glatte Lösung der inhomogenen Wärmeleitungsgleichung aus Definition 3.7 und $u_h^m \in S_{h,0}^k$, $m = 1, \dots, M$ die Approximationen des θ -Verfahrens aus Definition 3.20, wobei das zugrundeliegende Gitter \mathcal{T}_h regulär sei, d.h. $\sigma(T) \leq \sigma$, $\forall T \in \mathcal{T}_h$. Zusätzlich existiere ein Konstante $c_* > 0$, so dass für alle $g \in L^2(\Omega)$ ein schwache Lösung $v_g \in H_0^1(\Omega)$ des dualen Poissonproblems

$$\int_{\Omega} \nabla \varphi \cdot \nabla v_g = \int_{\Omega} g \varphi, \quad \forall \varphi \in H_0^1(\Omega)$$

existiert, die H^2 -regulär ist ($v_g \in H^2(\Omega)$) und die folgende uniforme a priori Abschätzung erfüllt

$$\|v_g\|_{H^2(\Omega)} \leq c_* \|g\|_{L^2(\Omega)}.$$

Ist $\theta \in [0, \frac{1}{2})$, so gelte zusätzlich die Stabilitätsbedingung aus Satz 3.23, d.h.

$$(1-\theta)\Delta t \leq \frac{h_{\min}^2}{c}.$$

Dann gilt für $m = 0, \dots, M$ die Fehlerabschätzung

$$\begin{aligned} \|u_h^m - u(t^m)\|_{L^2(\Omega)} &\leq \|u_h^0 - u_0\|_{L^2(\Omega)} + \frac{|2\theta - 1|}{2} \Delta t \max_{j=1}^m \|\partial_{tt} u(t^j)\|_{L^2(\Omega)} \\ &\quad + \frac{2(1-\theta) + \theta}{2} \Delta t^2 \max_{j=0}^{m-1} \frac{1}{\Delta t} \int_{t^j}^{t^{j+1}} \|\partial_t^3 u(s)\| ds \\ &\quad + Ch^{k+1} \left(\|u_0\|_{H^{k+1}(\Omega)} + \max_{j=0}^{m-1} \frac{1}{\Delta t} \int_{t^j}^{t^{j+1}} \|\partial_t u(s)\|_{H^{k+1}(\Omega)} ds \right). \end{aligned}$$

Beweis: Wie im Beweis der Fehlerabschätzung für die implizite Euler Diskretisierung machen wir auch hier von der Galerkin-Projektion Gebrauch. Sei also $P_h u(t) \in S_{h,0}^k$ definiert durch

$$(\nabla P_h u(t), \nabla \varphi_h) = (\nabla u(t), \nabla \varphi_h), \quad \forall \varphi_h \in S_{h,0}^k, t \in [0, T].$$

Definieren den Fehler $e_h^m := u_h^m - P_h u(t^m)$, so erhalten wir zusammen mit der Definition des θ -Verfahrens

$$\frac{1}{\Delta t} (e_h^{m+1} - e_h^m, \varphi_h) + (\nabla(\theta e_h^{m+1} + (1-\theta)e_h^m), \nabla \varphi_h) = (\varepsilon_h^m, \varphi_h), \quad \forall \varphi_h \in S_{h,0}^k,$$

wobei ε_h^m definiert ist durch

$$\begin{aligned} (\varepsilon_h^m, \varphi_h) &= (\theta f(\cdot, t^{m+1}) + (1-\theta)f(\cdot, t^m), \varphi_h) \\ &\quad - \frac{1}{\Delta t} P_h(u(t^{m+1}) - u(t^m)), \varphi_h \\ &\quad - (\nabla(\theta u(t^{m+1}) + (1-\theta)u(t^m)), \nabla \varphi_h), \quad \forall \varphi_h \in S_{h,0}^k. \end{aligned}$$

Verwenden wir die schwache Form der Wärmeleitungsgleichung für u , so gilt

$$\begin{aligned} (\theta f(\cdot, t^{m+1}) + (1-\theta)f(\cdot, t^m), \varphi_h) - (\nabla(\theta u(t^{m+1}) + (1-\theta)u(t^m)), \nabla \varphi_h) \\ = (\theta \partial_t u(t^{m+1}) + (1-\theta) \partial_t u(t^m), \varphi_h) \end{aligned}$$

und wir erhalten

$$\begin{aligned} (\varepsilon_h^m, \varphi_h) &= (\theta \partial_t u(t^{m+1}) + (1-\theta) \partial_t u(t^m), \varphi_h) \\ &\quad - \frac{1}{\Delta t} P_h(u(t^{m+1}) - u(t^m)), \varphi_h. \end{aligned}$$

Insbesondere folgt also

$$\varepsilon_h^m = \theta \partial_t u(t^{m+1}) + (1-\theta) \partial_t u(t^m) - \frac{1}{\Delta t} P_h(u(t^{m+1}) - u(t^m)).$$

Der Fehler e_h^m genügt also einem θ -Verfahren mit rechter Seite ε_h^m anstelle von $\theta f(\cdot, t^{m+1}) + (1-\theta)f(\cdot, t^m)$. Daher folgt analog zum Beweis des Stabilitätsresultats aus Satz 3.23

$$\|e_h^m\| \leq \|e_h^0\| + C \max_{j=0}^m \|\varepsilon_h^j\|.$$

Zur Abschätzung von $\|e_h^0\|_{L^2(\Omega)}$ gehen wir analog zum Beweis von Satz 3.18 vor und erhalten

$$\|e_h^0\| \leq \|u_h^0 - u_0\| + Ch^{k+1} \|u_0\|_{H^{k+1}(\Omega)}.$$

Bleibt also noch eine Anschätzung für $\|\varepsilon_h^m\|$ zu beweisen, wobei wir analog zum Beweis von Satz 3.18 vorgehen können. Zunächst erhalten wir

$$\begin{aligned}\varepsilon_h^m &= \theta \partial_t u(t^{m+1}) + (1 - \theta) \partial_t u(t^m) - \frac{1}{\Delta t} (u(t^{m+1}) - u(t^m)) \\ &\quad + \frac{1}{\Delta t} (Id - P_h) (u(t^{m+1}) - u(t^m)).\end{aligned}$$

Der letzte Summand auf der rechten Seite kann jetzt analog zum Beweis von Satz 3.18 abgeschätzt werden durch

$$\frac{1}{\Delta t} \int_{t^m}^{t^{m+1}} \|(Id - P_h) \partial_t u(s)\| ds \leq ch^{k+1} \frac{1}{\Delta t} \int_{t^{m-1}}^{t^m} \|\partial_t u(s)\|_{H^{k+1}(\Omega)} ds.$$

Betrachten wir also noch die verbleibenden Terme. Mit Taylorentwicklung für $u(t)$ folgt

$$\begin{aligned}u(t^{m+1}) &= u(t^m) + \Delta t \partial_t u(t^m) + \frac{1}{2} \Delta t^2 \partial_t^2 u(t^m) + \frac{1}{2} \int_{t^m}^{t^{m+1}} (t^{m+1} - s)^2 \partial_t^3 u(s) d\sigma(x), \\ u(t^m) &= u(t^{m+1}) + \Delta t \partial_t u(t^{m+1}) + \frac{1}{2} \Delta t^2 \partial_t^2 u(t^{m+1}) - \frac{1}{2} \int_{t^m}^{t^{m+1}} (t^m - s)^2 (\partial_t^3 u(s)) d\sigma(x).\end{aligned}$$

Aus diesen beiden Entwicklungen erhalten wir

$$\begin{aligned}\frac{1}{\Delta t} (u(t^{m+1}) - u(t^m)) &= \partial_t u(t^m) + \frac{1}{2} \Delta t \partial_t^2 u(t^m) + \frac{1}{2\Delta t} \int_{t^m}^{t^{m+1}} (t^{m+1} - s)^2 \partial_t^3 u(s) d\sigma(x), \\ \frac{1}{\Delta t} (u(t^{m+1}) - u(t^m)) &= \partial_t u(t^{m+1}) - \frac{1}{2} \Delta t \partial_t^2 u(t^{m+1}) + \frac{1}{2\Delta t} \int_{t^m}^{t^{m+1}} (t^m - s)^2 \partial_t^3 u(s) d\sigma(x).\end{aligned}$$

Multiplizieren wir die erste Gleichung mit $(1 - \theta)$, die zweite mit θ und addieren die resultierenden Gleichungen, so erhalten wir

$$\begin{aligned}\varepsilon_{h,1}^m &:= \theta \partial_t u(t^{m+1}) + (1 - \theta) \partial_t u(t^m) - \frac{1}{\Delta t} (u(t^{m+1}) - u(t^m)) \\ &= \frac{\Delta t}{2} (\theta \partial_t^2 u(t^{m+1}) - (1 - \theta) \partial_t^2 u(t^m)) - \frac{\theta}{2\Delta t} \int_{t^m}^{t^{m+1}} (t^m - s)^2 \partial_t^3 u(s) \\ &\quad - \frac{1 - \theta}{2\Delta t} \int_{t^m}^{t^{m+1}} (t^{m+1} - s)^2 \partial_t^3 u(s).\end{aligned}$$

Werwenden wir nun noch die Darstellung

$$\partial_t^2 u(t^m) = \partial_t^2 u(t^{m+1}) - \int_{t^m}^{t^{m+1}} \partial_t^3 u(s) ds,$$

so erhalten wir

$$\begin{aligned}\varepsilon_{h,1}^m &= \frac{\Delta t}{2}(2\theta - 1)\partial_t^2 u(t^{m+1}) + \frac{\Delta t}{2}(1 - \theta) \int_{t^m}^{t^{m+1}} \partial_t^3 u(s) ds \\ &\quad - \frac{\theta}{2\Delta t} \int_{t^m}^{t^{m+1}} (t^m - s)^2 \partial_t^3 u(s) ds - \frac{1 - \theta}{2\Delta t} \int_{t^m}^{t^{m+1}} (t^{m+1} - s)^2 \partial_t^3 u(s) ds.\end{aligned}$$

Also erhalten wir als Abschätzung in der Norm

$$\|\varepsilon_{h,1}^m\| \leq \frac{|2\theta - 1|}{2} \Delta t \|\partial_t^2(t^{m+1})\| + \frac{2(1 - \theta) + \theta}{2} \Delta t^2 \frac{1}{\Delta t} \int_{t^m}^{t^{m+1}} \|\partial_t^3 u(s)\| ds.$$

Damit folgt die Behauptung mit Hilfe der Dreiecksungleichung

$$\|u_h^m - u(t^m)\| \leq \|P_h u(t^m) - u(t^m)\| + \|e_h^m\|$$

analog zum Beweis von Satz 3.18. \square .

Bemerkung 3.35

Zunächst halten wir fest, dass die Fehlerabschätzung in Satz 3.24 zeigt, dass das θ -Verfahren von erster Ordnung in der Zeit ist und lediglich in dem Spezialfall $\theta = \frac{1}{2}$ zu einem Verfahren zweiter Ordnung wird. Allerdings bleibt auch festzuhalten, dass der Faktor $|2\theta - 1|$ in der Fehlerabschätzung klein wird, wenn θ nahe an $\frac{1}{2}$ gewählt wird.

Weiter wollen wir festhalten, dass der Beweis von Satz 3.24 auch zeigt, dass wir die Frage der Konvergenz und Fehlerabschätzung beantworten können, indem wir getrennt die Stabilität und Konsistenz des Verfahrens untersuchen.

3.6 A posteriori Fehlerabschätzung

Im letzten Abschnitt der Vorlesung wollen wir uns noch mit a posteriori Fehlerabschätzungen für die Wärmeleitungsgleichung beschäftigen. Das Ziel ist also eine Abschätzung der Form

$$\|u(t^m) - u_h^m\| \leq c\eta(u_h^m)$$

herzuleiten. Dabei werden wir uns auf die Betrachtung der impliziten Euler Diskretisierung mit linearen Finiten Elementen beschränken. Zunächst werden wir dazu eine kontinuierliche Gleichung für den Fehler herleiten.

Lemma 3.36 (Fehlergleichung)

Seien u eine genügend glatte Lösung der inhomogenen Wärmeleitungsgleichung aus Definition 3.7 und $u_h^m \in S_{h,0}^k$, $m = 0, \dots, M$ die impliziten Euler Diskretisierung aus Definition 3.16. Wir definieren zu den Funktionswerten $(u_h^m)_{m=0, \dots, M}$ eine in der Zeit stückweise konstante Funktion $\bar{u}_h^M : [0, T] \rightarrow S_{h,0}^k$ durch

$$\bar{u}_h^M|_{I_m} := u_h^{m+1}, \quad \text{für } m = 0, \dots, M-1.$$

und eine stetige und stückweise lineare Funktion $\tilde{u}_h^M : [0, T] \rightarrow S_{h,0}^k$ durch

$$\tilde{u}_h^M(t)|_{I_m} := \frac{t^{m+1} - t}{\Delta t} u_h^m + \frac{t - t^m}{\Delta t} u_h^{m+1}, \quad \text{für } t \in I_m \text{ und } m = 0, \dots, M-1.$$

Mit dieser Notation definieren wir die Fehlerfunktionen \tilde{e}, \bar{e} durch

$$\tilde{e} := u - \tilde{u}_h^M, \quad \bar{e} := u - \bar{u}_h^M$$

Dann gilt für genügend glatte Testfunktionen $\varphi : [0, T] \times \Omega \rightarrow \mathbb{R}$, d.h. $\varphi \in L^2(0, T; H_0^1(\Omega))$ mit schwacher Ableitung $\partial_t \varphi \in L^2(0, T; H^{-1}(\Omega))$, die Fehleridentität

$$\begin{aligned} \frac{d}{dt}(\tilde{e}(t), \varphi(t)) &= -(\nabla \tilde{e}(t), \nabla \varphi(t)) + \langle \tilde{e}, \partial_t \varphi(t) \rangle + \langle (f - \bar{f})(t), \varphi(t) \rangle \\ &\quad - (\nabla(\tilde{u}_h^M(t) - \bar{u}_h^M(t)), \nabla \varphi(t)) - \langle R(u_h^M)(t), \varphi(t) - v_h(t) \rangle, \end{aligned}$$

für beliebige $v_h(t) \in S_{h,0}^k$, wobei das Residuum $R(u_h^M)$ definiert ist durch

$$\langle R(u_h^M)(t), \varphi(t) \rangle := \langle \partial_t \tilde{u}_h^M(t), \varphi(t) \rangle + (\nabla \bar{u}_h^M(t), \nabla \varphi(t)) - \langle \bar{f}(t), \varphi(t) \rangle$$

und $\bar{f} : [0, T] \rightarrow L^2(\Omega)$ ist definiert durch $\bar{f}(t)|_{I_m} := f(t^{m+1}) \forall t \in I_m$.

Beweis: Es gilt

$$\begin{aligned} \frac{d}{dt}(\tilde{e}(t), \varphi(t)) &= \langle \partial_t \tilde{e}, \varphi(t) \rangle + \langle \tilde{e}, \partial_t \varphi(t) \rangle \\ &= \langle f(t), \varphi(t) \rangle - (\nabla \bar{e}(t), \nabla \varphi(t)) - (\nabla \bar{u}_h^M(t), \nabla \varphi(t)) \\ &\quad - \langle \partial_t \tilde{u}_h^M(t), \varphi(t) \rangle + \langle \tilde{e}, \partial_t \varphi(t) \rangle \\ &= -(\nabla \bar{e}(t), \nabla \varphi(t)) + \langle \tilde{e}, \partial_t \varphi(t) \rangle - \langle R(u_h^M)(t), \varphi(t) \rangle \\ &\quad + \langle (f - \bar{f})(t), \varphi(t) \rangle \\ &= -(\nabla \tilde{e}(t), \nabla \varphi(t)) - (\nabla (\tilde{u}_h^M(t) - \bar{u}_h^M(t))) \\ &\quad + \langle \tilde{e}, \partial_t \varphi(t) \rangle - \langle R(u_h^M)(t), \varphi(t) - v_h(t) \rangle \\ &\quad + \langle (f - \bar{f})(t), \varphi(t) \rangle, \end{aligned}$$

da aufgrund der Definition des impliziten Euler Verfahrens gilt $\langle R(u_h^M)(t), v_h(t) \rangle = 0$, für alle $v_h \in S_{h,0}^k$. \square

Bemerkung 3.37

Ausgehend von der Fehlergleichung in Lemma 3.26, kann man versuchen a posteriori Fehlerabschätzungen zu beweisen. Dazu kann man einerseits versuchen $\varphi(t) = \tilde{e}(t)$ zu wählen, andererseits kann φ allerdings auch als Lösung eines geeigneten dualen Problems gewählt werden. Um zu einer Fehlerabschätzung zu gelangen, muß dann noch $v_h(t)$ gewählt werden, z.B. als H_0^1 -Interpolation von $\varphi(t)$. Im folgenden Satz werden wir zunächst den ersten Ansatz wählen (vgl. hierzu auch [CF04]).

Satz 3.38 (A posteriori Fehlerabschätzung)

Es gelten die Notationen und Voraussetzungen aus Lemma 3.26. Sei $f, \partial_t u \in L^2(0, T; L^2(\Omega))$ und $k = 1$, dann gilt für den Fehler $\tilde{e}(t) := u(t) - \tilde{u}_h^M(t)$,

$t \in [0, T]$ die a posteriori Fehlerabschätzung

$$\begin{aligned} \|u(t^m) - u_h^m\|_{L^2(\Omega)}^2 + \int_0^{t^m} \|\nabla(u(s) - u_h^M(s))\|_{L^2(\Omega)}^2 ds &\leq \|u_0 - u_h^0\|_{L^2(\Omega)}^2 \\ &+ 4c_p \int_0^{t^m} \|(f - \bar{f})(s)\|^2 ds + \sum_{j=0}^{m-1} \Delta t \left(\eta_{\text{time}}^j + C \eta_{\text{element}}^j + C \eta_{\text{jump}}^j \right). \end{aligned}$$

Dabei hängt die Konstante C nur von der Regularität der Triangulierung ab und die Indikatoren η_{time}^j , η_{element}^j , und η_{jump}^j sind definiert durch

$$\begin{aligned} \eta_{\text{time}}^j &:= \frac{2}{3} \|\nabla(u_h^{j+1} - u_h^j)\|^2, \\ \eta_{\text{element}}^j &:= \sum_{T \in \mathcal{T}_h} \left\| h(T) \left(\frac{u_h^{j+1} - \bar{u}_h^j}{\Delta t} - f(t) \right) \right\|_{L^2(T)}^2, \\ \eta_{\text{jump}}^j &:= \sum_{S \in \mathcal{E}(\mathcal{T}_h)} \left\| h(S)^{1/2} [\nabla u_h^{j+1}(t) \cdot n] \right\|_{L^2(S)}^2, \end{aligned}$$

wobei $\mathcal{E}(\mathcal{T}_h)$ die Menge der Seitenflächen von \mathcal{T}_h bezeichnet.

Beweis: Wir wählen $\varphi(t) := \tilde{e}(t)$ in der Fehlergleichung aus Lemma 3.26. Mit

$$(\tilde{e}(t), \partial_t e(t)) = \frac{1}{2} \frac{d}{dt} \|\tilde{e}(t)\|^2$$

folgt dann nach Multiplikation mit 2

$$\begin{aligned} \frac{d}{dt} \|\tilde{e}(t)\|^2 + 2 \|\nabla \tilde{e}(t)\|^2 &= 2(\nabla(\tilde{u}_h^M(t) - \bar{u}_h^M(t)), \nabla \tilde{e}(t)) - 2\langle R(u_h^M)(t), \tilde{e}(t) - v_h(t) \rangle \\ &+ 2\langle (f - \bar{f})(t), \tilde{e}(t) \rangle. \end{aligned}$$

Mit noch zu wählenden Parametern α, β folgt dann

$$\begin{aligned}
\frac{d}{dt} \|\tilde{e}(t)\|^2 + 2\|\nabla \tilde{e}(t)\|^2 &= 2 \left((\nabla(\tilde{u}_h^M(t) - \bar{u}_h^M(t)), \nabla \tilde{e}(t)) \right. \\
&\quad \left. - (\partial_t \tilde{u}_h^M(t) - f(t), \tilde{e}(t) - v_h(t)) - (\nabla \bar{u}_h^M(t), \nabla(\tilde{e}(t) - v_h(t))) \right) \\
&\quad + 2\langle (f - \bar{f})(t), \tilde{e}(t) \rangle \\
&= 2 \left((\nabla(\tilde{u}_h^M(t) - \bar{u}_h^M(t)), \nabla \tilde{e}(t)) + \langle (f - \bar{f})(t), \tilde{e}(t) \rangle \right. \\
&\quad \left. - \sum_{T \in \mathcal{T}_h} (h(T)^\alpha (\partial_t \tilde{u}_h^M(t) - f(t)), h(T)^{-\alpha} (\tilde{e}(t) - v_h(t)))_{L^2(T)} \right. \\
&\quad \left. - \sum_{T \in \mathcal{T}_h} (h(T)^\beta \nabla \bar{u}_h^M(t), h(T)^{-\beta} \nabla(\tilde{e}(t) - v_h(t)))_{L^2(T)} \right).
\end{aligned}$$

Mit partieller Integration, $k = 1$ und Cauchy-Schwarz folgt nun die Abschätzung

$$\begin{aligned}
&\frac{d}{dt} \|\tilde{e}(t)\|^2 + 2\|\nabla \tilde{e}(t)\|^2 \\
&\leq 2\|\nabla(\tilde{u}_h^M(t) - \bar{u}_h^M(t))\| \|\nabla \tilde{e}(t)\| + 2\|(f - \bar{f})(t)\| \|\tilde{e}(t)\| \\
&\quad + 2 \sum_{T \in \mathcal{T}_h} \|h(T)^\alpha (\partial_t \tilde{u}_h^M(t) - f(t))\|_{L^2(T)} \|h(T)^{-\alpha} (\tilde{e}(t) - v_h(t))\|_{L^2(T)} \\
&\quad + 2 \sum_{S \in \mathcal{E}(\mathcal{T}_h)} \|h(S)^\beta [\nabla \bar{u}_h^M(t) \cdot n]\|_{L^2(S)} \|h(T)^{-\beta} \nabla(\tilde{e}(t) - v_h(t))\|_{L^2(S)}
\end{aligned}$$

Wählen wir nun $v_h(t) = I_h(\tilde{e}(t))$, wobei I_h die H_0^1 -Interpolation ist, so folgt mit $\alpha = 1$ und $\beta = \frac{1}{2}$ (vgl. Beweis von Satz 2.62)

$$\begin{aligned}
\|h(T)^{-1}(\tilde{e}(t) - I_h(\tilde{e}(t)))\|_{L^2(T)} &\leq c \|\nabla \tilde{e}(t)\|_{L^2(T)}, \\
\|h(T)^{-\frac{1}{2}} \nabla(\tilde{e}(t) - v_h(t))\|_{L^2(S)} &\leq c \|\nabla \tilde{e}(t)\|_{L^2(\omega(T))}
\end{aligned}$$

die Abschätzung

$$\begin{aligned}
&\frac{d}{dt} \|\tilde{e}(t)\|^2 + 2\|\nabla \tilde{e}(t)\|^2 \\
&\leq 2\|\nabla(\tilde{u}_h^M(t) - \bar{u}_h^M(t))\| \|\nabla \tilde{e}(t)\| + 2\|(f - \bar{f})(t)\| \|\tilde{e}(t)\| \\
&\quad + C \left[\left(\sum_{T \in \mathcal{T}_h} \|h(T)(\partial_t \tilde{u}_h^M(t) - f(t))\|_{L^2(T)}^2 \right)^{1/2} \right. \\
&\quad \left. + \left(\sum_{S \in \mathcal{E}(\mathcal{T}_h)} \|h(S)^{1/2} [\nabla \bar{u}_h^M(t) \cdot n]\|_{L^2(S)}^2 \right)^{1/2} \right] \|\nabla \tilde{e}(t)\|.
\end{aligned}$$

Mit der Cauchy-Schwarzschen Ungleichung und der Youngschen Ungleichung folgt

$$\begin{aligned}
& \frac{d}{dt} \|\tilde{e}(t)\|^2 + 2 \|\nabla \tilde{e}(t)\|^2 \\
& \leq 2 \|\nabla(\tilde{u}_h^M(t) - \bar{u}_h^M(t))\|^2 + \frac{1}{2} \|\nabla \tilde{e}(t)\|^2 \\
& \quad + 4c_p \|(f - \bar{f}(t))\|^2 + \frac{1}{4} \|\nabla \tilde{e}(t)\|^2 \\
& \quad + C \left[\left(\sum_{T \in \mathcal{T}_h} \|h(T)(\partial_t \tilde{u}_h^M(t) - f(t))\|_{L^2(T)}^2 \right) \right. \\
& \quad \left. + \left(\sum_{S \in \mathcal{E}(\mathcal{T}_h)} \|h(S)^{1/2} [\nabla \bar{u}_h^M(t) \cdot n]\|_{L^2(S)}^2 \right) \right] + \frac{1}{4} \|\nabla \tilde{e}(t)\|^2.
\end{aligned}$$

Integration über die Zeit ergibt dann

$$\begin{aligned}
& \|\tilde{e}(t^m)\|^2 + \int_0^{t^m} \|\nabla \tilde{e}(s)\|^2 \leq \|\tilde{e}(0)\|^2 \\
& \quad + 2 \int_0^{t^m} (\|\nabla(\tilde{u}_h^M(s) - \bar{u}_h^M(s))\|^2 + 4c_p \|(f - \bar{f}(s))\|^2) ds \\
& \quad + C \int_0^{t^m} \left[\left(\sum_{T \in \mathcal{T}_h} \|h(T)(\partial_t \tilde{u}_h^M(s) - f(s))\|_{L^2(T)}^2 \right) \right. \\
& \quad \left. + \left(\sum_{S \in \mathcal{E}(\mathcal{T}_h)} \|h(S)^{1/2} [\nabla \bar{u}_h^M(s) \cdot n]\|_{L^2(S)}^2 \right) \right].
\end{aligned}$$

Verwenden wir nun die Definitionen von \tilde{e} und \bar{e} , so folgt

$$\begin{aligned}
& \|u(t^m) - u_h^m\|^2 + \int_0^{t^m} \|\nabla(u(s) - u_h^M(s))\|^2 ds \leq \|u_0 - u_h^0\|^2 + 4c_p \int_0^{t^m} \|(f - \bar{f}(s))\|^2 ds \\
& \quad + 2 \sum_{j=0}^{m-1} \int_0^{t^j} \left(\frac{s - t^j}{\Delta t} \right)^2 ds \|\nabla(u_h^{j+1} - u_h^j)\|^2 \\
& \quad + C \sum_{j=0}^{m-1} \Delta t \left[\left(\sum_{T \in \mathcal{T}_h} \|h(T) \left(\frac{u_h^{j+1} - \bar{u}_h^j}{\Delta t} - f(t) \right)\|_{L^2(T)}^2 \right) \right. \\
& \quad \left. + \left(\sum_{S \in \mathcal{E}(\mathcal{T}_h)} \|h(S)^{1/2} [\nabla u_h^{j+1}(t) \cdot n]\|_{L^2(S)}^2 \right) \right].
\end{aligned}$$

Mit

$$\int_0^{t^j} \left(\frac{s - t^j}{\Delta t} \right)^2 ds = \frac{\Delta t}{3}$$

folgt dann die Behauptung. □

Bemerkung 3.39

Die a posteriori Fehlerabschätzung aus Theorem 3.28 ist nicht optimal bzgl. der $L^\infty(0, T; L^2(\Omega))$ -Norm. Insbesondere gilt dies für die Konvergenz des Sprungresiduums η_{jump} . Dies liegt vor allem daran, dass der Fehler \tilde{e} bzgl. des Ortes nur in $H_0^1(\Omega)$ liegt und wir daher bei den Interpolationsabschätzungen nur eine h -Potenz gewinnen können. Dieses Problem lässt sich umgehen, wenn man beispielsweise die Testfunktion φ in der Fehleridentität aus Lemma 3.26 als Lösung des folgenden dualen Problems wählt

$$\begin{aligned} -\langle \partial_t \varphi(t), \psi \rangle + (\nabla \varphi(t), \nabla \psi) &= 0, \quad \forall \psi \in H_0^1(\Omega), \\ \varphi(T) &= \tilde{e}(T). \end{aligned}$$

Um mit diesem Ansatz zu einer optimalen a posteriori Fehlerabschätzung in $L^\infty(0, T; L^2(\Omega))$ zu gelangen, benötigt man ein parabolisches Regularitätsresultat und eine Stabilitätsabschätzung der Form

$$\|\varphi(t)\|_{H^2(\Omega)} \leq C \|\varphi(T)\|_{L^2(\Omega)}.$$

Eine andere Möglichkeit besteht darin analog zum Beweis der a priori Fehlerabschätzung vorzugehen und als kontinuierliches Äquivalent zur Galerkin-Projektion eine sogenannte elliptische Rekonstruktion einzuführen. In diesem Fall kann dann die elliptische Regularität ausgenutzt werden. Diesen Zugang wollen wir nun im Detail darstellen (vgl. hierzu [LM06]).

Definition 3.40 (Elliptische Rekonstruktion)

Sei $v_h \in S_{h,0}^k$ und der diskrete Laplace-Operator $-\Delta_h : S_{h,0}^k \rightarrow S_{h,0}^k$ definiert durch

$$(-\Delta_h v_h, \varphi_h) := (\nabla v_h, \nabla \varphi_h), \quad \forall \varphi_h \in S_{h,0}^k.$$

Dann definieren wir den elliptischen Rekonstruktionsoperator $\mathcal{R} : S_{h,0}^k \rightarrow H_0^1(\Omega)$ durch

$$(\nabla \mathcal{R} v_h, \nabla \varphi) = (-\Delta_h v_h, \varphi), \quad \forall \varphi \in H_0^1(\Omega).$$

Bemerkung 3.41 (Galerkin Orthogonalität)

Ist $v_h \in S_{h,0}^k$ und $\mathcal{R} : S_{h,0}^k \rightarrow H_0^1(\Omega)$ der elliptischen Rekonstruktionsoperator, so gilt die Galerkin Orthogonalität

$$(\nabla(v_h - \mathcal{R}v_h), \nabla\varphi_h) = 0, \quad \forall \varphi_h \in S_{h,0}^k.$$

D.h. v_h ist die Galerkin-Projektion der elliptischen Rekonstruktion

$$v_h = P_h(\mathcal{R}v_h).$$

Definition 3.42 (Zeitabhängige Rekonstruktion und parabolischer Fehler)

Seien u eine genügend glatte Lösung der inhomogenen Wärmeleitungsgleichung aus Definition 3.7 und $u_h^m \in S_{h,0}^k$, $m = 0, \dots, M$ die impliziten Euler Diskretisierung aus Definition 3.16. Mit den Notationen aus Lemma 3.26 definieren wir die zeitabhängigen elliptischen Rekonstruktionen $\tilde{u} : [0, T] \rightarrow H_0^1(\Omega)$ und $\bar{u} : [0, T] \rightarrow H_0^1(\Omega)$ durch

$$\begin{aligned}\tilde{u}(t) &:= \mathcal{R}\tilde{u}_h^M(t), \quad \forall t \in [0, T], \\ \bar{u}(t) &:= \mathcal{R}\bar{u}_h^M(t), \quad \forall t \in [0, T],\end{aligned}$$

$\tilde{u}(t)$ erbt bzgl. der Zeit die Regularität von \tilde{u}_h^M und ist daher insbesondere schwach differenzierbar bezüglich t .

Mit Hilfe der Rekonstruktion $\tilde{u}(t)$ zerlegen wir den Fehler $\tilde{e}(t)$ in

$$\tilde{e}(t) = \varepsilon(t) - \rho(t)$$

mit dem Rekonstruktionsfehler

$$\varepsilon(t) := \tilde{u}(t) - \tilde{u}_h^M(t)$$

und dem parabolischen Fehler

$$\rho(t) := \tilde{u}(t) - u(t).$$

Bemerkung 3.43

Wir wollen bemerken, dass mit Hilfe des diskreten Laplace-Operators die definierende Gleichung des impliziten Euler Verfahrens punktweise geschrieben

werden kann, als

$$\partial_t \tilde{u}_h^M(x, t^{m+1}) - \Delta_h u_h^{m+1}(x) - P_0 f(x, t^{m+1}) = 0.$$

Dabei bezeichnet $P_0 : L^2(\Omega) \rightarrow S_{h,0}^k$ die L^2 -Projektion, definiert durch

$$(P_0 v, \varphi_h) = (v, \varphi_h), \quad \forall \varphi_h \in S_{h,0}^k.$$

Mit der Definition von \tilde{u} und der Galerkin-Orthogonalität der elliptischen Rekonstruktion folgt dann für alle $\varphi \in H_0^1(\Omega)$

$$(\partial_t \tilde{u}_h^M(x, t^{m+1}), \varphi) + (\nabla \tilde{u}(t^{m+1}), \nabla \varphi) - (P_0 f(t^{m+1}), \varphi) = 0. \quad (3.11)$$

Der Vorteil der Zerlegung des Fehlers in Rekonstruktionsfehler und parabolischen Fehler liegt darin, dass die Effekte der Orts- und Zeitdiskretisierung nun unabhängig voneinander untersucht werden können. Außerdem ist der parabolischen Fehler $\rho(t) \in H_0^1(\Omega)$ und hat automatisch bessere Regularitäts-eigenschaften, falls die Daten und die Regularität von Ω dies zulassen.

Weiter sei angemerkt, dass der elliptische Fehler ε aufgrund der Galerkin-Orthogonalität mit Hilfe einer a posteriori Fehlerabschätzung abgeschätzt werden kann, wie wir sie in Abschnitt 2.5 hergeleitet haben. Benutzt man zudem ein Dualitätsargument, so kann man auch in der $L^\infty(0, T; L^2(\Omega))$ -Norm eine optimale a posteriori Fehlerabschätzung erhalten.

Neu ist also insbesondere die Behandlung des parabolischen Fehlers. Dazu leiten wir zunächst eine Fehlergleichung für ρ her.

Lemma 3.44 (Fehlergleichung für ρ)

Es gelten die Voraussetzungen und Notationen aus Definition 3.32. Dann gilt für $m = 1, \dots, M$ und für alle $\varphi \in L^2(0, T; H_0^1(\Omega))$ mit schwacher Ableitung $\partial_t \varphi \in L^2(0, T; H^{-1}(\Omega))$ die Fehlergleichung

$$\begin{aligned} \frac{d}{dt}(\rho(t), \varphi(t)) &= -(\nabla \rho(t), \nabla \varphi(t)) + \langle \rho(t), \partial_t \varphi(t) \rangle \\ &\quad + (\nabla(\tilde{u}(t) - \bar{u}(t)), \nabla \varphi(t)) + \langle \partial_t \varepsilon(t) + (P_0 \bar{f} - f)(t), \varphi(t) \rangle. \end{aligned}$$

Dabei ist $P_0 : L^2(\Omega) \rightarrow S_{h,0}^k$ die L^2 -Projektion und $\bar{f} : [0, T] \rightarrow L^2(\Omega)$ ist definiert durch $\bar{f}(t)|_{I_m} := f(t^{m+1}) \forall t \in I_m$.

Beweis: Es gilt

$$\frac{d}{dt}(\rho, \varphi(t)) = \langle \partial_t \rho(t), \varphi(t) \rangle + \langle \rho(t), \partial_t \varphi(t) \rangle.$$

Betrachten wir also noch $\langle \partial_t \rho(t), \varphi(t) \rangle$. Mit der schwachen Form der inhomogenen Wärmeleitungsgleichung erhalten wir mit $u(t) = -\rho(t) + \tilde{u}(t)$ und

$$\langle \partial_t \rho(t), \varphi(t) \rangle = -\langle f(t), \varphi(t) \rangle - (\nabla \rho(t), \nabla \varphi(t)) + (\nabla \tilde{u}(t), \nabla \varphi(t)) + \langle \partial_t \tilde{u}, \varphi(t) \rangle.$$

Subtrahieren wir davon Gleichung (3.8), so erhalten wir

$$\begin{aligned} \langle \partial_t \rho(t), \varphi(t) \rangle &= \langle (P_0 \bar{f} - f)(t), \varphi(t) \rangle - (\nabla \rho(t), \nabla \varphi(t)) \\ &\quad + (\nabla(\tilde{u}(t) - \bar{u}(t)), \nabla \varphi(t)) + \langle \partial_t(\tilde{u} - \tilde{u}_h^M)(t), \varphi(t) \rangle. \end{aligned}$$

Mit $\varepsilon = \tilde{u} - \tilde{u}_h^M(t)$ folgt nun die Behauptung. \square

Bemerkung 3.45

Vergleicht man die Fehlergleichung für ρ mit der Fehlergleichung für \tilde{e} aus Lemma 3.26, so ist wichtig, dass in der Gleichung für ρ der Term mit den Kantensprüngen nicht mehr auftaucht. Im nächsten Satz werden wir aus der Fehlergleichung für ρ eine a posteriori Fehlerabschätzung für ρ ableiten.

Satz 3.46 (Abschätzung des parabolischen Fehlers)

Es gelten die Voraussetzungen und Notationen aus Lemma 3.34. Ausserdem sei Ω hinreichend glatt, so daß wir uniforme H^2 -Regularität für die Poisson-gleichung erwarten können. Dann gilt für alle $t^m, m = 1, \dots, M$ die Fehlerabschätzung

$$\begin{aligned} &\left(\max_{s \in [0, t^m]} \|\rho(s)\|^2 + 2 \int_0^{t^m} \|\nabla \rho(s)\|^2 ds \right)^{1/2} \\ &\leq \|\rho(0)\| + 4 \left[\left(\sum_{j=0}^{m-1} \Delta t \left(\eta_{\text{time}}^j + \eta_{\text{space}}^j + \eta_{\text{data,t}}^j \right) \right)^2 + \sum_{j=0}^{m-1} \Delta t \left(\eta_{\text{data,x}}^j \right)^2 \right]^{1/2}. \end{aligned}$$

Dabei sind die Fehlerschätzer definiert durch

$$\begin{aligned}
 \eta_{\text{time}}^j &:= \frac{1}{2} \|\Delta_h(u_h^{j+1} - u_h^j)\|, \\
 \eta_{\text{space}}^j &:= C_1 \left[\left(\sum_{T \in \mathcal{T}_h} \|h(T)^2(-\Delta_h u_{h,t}(s) + \Delta u_{h,t}(s))\|_{L^2(T)}^2 \right)^{1/2} \right. \\
 &\quad \left. + \left(\sum_{S \in \mathcal{E}(\mathcal{T}_h)} \|h(S)^{3/2}[\nabla u_{h,t}(s) \cdot n]\|_{L^2(S)}^2 \right)^{1/2} \right], \\
 \eta_{\text{data,t}}^j &:= \frac{1}{\Delta t} \int_{t^j}^{t^{j+1}} \|(\bar{f} - f)(s)\| ds, \\
 \eta_{\text{data,x}}^j &:= C_2 \left(\sum_{T \in \mathcal{T}_h} \|h(T)((P_0 - Id)f(t^{j+1})\|_{L^2(T)}^2 \right)^{1/2},
 \end{aligned}$$

wobei gilt $u_{h,t}(t^{j+1}) := \frac{1}{\Delta t}(u_h^{j+1} - u_h^j)$.

Beweis: Analog zur Vorgehensweise zum Beweis von Satz 3.28, wählen wir als Testfunktion $\varphi = \rho$ und erhalten aus der Fehlergleichung für ρ nach Integration in der Zeit

$$\begin{aligned}
 &\|\rho(t)\|^2 - \|\rho(0)\|^2 + 2 \int_0^t \|\nabla \rho(s)\|^2 ds \\
 &\leq 2 \int_0^t \left[|(\nabla(\tilde{u}(s) - \bar{u}(s)), \nabla \rho(s))| + |\langle \partial_t \varepsilon(s) + (P_0 \bar{f} - f)(s), \rho(s) \rangle| \right] ds.
 \end{aligned}$$

Ist $t_*^m \leq t^m$ der Zeitpunkt, für den gilt $\|\rho(t_*^m)\| = \max_{s \in [0, t^m]} \|\rho(s)\|$, so folgt durch die entsprechenden Ungleichungen für t^m und t_*^m

$$\begin{aligned}
 &\|\rho(t_*^m)\|^2 - \|\rho(0)\|^2 + 2 \int_0^{t_*^m} \|\nabla \rho(s)\|^2 ds \\
 &\leq 4 \sum_{j=0}^{m-1} \int_{t^j}^{t^{j+1}} \left[|(\nabla(\tilde{u}(s) - \bar{u}(s)), \nabla \rho(s))| + |\langle \partial_t \varepsilon(s) + (P_0 \bar{f} - f)(s), \rho(s) \rangle| \right] ds.
 \end{aligned}$$

Wir werden jetzt die Terme auf der rechten Seite nacheinander behandeln.

Für den ersten Term gilt

$$\begin{aligned}
\int_{t^j}^{t^{j+1}} |(\nabla(\tilde{u}(s) - \bar{u}(s)), \nabla \rho(s))| ds &= \int_{t^j}^{t^{j+1}} \frac{s - t^j}{\Delta t} |(-\Delta_h(u_h^{j+1} - u_h^j), \rho(s))| ds \\
&\leq \int_{t^j}^{t^{j+1}} \frac{s - t^j}{\Delta t} \| -\Delta_h(u_h^{j+1} - u_h^j) \| \| \rho(s) \| ds \\
&\leq \max_{s \in [0, t^m]} \| \rho(s) \| \frac{1}{2} \Delta t \| -\Delta_h(u_h^{j+1} - u_h^j) \|.
\end{aligned}$$

Betrachten wir nun den Term $\int_{t^j}^{t^{j+1}} |\langle \partial_t \varepsilon(s), \rho(s) \rangle| ds$. Mit der Definition von ε erhalten wir mit $u_{h,t} := \partial_t \tilde{u}_h^M$

$$\int_{t^j}^{t^{j+1}} |\langle \partial_t \varepsilon(s), \rho(s) \rangle| ds = \int_{t^j}^{t^{j+1}} |\langle (\mathcal{R}(u_{h,t}) - u_{h,t})(s), \rho(s) \rangle| ds.$$

Wir wenden nun ein Dualitätsargument an, um diesen Fehler optimal zu kontrollieren. Sei daher $\psi : [0, T] \rightarrow H_0^1(\Omega)$ definiert als die Lösung des dualen elliptischen Problems

$$(\nabla \varphi, \nabla \psi(s)) = (\rho(s), \varphi), \quad \forall \varphi \in H_0^1(\Omega), s \in [0, T].$$

Dann gilt mit $\varphi = \mathcal{R}(u_{h,t}) - u_{h,t}(s)$

$$\langle (\mathcal{R}(u_{h,t}) - u_{h,t})(s), \rho(s) \rangle = (\nabla(\mathcal{R}(u_{h,t}) - u_{h,t})(s), \nabla \psi(s)).$$

Dies Umformulierung erlaubt uns jetzt die Galerkin-Orthogonalität der elliptischen Rekonstruktion zu nutzen. damit erhalten wir insbesondere

$$\langle (\mathcal{R}(u_{h,t}) - u_{h,t})(s), \rho(s) \rangle = (\nabla(\mathcal{R}(u_{h,t}) - u_{h,t})(s), \nabla(\psi(s) - I_h(\psi(s))).$$

Dies wiederum ergibt mit der Definition der elliptischen Rekonstruktion und elementweiser partieller Integration

$$\begin{aligned}
\langle (\mathcal{R}(u_{h,t}) - u_{h,t})(s), \rho(s) \rangle &= \sum_{T \in \mathcal{T}_h} (-\Delta_h u_{h,t}(s) + \Delta u_{h,t}(s), \psi(s) - I_h(\psi(s)))_{L^2(T)} \\
&\quad + \sum_{S \in \mathcal{E}(\mathcal{T}_h)} ([\nabla u_{h,t}(s) \cdot n], \psi(s) - I_h(\psi(s)))_{L^2(S)}
\end{aligned}$$

Verwenden wir nun die optimalen Interpolationsabschätzungen für $\psi(s) - I_h(\psi(s))$, d.h.

$$\begin{aligned}
\| \psi(s) - I_h(\psi(s)) \|_{L^2(T)} &\leq Ch(T)^\alpha |\psi(s)|_{H^\alpha(T)}, \\
\| \psi(s) - I_h(\psi(s)) \|_{L^2(S)} &\leq Ch(S)^{\alpha-1/2} |\psi(s)|_{H^\alpha(\omega(S))},
\end{aligned}$$

so folgt mit $\alpha = 2$

$$\begin{aligned} |\langle (\mathcal{R}(u_{h,t}) - u_{h,t})(s), \rho(s) \rangle| &\leq C \left[\left(\sum_{T \in \mathcal{T}_h} \|h(T)^2(-\Delta_h u_{h,t}(s) + \Delta u_{h,t}(s))\|_{L^2(T)}^2 \right)^{1/2} \right. \\ &\quad \left. + \left(\sum_{S \in \mathcal{E}(\mathcal{T}_h)} \|h(S)^{3/2}[\nabla u_{h,t}(s) \cdot n]\|_{L^2(S)}^2 \right)^{1/2} \right] |\psi(s)|_{H^2(\Omega)}. \end{aligned}$$

Mit der H^2 -Regularität des dualen Problems gilt

$$|\psi(s)|_{H^2(\Omega)} \leq C \|\rho(s)\|$$

und wir erhalten

$$\begin{aligned} |\langle (\mathcal{R}(u_{h,t}) - u_{h,t})(s), \rho(s) \rangle| &\leq C \max_{s \in [0, t^m]} \|\rho(s)\| \\ &\quad \left[\left(\sum_{T \in \mathcal{T}_h} \|h(T)^2(-\Delta_h u_{h,t}(s) + \Delta u_{h,t}(s))\|_{L^2(T)}^2 \right)^{1/2} \right. \\ &\quad \left. + \left(\sum_{S \in \mathcal{E}(\mathcal{T}_h)} \|h(S)^{3/2}[\nabla u_{h,t}(s) \cdot n]\|_{L^2(S)}^2 \right)^{1/2} \right]. \end{aligned}$$

Insgesamt haben wir also gezeigt

$$\begin{aligned} \int_{t^j}^{t^{j+1}} |\langle \partial_t \varepsilon(s), \rho(s) \rangle| ds &\leq C \max_{s \in [0, t^m]} \|\rho(s)\| \Delta t \\ &\quad \left[\left(\sum_{T \in \mathcal{T}_h} \|h(T)^2(-\Delta_h u_{h,t}(t^{j+1}) + \Delta u_{h,t}(t^{j+1}))\|_{L^2(T)}^2 \right)^{1/2} \right. \\ &\quad \left. + \left(\sum_{S \in \mathcal{E}(\mathcal{T}_h)} \|h(S)^{3/2}[\nabla u_{h,t}(t^{j+1}) \cdot n]\|_{L^2(S)}^2 \right)^{1/2} \right], \end{aligned}$$

wobei nach Definition gilt $u_{h,t}(t^{j+1}) = \frac{1}{\Delta t}(u_h^{j+1} - u_h^j)$.

Schließlich bleibt nur noch die Abschätzung des Datenapproximationsfehlers.

Wir erhalten

$$\begin{aligned}
\int_{t^j}^{t^{j+1}} |\langle (P_0 \bar{f} - f)(s), \rho(s) \rangle| ds &= \int_{t^j}^{t^{j+1}} |\langle (P_0 \bar{f} - \bar{f})(s) + (\bar{f} - f)(s), \rho(s) \rangle| ds \\
&\leq \int_{t^j}^{t^{j+1}} |\langle (P_0 \bar{f} - \bar{f})(s), \rho(s) - I_h(\rho(s)) \rangle| ds \\
&\quad + \int_{t^j}^{t^{j+1}} |\langle (\bar{f} - f)(s), \rho(s) \rangle| ds \\
&\leq C \sqrt{\Delta t} \left(\sum_{T \in \mathcal{T}_h} \|h(T)((P_0 - Id)f(t^{j+1})\| \right)^{1/2} \\
&\quad \left(\int_{t^j}^{t^{j+1}} \|\nabla \rho(s)\|^2 ds \right)^{1/2} \\
&\quad + \max_{s \in [0, t^m]} \|\rho(s)\| \int_{t^j}^{t^{j+1}} \|(\bar{f} - f)(s)\| ds.
\end{aligned}$$

Wir fassen nun die erzielten Abschätzungen zusammen, indem wir die Notationen aus dem Satz verwenden

$$\begin{aligned}
&\|\rho(t^m_*)\|^2 - \|\rho(0)\|^2 + 2 \int_0^{t^m} \|\nabla \rho(s)\|^2 ds \\
&\leq 4 \|\rho(t^m_*)\| \sum_{j=0}^{m-1} \Delta t \left(\eta_{\text{time}}^j + \eta_{\text{space}}^j + \eta_{\text{data,t}}^j \right) \\
&\quad + 4 \sum_{j=0}^{m-1} \left(\int_{t^j}^{t^{j+1}} \|\nabla \rho(s)\|^2 ds \right)^{1/2} \sqrt{\Delta t} \eta_{\text{data,x}}^j.
\end{aligned}$$

Da für Vektoren $a, b \in \mathbb{R}^{m+1}$ und $c \in R$ gilt

$$\left[|a|^2 \leq c^2 + a \cdot b \right] \implies \left[|a| \leq |c| + |b| \right],$$

erhalten wir mit $c = \|\rho(0)\|$, $a_m = \|\rho(t^m_*)\|$, $a_j = \left(\int_{t^j}^{t^{j+1}} \|\nabla \rho(s)\|^2 ds \right)^{1/2}$, $j = 0, \dots, m-1$, $b_m = 4 \sum_{j=0}^{m-1} \Delta t \left(\eta_{\text{time}}^j + \eta_{\text{space}}^j + \eta_{\text{data,t}}^j \right)$ und $b_j = 4 \sqrt{\Delta t} \eta_{\text{data,x}}^j$, $j = 0, \dots, m-1$ die Behauptung. \square

Satz 3.47 (Fehlerabschätzung für ε)

Es gelten die Voraussetzungen und Notationen aus Satz 3.36. Dann gilt für alle $s \in [0, T]$ die Fehlerabschätzung

$$\|\varepsilon(s)\| \leq \eta_{\text{space}}(\tilde{u}_h^M(s))$$

mit

$$\begin{aligned}\eta_{\text{space}}(\tilde{u}_h^M(s)) := & C \left[\left(\sum_{T \in \mathcal{T}_h} \|h(T)^2(-\Delta_h \tilde{u}_h^M(s) + \Delta \tilde{u}_h^M(s))\|_{L^2(T)}^2 \right)^{1/2} \right. \\ & \left. + \left(\sum_{S \in \mathcal{E}(\mathcal{T}_h)} \|h(S)^{3/2}[\nabla \tilde{u}_h^M(s) \cdot n]\|_{L^2(S)}^2 \right)^{1/2} \right].\end{aligned}$$

Beweis: Der Beweis folgt analog zur Abschätzung des zweiten Terms im Beweis von Satz 46. \square

Mit Hilfe der Sätze 3.36 und 3.37 erhalten wir nun mit Dreiecksungleichung eine optimale Fehlerabschätzung für den Fehler \tilde{e} .

Satz 3.48 (A posteriori Fehlerabschätzung)

Es gelten die Voraussetzungen und Notationen aus den Sätzen 3.36 und 3.37. Dann gilt für alle $t^m, m = 1, \dots, M$ die Fehlerabschätzung

$$\begin{aligned}\max_{s \in [0, t^m]} \|u(s) - \tilde{u}_h^M(s)\|_{L^2(\Omega)} & \leq \|\mathcal{R}u_h^0 - u_0\|_{L^2(\Omega)} + \max_{s \in [0, t^m]} \eta_{\text{space}}(\tilde{u}_h^M(s)) \\ & + 4 \left[\left(\sum_{j=0}^{m-1} \Delta t \left(\eta_{\text{time}}^j + \eta_{\text{space}}^j + \eta_{\text{data,t}}^j \right) \right)^2 + \Delta t \left(\eta_{\text{data,x}}^j \right)^2 \right]^{1/2}.\end{aligned}$$

Beweis: Der Beweis folgt mit $\|\tilde{e}(s)\| \leq \|\varepsilon(s)\| + \|\rho(s)\|$ und den Abschätzungen aus den Sätzen 3.36 und 3.37. \square

Kapitel 4

Das CG-Verfahren und die Kondition der Steifigkeitsmatrix

In Abschnitt 2 und 3 haben wir gesehen, dass Finite Elemente Verfahren auf die Lösung eines linearen Gleichungssystems führen. Da die zugehörigen Matrizen dünn besetzt und oft symmetrisch und positiv definit sind, eignet sich zur Berechnung der Lösungen der resultierenden linearen Gleichungssysteme das konjugierte Gradientenverfahren. Betrachten wir aber zunächst allgemein Gradientenverfahren zur iterativen Lösung linearer Gleichungssysteme.

4.1 Gradientenverfahren

Generalvereinbarung 4.1

In diesem Abschnitt gelte stets

- 1) $A \in \mathbb{R}^{m \times m}$ mit $A = A^T$,
- 2) A sei positiv definit und
- 3) $b \in \mathbb{R}^m$.

Ziel: Löse das lineare Gleichungssystem $Ax = b$.

Dazu wollen wir das Problem zunächst in ein Minimierungsproblem überführen.

Definition 4.2 (Minimierungsaufgabe)

Sei $F(x) := \langle A^{-1}(Ax - b), Ax - b \rangle, x \in \mathbb{R}^m$.

x heißt Lösung der Minimierungsaufgabe (M) , g.d.w.

$$(M) \quad F(x) = \min_{y \in \mathbb{R}^m} F(y).$$

Lemma 4.3

A, b seien mit den Eigenschaften der Generalvereinbarung 2.1 gegeben.

Dann sind äquivalent:

- 1) $x \in \mathbb{R}^m$ löst $Ax = b$,
- 2) $x \in \mathbb{R}^m$ löst (M) .

Beweis: A positiv definit $\implies A^{-1}$ positiv definit. Also gilt $F(y) \geq 0 \quad \forall y \in \mathbb{R}^m$.

1) \Rightarrow 2): Gilt $Ax = b$, so folgt $F(x) = \langle A^{-1}0, 0 \rangle = 0$.

Folglich löst x die Minimierungsaufgabe (M) .

2) \Rightarrow 1): x löst $(M) \implies \nabla F(x) = 0 \iff 2(Ax - b) = 0 \iff Ax - b = 0 \iff 1)$ \square

Idee der Gradientenverfahren 4.4

Aus Lemma 2.3 folgt: $Ax = b \iff F(x) = 0$.

Idee: Sei $(z_n)_{n \in \mathbb{N}}$ eine Folge im \mathbb{R}^m definiert durch

$$z_{n+1} := z_n + \alpha_n t_n \quad n = 1, 2, \dots$$

mit Koeffizienten $\alpha_n \in \mathbb{R}$ und Richtungsvektoren $t_n \in \mathbb{R}^m \setminus \{0\}$.

Wähle α_n, t_n so, dass $F(z_n) \rightarrow 0$ ($n \rightarrow \infty$).

Ansatz für die Wahl von α_n :

$$\alpha_n := \beta_n \frac{\langle t_n, r_n \rangle}{\langle At_n, t_n \rangle}$$

mit $\beta_n \in \mathbb{R}$ und $r_n := b - Az_n$ der "Residuenvektor".

Dann gilt:

$$\begin{aligned}
 (*) \quad F(z_n) - F(z_{n+1}) &= \langle A^{-1}r_n, r_n \rangle - \langle A^{-1}(r_n - A\alpha_n t_n), r_n - A\alpha_n t_n \rangle \\
 &= 2\alpha_n \langle t_n, r_n \rangle - \alpha_n^2 \langle At_n, t_n \rangle \\
 &= 2\beta_n \frac{\langle t_n, r_n \rangle^2}{\langle At_n, t_n \rangle} - \beta_n^2 \frac{\langle t_n, r_n \rangle^2}{\langle At_n, t_n \rangle} \\
 &= (2 - \beta_n)\beta_n \underbrace{\frac{\langle t_n, r_n \rangle^2}{\langle At_n, t_n \rangle}}_{\geq 0} \stackrel{0 \leq \beta_n \leq 2}{\geq} 0.
 \end{aligned}$$

\implies Für $0 \leq \beta_n \leq 2$ ist $F(z_n) \leq F(z_{n+1})$.

$\implies (F(z_n))_{n \in \mathbb{N}}$ ist monoton fallend.

Also folgt: $\lim_{n \rightarrow \infty} F(z_n) = \lambda \geq 0$ existiert.

$\implies (F(z_n) - F(z_{n-1}))_{n \in \mathbb{N}}$ ist Nullfolge.

Man nennt β_n den Relaxationsparameter der Gradientenverfahren.

Die Gleichung (*) zeigt:

Für $\beta_n = 1$ wird $F(z_n) - F(z_{n+1})$ maximal und F nimmt auf der Geraden $z_n + \alpha_n t_n$ ein Minimum an.

Definition 4.5 (Allgemeines Gradientenverfahren)

Seien $(\beta_n)_{n \in \mathbb{N}}$ und $(t_n)_{n \in \mathbb{N}}$ gegeben mit $\beta_n \in [0, 2]$ und $t_n \in \mathbb{R}^m$. Dann heißt die Folge $(z_n)_{n \in \mathbb{N}}$, $z_n \in \mathbb{R}^m$ Lösung des Gradientenverfahren mit Startwert $z_1 \in \mathbb{R}^m$ wenn gilt:

$$r_1 = b - Az_1$$

und für $n = 1, 2, \dots$ gilt

$$\alpha_n = \beta_n \frac{\langle t_n, r_n \rangle}{\langle At_n, t_n \rangle},$$

$$z_{n+1} = z_n + \alpha_n t_n,$$

$$r_{n+1} = b - Az_{n+1} = r_n - \alpha_n At_n.$$

Definition 4.6 (Konvergenz)

Ein Gradientenverfahren heißt konvergent, falls gilt

$$r_n \rightarrow 0 \quad (n \rightarrow \infty)$$

Dies ist äquivalent zu

$$\begin{aligned} z_n &\rightarrow A^{-1}b \quad (n \rightarrow \infty) \\ \iff z_n &\rightarrow x \quad (n \rightarrow \infty) \text{ mit } Ax = b. \end{aligned}$$

4.2 CG-Verfahren

Das CG-Verfahren erhält man durch eine geeignete Wahl von Abstiegsrichtungen t_n , die ein A-orthogonales System in folgenden Sinne bilden.

Definition 4.7 (A-orthogonal)

Ein System von k Vektoren $q_1, \dots, q_k \in \mathbb{R}^m$ ($k \leq m$) heißt A-orthogonal oder A-konjugiert, wenn gilt:

$$\langle Aq_i, q_j \rangle = 0 \quad i \neq j; \quad i, j = 1, \dots, k,$$

$$\langle Aq_i, q_i \rangle \neq 0 \quad \forall i = 1, \dots, k.$$

Für $k = m$ bilden A-orthogonale Systeme eine Basis des \mathbb{R}^m . Wir setzen $\langle x, y \rangle_A = \langle Ax, y \rangle$, $\|x\|_A = \sqrt{\langle x, x \rangle_A} = \sqrt{\langle Ax, x \rangle}$.

Definition 4.8 (cg-Verfahren) "conjugate gradient"

Wähle im allgemeinen Gradientenverfahren $\beta_n = 1 \quad \forall n = 1, \dots, m+1$, $t_1 = r_1$ und

$$t_n = r_n + \gamma_{n-1} t_{n-1}$$

mit $\gamma_{n-1} := -\frac{\langle Ar_n, t_{n-1} \rangle}{\langle At_{n-1}, t_{n-1} \rangle}$ für $n = 2, 3, \dots$

Idee: Das A-orthogonale System wird mit Hilfe des Schmidt'schen Orthogonalisierungsverfahrens bzgl. $\langle \cdot, \cdot \rangle_A$ schrittweise aufgebaut.

Satz 4.9 (Algorithmus und Konvergenz des cg-Verfahrens)

Sei $z_1 \in \mathbb{R}^m$ und $t_1 = r_1 = b - Az_1$. Dann lautet das cg-Verfahren für $n = 1, \dots, l$ mit $l \leq m$:

$$\begin{cases} a_n = \frac{\langle t_n, r_n \rangle}{\langle At_n, t_n \rangle} \\ z_{n+1} = z_n + \alpha_n t_n \\ r_{n+1} = b - Az_{n+1} \\ \gamma_n = -\frac{\langle Ar_{n+1}, t_n \rangle}{\langle At_n, t_n \rangle} \\ t_{n+1} = r_{n+1} + \gamma_n t_n \end{cases}$$

Es erfüllt die Gleichungen

a) $\langle At_i, t_j \rangle = 0 \quad 1 \leq j \leq i - 1$

b) $\langle r_i, r_j \rangle = 0 \quad 1 \leq j \leq i - 1$

c) $\langle t_i, r_j \rangle = \langle r_j, r_i \rangle \quad 1 \leq j \leq i$

und $l \leq m$ ist so gewählt, dass gilt $r_{l+1} = 0$.

Beweis: (siehe Numerische Lineare Algebra)

Bemerkung 4.10

Für Gleichungssysteme resultierend aus Diskretisierungsverfahren, wie z.B. der Finite Elemente Methode, ist m so groß, dass man in der Praxis weniger als m Schritte iterieren wird. Dass dies Sinn macht zeigt die folgende Fehlerabschätzung.

Satz 4.11

Sei $\kappa(A) = \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}$ die Kondition von $A \in \mathbb{R}^{n \times n}$. Dann gilt für das cg-Verfahren für $Ax = b$:

$$\|z_n - x\|_A \leq 2 \left(\frac{\sqrt{\kappa(A)} - 1}{\sqrt{\kappa(A)} + 1} \right)^n \|z_1 - x\|_A$$

wobei $\|y\|_A := \sqrt{\langle Ay, y \rangle}$ definiert ist.

(ohne Beweis)

4.3 Kondition der Steifigkeitsmatrix

Im Folgenden wollen wir untersuchen, wie die Kondition der Steifigkeitsmatrix von Finite Elemente Verfahren von der Gitterweite h abhängt. Die Antwort auf diese Frage liefert der folgende Satz.

Satz 4.12 (Kondition der Steifigkeitsmatrix)

Sei $\Omega \subset \mathbb{R}^d$ ein Gebiet und \mathcal{T}_h ein zulässige Triangulierung mit $\sigma(T) \leq \sigma \forall T \in \mathcal{T}_h$. Weiter sei $X := H_0^m(\Omega)$ und $X_h \subset X$ ein endlichdimensionaler Finite Elemente Teilraum mit Basis $\varphi_1, \dots, \varphi_N$. Sei $B : X \times X \rightarrow \mathbb{R}$ eine symmetrische, koerzive (c_0) und stetige (c_1) Bilinearform. Dann gilt für die Kondition der Steifigkeitsmatrix $S \in \mathbb{R}^{N \times N}$, definiert durch

$$S_{ij} := B(\varphi_i, \varphi_j)$$

die Konditionsabschätzung

$$\kappa(S) \leq ch^{-2m}.$$

Beweis: Es ist nach Definition $\kappa(S) := \frac{\lambda_{\max}(S)}{\lambda_{\min}(S)}$, daher schätzen wir zunächst $\lambda_{\max}(S)$ nach oben und dann $\lambda_{\min}(S)$ nach unten ab.

1. Schritt: Abschätzung von $\lambda_{\max}(S)$.

Es ist

$$\lambda_{\max}(S) = \sup_{\xi \in \mathbb{R}^N \setminus \{0\}} \frac{(S\xi, \xi)}{|\xi|^2}.$$

Ist $\xi \in \mathbb{R}^N \setminus \{0\}$, so gilt

$$(S\xi, \xi) = \sum_{i,j=1}^N B(\varphi_i, \varphi_j) \xi_i \xi_j = B(v_h, v_h),$$

falls $v_h \in X_h$ definiert ist durch

$$v_h(x) = \sum_{i=1}^N \xi_i \varphi_i(x).$$

Mit der Stetigkeit von B folgt also

$$\begin{aligned}
 (S\xi, \xi) &\leq c_1 \|v_h\|_X^2 \\
 &\leq c|v_h|_X^2 = c \sum_{T \in \mathcal{T}_h} |v_h|_{H^m(T)}^2 \\
 &= c \sum_{T \in \mathcal{T}_h} (v_h, v_h)_{H^m(T)} \\
 &= c \sum_{T \in \mathcal{T}_h} \left(\sum_{i,j=1}^N (\xi_i \chi_T)(\xi_j \chi_T) (\varphi_i, \varphi_j)_{H^m(T)} \right)
 \end{aligned}$$

Wir definieren die Matrix S^T durch

$$S_{ij}^T := (\varphi_i, \varphi_j)_{H^m(T)}$$

und ξ^T durch

$$\xi^T := \xi \chi_T,$$

wobei χ_T die charakteristische Funktion auf T ist. Dann folgt weiter

$$\begin{aligned}
 (S\xi, \xi) &\leq c \sum_{T \in \mathcal{T}_h} (S^T \xi^T, \xi^T)_{\mathbb{R}^N} \\
 &= c \sum_{T \in \mathcal{T}_h} \frac{(S^T \xi^T, \xi^T)_{\mathbb{R}^N}}{|\xi|^2} |\xi^T|^2 \\
 &\leq c \sum_{T \in \mathcal{T}_h} \sup_{\eta \in \mathbb{R}^N} \frac{(S^T \eta, \eta)_{\mathbb{R}^N}}{|\eta|^2} |\xi^T|^2 \\
 &\leq c \sum_{T \in \mathcal{T}_h} \lambda_{\max}(S^T) |\xi^T|^2 \\
 &\leq c \max_{T \in \mathcal{T}_h} \lambda_{\max}(S^T) \sum_{T \in \mathcal{T}_h} |\xi^T|^2 \\
 &\leq c \max_{T \in \mathcal{T}_h} \lambda_{\max}(S^T) \sum_{i=1}^N \xi_i^2 d_{\max},
 \end{aligned}$$

wobei d_{\max} die maximale Anzahl von Simplizes ist, auf denen eine Basisfunktion ungleich Null ist. Somit haben wir gezeigt, dass gilt

$$\lambda_{\max}(S) \leq c \max_{T \in \mathcal{T}_h} \lambda_{\max}(S^T).$$

Mit Cauchy-Schwarz und Skalierungsargument folgt nun weiter

$$\begin{aligned} |S_{ij}^T| &= |(\varphi_i, \varphi_j)_{H^m(T)}| \leq |\varphi_i|_{H^m(T)} |\varphi_j|_{H^m(T)} \\ &\leq c |\nabla F_T^{-1}|^{2m} |\det \nabla F_T| |\varphi_i \circ F|_{H^m(T_0)} |\varphi_j \circ F|_{H^m(T_0)} \\ &\leq ch^{-2m+d}. \end{aligned}$$

Insbesondere folgt also

$$\lambda_{\max}(S) \leq h^{-2m+d}.$$

2. Schritt: Abschätzung von $\lambda_{\min}(S)$.

Mit der Koerzivität von B folgt

$$\begin{aligned} (S\xi, \xi)_{\mathbb{R}^N} &= B(v_h, v_h) \geq c_0 \|v_h\|_{H^m(\Omega)}^2 \geq c_0 \|v_h\|_{L^2(\Omega)}^2 \\ &= c_0 \sum_{T \in \mathcal{T}_h} \sum_{i,j=1}^N \xi_i^T \xi_j^T (\varphi_i, \varphi_j)_{L^2(T)}. \end{aligned}$$

Wir definieren die Matrix M^T durch

$$M_{ij}^T := (\varphi_i, \varphi_j)_{L^2(T)},$$

so folgt weiter analog zu Schritt 1

$$(S\xi, \xi)_{\mathbb{R}^N} \geq c_0 d_{\min} \min_{T \in \mathcal{T}_h} \lambda_{\min}(M^T) |\xi|^2$$

und somit

$$\lambda_{\min}(S) \geq c \min_{T \in \mathcal{T}_h} \lambda_{\min}(M^T).$$

Mit einem Skalierungsargument folgt nun

$$\begin{aligned} (M^T \xi, \xi)_{\mathbb{R}^N} &= \sum_{i,j=1}^N \xi_i^T \xi_j^T (\varphi_i, \varphi_j)_{L^2(T)} \\ &= |\det F_T| \sum_{i,j=1}^N \xi_i^T \xi_j^T (\varphi_i \circ F, \varphi_j \circ F)_{L^2(T_0)} \\ &\geq ch^d |\xi|^2. \end{aligned}$$

Somit haben wir

$$\lambda_{\min}(S) \geq ch^d.$$

3. Schritt: Abschätzung von $\kappa(S)$.

Mit den Abschätzungen aus den Schritten 1. und 2. folgt für die Kondition

$$\kappa(S) := \frac{\lambda_{\max}(S)}{\lambda_{\min}(S)} \geq c \frac{h^{-2m+d}}{h^d} = ch^{-2m}.$$

□

Folgerung 4.13 (Konvergenz des CG-Verfahren für Finite Elemente Verfahren)

Betrachten wir das Poissonproblem, so erhalten wir aus Satz 2.77 als Abschätzung für die Kondition der Steifigkeitsmatrix eines Lagrange Finite Elemente Verfahrens ($m = 1$):

$$\kappa(S) \leq ch^{-2}.$$

Verwendet wir zur Lösung des resultierenden Gleichungssystems $Su = f$ das Konjugierte Gradientenverfahren (CG -Verfahren), so gilt für die Konvergenzrate des Verfahrens (vgl. Numerik II)

$$\rho_{CG} = \frac{\sqrt{\kappa(S)} - 1}{\sqrt{\kappa(S)} + 1} \approx \frac{ch^{-1} - 1}{ch^{-1} + 1} \approx 1 - \frac{2}{c}h + \mathcal{O}(h^2).$$

Dies bedeutet, dass eine Halbierung der Gitterweite h zu einer Verdopplung der CG-Iterationen führen wird.

Zur Verbesserung der Kondition der Steifigkeitsmatrix können Methoden zur Vorkonditionierung angewendet werden. Für mehr Details verweisen wir auf die Literatur, z.B. [Bra97, Hac91, Saa03].

Literaturverzeichnis

- [Alt92] Hans Wilhelm Alt. *Lineare Funktionalanalysis. Eine anwendungsorientierte Einführung*. Hochschultext. Berlin etc.: Springer-Verlag., 1992.
- [BO04] T. Barth and M. Ohlberger. Finite volume methods: foundation and analysis. In T.J.R. Hughes E. Stein, R. de Borst, editor, *Encyclopedia of Computational Mechanics*, volume 1, chapter 15. John Wiley & Sons, Ltd, 2004.
- [Bra97] D. Braess. *Finite Elemente*. Springer, Berlin, 1997.
- [BS08] Susanne C. Brenner and L. Ridgway Scott. *The mathematical theory of finite element methods*, volume 15 of *Texts in Applied Mathematics*. Springer, New York, third edition, 2008.
- [CF04] Zhiming Chen and Jia Feng. An adaptive finite element algorithm with reliable and efficient error control for linear parabolic problems. *Math. Comp.*, 73(247):1167–1193, 2004.
- [Cia87] P.G. Ciarlet. *The finite element methods for elliptic problems*. North-Holland, Amsterdam, 1987.
- [dun] DUNE – Distributed and Unified Numerics Environment. <http://dune-project.org/>.
- [Dzi] G. Dziuk. Skript zur Vorlesung *Theorie und Numerik partieller Differentialgleichungen I*, WS 06/07, Universität Freiburg. <http://www.mathematik.uni-freiburg.de/IAM/homepages/moessner/Lehre/TNPDE1/>.
- [Eva98] Lawrence C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1998.

- [GR94] Christian Großmann and Hans-Görg Roos. *Numerik partieller Differentialgleichungen*. Teubner Studienbücher Mathematik. [Teubner Mathematical Textbooks]. B. G. Teubner, Stuttgart, second edition, 1994.
- [GT77] David Gilbarg and Neil S. Trudinger. *Elliptic partial differential equations of second order*. Springer-Verlag, Berlin, 1977. Grundlehren der Mathematischen Wissenschaften, Vol. 224.
- [Hac85] Wolfgang Hackbusch. *Multigrid methods and applications*, volume 4 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 1985.
- [Hac91] Wolfgang Hackbusch. *Iterative Lösung großer schwachbesetzter Gleichungssysteme*, volume 69 of *Leitfäden der Angewandten Mathematik und Mechanik [Guides to Applied Mathematics and Mechanics]*. B. G. Teubner, Stuttgart, 1991. Teubner Studienbücher Mathematik. [Teubner Mathematical Textbooks].
- [Hac96] Wolfgang Hackbusch. *Theorie und Numerik elliptischer Differentialgleichungen*. Teubner Studienbücher Mathematik. [Teubner Mathematical Textbooks]. B. G. Teubner, Stuttgart, second edition, 1996.
- [Krö97] Dietmar Kröner. *Numerical schemes for conservation laws*. Wiley-Teubner Series Advances in Numerical Mathematics. John Wiley & Sons Ltd., Chichester, 1997.
- [LM06] Omar Lakkis and Charalambos Makridakis. Elliptic reconstruction and a posteriori error estimates for fully discrete linear parabolic problems. *Math. Comp.*, 75(256):1627–1658, 2006.
- [Ran] R. Rannacher. Skript zur Vorlesung *Numerische Mathematik 2: Numerik partieller Differentialgleichungen*, WS 07/08, Universität Heidelberg. <http://numerik.iwr.uni-heidelberg.de/~lehre/notes>.
- [Saa03] Yousef Saad. *Iterative methods for sparse linear systems*. Society for Industrial and Applied Mathematics, Philadelphia, PA, second edition, 2003.
- [Sch91] Hans-Rudolf Schwarz. *Methode der finiten Elemente*, volume 47 of *Leitfäden der Angewandten Mathematik und Mechanik [Guides to Applied Mathematics and Mechanics]*. B. G. Teubner, Stuttgart,

third edition, 1991. Eine Einführung unter besonderer Berücksichtigung der Rechenpraxis. [An introduction with special reference to computational practice], Teubner Studienbücher Mathematik. [Teubner Mathematical Textbooks].

- [Ver96] R. Verfürth. *A review of a posteriori error estimation and adaptive mesh-refinement techniques*. Wiley-Teubner Series Advances in Numerical Mathematics. Teubner, Stuttgart, 1996.

Index

- $C(0, T; X)$, 117
 H^1 -Interpolation, 87
 $H^{-m,p'}(\Omega)$, 28
 $L^2(0, T; H_0^1(\Omega))$, 116
 $L^2(0, T; H^{-1}(\Omega))$, 116
 L^p -Räume, 13
- A posteriori Fehlerabschätzung, 146
A priori Abschätzungen, 125
A priori Fehlerabschätzung, 60
A-konjugiert, 162
A-orthogonal, 162
Abstrakte Fehlerabschätzung, 38
Adaptive Verfahren, 99
Anfangs-Randwertproblem, 4
- Banachraum, 12, 13
Banachscher Fixpunktsatz, 13
Baryzentrische Koordinaten, 43
Bramble-Hilbert, 55
Bramble-Hilbert Lemma, 108
- cg-Verfahren, 162
- Diffusionsgleichung
 instationär, 4
 stationär, 5
- Distanzfunktion, 67
- Divergenz, 11
- Dualraum, 12
- Element-Geometrie, 50
- Elliptische Rekonstruktion, 150
- Energieminimierung, 5
- Erhaltungsgleichungen, 2
Erhaltungsprinzip, 2
- Fehlerabschätzung mit Quadratur, 111
Fehlergleichung, 145
Fehlerschätzer, 80
Finite Elemente Semi-Diskretisierung, 124
Finites Element, 50
Formfunktionen, 50
- Galerkin-Orthogonalität, 38
Galerkin-Projektion, 129
Gleichverteilungsstrategie, 100
Gradient, 11
Gradientenverfahren, 159
- Hölder-Konstante, 12
Hölderräume, 12
Höldersche Ungleichung, 14
Hilbertraum, 13
- Implizite Euler Diskretisierung, 132
Inhomogene Wärmeleitungsgleichung, 124
- Interpolationsfehlerabschätzung, 59
- Kettenregel, 26
- Kondition, 163
- Kondition der Steifigkeitsmatrix, 164
- Langzeitverhalten, 5
- Laplace Operator, 11, 81
- Lax-Milgram, 21
- Lemma von Aubin-Nitsche, 62

- Lemma von Cea, 38
Lineares simpliziales Lagrange Element, 44
Lipschitz-Gebiet, 14
Lipschitz-stetig, 12, 13
Menge der lokalen Freiheitsgrade, 50
Minimierungsaufgabe, 160
nodale Basis, 50
nodale Variablen, 50
normierter Raum
 Hilbertraum, 13
 Prähilbertraum, 13
Operator
 Lipschitz-stetig, 13
 Stetigkeit, 13
Orientierte Distanzfunktion, 67
Partielle Integration, 15
Poisson-Gleichung, 1
Poisson-Problem, 17
Prähilbertraum, 13
Produktregel, 26
Quadraturen für Simplizes, 104
Quadraturformel, 103
Quadraturformeln, 101
Randwertproblem
 in Divergenzform, 18
Relaxationsparameter, 161
Residuenvektor, 161
Residuum, 83
Residuumsfehlerschätzer, 83
Rieszscher Darstellungssatz, 19
Ritz-Galerkin Verfahren, 38
Ritz-Projektion, 129
Rotation, 11
Satz von Gauß, 15
Schwache Ableitung, 24
schwache Lösung, 31, 33
Schwache Nullrandwerte, 27
Schwache Zeitableitung, 117
Simplex, 42
Sobolevräume, 25
Spektralverfahren, 39
Spursatz, 27, 89
Stabilität, 138
Stabilität des θ -Verfahrens, 138
Variationsgleichungen, 5
Variationsrechnung
 Fundamentalsatz, 14
Wärmeleitungsgleichung, 1
Wellengleichung, 1
Youngsche Ungleichung, 14
Zeitabhängige Rekonstruktion, 151
Zeitschrittverfahren, 132
Zulässige Triangulierung, 44