

# Finite Elemente höherer Ordnung für elliptische Randwertprobleme

Diplomarbeit

vorgelegt von  
**Jan Hendrik Starcke**  
aus  
**Wolfenbüttel**

angefertigt im  
**Institut für Numerische und Angewandte Mathematik**  
der Georg-August-Universität zu Göttingen  
**2003**



# Inhaltsverzeichnis

<b>Einleitung</b>	<b>1</b>
<b>1 Grundlagen</b>	<b>3</b>
1.1 Funktionenräume . . . . .	3
1.2 Fortsetzbarkeit von Funktionen auf den Gebietsrand . . . . .	8
1.3 Lösbarkeitstheorie von Lax–Milgram . . . . .	9
1.4 Konforme Galerkin–Approximation . . . . .	11
<b>2 Das kontinuierliche Problem</b>	<b>15</b>
2.1 Klassische und schwache Lösung . . . . .	15
2.2 Regularität der schwachen Lösung . . . . .	17
2.3 Singulär gestörte Probleme für $0 < \varepsilon \ll 1$ . . . . .	19
<b>3 Das diskrete Problem</b>	<b>23</b>
3.1 Gebiete und Gebietszerlegungen . . . . .	23
3.2 Finite–Elemente–Ansätze . . . . .	28
3.3 Interpolationsfehler–Abschätzungen . . . . .	34
3.4 Fehler der Galerkin–Finite–Elemente–Methode . . . . .	40
3.5 Numerische Experimente . . . . .	43
<b>4 Einarbeitung von Randbedingungen in schwacher Form</b>	<b>51</b>
4.1 Das Dirichlet–Randwert–Problem . . . . .	51
4.2 Stabilisierte Methoden . . . . .	56
4.3 Gemischte Randwert–Probleme . . . . .	63
<b>5 Isoparametrische Elemente</b>	<b>67</b>
5.1 Theorie . . . . .	67
5.2 Numerische Experimente . . . . .	71

<b>6 Singulär gestörte Probleme</b>	<b>75</b>
6.1 Stabilisierung in Strömungsrichtung („Streamline Diffusion“)	75
6.2 Numerische Experimente	80
<b>Zusammenfassung und Ausblick</b>	<b>85</b>
<b>Literaturverzeichnis</b>	<b>87</b>

# Einleitung

Um lineare, elliptische Differentialgleichungen zweiter Ordnung effektiv zu lösen, gibt es verschiedene Möglichkeiten. Numerisch bietet sich die Methode der finiten Elemente (FEM) an, denn sie ist der Gebietsstruktur gegenüber flexibel. Außerdem ist sie relativ einfach zu implementieren und führt uns die Differentialgleichung auf ein großes, lineares Gleichungssystem zurück, mit dessen Lösung sich die Numerische Mathematik schon sehr lange beschäftigt und deshalb schon diverse, effektive Verfahren entwickelt hat.

Lineare finite Elemente wurden gegenüber finiten Elementen höherer Ordnung bevorzugt, weil Elemente höherer Ordnung wegen des großen Rechenaufwandes langsamer waren und wegen ihrer geringen Flexibilität gegenüber Störungen, wie z.B. Singularitäten, die Lösung oft schlechter approximierten. Durch fortschreitende Computertechnik mit größeren Speichern und schnelleren Prozessoren eröffnet sich jedoch die Möglichkeit des direkten Vergleichs mit linearen, finiten Elementen.

Ziel dieser Arbeit ist es, einerseits einen Überblick über die Theorie linearer, elliptischer Differentialgleichungen zweiter Ordnung zu geben und finite Elemente höherer Ordnung zur Lösung solcher Probleme heran zu ziehen. Andererseits werden die Möglichkeiten und Grenzen des Programmes FEMLAB in Bezug auf verschiedene Problemdata und ihre Auswirkungen auf die Lösung erörtert. Dabei werden Konvergenzresultate für die verschiedenen Problemklassen aus Theorie und Praxis miteinander verglichen, nachdem die notwendige Anpassung FEMLABs vorgenommen worden ist.

Zunächst formuliere ich das Problem allgemein.

## Problemstellung

**Definition 0.1** Sei  $\Omega$  ein beschränktes Gebiet. Die lineare, elliptische Differentialgleichung zweiter Ordnung

$$Lu := -\varepsilon \Delta u(x) + \vec{b}(x) \nabla u(x) + c(x)u(x) = f(x) \quad \text{in } \Omega \in \mathbb{R}^n \text{ mit } x = (x_1, \dots, x_n) \quad (0.1)$$

heißt **Diffusions-Konvektions-Reaktions-Gleichung**, wobei

- $u(x)$  die Lösung der Differentialgleichung,
- $-\varepsilon \Delta u(x)$  der **Diffusionsterm**,
- $\vec{b}(x) \nabla u(x)$  der **Konvektionsterm**,

- $c(x)u(x)$  der **Reaktionsterm**,
- $f(x)$  die Funktion der rechten Seite und
- $Lu$  der **Differentialoperator** sind.

**Definition 0.2**  $\Omega$  besitze einen Lipschitz-stetigen (z.T. polyedrischen) Rand  $\partial\Omega =: \Gamma$ . Die Gleichung

$$u(x) = g(x) \quad \text{auf } \Gamma. \quad (0.2)$$

heißt **Randbedingung erster Art** oder **Dirichlet-Randbedingung** mit  $g(x)$  als Funktion der Randwerte von  $u$ . Für  $g(x) \equiv 0$  bezeichnet man die Randbedingungen als **homogen**. Andernfalls werden sie als **inhomogen** bezeichnet.

**Definition 0.3** In Kombination heißen die beiden Gleichungen (0.1), (0.2) **Dirichlet-Randwertproblem**.

**Bemerkung 0.4** In den theoretischen Betrachtungen taucht der Reaktionsterm  $c(x)u(x)$  mit  $c(x) \geq 0$  der Vollständigkeit halber auf. Alle Anwendungsbeispiele werden aber mit  $c(x) = 0$  gerechnet.

Im ersten Kapitel sind wichtige mathematische Grundlagen kurz zusammengestellt. Für die Beweise der meisten Sätze wird auf die Standardliteratur zu diesem Thema verwiesen.

Im zweiten Kapitel wird für unterschiedliche Problemdaten das Lösungsverhalten des Randwertproblems (0.1), (0.2) betrachtet. Wir überführen es dazu in eine schwache Formulierung und untersuchen die Regularität der Lösung des äquivalenten, schwachen Problems.

Kapitel 3 befasst sich mit Finite-Elemente-Methoden für lineare, elliptische Differentialgleichungen. Dabei wird vor allem auf die Einsetzbarkeit finiter Elemente höherer Ordnung für die Problemstellung eingegangen. **FEMLAB**, als Erweiterung von **Matlab** speziell für die Lösung von partiellen Differentialgleichungen (PDE) mit Finite-Elemente-Methoden entwickelt, wird zur numerischen Lösung der Probleme genutzt.

In den Kapiteln 4, 5 und 6 werden speziell erweiterte Randwertaufgaben aufgegriffen. In Kapitel 4 wird der Fall homogener Dirichlet-Randbedingungen auf den Fall inhomogener Dirichlet-Randbedingungen und gemischter Randbedingungen erweitert. Die Randbedingungen werden dabei in schwacher Form in die Variationsformulierung eingearbeitet. Außerdem wird das Diffusions-Konvektions-Reaktions-Problem auf das Poisson-Problem eingeschränkt.

In Kapitel 5 wird der stückweise polyedrische Rand auf Lipschitz-stetige Kurvenstücke erweitert und die Finite-Elemente-Methode mittels isoparametrischer, finiter Elemente daran angepasst.

Das singular gestörte Problem aus Kapitel 2 wird in Kapitel 6 numerisch erfasst, stabilisiert und mit FEMLAB berechnet und visualisiert.

# Kapitel 1

## Grundlagen

Im ersten Teil dieses Kapitels werden in zwei Abschnitten Funktionenräume und Normen für die Finite-Elemente-Theorie bereitgestellt. Im zweiten Teil werden die Lösbarkeit von Variationsgleichungen mittels der Lax-Milgram-Theorie und eine konforme Approximation mittels Galerkin-Methode vorgestellt.

### 1.1 Funktionenräume

Hier werden Räume zur Lösung von linearen, elliptischen Differentialgleichungen zweiter Ordnung eingeführt. Es hat sich gezeigt, dass Lebesgue-Räume bzw. Sobolev-Räume dafür geeignet sind.

#### Räume stetig differenzierbarer Funktionen

Wir betrachten stets ein beschränktes Gebiet  $\Omega$  mit Lipschitz-stetigem Rand  $\partial\Omega =: \Gamma$ . Die Menge der stetigen Funktionen auf  $\Omega$  ist  $\mathbf{C}(\Omega)$ .

**Definition 1.1** Sei  $m \in \mathbb{N}_0$ . Die Menge der  $m$ -fach auf  $\Omega$  stetig differenzierbaren Funktionen heißt

$$\mathbf{C}^m(\Omega) := \{v : \Omega \rightarrow \mathbb{R} : D^\alpha v \in \mathbf{C}(\Omega) \text{ für alle } \alpha \text{ mit } |\alpha| \leq m\}. \quad (1.1)$$

Dabei ist  $\alpha = (\alpha_1, \dots, \alpha_n)$  ein **Multiindex** der Länge  $|\alpha| = \alpha_1 + \dots + \alpha_n$ .

Dann ist

$$\mathbf{C}^m(\bar{\Omega}) := \{v \in \mathbf{C}^m(\Omega) : D^\alpha v \text{ stetig fortsetzbar auf } \partial\Omega \text{ für alle } \alpha \text{ mit } |\alpha| \leq m\} \quad (1.2)$$

mit der Norm

$$\|v\|_{\mathbf{C}^m(\bar{\Omega})} := \max_{|\alpha| \leq m} \max_{x \in \bar{\Omega}} |D^\alpha v(x)| \quad \text{für } v \in \mathbf{C}^m(\bar{\Omega}) \quad (1.3)$$

ein Banach-Raum.

**Definition 1.2** Seien  $0 < s \leq 1$  und  $m \in \mathbb{N}_0$ . Der Hölder-Raum  $\mathbf{C}^{m,s}(\overline{\Omega})$  ist dann die Menge der Funktionen  $v \in \mathbf{C}^m(\overline{\Omega})$ , die

$$\|v\|_{\mathbf{C}^{m,s}(\overline{\Omega})} := \|v\|_{\mathbf{C}^m(\overline{\Omega})} + \sum_{|\alpha|=m} \sup_{\substack{x,y \in \overline{\Omega} \\ x \neq y}} \frac{|D^\alpha v(x) - D^\alpha v(y)|}{|x-y|^s} < \infty \quad (1.4)$$

erfüllen.

Ist  $\overline{\Omega}$  kompakt, so ist  $\mathbf{C}^{m,s}(\overline{\Omega})$  mit der Norm (1.4) ein Banach-Raum.

**Definition 1.3** Die Menge der unendlich oft differenzierbaren Funktionen mit kompaktem Träger in  $\Omega$  wird definiert als

$$\mathbf{C}_0^\infty(\Omega) = \{v \in \mathbf{C}^\infty(\Omega) : \text{supp } v \Subset \Omega\}. \quad (1.5)$$

Dabei ist  $\text{supp}$  die Bezeichnung für einen kompakten Träger. Ein abgeschlossener Unterraum wird mit  $\Subset$  bezeichnet.

## Lebesgue-Räume

Im folgenden Abschnitt seien  $p, q \in \mathbb{R} \cup \{\infty\}$  mit  $\frac{1}{p} + \frac{1}{q} = 1$  und  $1 \leq p, q \leq \infty$ .

**Definition 1.4** Sei  $1 \leq p < \infty$ . Dann definiert man **Lebesgue-Räume**

$$\mathbf{L}^p(\Omega) := \{v : \Omega \rightarrow \mathbb{R} : \int_{\Omega} |v(x)|^p dx < \infty\} \quad (1.6)$$

als die Menge aller Äquivalenzklassen messbarer Funktionen, wobei

$$\|v\|_{0,p,\Omega} := \|v\|_{\mathbf{L}^p(\Omega)} := \left( \int_{\Omega} |v(x)|^p dx \right)^{\frac{1}{p}} < \infty \quad (1.7)$$

die zugehörige Norm darstellt. Für  $p = \infty$  wird

$$\mathbf{L}^\infty(\Omega) := \{v : \Omega \rightarrow \mathbb{R} \text{ messbar} : \exists M < \infty \text{ mit } |v(x)| \leq M \text{ fast überall in } \Omega\}$$

mit der Norm

$$\|v\|_{\mathbf{L}^\infty(\Omega)} := \text{ess max}_{x \in \Omega} |v(x)| := \inf M$$

die Menge aller Äquivalenzklassen wesentlich beschränkter Funktionen definiert.

**Satz 1.5** Die  $\mathbf{L}^p(\Omega)$ -Räume,  $1 \leq p \leq \infty$ , sind mit der zu  $p$  passenden Norm **Banach-Räume**.

**Beweis :** vgl. [Alt99], Satz 1.17 bzw. Lemma 1.13. □

**Bemerkung 1.6** Durch

$$(v, w) := \int_{\Omega} v(x)w(x)dx \quad (1.8)$$

wird ein Skalarprodukt definiert, das die  $\|\cdot\|_{0,2,\Omega}$ -Norm induziert. Damit ist der Banachraum  $\mathbf{L}^2(\Omega)$  ein Hilbertraum.

## Sobolev-Räume

Um partielle Differentialgleichungen behandeln zu können, ist die Definition verallgemeinerter Ableitungen und die Einführung von Sobolev-Räumen wichtig.

**Definition 1.7** Eine Funktion  $w$  heißt **verallgemeinerte Ableitung**  $D^\alpha v$  von  $v$ , wenn

$$\int_{\Omega} w(x)\varphi(x)dx = (-1)^{|\alpha|} \int_{\Omega} v(x)D^\alpha\varphi(x)dx \quad \text{für alle } \varphi \in \mathbf{C}_0^\infty(\Omega)$$

gilt.

**Definition 1.8** Die Räume

$$\mathbf{W}^{l,p}(\Omega) := \{v \in \mathbf{L}^p(\Omega) : \|v\|_{l,p,\Omega} < \infty\}, \quad (1.9)$$

mit  $l \in \mathbb{N}_0$  und der Norm

$$\|v\|_{l,p,\Omega} := \|v\|_{\mathbf{W}^{l,p}(\Omega)} := \left( \sum_{|\alpha| \leq l} \|(D^\alpha v)(x)\|_{\mathbf{L}^p(\Omega)}^p \right)^{\frac{1}{p}} \quad \text{für } 1 \leq p < \infty \quad (1.10)$$

bzw.

$$\|v\|_{\mathbf{W}^{l,\infty}(\Omega)} := \sum_{|\alpha| \leq l} \|(D^\alpha v)(x)\|_{\mathbf{L}^\infty(\Omega)} \quad \text{für } p = \infty$$

heißen **Sobolev-Räume**. Für  $l = 0$  ist es der Raum  $\mathbf{L}^p(\Omega)$ .

**Bemerkung 1.9** Ist  $p = 2$ , werden die Räume  $\mathbf{W}^{l,2}(\Omega)$  auch mit  $\mathbf{H}^l(\Omega)$  bezeichnet. Mit dem Skalarprodukt

$$(v, w) := \int_{\Omega} \left( \sum_{|\alpha| \leq l} (D^\alpha v)(x)(D^\alpha w)(x) \right) dx \quad (1.11)$$

werden sie zu Hilberträumen. Oft lässt man bei  $p = 2$  den Index in der Norm-Bezeichnung weg ( $\|\cdot\|_{l,2,\Omega} =: \|\cdot\|_{l,\Omega}$ ).

Sobolev-Räume lassen sich unter bestimmten Voraussetzungen ineinander einbetten. Dazu ist folgende allgemeine Definition hilfreich.

**Definition 1.10** Seien  $X$  und  $Y$  normierte Räume. Dann heißt die **Einbettung**  $X \hookrightarrow Y$  **stetig**, wenn eine Konstante  $C > 0$  existiert mit  $\|v\|_Y \leq C \|v\|_X$  für alle Funktionen  $v \in X$ . Die Einbettung ist zusätzlich **kompakt**, wenn der Einbettungsoperator  $\mathbb{I} \in \mathcal{L}(X, Y)$  mit  $\mathbb{I}v = v$  für alle  $v \in X$  kompakt ist.

Für Sobolev-Räume gilt nach [Ada75], Theorem 5.4, der **Sobolev'sche Einbettungssatz**:

**Satz 1.11** Sei  $\Omega \subset \mathbb{R}^n$ ,  $n = 2, 3$ , ein Gebiet mit Lipschitz-stetigem Rand,  $1 \leq p \leq \infty$  und  $l \in \mathbb{N}_0$  gegeben. Dann gibt es folgende stetige Einbettungen:

1. Wenn  $0 \leq lp < n$  ist, dann gilt für  $p^* = \frac{np}{n-lp}$

$$\mathbf{W}^{l,p}(\Omega) \hookrightarrow \mathbf{L}^{p^*}(\Omega). \quad (1.12)$$

2. Wenn  $lp = n$  ist, dann gilt für  $q$  mit  $p \leq q \leq \infty$

$$\mathbf{W}^{l,p}(\Omega) \hookrightarrow \mathbf{L}^q(\Omega). \quad (1.13)$$

3. Wenn  $lp > n$  ist, dann gilt

$$\mathbf{W}^{l,p}(\Omega) \hookrightarrow \mathbf{C}^0(\bar{\Omega}). \quad (1.14)$$

**Bemerkung 1.12** Der Abschluss des Raumes  $\mathbf{C}_0^\infty(\Omega)$  bezüglich der  $\mathbf{W}^{1,2}$ -Norm ist der Raum  $\mathbf{W}_0^{1,2}(\Omega)$  bzw.  $\mathbf{H}_0^1(\Omega)$ .

Neben den Sobolev-Normen  $\|\cdot\|_{l,p,\Omega}$  benötigt man außerdem die **Sobolev-Halbnormen**

$$|v|_{l,p,\Omega} := \int_{\Omega} \left( \sum_{|\alpha|=l} |(D^\alpha v)(x)|^p dx \right)^{\frac{1}{p}} \quad \text{für } 1 \leq p < \infty \quad (1.15)$$

und

$$|v|_{l,p,\Omega} := \sum_{\|\alpha\|_1=l} \|D^\alpha v\|_{\mathbf{L}^\infty(\Omega)} \quad \text{für } p = \infty. \quad (1.16)$$

**Definition 1.13 Sobolev-Slobodeckij-Räume** sind Sobolev-Räume  $\mathbf{W}^{s,p}(\Omega)$  mit einem rationalen  $s = k + \beta$  mit  $k \in \mathbb{N}_0$  und  $\beta \in (0, 1) \subset \mathbb{R}$ . Sie werden definiert als

$$\mathbf{W}^{s,p}(\Omega) := \{v \in \mathbf{W}^{k,p}(\Omega) : \|v\|_{\mathbf{W}^{s,p}(\Omega)} < \infty\} \quad (1.17)$$

mit den Normen

$$\begin{aligned} \|v\|_{\mathbf{W}^{s,p}(\Omega)} &:= \left( \|v\|_{\mathbf{W}^{k,p}(\Omega)}^p + |v|_{\mathbf{W}^{s,p}(\Omega)}^p \right)^{\frac{1}{p}} \\ |v|_{\mathbf{W}^{s,p}(\Omega)} &:= \left( \sum_{|\alpha|=k} \int_{\Omega} \int_{\Omega} \frac{|D^\alpha u(x) - D^\alpha u(y)|^p}{|x-y|^{n+\beta p}} dx dy \right)^{\frac{1}{p}} \end{aligned}$$

für  $1 \leq p < \infty$ . Analog erfolgt die Definition für  $p = \infty$ . Dabei ist  $n$  die Dimension des Raumes  $\mathbb{R}^n$ .

**Bemerkung 1.14** Die Sobolev-Slobodeckij-Räume sind vollständig bezüglich ihrer Norm. Außerdem ist  $\mathbf{W}^{s,p}(\Omega)$  stetig in  $\mathbf{C}(\bar{\Omega})$  eingebettet, falls  $s > n/p$  ist (vgl. [Ada75], 7.36 und 7.52).

**Norm–Abschätzungen**

Es gilt nach Definition der Halbnorm

$$|v|_{k,p,\Omega} \leq \|v\|_{l,p,\Omega} \quad \text{für alle } v \in \mathbf{W}^{l,p}(\Omega) \text{ und } k, l \in \mathbb{N}_0 \text{ mit } 0 \leq k \leq l. \quad (1.18)$$

Nach unten lässt sich die Halbnorm für  $k = 1$  mit der **Friedrichs'schen Ungleichung** abschätzen:

**Lemma 1.15** *Sei  $\Omega \subset \mathbb{R}^n$  ein beschränktes Gebiet. Dann existiert eine Konstante  $c_F > 0$ , so dass*

$$\|v\|_{0,\Omega} \leq c_F |v|_{1,\Omega} \quad \text{für alle } v \in \mathbf{W}_0^{1,2}(\Omega) \quad (1.19)$$

*gilt.*

**Beweis :** Da der Raum  $\mathbf{C}_0^\infty(\Omega)$  nach Bemerkung 1.12 dicht im  $\mathbf{W}_0^{1,2}(\Omega)$  liegt, reicht es, den Beweis für Funktionen  $v \in \mathbf{C}_0^\infty(\Omega)$  durchzuführen. Wegen der Beschränktheit kann man  $\Omega$  als Teilmenge eines  $n$ -dimensionalen, offenen Würfels  $W := \{x = (x_1, \dots, x_n) : -s < x_i < s\}$  mit der Kantenlänge  $2s$  annehmen. Außerhalb von  $\Omega$ , also in  $W \setminus \Omega$ , sei  $v \equiv 0$ . Dann folgt

$$\begin{aligned} v(x_1, x_2, \dots, x_n) &= \underbrace{v(-s, x_2, \dots, x_n)}_{=0, \text{ da } v \in \mathbf{C}_0^\infty(\Omega)} + \int_{-s}^{x_1} \partial_1 v(t, x_2, \dots, x_n) dt \\ &= \int_{-s}^{x_1} \partial_1 v(t, x_2, \dots, x_n) dt. \end{aligned}$$

Mit der Cauchy–Schwarzschen Ungleichung erhält man

$$|v(x)|^2 \leq \int_{-s}^{x_1} 1^2 dt \int_{-s}^{x_1} |\partial_1 v(t, x_2, \dots, x_n)|^2 dt.$$

Jetzt vergrößert man die rechte Seite, indem man nicht nur bis  $x_1$ , sondern bis zum Rand  $s$  in  $x_1$ -Richtung integriert. Dann ist

$$|v(x)|^2 \leq (x_1 + s) \int_{-s}^s |\partial_1 v(t, x_2, \dots, x_n)|^2 dt.$$

Anschließend wird die ganze Gleichung nach  $x_1$  integriert. Da das Integral der rechten Seite von  $x_1$  unabhängig ist, gilt

$$\int_{-s}^s |v(x)|^2 dx_1 \leq \underbrace{\int_{-s}^s (x_1 + s) dx_1}_{=2s^2} \int_{-s}^s |\partial_1 v(x)|^2 dx_1.$$

Schließlich integriert man über die anderen Koordinaten  $x_2, \dots, x_n$  und erhält

$$\|v\|_{0,\Omega}^2 = \int_W |v(x)|^2 \leq 2s^2 \int_W |\partial_1 v(x)|^2 dx \leq c_F^2 |v|_{1,\Omega}^2.$$

Jetzt zieht man noch die Wurzel und erhält die Behauptung.  $\square$

Die Sobolev-Halbnorm ist also äquivalent zur Sobolevnorm. Es gilt somit

$$\|v\|_{0,\Omega} \leq \|v\|_{1,\Omega} \leq \sqrt{1 + c_F^2} |v|_{1,\Omega} \leq \sqrt{1 + c_F^2} \|v\|_{1,\Omega}. \quad (1.20)$$

## 1.2 Fortsetzbarkeit von Funktionen auf den Gebietsrand

Als nächstes betrachtet man die Fortsetzbarkeit von Funktionen  $v \in \mathbf{W}^{1,p}(\Omega)$  auf den Rand  $\Gamma$ . Die Werte der Funktionen auf dem Rand können bei geeigneten Voraussetzungen sinnvoll sein. Dazu definiert man eine Spurabbildung.

**Satz 1.16** *Sei  $\Gamma$  Lipschitz-stetiger Rand von  $\Omega$  und  $1 \leq p \leq \infty$ . Dann gibt es genau eine stetige lineare Abbildung*

$$\hat{\gamma} : \mathbf{W}^{1,p}(\Omega) \rightarrow \mathbf{L}^p(\Gamma), \quad (1.21)$$

so dass  $\hat{\gamma}(v) = v|_{\Gamma}$  für alle  $v \in \mathbf{W}^{1,p}(\Omega) \cap \mathbf{C}^0(\bar{\Omega})$ .  $\hat{\gamma}$  heißt **Spuroperator** oder **Spurabbildung**. Außerdem existiert für  $p = 2$  eine Konstante  $c > 0$ , so dass

$$\|\hat{\gamma}(v)\|_{0,\Gamma} \leq c \|v\|_{1,\Omega} \quad \text{für alle } v \in \mathbf{W}^{1,2}(\Omega) \quad (1.22)$$

gilt.

**Beweis :** siehe [Ada75], 5.22.  $\square$

Es lassen sich Funktionenräume über den Rand  $\Gamma$  definieren. Der wichtigste ist

$$\mathbf{H}^{\frac{1}{2}}(\Gamma) := \{w \in \mathbf{L}^2(\Gamma) : \exists v \in \mathbf{H}^1(\Omega) \text{ mit } w = \hat{\gamma}(v)\} \quad (1.23)$$

mit der Norm

$$\|w\|_{\frac{1}{2},\Gamma} := \|w\|_{\mathbf{H}^{\frac{1}{2}}(\Gamma)} := \inf\{\|v\|_{1,\Omega} : v \in \mathbf{H}^1(\Omega), w = \hat{\gamma}(v)\}. \quad (1.24)$$

Der zu  $\mathbf{H}^{\frac{1}{2}}(\Gamma)$  gehörende Dualraum heißt  $\mathbf{H}^{-\frac{1}{2}}(\Gamma)$ , das ist die Menge der stetigen linearen Funktionale auf  $\mathbf{H}^{\frac{1}{2}}(\Gamma)$ , mit der Norm

$$\|g\|_{-\frac{1}{2},\Gamma} := \|g\|_{\mathbf{H}^{-\frac{1}{2}}(\Gamma)} := \sup_{w \in \mathbf{H}^{\frac{1}{2}}(\Gamma)} \frac{\langle g, w \rangle}{\|w\|_{\frac{1}{2},\Gamma}} \quad \text{für alle } g \in \mathbf{H}^{-\frac{1}{2}}(\Gamma), \quad (1.25)$$

wobei

$$\langle \cdot, \cdot \rangle : \mathbf{H}^{-\frac{1}{2}}(\Gamma) \times \mathbf{H}^{\frac{1}{2}}(\Gamma) \rightarrow \mathbb{R} \quad (1.26)$$

das duale Produkt ist.

Für Funktionen  $v \in \mathbf{H}^1(\Omega)$  mit  $\Delta v \in \mathbf{L}^2(\Omega)$  gilt:  $\frac{\partial v}{\partial n} \in \mathbf{H}^{-\frac{1}{2}}(\Gamma)$  und

$$\left\| \frac{\partial v}{\partial n} \right\|_{-\frac{1}{2}, \Gamma} \leq C \left( \|v\|_{1, \Omega} + \|\Delta v\|_{0, \Omega} \right). \quad (1.27)$$

Eine ausführlichere Behandlung der Fortsetzung von Funktionen auf den Gebietsrand findet man in [Lio68], Abschnitt 9.

### 1.3 Lösbarkeitstheorie von Lax–Milgram

In diesem Teil der Arbeit wird die für die Lösbarkeit von elliptischen Variationsgleichungen fundamentale Theorie vorgestellt. Das Dirichlet–Randwertproblem (0.1), (0.2) wird mit Hilfe dieser Theorie später auf Lösbarkeit untersucht.

Im folgenden sei  $\mathbf{X}$  ein Hilbertraum mit dem Skalarprodukt  $(\cdot, \cdot)$  und der induzierten Norm  $\|\cdot\|_{\mathbf{X}} = \sqrt{(\cdot, \cdot)}$ . Dann heißt  $\mathbf{X}^* = \mathcal{L}(\mathbf{X}, \mathbb{R})$  der zugehörige Dualraum mit der Schreibweise

$$\langle f, v \rangle := f(v) \quad \text{für } f \in \mathbf{X}^* \text{ und alle } v \in \mathbf{X}. \quad (1.28)$$

Aus der Symmetrie des Skalarproduktes erhält man die folgende Form des **Riesz’schen Darstellungssatzes**.

**Lemma 1.17** *Sei  $\mathbf{X}$  ein reeller Hilbertraum. Dann existiert zu jedem Funktional  $f \in \mathbf{X}^*$  ein eindeutig bestimmtes Element  $u \in \mathbf{X}$ , so dass*

$$f(v) = \langle f, v \rangle = (u, v) \quad \text{für alle } v \in \mathbf{X} \quad (1.29)$$

*gilt. Der Riesz’sche Darstellungs–Operator  $R : \mathbf{X}^* \rightarrow \mathbf{X}$  mit  $f \mapsto u$  ist linear, bijektiv und isometrisch.*

**Beweis :** siehe [Lub02], Lemma 6.1. □

Wie für Linearformen ergibt sich analog eine Darstellung für Bilinearformen als Folgerung aus Lemma 1.17.

**Lemma 1.18** *Es gibt genau einen beschränkten Operator  $A = \mathcal{L}(\mathbf{X}, \mathbf{X}^*)$ , so dass*

$$a(u, v) = \langle Au, v \rangle \quad \text{für alle } u, v \in \mathbf{X}. \quad (1.30)$$

*gilt.*

Im folgenden betrachtet man die Variationsgleichung:

Finde  $u \in \mathbf{X}$ , so dass für die Bilinearform  $a(\cdot, \cdot)$  und die Linearform  $f(\cdot)$

$$a(u, v) = f(v) \quad \text{für alle } v \in \mathbf{X} \quad (1.31)$$

*gilt.*

Es gibt nach dem folgenden Lemma 1.21 für die Variationsgleichung eine eindeutige Lösung im Raum  $\mathbf{X}$ .

Dazu werden Bedingungen an  $a(\cdot, \cdot)$  und  $f(\cdot)$  gestellt:

**Definition 1.19** (i) Die Bilinearform  $a(\cdot, \cdot) : \mathbf{X} \times \mathbf{X} \rightarrow \mathbb{R}$  ist **stetig**, wenn eine Konstante  $M > 0$  existiert mit

$$|a(u, v)| \leq M \|u\|_{\mathbf{X}} \|v\|_{\mathbf{X}} \quad \text{für alle } u, v \in \mathbf{X}. \quad (1.32)$$

(ii) Die Linearform  $f(\cdot) : \mathbf{X} \rightarrow \mathbb{R}$  ist **stetig**, wenn eine Konstante  $C > 0$  existiert mit

$$|f(v)| \leq C \|v\|_{\mathbf{X}} \quad \text{für alle } v \in \mathbf{X}. \quad (1.33)$$

(iii)  $a(\cdot, \cdot)$  ist  **$\mathbf{X}$ -elliptisch**, wenn eine Konstante  $\gamma > 0$  existiert mit

$$a(v, v) \geq \gamma \|v\|_{\mathbf{X}}^2 \quad \text{für alle } v \in \mathbf{X}. \quad (1.34)$$

**Bezeichnung 1.20** Ist die Bilinearform  $(\cdot, \cdot)$   **$\mathbf{X}$ -elliptisch**, wird die Variationsgleichung (1.31) auch als **elliptische Variationsgleichung** bezeichnet.

**Lemma 1.21 (Lax–Milgram)** Die Bilinearform  $a(\cdot, \cdot)$  und die Linearform  $f(\cdot)$  seien im Hilbertraum  $\mathbf{X}$  stetig. Außerdem sei  $a(\cdot, \cdot)$   **$\mathbf{X}$ -elliptisch**. Dann besitzt die elliptische Variationsgleichung (1.31) genau eine Lösung.

**Beweis :** Wegen der Stetigkeit von  $a(\cdot, \cdot)$  existiert ein Funktional  $A : \mathbf{X} \rightarrow \mathbf{X}^*$  mit  $Au(v) = a(u, v)$  für alle  $v \in \mathbf{X}$  und ein beliebiges  $u \in \mathbf{X}$ . Das Funktional ist linear und stetig, denn

$$\begin{aligned} Au(\alpha v_1 + \beta v_2) &= a(u, \alpha v_1 + \beta v_2) \\ &= \alpha a(u, v_1) + \beta a(u, v_2) \\ &= \alpha Au(v_1) + \beta Au(v_2) \quad \text{für alle } v_1, v_2 \in \mathbf{X} \text{ und } \alpha, \beta \in \mathbb{R}, \\ |Au(v)| &= |a(u, v)| \\ &\stackrel{(1.32)}{\leq} M \|u\|_{\mathbf{X}} \|v\|_{\mathbf{X}}. \end{aligned}$$

Daraus ergibt sich

$$\|Au\|_{\mathbf{X}^*} = \sup_{v \neq 0} \frac{|Au(v)|}{\|v\|_{\mathbf{X}}} \leq M \|u\|_{\mathbf{X}} \leq \infty.$$

Also ist  $Au$  wohldefiniert.

Nach dem Riesz'schen Darstellungssatz (1.17) existiert für ein  $\phi \in \mathbf{X}^*$  genau ein  $R\phi \in \mathbf{X}$ , so dass  $\phi(v) = (R\phi, v)$  für ein  $v \in \mathbf{X}$  gilt. Gesucht ist damit ein  $u \in \mathbf{X}$  mit

$$Au = f \quad \text{in } \mathbf{X}^* \quad (1.35)$$

$$RAu = Rf \quad \text{in } \mathbf{X}. \quad (1.36)$$

Um letztere Gleichung zu lösen, betrachtet man die Abbildung  $T : \mathbf{X} \rightarrow \mathbf{X}$  mit

$$Tv := v - \rho(RAv - Rf) \quad (1.37)$$

und einem  $\rho \neq 0$ , das so gewählt wird, dass  $T$  kontrahierend ist. Dann gibt es nach dem Banach'schen Fixpunktsatz für kontrahierende Abbildungen genau ein  $u \in \mathbf{X}$  mit

$$Tu = u - \rho(RAu - Rf) = u, \quad (1.38)$$

d. h.  $RAu - Rf = 0$ .

Es muss also noch die Existenz des  $\rho \neq 0$  gezeigt werden. Dazu betrachtet man ein  $v \in \mathbf{X}$  mit  $v := v_1 - v_2$ . Dann gilt

$$\begin{aligned}
\|Tv_1 - Tv_2\|_{\mathbf{X}}^2 &= \|v_1 - \rho(RAv_1 - Rf) - v_2 + \rho(RAv_2 - Rf)\|_{\mathbf{X}}^2 \\
&\stackrel{A,R \text{ linear}}{=} \|v - \rho(RAv)\|_{\mathbf{X}}^2 \\
&= \|v\|_{\mathbf{X}}^2 - 2\rho(RAv, v) + \rho^2 \|RAv\|_{\mathbf{X}}^2 \\
&\stackrel{\text{Def. von } A \text{ und } R}{=} \|v\|_{\mathbf{X}}^2 - 2\rho a(v, v) + \rho^2 a(v, RAv) \\
&\stackrel{(1.32)(1.34)}{\leq} \|v\|_{\mathbf{X}}^2 - 2\rho\gamma \|v\|_{\mathbf{X}}^2 + \rho^2 M \|v\|_{\mathbf{X}} \underbrace{\|RAv\|_{\mathbf{X}}}_{\leq \|Av\|_{\mathbf{X}^*} \leq M \|v\|_{\mathbf{X}}} \\
&\leq \|v\|_{\mathbf{X}}^2 (1 - 2\rho\gamma + \rho^2 M^2) \\
&= L^2 \|v_1 - v_2\|_{\mathbf{X}}^2.
\end{aligned}$$

Bestimme  $\rho > 0$  so, dass  $L < 1$  ist, denn dann ist  $T$  kontrahierend.

$$\begin{aligned}
1 - 2\rho\gamma + \rho^2 M^2 &< 1 \\
\Leftrightarrow \rho(\rho M^2 - 2\gamma) &< 0 \\
\Leftrightarrow 0 &< \rho < \frac{2\gamma}{M^2}.
\end{aligned}$$

Damit ist das Lemma bewiesen. □

## 1.4 Konforme Galerkin-Approximation

In diesem Abschnitt wird zur numerischen Lösung von elliptischen Variationsgleichungen die Variationsgleichung (1.31) im unendlich dimensionalen Hilbertraum  $\mathbf{X}$  durch ein lineares Gleichungssystem im endlich dimensionalen Raum  $\mathbf{X}_h$  approximiert, das sich numerisch auf viele Arten effizient lösen lässt. Man spricht von einer **konformen Approximation**, wenn  $\mathbf{X}_h \subset \mathbf{X}$  gilt.

**Definition 1.22** Seien  $a(\cdot, \cdot)$  und  $f(\cdot)$  aus (1.31) gegeben. Dann bezeichnet man die elliptische Variationsgleichung

Finde  $u_h \in \mathbf{X}_h$ , so dass

$$a(u_h, v) = f(v) \quad \text{für alle } v \in \mathbf{X}_h \text{ gilt.} \quad (1.39)$$

als **Ritz-Galerkin-Verfahren** der elliptischen Variationsgleichung (1.31).

Eine Abschätzung des Fehlers zwischen der Lösung auf dem Raum  $\mathbf{X}$  und der Lösung auf dem Raum  $\mathbf{X}_h \subset \mathbf{X}$  liefert das nächste Lemma.

**Lemma 1.23 (Cea)** Sei  $a(\cdot, \cdot)$  die stetige und  $\mathbf{X}$ -elliptische Bilinearform aus (1.31). Dann sind für jedes  $f(\cdot)$  aus (1.31) die elliptischen Variationsgleichungen (1.31) und (1.39) eindeutig lösbar

nach Lemma 1.21. Für die Lösungen  $u \in \mathbf{X}$  bzw.  $u_h \in \mathbf{X}_h$  gilt dann die Abschätzung

$$\|u - u_h\| \leq \frac{M}{\gamma} \inf_{v \in \mathbf{X}_h} \|u - v\| \quad (1.40)$$

mit den Konstanten  $M$  aus (1.32) und  $\gamma$  aus (1.34).

**Beweis :** Es gilt nach Voraussetzung

$$a(u, v) = f(v) = a(u_h, v) \quad \text{für alle } v \in \mathbf{X}_h,$$

denn  $\mathbf{X}_h \subset \mathbf{X}$ . Umgeformt ergibt das

$$a(u - u_h, v) = 0 \quad \text{für alle } v \in \mathbf{X}_h, \quad (1.41)$$

die sogenannte **Galerkin-Orthogonalität**. Es folgt

$$\begin{aligned} \gamma \|u - u_h\|^2 &\leq a(u - u_h, u - u_h) \\ &= a(u - u_h, u - v) + \underbrace{a(u - u_h, v - u_h)}_{=0 \text{ nach (1.41)}} \\ &\leq M \|u - u_h\| \|u - v\| \quad \text{für alle } v \in \mathbf{X}_h. \end{aligned}$$

Nach Division durch  $\gamma \|u - u_h\| \neq 0$  erhält man

$$\|u - u_h\| \leq \frac{M}{\gamma} \|u - v\| \quad \text{für alle } v \in \mathbf{X}_h$$

und damit die Behauptung. □

Man kann eine Beziehung zwischen (1.39) und einem linearen Gleichungssystem herstellen. Sei  $\mathcal{B} := \{\phi_i\}_{i=1}^n$  eine Basis des Raumes  $\mathbf{X}_h$ . Dann lässt sich jedes  $v \in \mathbf{X}_h$  als eine Linearkombination

$$v = \sum_{i=1}^n v_i \phi_i \quad \text{mit } \vec{v} = (v_1, \dots, v_n)^T \in \mathbb{R}^n \quad (1.42)$$

darstellen. Man definiert

$$a(\phi_j, \phi_i) =: a_{ij}, \quad (1.43)$$

$$\mathcal{A} = (a_{ij})_{i,j=1}^n \in \mathbb{R}^{n \times n}, \quad (1.44)$$

$$u_h = \sum_{i=1}^n u_{hi} \phi_i \quad \text{mit } \vec{u}_h = (u_{h1}, \dots, u_{hn})^T \in \mathbb{R}^n \quad \text{und} \quad (1.45)$$

$$f_i = f(\phi_i) \quad \text{mit } \vec{f} = (f_1, \dots, f_n)^T \in \mathbb{R}^n. \quad (1.46)$$

**Lemma 1.24** *Das Ritz-Galerkin-Verfahren (1.39) ist äquivalent zu dem linearen Gleichungssystem*

$$\mathcal{A} \vec{u}_h = \vec{f}. \quad (1.47)$$

**Beweis :** Da  $a(\cdot, \cdot)$  und  $f(\cdot)$  linear sind, folgt die Äquivalenz, wenn man fordert, dass

$$a(u_h, \phi_i) = f(\phi_i) \quad \text{für alle } \phi_i \in \mathcal{B} \quad (1.48)$$

gilt. Mit (1.43)-(1.46) erhält man

$$a(u_h, \phi_i) = \sum_{j=1}^n u_{hj} a(\phi_j, \phi_i) = \sum_{j=1}^n a_{ij} u_{hj} = f(\phi_i) \quad \text{für alle } i \in \{1, \dots, n\}.$$

Damit ist alles gezeigt. □

**Bezeichnungen 1.25** *Die Matrix  $\mathcal{A}$  heißt Steifigkeitsmatrix. Die Funktionen in der ersten Komponente der Bilinearform  $a(\cdot, \cdot)$  für  $u_h$  heißen Ansatzfunktionen, in der zweiten Komponente von  $a(\cdot, \cdot)$  stehen Funktionen für  $v$ , die Testfunktionen. Die Matrix  $\mathcal{A}$  ist für Ansatz- und Testfunktionen mit kleinem kompaktem Träger dünnbesetzt, d.h. sie hat wenige von 0 verschiedene Einträge.*

Das lineare Gleichungssystem ist lösbar, denn die Bilinearform  $a(\cdot, \cdot)$  vererbt mit ihrer  $\mathbf{X}$ -Elliptizität die wichtige Regularitätseigenschaft der positiven Definitheit auf die Steifigkeitsmatrix  $\mathcal{A}$ . Diese garantiert die eindeutige Lösbarkeit von linearen Gleichungssystemen.



# Kapitel 2

## Das kontinuierliche Problem

Hier wird zuerst aus dem klassischen der schwache Lösungsbegriff für lineare, elliptische Differentialgleichungen zweiter Ordnung hergeleitet. Danach überführt man diesen auf elliptische Variationsgleichungen des Typs (1.31). Anschließend wird die Regularität der Lösung erörtert. Zum Abschluss des Kapitels wird das bei  $0 < \varepsilon \ll 1$  auftretende, singulär gestörte Problem der linearen, elliptischen Differentialgleichung zweiter Ordnung betrachtet.

### 2.1 Klassische und schwache Lösung

**Definition 2.1** Für eine lineare, elliptische Differentialgleichung zweiter Ordnung heißt die **klassische Formulierung** des Randwertproblems mit homogenen Dirichlet-Randbedingungen im Gebiet  $\Omega \subset \mathbb{R}^n$ ,  $n = 2, 3$ :

Finde  $u \in \mathbf{C}^2(\Omega) \cap \mathbf{C}(\overline{\Omega})$ , so dass

$$-\varepsilon \Delta u + \vec{b} \nabla u + cu = f \quad \text{in } \Omega \tag{2.1}$$

$$u = 0 \quad \text{auf } \Gamma \tag{2.2}$$

gilt.

Dabei sind  $\varepsilon \in \mathbb{R}_+$ ,  $\vec{b}(x) = (b_1(x), b_2(x), \dots, b_n(x))^T \in \mathbb{R}^n$ , und  $c(x) \geq 0$  für alle  $x = (x_1, x_2, \dots, x_n) \in \Omega$  und  $b_i, c, f \in \mathbf{C}(\Omega)$  für alle  $i \in \{1, \dots, n\}$ .

Im folgenden wird eine Umformung der klassischen Formulierung in eine schwache Formulierung durchgeführt.

Zuerst multipliziert man die Gleichung (2.1) mit einer Funktion  $v$  aus dem Raum  $\mathbf{C}_0^\infty(\Omega)$  und integriert die Gleichung über  $\Omega$ . Dann erhält man

$$-\varepsilon \int_{\Omega} \Delta u(x)v(x)dx + \int_{\Omega} \left( \vec{b}(x) \nabla u(x)v(x) + c(x)u(x)v(x) \right) dx = \int_{\Omega} f(x)v(x)dx.$$

In dieser Gleichung integriert man das erste Integral partiell und hat

$$\varepsilon \int_{\Omega} \nabla u(x) \nabla v(x) dx - \varepsilon \underbrace{\int_{\Gamma} \frac{\partial u}{\partial n} v ds}_{=0} + \int_{\Omega} \vec{b}(x) \nabla u(x)v(x) + c(x)u(x)v(x) dx = \int_{\Omega} f(x)v(x) dx.$$

Das Integral über den Rand  $\Gamma$  des Gebietes  $\Omega$  verschwindet, da  $v = 0$  auf  $\Gamma$  ist. Man erweitert den Lösungsraum, indem man für die Funktionen  $u, v$  von  $\mathbf{C}_0^\infty(\Omega)$  auf  $\mathbf{X} := \mathbf{W}_0^{1,2}(\Omega)$  übergeht, der den Raum  $\mathbf{C}_0^\infty(\Omega)$  nach Bemerkung 1.12 dicht enthält. Dann definiert man

$$a(u, v) := \int_{\Omega} \left( \varepsilon \nabla u(x) \nabla v(x) + \vec{b}(x) \nabla u(x) v(x) + c(x) u(x) v(x) \right) dx \quad \text{für alle } u, v \in \mathbf{X} \quad (2.3)$$

und

$$f(v) := \int_{\Omega} f(x) v(x) dx \quad \text{für alle } v \in \mathbf{X}. \quad (2.4)$$

Daraus ergibt sich die **Variationsgleichung** (vgl. (1.31)).

**Definition 2.2** Die schwache oder verallgemeinerte Formulierung von (2.1) mit homogenen Dirichlet-Randbedingungen (2.2) lautet:

Finde  $u \in \mathbf{X}$ , so dass

$$a(u, v) = f(v) \quad \text{für alle } v \in \mathbf{X} \quad (2.5)$$

gilt. Dabei ist  $u \in \mathbf{X}$  die **schwache Lösung** von (2.1).

Dass  $a(\cdot, \cdot)$  eine Bilinearform und  $f(\cdot)$  eine Linearform sind, lässt sich leicht nachrechnen. Im Fall  $\vec{b}(x) \equiv 0$  für alle  $x \in \Omega$  ist  $a(\cdot, \cdot)$  sogar symmetrisch. Notwendige Voraussetzungen für  $a(\cdot, \cdot)$  und  $f(\cdot)$  zur Anwendbarkeit der Theorie von Lax–Milgram aus Kapitel 1 für elliptische Variationsgleichungen werden im nächsten Lemma bewiesen.

### Lemma 2.3

(i) Die Bilinearform  $a(\cdot, \cdot) : \mathbf{X} \times \mathbf{X} \rightarrow \mathbb{R}$  ist stetig,

(ii) Die Linearform  $f(\cdot) : \mathbf{X} \rightarrow \mathbb{R}$  ist stetig,

(iii)  $a(\cdot, \cdot)$  ist  $\mathbf{X}$ -elliptisch, wenn

$$\nabla \cdot \vec{b} = \sum_{i=1}^n \frac{\partial b_i}{\partial x_i} \in \mathbf{L}^\infty(\Omega) \quad (2.6)$$

und

$$c - \frac{1}{2} \nabla \cdot \vec{b} \geq c_0 \quad \text{fast überall in } \Omega \quad (2.7)$$

mit einer Konstanten  $c_0 \geq 0$  gelten.

**Beweis :**

(i)  $a(\cdot, \cdot)$  ist stetig, denn mit Hölder- und Cauchy–Schwarz–Ungleichung gilt

$$|a(u, v)| = \left| \int_{\Omega} \varepsilon \nabla u(x) \nabla v(x) + \vec{b}(x) \nabla u(x) v(x) + c(x) u(x) v(x) dx \right|$$

$$\begin{aligned}
&\leq \varepsilon |u|_{1,\Omega} |v|_{1,\Omega} + \sum_{i=1}^n \left( \|b_i\|_{\mathbf{L}^\infty(\Omega)} \left\| \frac{\partial u}{\partial x_i} \right\|_{0,\Omega} \right) \|v\|_{0,\Omega} + \|c\|_{\mathbf{L}^\infty(\Omega)} \|u\|_{0,\Omega} \|v\|_{0,\Omega} \\
&\leq (\varepsilon + \|c\|_{\mathbf{L}^\infty(\Omega)}) \|u\|_{1,\Omega} \|v\|_{1,\Omega} + \|v\|_{0,\Omega} \underbrace{\sqrt{\sum_{i=1}^n \|b_i\|_{\mathbf{L}^\infty(\Omega)}^2}}_{=:\|\vec{b}\|_2} \underbrace{\sqrt{\sum_{i=1}^n \left\| \frac{\partial u}{\partial x_i} \right\|_{0,\Omega}^2}}_{\leq |u|_{1,\Omega} \leq \|u\|_{1,\Omega}} \\
&\leq \underbrace{(\varepsilon + \|\vec{b}\|_2 + \|c\|_{\mathbf{L}^\infty(\Omega)})}_{=:M} \|u\|_{\mathbf{X}} \|v\|_{\mathbf{X}}.
\end{aligned}$$

(ii)  $f(\cdot)$  ist stetig, denn mit Hölder- und Friedrichs-Ungleichung gilt

$$\begin{aligned}
|f(v)| &\leq \|f\|_{0,\Omega} \|v\|_{0,\Omega} \\
&\leq \underbrace{C_1 \|f\|_{0,\Omega}}_{=:C} \|v\|_{1,\Omega} \\
&\leq C \|v\|_{\mathbf{X}}.
\end{aligned}$$

(iii)  $a(\cdot, \cdot)$  ist  $\mathbf{X}$ -elliptisch, denn mit partieller Integration des Konvektionsterms und den Voraussetzungen (2.6) und (2.7) folgt

$$\begin{aligned}
a(v, v) &= \int_{\Omega} \varepsilon (\nabla v(x))^2 + \vec{b}(x) \nabla v(x) v(x) + c(x) v(x)^2 dx \\
&= \int_{\Omega} \varepsilon (\nabla v(x))^2 + c(x) v(x)^2 dx - \frac{1}{2} \int_{\Omega} \nabla \cdot \vec{b}(x) v(x)^2 dx + \frac{1}{2} \underbrace{\int_{\Gamma} \vec{b}(x) v(x)^2 ds}_{=0} \\
&\geq \varepsilon \underbrace{|v|_{1,\Omega}^2}_{\geq \frac{1}{1+c_F^2} \|v\|_{1,\Omega}^2} + \int_{\Omega} \underbrace{(c(x) - \frac{1}{2} \nabla \cdot \vec{b}(x)) v(x)^2 dx}_{\geq c_0} \\
&\geq \frac{\varepsilon}{1+c_F^2} \|v\|_{\mathbf{X}}^2 + c_0 \|v\|_{0,\Omega}^2 \\
&\geq \gamma \|v\|_{\mathbf{X}}^2
\end{aligned} \tag{2.8}$$

mit  $\gamma = \min\{\frac{\varepsilon}{1+c_F^2}, c_0\}$ . □

Aus diesen Ergebnissen folgt die eindeutige Existenz einer Lösung der nachweislich  $\mathbf{X}$ -elliptischen Variationsgleichung (2.5) mit der Lax-Milgram-Theorie. Es fehlt noch eine Aussage über die Regularität dieser Lösung in  $\mathbf{X} = \mathbf{W}_0^{1,2}(\Omega)$ .

## 2.2 Regularität der schwachen Lösung

In diesem Abschnitt werden die nach dem Lax-Milgram-Lemma (1.21) existierenden schwachen Lösungen von (2.5) genauer untersucht. Dabei wird die Frage aufgegriffen, wie regulär diese Lösungen eigentlich sind.

Man betrachtet dazu das homogene Dirichlet–Randwertproblem (2.1), (2.2). Es gelte für die Koeffizienten

$$\begin{aligned} \varepsilon &\in \mathbb{R}_+, \\ b_j, c &\in \mathbf{L}^\infty(\Omega) \quad \text{für } j = 1, \dots, n \\ \text{und} \quad f &\in \mathbf{L}^2(\Omega). \end{aligned}$$

Dann folgt für konvexe Gebiete  $\Omega$ , dass  $u \in \mathbf{W}^{2,2}(\Omega)$  ist. Das gilt im allgemeinen nicht für nicht konvexe Gebiete, wie das folgende Beispiel zeigt.

**Beispiel 2.4** Die Funktion  $u(x, y) \equiv u(r, \phi) := r^\beta \sin(\beta\phi)$  mit  $\beta := \pi/\phi_0$ ,  $\phi_0 \in ]0, 2\pi]$  und den Polarkoordinaten  $r = \sqrt{x^2 + y^2}$  und  $\phi = \arctan \frac{y}{x}$  ist Lösung des Problems:  
Finde  $u \in \mathbf{X}$ , so dass

$$\begin{aligned} -\Delta u &= 0 && \text{in } \Omega, \\ u(r, 0) &= u(r, \phi_0) = 0, \\ u(0, \phi) &= 0, \\ u(R, \phi) &= R^\beta \sin(\beta\phi) && \text{auf } \Gamma \end{aligned}$$

gilt.

Dabei ist  $\Omega = \{(r, \phi) : 0 \leq r < R \text{ und } 0 < \phi < \phi_0\} \subset \mathbb{R}^2$ .  $u$  hat in der zweiten schwachen Ableitung bei  $r = 0$  eine Singularität für  $\pi < \phi_0 \leq 2\pi$  und liegt deshalb nicht in  $\mathbf{W}^{2,2}(\Omega)$  (vgl. Abbildung 2.1 mit  $\phi_0 = \frac{3}{2}\pi$ ).

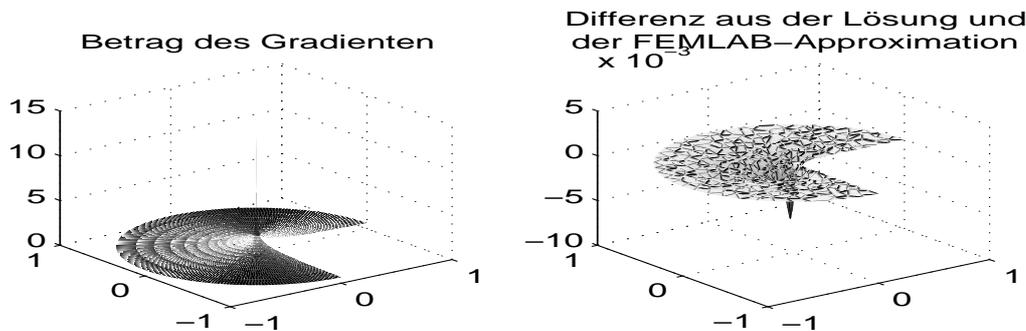


Abbildung 2.1: Der Betrag des Gradienten von  $u$  und der Approximationsfehler bei  $\beta = \frac{2}{3}$

Im folgenden werden Aussagen über die Regularität der Lösung  $u$  auch für nicht konvexe Gebiete getroffen. Dabei müssen stärkere Bedingungen an die Problemdata gestellt werden. Der folgende Satz zeigt zunächst, dass die Lösung des elliptischen Randwertproblems (2.1), (2.2) in der Regel nur lokale Singularitäten hat. In Beispiel 2.4 befindet sich die Singularität an der „einspringenden“ Ecke im Punkt  $(0|0)$ , ist also lokal.

**Satz 2.5** Sei  $\Omega$  offen und  $u \in \mathbf{W}^{1,2}(\Omega)$  schwache Lösung des homogenen Dirichlet-Problems aus (2.1), (2.2). Sei  $m \geq 0$  und für die Problemdaten gelte

$$\begin{aligned} f &\in \mathbf{W}^{m,2}(\Omega) && (\in \mathbf{L}^2(\Omega) \text{ für } m = 0), \\ b_j, c &\in \mathbf{W}^{m+1,2}(\Omega) && \text{für } j = 1, \dots, n. \end{aligned} \quad (2.9)$$

Dann ist  $u \in \mathbf{W}_{loc}^{m+2,2}(\Omega)$ , d.h.  $u \in \mathbf{W}^{m+2,2}(G)$  für jede offene Teilmenge  $G \subset \Omega$ , dessen Abschluss  $\bar{G}$  kompakt ist.

**Beweis :** siehe [Alt99], Satz A 10.2. □

Die lokalen Singularitäten können unter den Voraussetzungen des letzten Satzes also höchstens auf dem Rand  $\Gamma$  vom Gebiet  $\Omega$  liegen. Bekommt der Rand eine zusätzliche Regularität wie im nächsten Satz, ist die Lösung sogar bis zum Rand glatt.

**Satz 2.6** Seien  $\Omega$  offen und beschränkt,  $\Gamma \in \mathbf{C}^{m+1,1}(\bar{\Omega})$  der Rand von  $\Omega$  und  $u \in \mathbf{W}^{1,2}(\Omega)$  die schwache Lösung des homogenen Dirichlet-Randwert-Problems aus (2.1), (2.2) mit den Voraussetzungen (2.9) gegeben. Dann folgt

$$u \in \mathbf{W}^{m+2,2}(\Omega). \quad (2.10)$$

**Beweis :** siehe [Alt99], Satz A 10.3. □

## 2.3 Singulär gestörte Probleme für $0 < \varepsilon \ll 1$

Die Lösbarkeit des Problems (2.5) ist für  $0 < \varepsilon \ll 1$  noch immer mit der Theorie von Lax-Milgram gegeben. Die Bilinearform  $a(\cdot, \cdot)$  aus (2.3) ist für  $\varepsilon \rightarrow 0$  jedoch nur  $\mathbf{X}$ -elliptisch, wenn für die Konstante aus (2.7)  $c_0 > 0$  gilt. Unter dieser Voraussetzung lässt sich dann das Lemma 1.21 anwenden.

In der numerischen Approximations-Lösung können bei einer Elliptizitätskonstanten nahe 0 Oszillationen auftreten, d.h. die numerische Lösung wird ungenau in der Nähe von Sprüngen der Lösung. Sie ist nicht mehr durch den fast verschwundenen Diffusionsterm geglättet. Es ist also eine numerische Stabilisierung erforderlich. Eine mögliche Variante wird in Kapitel 6 vorgestellt.

Insgesamt beobachtet man für  $0 < \varepsilon \ll 1$  ein Lösungsverhalten des Problems (2.1), (2.2), das der Lösung der linearen, hyperbolischen Differentialgleichung erster Ordnung

$$\vec{b}(x)\nabla w(x) + c(x)w(x) = f(x) \quad (2.11)$$

ähnelt. Man nennt diese Gleichung auch **reduzierte Gleichung**. Sie ist linear, weil weder  $\vec{b}$  noch  $c$  von der Lösung  $w$  abhängen.

Die Theorie zu diesem Problem kann auf die Untersuchung geeigneter Systeme gewöhnlicher Differentialgleichungen zurückgeführt werden.

**Definition 2.7** Das System gewöhnlicher Differentialgleichungen

$$\frac{d}{ds} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1(x_1, \dots, x_n) \\ \vdots \\ b_n(x_1, \dots, x_n) \end{pmatrix} \quad (2.12)$$

und

$$\frac{dw}{ds} = f(x(s)) - c(x(s))w(x(s)) \quad (2.13)$$

heißt **charakteristisches System von (2.11)**. Die Lösungskurven  $(x_1, \dots, x_n, w)^T(s)$  im  $\mathbb{R}^n \times \mathbb{R}$  heißen **Charakteristiken** von (2.11). Die Projektion  $(x_1, \dots, x_n)^T(s)$  einer Charakteristik von (2.11) in den  $(x_1, \dots, x_n)$ -Raum heißt **Grundcharakteristik**.

Für die vorliegende lineare, hyperbolische Differentialgleichung sind die Grundcharakteristiken unabhängig von der Lösung  $w$ . Das vereinfacht die Lösung von (2.12), wie in den meisten Fällen. Mehr zum Thema hyperbolischer Differentialgleichungen erster Ordnung findet man in [Lub03], Kapitel 16. Dort wird auch die Lösbarkeit theoretisch abgehandelt.

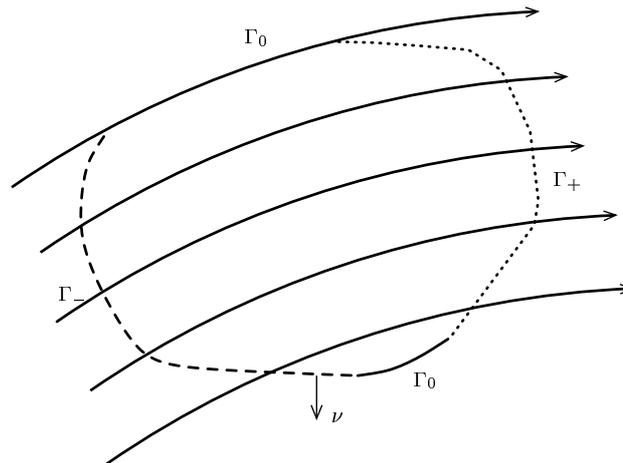


Abbildung 2.2:  $\Omega$  mit Grundcharakteristiken und Teilrändern

Um auf das dem Randwertproblem (2.1),(2.2) entsprechende Problem zurück zu kommen, braucht man noch geeignete Randbedingungen. In der Regel kann man diese nur einem Teil des Randes  $\Gamma$  auferlegen. Es ist daher sinnvoll, den Rand bezüglich dem Verlauf der Charakteristiken von (2.11) in die Abschnitte

$$\begin{aligned} \Gamma_+ &:= \{x \in \Gamma : \vec{b} \cdot \nu > 0\}, \\ \Gamma_- &:= \{x \in \Gamma : \vec{b} \cdot \nu < 0\} \quad \text{und} \\ \Gamma_0 &:= \{x \in \Gamma : \vec{b} \cdot \nu = 0\} \end{aligned}$$

zu trennen, wobei  $\nu$  der nach außen zeigende Einheitsnormalenvektor des Randes  $\Gamma$  ist (vgl. Abbildung 2.2).

Physikalisch bedeutet diese Randeinteilung, dass Charakteristiken bei  $\Gamma_-$  in das Gebiet  $\Omega$  „einströmen“ und das Gebiet bei  $\Gamma_+$  verlassen. Die Randstücke  $\Gamma_0$  sind selbst „charakteristisch“, d.h. sie werden von Grundcharakteristiken von (2.12) gebildet. Gilt in einem Punkt  $x \in \Gamma_0$  die Bedingung  $\vec{b}(x) \neq \vec{0}$ , so laufen die Grundcharakteristiken in einer Umgebung von  $x$  parallel zu  $\Gamma_0$ .

Man beschränkt sich darauf, Randbedingungen auf  $\Gamma_-$  vorzuschreiben. Dann lautet das **reduzierte** lineare, hyperbolische **Randwertproblem** erster Ordnung mit homogenen Dirichlet-Randbedingungen.

Finde  $w \in \mathbf{X}_- := \{w \in \mathbf{L}^2(\Omega) : \vec{b} \cdot \nabla w \in \mathbf{L}^2(\Omega), \gamma(w)|_{\Gamma_-} = w\}$ , so dass

$$\vec{b} \nabla w + cw = f \quad \text{in } \Omega, \tag{2.14}$$

$$w = 0 \quad \text{auf } \bar{\Gamma}_- \tag{2.15}$$

gilt.



# Kapitel 3

## Das diskrete Problem

In diesem Kapitel werden zur approximativen Lösung des Dirichlet–Randwertproblems (0.1), (0.2) Näherungsverfahren in endlich dimensionalen Räumen  $\mathbf{X}_h$ , Unterräumen des Lösungsraumes  $\mathbf{X}$ , aus Kapitel 2 behandelt. Diese numerische Approximation in  $\mathbf{X}_h$  wird als **diskretes Problem** bezeichnet. Die dafür genutzte Finite–Elemente–Methode, ein spezielles Ritz–Galerkin–Verfahren (vgl. 1.39), eignet sich besonders auf komplizierteren Geometrien  $\Omega \subset \mathbb{R}^n$ , da sie sich der Struktur des Gebietes gut anpasst. Dies zeigt sich auch in der Güte der resultierenden Fehlerabschätzungen.

In diesem Abschnitt der Arbeit beginnt eine parallel eingefügte Anwendung von FEMLAB, d.h. ich werde an gegebener Stelle geeignete Befehle von FEMLAB für das entsprechende Problem, gesondert in einem Kasten, angeben und mit durch %-Zeichen gekennzeichneten Kommentaren erläutern.

### 3.1 Gebiete und Gebietszerlegungen

#### Gebiete

Das Gebiet  $\Omega$  sei in diesem Kapitel polyedrisch berandet. Die Verallgemeinerung auf krummlinige Ränder wird in Kapitel 5 behandelt.

Polyedrisch berandete Gebiete können in FEMLAB mit folgenden Befehlen erzeugt werden.

```
%=====
% I.   Geometrie des Gebiets
%-----
% Einfache geometrische Objekte erzeugen

% Beispiele in 2D:
Rechteck = rect2(0,0.5,0,0.5);
% Parameter: Begrenzung in x- und y-Richtung (xMin,xMax,yMin,yMax)

Quadrat = square2(0,0,1);
% Parameter: linke untere Ecke (x,y), Seitenlänge
```

```

% Andere 2D-Geometrien: circ2, arc2, ellip2, solid2
% Beispiele in 3D:
Quader = block3(1,2,1,'corner',[0 0 0]);
% Parameter: Seitenlängen (x,y,z), Bezugspunkt P ist die linke untere Ecke
%           ('corner'), Koordinaten von P(x|y|z)

Tetraeder = tetrahedron3([0 0 1 0; 0 1 0 0; 0 0 0 1/2]);
% Parameter: eine 4x3-Matrix mit den Eckpunkten

% Andere 3D-Geometrien: cylinder3, gencyl3, cone3, econe3, sphere3, torus3
%-----
% Erzeugen komplizierter Geometrien

% durch Mengenoperationen (Verlängert die Rechenzeit, da recht zeitaufwen-
% dige Routinen dahinter verborgen sind.)
Bereich2 = Quadrat - Rechteck;
Bereich3 = Quader - Tetraeder;

% durch Eingabe der Eckpunkt-Koordinaten eines n-Ecks
Polygon = line2([0.5 0.5 0 0 1 1],[0 0.5 0.5 1 1 0]);
% 'Polygon' hat die gleiche Form wie 'Bereich2'
% Parameter: Vektoren mit x- und y-Koordinaten der Eckpunkte
% genauso: poly2

```

Die eben erzeugten geometrischen Objekte werden im nächsten Programmabschnitt in der Datenstruktur 'fem2D' bzw. 'fem3D' gespeichert und visualisiert (vgl. Abbildung 3.1).

```

%-----
% Grafische Darstellung von geometrischen Objekten

% Der 'fem'-Struktur wird eine Geometrie zugewiesen und mit 'geomplot'
% ausgegeben.
fem2D.geom = Bereich2;
fem3D.geom = Bereich3;
subplot(1,2,1),
geomplot(fem2D,'edgelabels','numeric'),
axis([-0.2 1.2 -0.2 1.2]),
title({'Gebiet in 2D mit' 'Randnummerierung'});

% Alternative: die Geometrie wird ohne Speichern in der 'fem'-Struktur ge-
% zeichnet.
subplot(1,2,2),
geomplot(fem3D,'facelabels','on','transparency',0.5),
axis([-0.2 1.2 -0.2 2.2 -0.2 1.2]),
title('Gebiet in 3D mit' 'nummerierten Oberflächen');

```

```

% Optionen:
% 'edgelabels': bewirkt die Angabe der Nummern der Teilränder des Gebietes
%               in 2D in Zahlen ('numeric' oder 'on'). Diese Information wird
%               für eine detaillierte Angabe der Randbedingungen benötigt.
%               Bei einer 3D-Geometrie heißt die Option 'facelabels' für die
%               Randflächen, die Kantenummerierung erfolgt weiterhin mit
%               der Option 'edgelabels'.
% 'transparency': (nur bei 3D-Objekten) macht die Randflächen des Objekts
%                transparent mit dem der Option folgenden Faktor, wobei 1 in-
%                transparent bedeutet und 0 die vollständige Durchsichtigkeit
%                der Randflächen gewährleistet.
% Befehle für die Grafikausgabe:
% subplot(a,b,c): teilt die Grafik in (axb) Untergrafiken ein und schreibt
%                alle nachfolgenden Grafikausgaben an Stelle c in der Grafi-
%                kenmatrix, bis eine neue Stelle mit 'subplot' zugewiesen
%                wird (c darf also Werte von 1 bis a*b annehmen).
% axis([xMin xMax yMin yMax {zMin zMax}]): schreibt die Achsenbegrenzungen
%                der Grafik vor. Die dritte Dimension ist optional.
% title('text'): gibt der Grafik den Titel text, wobei jede Zeile des Ti-
%                tels in ein neues Paar Anführungsstriche geschrieben wird.

```

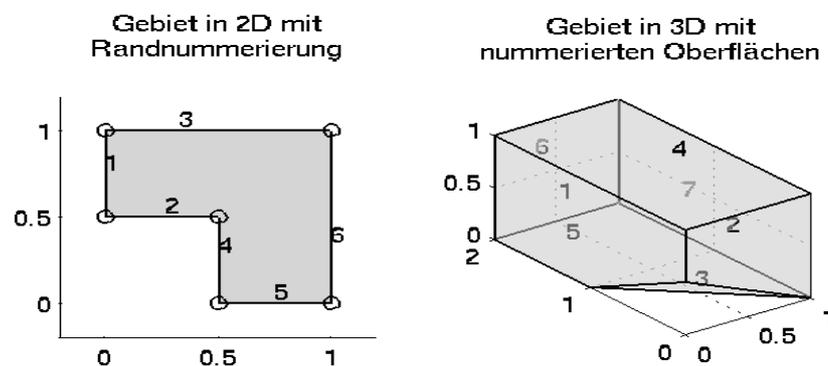


Abbildung 3.1: Die beiden Gebiete aus dem Programmcode

## Zerlegungen

Als nächstes betrachtet man eine Zerlegung des Gebietes  $\Omega \subset \mathbb{R}^n$  in Teilgebiete. Hier beschränke ich mich auf simpliziale Teilgebiete. Für  $n = 2$  sind das Dreiecke und für  $n = 3$  Tetraeder.

**Definition 3.1** Eine Zerlegung  $\mathcal{T} = \{K_j\}_{j=1}^M$  eines Gebietes  $\Omega$  heißt *zulässig*, falls jeweils zwei verschiedene abgeschlossene Teilgebiete  $\overline{K}_i$  und  $\overline{K}_j$ ,  $i \neq j$ , entweder

- genau eine vollständige gemeinsame Fläche ( $n = 3$ ) haben,
- genau eine vollständige gemeinsame Kante ( $n \geq 2$ ) haben,
- genau einen gemeinsamen Punkt ( $n \geq 1$ ) haben

oder paarweise durchschnittsfremd sind.

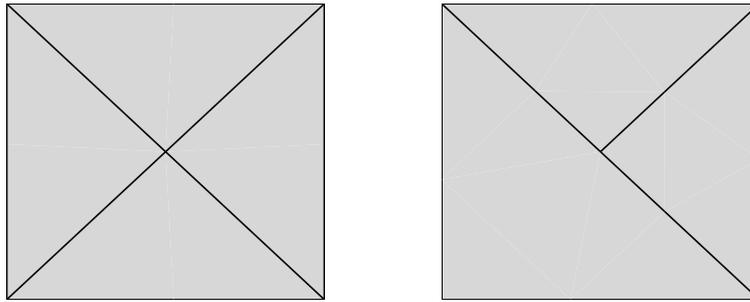


Abbildung 3.2: Zulässige Zerlegung eines Gebietes und unzulässige Zerlegung bzw. Zerlegung mit hängendem Knoten

In FEMLAB werden Zerlegungen automatisch zulässig erzeugt.

```

%=====
% II. Zerlegung erzeugen
%-----
% zulässige Zerlegungen
Zerlegung = meshinit(Bereich3, 'hmax', 1/2);

% Für die in 'fem2D' gespeicherte Geometrie 'fem2D.geom' wird die Zerlegung
% in dem Datenfeld 'fem2D.mesh' gespeichert.
fem2D.mesh = meshinit(fem2D, 'hmax', 1/8);

% Optionen:
% 'hmax': Der maximale Elementdurchmesser wird durch die folgende Zahl
% vorgegeben. Die Option erzeugt eine Zerlegung mit dem unten
% in (3.1) definierten h.
%-----
% Ausgabe der Zerlegung 'Zerlegung' des Bereich3-Gebietes
subplot(1,2,2)
meshplot(Zerlegung, 'transparency', 0.3);
title('Erste grobe Zerlegung in 3D');

% bzw. für die 'fem2D'-Zerlegung (vgl. Abbildung 3.3)
subplot(1,2,1)
meshplot(fem2D),
title('Erste grobe Zerlegung in 2D');

```

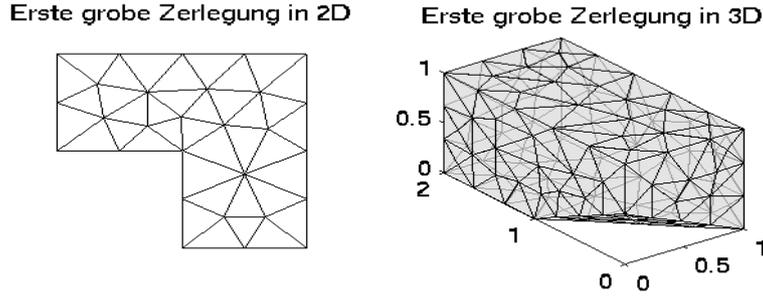


Abbildung 3.3: Zerlegungen des Beispiel-Programmcodes

Nachdem man das Gebiet  $\Omega$  exakt mittels einer zulässigen Zerlegung  $\mathcal{T} = \{K_j\}_{j=1}^M$  in Teilgebiete  $K_j$  zerlegt hat, definiert man für ein beliebiges Element  $K$  die Größe  $h_K$  als Radius der kleinsten Kugel, in die  $K$  eingeschrieben werden kann, und  $\rho_K$  als Radius der größten Kugel, die in  $K$  eingeschrieben werden kann. Jeder Zerlegung  $\mathcal{T}$  ordnet man einen Index

$$h := \max_{K \in \mathcal{T}} h_K \quad (3.1)$$

zu und erhält für jedes Gebiet  $\Omega$  eine Familie  $\{\mathcal{T}_h\}_h$  von zulässigen Zerlegungen. Diese Familie  $\{\mathcal{T}_h\}_h$  zulässiger Zerlegungen von  $\Omega$  wird in verschiedene Klassen eingeteilt.

### Definition 3.2

- (i)  $\mathcal{T}_h$  heißt **isotrop**, falls es gleichmäßig für  $0 < h \leq h_0$  eine Konstante  $c_0 > 0$  gibt, so dass  $\max_{K \in \mathcal{T}_h} \frac{h_K}{\rho_K} \leq c_0$  gilt.
- (ii)  $\mathcal{T}_h$  heißt **quasiuniform**, wenn es gleichmäßig für  $0 < h \leq h_0$  Konstanten  $0 < c_1 < c_2$  gibt, so dass jedem Element  $K$  eine Kugel mit Radius  $c_1 h$  eingeschrieben und jedes Element  $K$  von einer Kugel mit Radius  $c_2 h$  umschrieben werden kann.

Die Isotropie einer Zerlegung schreibt geometrisch nicht entartete Elemente vor (Dreiecke mit kleinstem Innenwinkel größer als ein Minimalwinkel  $\gamma_0$ ). Quasiuniforme Zerlegungen sind auch isotrop, aber zusätzlich sind alle Elemente von gleicher Größenordnung.

Die folgende Ungleichung ist von Bedeutung für spätere Abschätzungen.

**Lemma 3.3 (inverse Ungleichung)** Sei  $\mathcal{T}_h$  eine quasiuniforme Zerlegung von  $\Omega$ . Dann existiert eine Konstante  $c_{inv} > 0$ , so dass

$$\|\nabla v\|_{0,\Omega}^2 \leq c_{inv} h^{-2} \|v\|_{0,\Omega}^2 \quad \text{für alle } v \in \mathbf{X} \quad (3.2)$$

gilt mit  $h$  wie in (3.1).

**Beweis** : siehe [Qua94], Proposition 6.3.2. □

Eine Zerlegung sollte man **verfeinern**, wenn z.B. die Genauigkeit der approximativen Lösung nicht ausreicht. Dazu teilt man Dreiecke einer vorhandenen Zerlegung in zwei oder mehr kleinere Dreiecke auf. Bei der **roten Verfeinerung** einer Zerlegung wird ein Dreieck in 4 kongruente Dreiecke zerlegt, die ähnlich zum Ursprungsdreieck sind. Die Anzahl der Teilgebiete  $K$  steigt um den Faktor 4 und die Zahl der Dreieckspunkte verdoppelt sich, denn zu den drei Eckpunkten eines jeden Dreiecks kommen seine Seitenmittelpunkte als neue Eckpunkte hinzu. Eine komplette rote Verfeinerung erhält die Regelmäßigkeit der Zerlegung, weil keine entarteten oder spitzwinkligeren Dreiecke entstehen.

In **FEMLAB** sind die automatischen Zerlegungen isotrop, aber nicht zwingend quasiuniform. FEMLAB kann eine rote Verfeinerung erzeugen. Dabei sollte schon eine Zerlegung in **fem2D.mesh** vorhanden sein. Die nötige Programm-Ergänzung kann man am Ende des Abschnitts II im Beispielprogrammcode einfügen. Den Effekt dieser Ergänzung sieht man in Abbildung 3.4.

```
fem2D.mesh=meshrefine(fem2D); % rote Verfeinerung überspeichert die grobe
                               % Zerlegung
meshplot(fem2D);              % Grafik der verfeinerten Zerlegung (vgl. Ab-
                               % bildung 3.4)
```

Rote Verfeinerung der groben 2D-Zerlegung

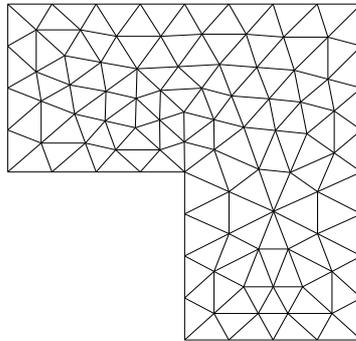


Abbildung 3.4: Rote Verfeinerung der in ‚fem2D.mesh‘ gespeicherten Zerlegung

## 3.2 Finite-Elemente-Ansätze

Finite Elemente definiert man auf den Teilgebieten  $K \subset \mathcal{T}_h$  mittels lokaler Ansatz- und Testfunktionen.

**Definition 3.4** *Ein finites Element ist ein Tripel  $(K, \mathcal{P}, \Sigma)$  mit den Eigenschaften:*

- $K \subset \mathbb{R}^n$  ist ein konvexes polyedrisches Gebiet im  $\mathbb{R}^n$ .
- Der Raum  $\mathcal{P}$  der Formfunktionen ist ein auf  $K$  definierter endlich dimensionaler linearer Funktionenraum der Dimension  $d$ .
- Die Menge  $\Sigma$  der Freiheitsgrade besteht aus  $d$  linear unabhängigen Funktionalen  $N$  über  $\mathcal{P}$ . Dabei ist jede Funktion  $p \in \mathcal{P}$  durch die Werte der  $d$  Funktionale  $N \in \Sigma$  eindeutig bestimmt.

## Finite-Elemente-Räume

Gesucht werden jetzt passende Basen und Basisfunktionen der lokalen Elemente.

**Definition 3.5** Sei  $(K, \mathcal{P}, \Sigma)$  ein finites Element. Eine Basis  $\{\Phi_1, \dots, \Phi_d\}$  von  $\mathcal{P}$  mit  $N_i(\Phi_j) = \delta_{ij}$ ,  $1 \leq i, j \leq d$ , heißt **nodale Basis** von  $\mathcal{P}$ . Hierbei ist  $\delta_{ij}$  das Kronecker-Symbol.

Auf einer nodalen Basis werden jeder Knotenvariablen ihr Funktionswert und die Werte der partiellen Ableitungen  $m$ -ter Ordnung zugewiesen. Ich beschränke mich in dieser Arbeit auf **Lagrange-Elemente** mit  $\mathbf{m} = \mathbf{0}$ .

Ist  $\mathcal{P}'$  der algebraische Dualraum des linearen Raumes  $\mathcal{P}$ , kann man eine Basis  $\{N_1, \dots, N_d\}$  von  $\mathcal{P}'$ , die Menge der Knotenvariablen, mit  $\Sigma$  identifizieren.

Geeignete lokale Beschreibungen von Ansatz- und Testfunktionen eines Finite-Elemente-Raumes über einem Teilgebiet  $K$  sind durch lokale Interpolation möglich. Zunächst werden der notwendige lokale Interpolant auf  $K \in \mathcal{T}$  und der globale Interpolant auf  $\mathcal{T}$  eingeführt.

**Definition 3.6** Seien  $(K, \mathcal{P}, \Sigma)$  ein finites Element und  $\{\Phi_1, \dots, \Phi_d\}$  eine nodale Basis von  $\mathcal{P}$ . Sei  $v$  eine Funktion, für die alle  $N_i \in \Sigma$ ,  $i = 1, \dots, d$ , definiert sind. Als **lokalen Interpolanten** von  $v$  bezeichnet man eine Funktion  $\Pi_K v \in \mathcal{P}$  mit:

$$v \rightarrow \Pi_K v := \sum_{i=1}^d N_i(v) \Phi_i. \quad (3.3)$$

### Bemerkung 3.7

- Die Abbildung  $v \rightarrow \Pi_K v$  ist linear.
- Es gilt  $N_i(\Pi_K v) = N_i(v)$  für alle  $i \in \{1, \dots, d\}$ .
- $\Pi_K$  ist ein Projektor, das heißt es gilt  $\Pi_K v = v$  für alle  $v \in \mathcal{P}$ .

Für ein Dreieck  $K$  ( $d = 3$ ,  $n = 2$ ) aus einer Zerlegung  $\mathcal{T}$  mit den Eckpunkten  $p_1, p_2, p_3$  werden den Punkten  $x \in \overline{K}$  eineindeutig die **baryzentrischen Koordinaten**  $\lambda_1, \lambda_2, \lambda_3$  mit den Gleichungen

$$x = \sum_{i=1}^3 \lambda_i p_i, \quad \sum_{i=1}^3 \lambda_i = 1 \quad (3.4)$$

zugeordnet. Damit kann man explizit Ansatz- und Testfunktionen über  $K$  angeben. So gilt für die Funktionen der nodalen Basis  $\{\Phi_1, \Phi_2, \Phi_3\}$ :

$$\Phi_j(x) = \lambda_j(x), \quad j = 1, 2, 3 \quad (3.5)$$



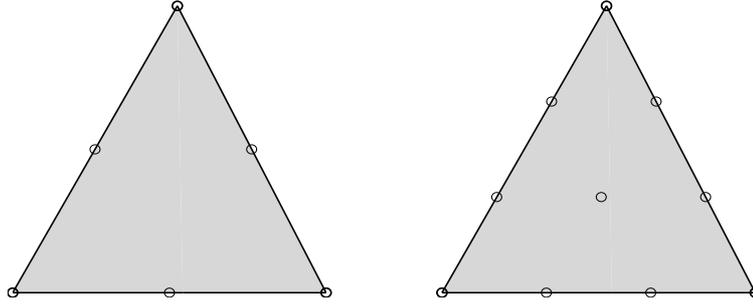


Abbildung 3.6: Dreieckelemente zweiter und dritter Ordnung

```
fem2D.xmesh = meshextend(fem2D); % Erweiterung der Datenstruktur auf Elemen-
                                % te der in 'fem2D.shape' geforderten Ord-
                                % nung.
```

**Definition 3.9** Sei  $\mathcal{T}$  eine zulässige Zerlegung des Gebietes  $\Omega$ . Sei außerdem  $(K, \mathcal{P}, \Sigma)$  finites Element auf jedem Teilgebiet  $K \subset \mathcal{T}$ . Ferner sei die Ordnung der höchsten partiellen Ableitungen auf der nodalen Basis null. Für  $v \in \mathbf{C}^0(\bar{\Omega})$  ist der **globale Interpolant** definiert durch

$$\Pi_{\mathcal{T}}v|_K = \Pi_K v \quad (3.7)$$

für alle  $K \in \mathcal{T}$ .

Der globale Interpolant ist im allgemeinen nicht stetig, für Lagrange-Elemente ( $m=0$ ) kann man jedoch einen stetigen, globalen Interpolanten angeben.

**Satz 3.10** Für eine zulässige Zerlegung  $\mathcal{T} \subset \Omega \subset \mathbb{R}^2$  können die Knotenpunkte von Lagrange-Elementen auf  $K \subset \mathcal{T}$  so gewählt werden, dass der globale Interpolant  $\Pi_{\mathcal{T}}v$  in  $\mathbf{C}^0(\bar{\Omega})$  liegt für  $v \in \mathbf{C}^0(\bar{\Omega})$ . Außerdem ist  $\Pi_{\mathcal{T}}v \in \mathbf{W}^{1,\infty}(\Omega)$ .

**Beweis :** Es reicht, für jede Kante  $e_l = \overline{p_i p_j}$ ,  $1 \leq i, j \leq 3$ ,  $i \neq j$ , die Knoten nach der Vorschrift  $\xi_t(p_j - p_i) + p_i$  mit  $0 < \xi_t < 1$  zu wählen, wobei die  $\xi_t$ ,  $t = 1, \dots, d-1$ , fest und symmetrisch zum Kantenmittelpunkt  $\frac{1}{2}(p_j - p_i) + p_i$  sind. Der Kantenmittelpunkt ist nicht unbedingt in der Menge der  $\xi_t$  enthalten. Es bleibt zu zeigen, dass die verlangte Stetigkeit über jede Kante hinweg erhalten bleibt.

Seien  $K_1, K_2$  zwei Dreiecke der Zerlegung  $\mathcal{T}$  mit einer gemeinsamen Kante  $e$ . Durch die symmetrische und koordinatenfreie Wahl der Knoten auf  $e$  haben sie die gleiche Lage im Raum. Sei  $w := \Pi_{K_1}v - \Pi_{K_2}v$ , wobei  $\Pi_{K_i}v, i = 1, 2$ , Polynome vom Grad  $d$  sind und außerhalb von  $K_i, i = 1, 2$ , als Polynome fortgesetzt werden. Dann ist  $w$  ein Polynom vom Grad  $d$  und dessen Einschränkung auf die Kante  $e$  hat  $d+1$  Stützstellen mit dem Wert 0. Also verschwindet  $w|_e$ , d. h. der Interpolant ist stetig über jede Kante  $e$  hinweg.

Dass  $\Pi_{\mathcal{T}}v \in \mathbf{W}^{1,\infty}(\Omega)$  ist, folgt aus der Existenz schwacher Ableitungen erster Ordnung.

$$(D_{(w)}^\alpha \Pi_{\mathcal{T}}v)|_K = D^\alpha \Pi_{\mathcal{T}}v \quad \forall K \in \mathcal{T} \text{ und Multiindex } \alpha \text{ der Länge } |\alpha| \leq 1.$$

$D^\alpha \Pi_{\mathcal{T}} v \in \mathbf{L}^\infty(\Omega)$  ist schwache Ableitung von  $\Pi_{\mathcal{T}} v$ , denn

$$\begin{aligned} \int_{\Omega} (D^\alpha \phi)(x) (\Pi_{\mathcal{T}} v)(x) dx &= \sum_{K \in \mathcal{T}} \int_K (D^\alpha \phi)(x) (\Pi_K v)(x) dx \\ &\stackrel{p.I.}{=} (-1)^{|\alpha|} \left( \sum_{K \in \mathcal{T}} \int_K \phi(x) (D^\alpha \Pi_K v)(x) dx \right. \\ &\quad \left. - \sum_{e \in \mathcal{T}} \int_e \overbrace{\phi(x) \frac{\partial}{\partial n} (\Pi_{\mathcal{T}} v)(x) ds}^{=0} \right) \\ &= (-1)^{|\alpha|} \int_{\Omega} \phi(x) \sum_{K \in \mathcal{T}} \chi_K (D^\alpha \Pi_K v)(x) dx \end{aligned}$$

für alle  $\phi \in \mathbf{C}_0^\infty(\Omega)$ , wobei  $\chi_K$  die charakteristische Funktion auf  $K$  ist. Bei der partiellen Integration (p.I.) fallen die Randintegrale aufgrund der Stetigkeitseigenschaften des Interpolanten weg.  $\square$

Im folgenden Satz wird die gerade bewiesene Aussage auf den  $n$ -dimensionalen Raum verallgemeinert.

**Satz 3.11** Sei  $\mathcal{T} = \{K_j\}_{j=1}^M$  eine zulässige Zerlegung des Gebietes  $\Omega \subset \mathbb{R}^n$ . Sei  $v : \overline{\Omega} \rightarrow \mathbb{R}$  eine Funktion mit  $v|_{K_j} \in \mathbf{C}^1(\overline{K_j})$ ,  $j = 1, \dots, M$ . Dann gilt  $v \in \mathbf{W}^{1,2}(\Omega)$  genau für  $v \in \mathbf{C}^0(\overline{\Omega})$ .

**Beweis :** siehe [Bra97], Kapitel II, Satz 5.2.  $\square$

Bisher wurde nur die Theorie der finiten Elemente behandelt. Jetzt soll diese Theorie in FEM-LAB realisiert werden. Dazu wird beispielhaft das Dirichlet–Randwert–Problem der Poisson–Gleichung mittels Angabe der Koeffizienten und Randbedingungen in die Datenstruktur von FEMLAB eingegeben. Die Umsetzung dieser Daten auf finite Elemente erledigt FEMLAB selbst.

```
%=====
% III. RWP angeben
%-----
% Unbekannte und Freiheitsgrade benennen.

% Problem im 2D
fem2D.dim = {'u'};
fem2D.sdim = {'x', 'y'};

% Problem im 3D
fem3D.dim = {'u'};
fem3D.sdim = {'x', 'y', 'z'};
%-----
% Koeffizienten des elliptischen RWP angeben
```

```

% Das Schema für die Koeffizienten in FEMLAB ist
%   - nabla ( c nabla u) + be nabla u + a u = f.
% Für jedes Teilgebiet eines Gebietes werden die Koeffizienten angegeben,
% z.B. bei einem Gebiet mit zwei Teilgebieten: fem.equ.c = {{{'1' '0.1'}}}.
% Die Koeffizienten für das Dirichlet-Randwert-Problem der Poisson-Glei-
% chung lauten:
fem2D.equ.c = {{{'1'}}};           % Bezeichnung außerhalb von FEMLAB:
                                   % Skalar epsilon
fem2D.equ.be = {{{'0' '0'}}};     % Vektorfunktion b
fem2D.equ.a = {{{'0'}}};         % Funktion c
fem2D.equ.f = {{{'exp(-x.*y).*((2*pi^2-y.^2-x.^2).*sin(pi*x).*sin(pi*y)+...
                2*pi*(x.*sin(pi*x).*cos(pi*y)+y.*cos(pi*x).*sin(pi*y))' }}}};
                                   % Funktion f der rechten Seite

%-----
% Randbedingungen angeben

% nach dem Schema
%   h u = r                               (DIRICHLET-Randbedingung)
% sind die verschiedenen Randbedingungen als Spalten anzugeben:
fem2D.bnd.h = {1 1 1};
fem2D.bnd.r = {0 'sin(pi*y).*exp(-0.5*y)' 'sin(pi*x).*exp(-0.5*x)'};

% Hier sind die Randbedingungen (1 0) (homogene Dirichlet-Randbedingung)
% und (1 'sin(pi*y).*exp(-0.5*y)') sowie (1 'sin(pi*x).*exp(-0.5*x)') (in-
% homogene Dirichlet-Randbedingungen) festgelegt. Die genauere Beschrei-
% bung inhomogener Dirichlet-Randbedingungen und allgemeinerer Randbedin-
% gungen folgt in Kapitel 4.
% Das Gebiet im 2D (vgl. Abbildung 3.1) hat 6 Teilränder. Durch einen Vek-
% tor der Länge 6 wird festgelegt, welche der Randbedingungen auf welchem
% Teil des Randes gelten sollen:
fem2D.bnd.ind = [1 3 1 2 1 1];

% Randbedingung 1 gilt also auf den Teilstücken des Randes mit den Nummern
% 1,3,5 und 6.
% Achtung: Alle definierten Randbedingungen müssen auch verwendet werden,
% sonst bricht das Programm mit einer Fehlermeldung ab.

```

Die eingegebenen Daten werden in FEMLAB zum Aufstellen des zur schwachen Formulierung des Problems äquivalenten linearen Gleichungssystems (1.47) verwendet. Dieses Gleichungssystem kann mit verschiedenen Lösungsroutinen in FEMLAB gelöst werden. Zur Verfügung stehen wahlweise direkte und iterative Lösungsverfahren.

```

%=====
% IV. Gleichungssystem lösen
%-----

```

```

% Direkter Löser
sol1 = femlin(fem2D);
%-----
% Iterativer Löser
sol2 = femiter(fem2D,'itsolv','gbit','prefun','gmg');
% Optionen:
% 'itsolv':      Name des in Matlab vorhandenen iterativen Löasers folgt.
% 'prefun':     Name des in Matlab vorhandenen Vorkonditionierers folgt.

```

Die Ausgabe der Lösung erfolgt dann mit dem FEMLAB-Befehl `postplot`, der z.B. die folgende Möglichkeit bietet.

```

%=====
% V. Darstellung der Lösung
%-----
% 2D-Oberflächen-Grafik (3D-Darstellung der Lösung)
figure(2);
for i=1:2
    if i==1
        fem2D.sol = sol1;
    else
        fem2D.sol = sol2;
    end
    % Grafik zu finden in Kapitel 4, Abbildung 4.4
    subplot(2,2,i),
    postplot(fem2D,'grid','on','tridata','u','triz','u'),
    title(strcat('FEM-Approximation mit Löser Nr.',num2str(i)));
end

% Optionen:
% 'grid'      zeichnet Hilfslinien im Koordinatengitter der Grafik ein.
% 'tridata':  Die Lösung soll als 3D-Gebirge gezeichnet werden.
% 'triz':     Auftragen der Werte der folgenden Lösungskomponente in z-
%             Achsen-Richtung.
% Befehle:
% figure(d)   gibt Zugriff auf das Grafikfenster d, das mit 'subplot()'
%             wiederum in mehrere Grafiken zerlegt werden kann.
% strcat()    hängt zwei oder mehr Strings hintereinander.
% num2str()   wandelt eine Zahl in einen String um.

```

### 3.3 Interpolationsfehler–Abschätzungen

Interpolationsfehler–Abschätzungen werden mit Hilfe einer Transformation auf ein Referenzelement vorgenommen. Dabei nimmt man an, dass jede Familie  $\{\mathcal{T}_h\}_h$  sich auf ein einheitliches

Referenzelement  $\tilde{K}$  transformieren lässt. Das Standard-Referenzelement ist in Abbildung 3.5 zu sehen. Die folgenden Resultate gelten für jegliche simpliziale Zerlegungen im  $\mathbb{R}^n$ . Exemplarisch wird auf eine Dreieckszerlegung des Gebietes  $\Omega \subset \mathbb{R}^2$  zurückgegriffen.

### Transformation auf das Referenzelement

Sei  $K \in \mathcal{T}_h$  ein nichtentartetes Dreieck mit den Eckpunkten  $p_i = (x_i, y_i)^T$  und den zugehörigen baryzentrischen Koordinaten (3.4). Dann findet man mit  $\tilde{\lambda} := (\tilde{\lambda}_1, \tilde{\lambda}_2)^T \in \mathbb{R}^2$  eine affine inverse Abbildung  $\tilde{\lambda} = Bx + b$ , die das Dreieck  $K$  allgemeiner Lage in das Einheitsdreieck  $\tilde{K} := \{\tilde{\lambda} : 0 < \tilde{\lambda}_1, \tilde{\lambda}_2 < 1 \text{ und } \tilde{\lambda}_1 + \tilde{\lambda}_2 < 1\}$  überführt. Diese affin lineare Abbildung nennt man **Transformation auf das Referenzdreieck  $\tilde{K}$**  (vgl. Abbildung 3.5).

Jeder Funktion  $u : K \rightarrow \mathbb{R}$  über einem allgemeinen Dreieck  $K$  der Zerlegung  $\mathcal{T}_h$  wird eine Funktion  $v : \tilde{K} \rightarrow \mathbb{R}$  über dem Referenzelement  $\tilde{K}$  mit

$$v(p) = u(F(p)) \quad (3.8)$$

zugeordnet, wobei die Abbildung  $F : \tilde{K} \rightarrow K$  **affin-linear** ist, also von der Form

$$\begin{pmatrix} x \\ y \end{pmatrix} = F(p) = \begin{pmatrix} x_2 - x_1 & x_3 - x_1 \\ y_2 - y_1 & y_3 - y_1 \end{pmatrix} \begin{pmatrix} \xi \\ \eta \end{pmatrix} + \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \quad (3.9)$$

mit  $p := (\xi, \eta)^T$  und den Eckpunkten  $p_i = (x_i, y_i)^T$ ,  $i = 1, 2, 3$  des allgemeinen Dreiecks  $K \subset \mathcal{T}_h$ .

**Lemma 3.12** *Ein Teilgebiet  $K$  und das Referenzelement  $\tilde{K}$  seien durch die affin-lineare Abbildung*

$$x = F(p) = Bp + b, \quad p \in \tilde{K} \quad (3.10)$$

*eineindeutig aufeinander abgebildet. Dann folgt für Transformationen nach (3.8)*

$$(i) \quad u \in \mathbf{W}^{l,2}(K) \quad \Leftrightarrow \quad v \in \mathbf{W}^{l,2}(\tilde{K})$$

(ii) *Für die Halbnorm aus (1.15) gilt*

$$|v|_{l,\tilde{K}} \leq C \|B\|^l |\det B|^{-\frac{1}{2}} |u|_{l,K} \quad (3.11)$$

$$|u|_{l,K} \leq C \|B^{-1}\|^l |\det B|^{\frac{1}{2}} |v|_{l,\tilde{K}} \quad (3.12)$$

*Dabei sind  $\|B\| := \sup_{\|p\|=1} \|Bp\|$  mit der Norm  $\|\cdot\|_\infty$ .*

**Beweis :**

(i) folgt aus (ii).

(ii) Sei  $u \in \mathbf{C}^l(K)$ . Aus (3.8) und (3.10) folgt  $v(p) = u(Bp + b)$ , nach Ableitung nach  $p$

$$\frac{\partial v}{\partial p_j} = \sum_{i=1}^2 \frac{\partial u}{\partial F} \frac{\partial F}{\partial p_j}$$

und mit (3.10) die Abschätzung

$$\left| \frac{\partial v}{\partial p_j} \right| \leq \|B\| \max_i \left| \frac{\partial u}{\partial x_i} \right|.$$

Dann resultiert die Abschätzung für die  $\alpha$ -te Ableitung rekursiv

$$|(D^\alpha v)(p)| \leq \|B\|^{|\alpha|} \max_{\beta: |\beta|=|\alpha|} |(D^\beta u)(x(p))|, \quad p \in \tilde{K}.$$

Daraus ergibt sich

$$\sum_{|\alpha|=l} |(D^\alpha v)(p)|^2 \leq C \|B\|^{2l} \sum_{|\beta|=l} |(D^\beta u)(x(p))|^2, \quad p \in \tilde{K}$$

und mit Variablentransformation im Integral von  $p$  nach  $x$  folgt schließlich

$$\begin{aligned} |v|_{l, \tilde{K}}^2 &\leq C \|B\|^{2l} \int_{\tilde{K}} \sum_{|\beta|=l} |(D^\beta u)(x(p))|^2 dp \\ &\leq C \|B\|^{2l} |\det B|^{-1} \int_K \sum_{|\beta|=l} |(D^\beta u)(x)|^2 dx \\ &= C \|B\|^{2l} |\det B|^{-1} |u|_{l, K}^2. \end{aligned}$$

Durch Ziehen der Wurzel erhält man (3.11), da  $\mathbf{C}^l(\bar{K})$  dicht in  $\mathbf{W}^{l,2}(K)$  liegt. Die Aussage (3.12) folgt, wenn man  $B^{-1}x - B^{-1}b$  für  $p$  einsetzt, also von  $\tilde{K}$  auf  $K$  abbildet.  $\square$

Im nächsten Lemma werden jetzt die Normen von  $B$  und  $B^{-1}$  bezüglich der Form und Größe der Teilgebiete  $K \in \mathcal{T}_h$  abgeschätzt.

**Lemma 3.13** *Seien für ein Element  $K$  die Voraussetzungen von Lemma 3.12 erfüllt. Außerdem seien  $h_K$  und  $\rho_K$  wie am Beginn des Abschnitts „Zerlegungen“ definiert. Ferner sei  $\tilde{K}$  ein festes und von der Zerlegung unabhängiges Referenzelement. Dann gilt*

$$\|B\| \leq Ch_K \tag{3.13}$$

$$\|B^{-1}\| \leq C\rho_K^{-1} \tag{3.14}$$

**Beweis :** Für das Referenzelement  $\tilde{K}$  gibt es Kugeln mit den Radien  $\tilde{\rho}$  und  $\tilde{h}$ . Dann existiert ein Punkt  $p_0 \in \tilde{K}$  mit  $p_0 + p \in \tilde{K}$  für beliebige  $p$  mit  $\|p\| = \tilde{\rho}$ . Die mit Hilfe der Abbildung (3.10) zugeordneten Punkte  $x_0 = Bp_0 + b$  und  $x = B(p_0 + p) + b$  gehören zu  $K$ . Für sie ist dann  $\|x - x_0\| \leq 2h_K$ . Dann folgt

$$\|B\| = \frac{1}{\tilde{\rho}} \sup_{\|p\|=\tilde{\rho}} \|Bp\| \leq \tilde{\rho}^{-1} \|x - x_0\| \leq 2h_K \tilde{\rho}^{-1} = Ch_K$$

Vertauscht man jetzt die Rollen von  $K$  und  $\tilde{K}$ , erhält man mit  $\|p - p_0\| \leq 2\tilde{h}$  folgende Abschätzung

$$\|B^{-1}\| = \frac{1}{\rho_K} \sup_{\|x\|=\rho_K} \|B^{-1}x\| \leq \rho_K^{-1} \|p - p_0\| \leq 2\rho_K^{-1}\tilde{h} = C\rho_K^{-1}$$

□

### Der Interpolationsfehler

Der lokale Interpolationsoperator  $\Pi_K$  aus Definition 3.6 ist sowohl auf  $K$  als auch auf dem Referenzelement  $\tilde{K}$  definiert. Der globale Interpolationsoperator  $\Pi_{\mathcal{T}}$  wird stückweise durch Hintereinanderausführung der Interpolation  $\Pi_K$  auf  $\tilde{K}$  und der Abbildung  $F_j : \tilde{K} \rightarrow K_j$ ,  $j = 1, \dots, M$ , definiert, wobei  $M = |\mathcal{T}_h|$  ist. Also  $\Pi_{\mathcal{T}}$  ist ein Projektor in den diskreten Raum  $\mathbf{X}_h$ .

**Lemma 3.14 (Bramble–Hilbert)** Sei  $\gamma : \mathbf{W}^{l,2}(\tilde{K}) \rightarrow \mathbf{W}^{r,2}(\tilde{K})$  mit  $l, r \geq 0$  eine stetige, lineare Abbildung, so dass

$$\gamma(\phi) = 0 \quad \forall \phi \in \mathbb{P}_k, k \geq 0. \quad (3.15)$$

Dann gilt für jedes  $v \in \mathbf{W}^{l,2}(\tilde{K})$

$$|\gamma|_{r,\tilde{K}} \leq \|\gamma\|_{\mathcal{L}(\mathbf{W}^{l,2}(\tilde{K}), \mathbf{W}^{r,2}(\tilde{K}))} \inf_{\phi \in \mathbb{P}_k} \|v + \phi\|_{l,\tilde{K}}. \quad (3.16)$$

Im einzelnen gilt für jedes  $v \in \mathbf{W}^{l,2}(\tilde{K})$ ,  $l \geq 2$

$$|v - \Pi_K v|_{r,\tilde{K}} \leq \|I - \Pi_K\|_{\mathcal{L}(\mathbf{W}^{l,2}(\tilde{K}), \mathbf{W}^{r,2}(\tilde{K}))} \inf_{\phi \in \mathbb{P}_k} \|v + \phi\|_{l,\tilde{K}} \quad (3.17)$$

für  $r, l, k \in \mathbb{N}_0$  und  $l \geq r$ . Außerdem ist  $I : v \rightarrow v$  die Identitätsabbildung.

**Beweis :** Sei  $v \in \mathbf{W}^{l,2}(\tilde{K})$ . Für ein  $\phi \in \mathbb{P}_k$  folgt aus (3.15)

$$|\gamma(v)|_{r,\tilde{K}} = |\gamma(v + \phi)|_{r,\tilde{K}} \leq \|\gamma\|_{\mathcal{L}(\mathbf{W}^{l,2}(\tilde{K}), \mathbf{W}^{r,2}(\tilde{K}))} |v + \phi|_{l,\tilde{K}}. \quad (3.18)$$

Daraus folgt (3.16), da  $\phi$  beliebig gewählt war.

Der Interpolationsoperator  $\Pi_K$  ist in  $\mathbf{W}^{l,2}(\tilde{K})$  definiert, denn mit dem Sobolev'schen Einbettungssatz (1.11) folgt  $\mathbf{W}^{l,2}(\tilde{K}) \subset \mathbf{C}^0(\tilde{K})$  für  $l \geq 2$ . Der Operator  $I - \Pi_K$  genügt also der Behauptung des Lemmas. □

Im folgenden Satz wird die Norm des Operators  $I - \Pi_K$  unabhängig von  $\mathcal{T}_h$  ermittelt, da  $\tilde{K}$  unabhängig von  $\mathcal{T}_h$  ist. Dabei wird benutzt, dass bei affin-linearen Transformationen  $F$  die Polynome über dem Referenzelement  $\tilde{K}$  in Polynome über  $K$  übergehen. Außerdem sei  $\mathbf{X}_h$  ausschließlich von einem einzigen Referenzelement  $(\tilde{K}, \tilde{\mathcal{P}}, \tilde{\Sigma})$  erzeugt.

**Satz 3.15** Die zulässige Zerlegung  $\mathcal{T}_h$  des Gebietes  $\Omega \subset \mathbb{R}^2$  sei durch affin-lineare Transformation von einem Referenzelement  $\tilde{K}$  erzeugt. Seien  $u \in \mathbf{W}^{k+1,2}(\Omega)$  und  $\Pi_{\mathcal{T}} : \mathbf{W}^{k+1,2}(\Omega) \rightarrow \mathbb{P}_k^T(\Omega) \subset \mathbf{W}^{r,2}(\Omega)$  der über  $\mathcal{T}_h$  definierte globale Projektor in die Menge  $\mathbb{P}_k^T(\Omega)$  der stückweise polynomialen Funktionen vom Grad  $k$ . Ferner seien  $r \in \mathbb{N}_0, k \in \mathbb{N}$  mit  $r \leq k$ .

- (i) Für die lokale Interpolationsaussage gilt:  
Es existiert eine Konstante  $C > 0$ , so dass

$$\|u - \Pi_{\mathcal{T}} u\|_{r,K} \leq Ch_K^{k+1} \rho_K^{-r} |u|_{k+1,K}. \quad (3.19)$$

- (ii) Für eine isotrope Zerlegung  $\mathcal{T}_h$  gilt die globale Interpolationsaussage:

$$\|u - \Pi_{\mathcal{T}} u\|_{r,\Omega} \leq C \left( \sum_K h_K^{2(k+1-r)} |u|_{k+1,\Omega}^2 \right)^{-\frac{1}{2}}. \quad (3.20)$$

- (iii) Für eine quasi-uniforme Zerlegung  $\mathcal{T}_h$  gilt die globale Interpolationsaussage:

$$\|u - \Pi_{\mathcal{T}} u\|_{r,\Omega} \leq Ch^{k+1-r} |u|_{k+1,\Omega}. \quad (3.21)$$

**Beweis :** Man betrachte die Zerlegung  $\mathcal{T}_h = \{K_j\}_{j=1}^M$  des Gebietes  $\Omega$ . Wegen

$$\|u - \Pi_{\mathcal{T}} u\|_{r,\Omega}^2 = \sum_{j=1}^M \|u - \Pi_{K_j} u\|_{r,K_j}^2 \quad (3.22)$$

kann man den Fehler für jedes Teilgebiet  $K = K_j$  abschätzen. Ferner sei  $\tilde{K}$  das einheitliche Referenzelement der Zerlegung. Unter Verwendung von (3.12) und (3.14) folgt

$$\|u - \Pi_{\mathcal{T}} u\|_{r,K} \leq C \rho_K^{-r} |\det B|^{\frac{1}{2}} |v - \Pi_{\tilde{K}} v|_{r,\tilde{K}}.$$

Da  $(I - \Pi_{\tilde{K}})(p) = 0$  für alle  $p \in \mathbb{P}_k$ , gilt mit dem Lemma 3.14 von Bramble–Hilbert

$$|v - \Pi_{\tilde{K}} v|_{r,\tilde{K}} \leq C |v|_{k+1,\tilde{K}}.$$

Abschließend verwendet man (3.11) und (3.13) und erhält die Behauptung (i). Die Behauptungen (ii) für die isotrope und (iii) für die quasi-uniforme Zerlegung ergeben sich aus (3.22).  $\square$

Für die Konstante  $C$  in der Ungleichung (3.19) erhalten Arcangeli und Gout (vgl.[Arc76]):

**Lemma 3.16** *Es gilt*

$$C(m, n, k, p) = \frac{1}{k+1 - \frac{n}{p}} \frac{1}{k!} \sum_{i=1}^N \max_{x \in K} \left( \sum_{|\alpha|=m} \frac{m!}{\alpha!} |\partial^\alpha \Phi_i(\lambda_1(x), \dots, \lambda_{n+1}(x))| \right) \quad (3.23)$$

mit Multiindex  $\alpha$  wie in (1.1) mit  $\alpha! = \alpha_1! \cdots \alpha_n!$  und  $N = \binom{n+k}{k}$ , der Anzahl der Stützstellen auf einem Element  $K \in \mathcal{T}_h$ .

**Beweis :** Seien  $k+1 > \frac{n}{p}$  und  $\Omega$  ein konvexes Gebiet mit Lipschitz-stetigem Rand. Seien  $u \in \mathbf{C}^{k+1}(\bar{\Omega})$ ,  $a$  ein beliebiger, fest gewählter Punkt in  $\bar{\Omega}$  und  $x \in \Omega$ . Es sei

$$J(u, a)(x) := \int_0^1 (1-t)^k D^{k+1} u(x + t(a-x))(a-x)^{k+1} dt.$$

Dann folgt für alle  $u$  mit Taylor-Entwicklung bis zum Grad  $k$  um  $x$

$$u(a_i) = \sum_{l=0}^k \frac{D^l u(x)(a_i - x)^l}{l!} + \frac{1}{k!} J(u, a_i)(x), \quad i = 1, \dots, N, \{a_j\}_{j=1}^N = \Sigma$$

mit  $\Sigma$  aus Definition 3.4. Dann gilt nach [Arc76], Proposition 1-1,

$$\|J(u, a_i)\|_{\mathbf{L}^p(\Omega)} \leq \frac{1}{k+1 - \frac{n}{p}} h^{k+1} |u|_{k+1,p,\Omega}. \quad (3.24)$$

Seien  $\Pi_{\mathcal{T}}$  der lineare, stetige, globale Projektor aus Satz 3.15 und  $\{\Phi_i\}_{i=1}^N$  die Menge der Basisfunktionen von  $\Pi_{\mathcal{T}}$ . Man betrachtet jetzt die Interpolationsfehler-Abschätzung

$$\|\Pi_{\mathcal{T}}u - u\|_{m,p,\Omega} \leq C(m, n, k, p) h^{k+1} |u|_{k+1,p,\Omega}$$

und berechnet die Konstante  $C$  genauer. Die Interpolation mit  $\Pi_{\mathcal{T}}$  ist bis zum Grad  $m = k$  exakt. Daher folgt mit Taylor-Entwicklung

$$D^m(\Pi_{\mathcal{T}}u)(x) - D^m u(x) = \frac{1}{k!} \sum_{i=1}^N J(u, a_i)(x) D^m(\Phi_i)(x)$$

und daraus mit (3.24)

$$\|u - \Pi_{\mathcal{T}}u\|_{m,p,\Omega} \leq \frac{1}{k+1 - \frac{n}{p}} \frac{1}{k!} \left( \sum_{i=1}^N D^m \Phi_i(x) \right) h^{k+1} |u|_{k+1,p,\Omega}.$$

Man setzt schließlich

$$\Phi_i(x) := f_i(\lambda_1(x), \dots, \lambda_{n+1}(x)), \quad i = 1, \dots, d$$

mit  $f \in \mathbf{C}^k$  und  $\lambda_1(x), \dots, \lambda_{n+1}(x)$  als baryzentrische Koordinaten von  $x$  bzgl. der Knoten von  $K \in \mathcal{T}_h$ . Dann gilt

$$\|D^m \Phi_i(x)\| \leq \left( \max_{j=1, \dots, n+1} \|D\lambda_j\| \right)^m \sum_{|\alpha|=m} \frac{m!}{\alpha!} |\partial^\alpha f_i(\lambda_1(x), \dots, \lambda_{n+1}(x))|.$$

Außerdem gilt

$$\|D\lambda_j\| \leq \rho^{-1}.$$

Damit folgt die Behauptung.  $\square$

Die Konstante  $C$  aus (3.23) lässt sich für den Fall  $m = 0$  (Lagrange-Elemente),  $p = 2$  und  $n = 2$  vereinfachen, abhängig vom Grad  $k$  der polynomialen Ansatzfunktionen  $\mathbb{P}_k$ .

$$C(0, 2, k, 2) = \frac{1}{k+1 - \frac{2}{2}} \frac{1}{k!} \sum_{i=1}^N \max_{x \in K} \left( \sum_{|\alpha|=0} \frac{0!}{\alpha!} |\partial^\alpha \Phi_i(\lambda_1(x), \lambda_2(x), \lambda_3(x))| \right)$$

$$\begin{aligned}
&= \frac{1}{k!k} \sum_{i=1}^N \max_{x \in K} (|\Phi_i(\lambda_1(x), \lambda_2(x), \lambda_3(x))|) \\
&= \frac{1}{k!k} \sum_{i=1}^N 1 = \frac{1}{k!k} \binom{2+k}{k} \\
&= \frac{(k+1)(k+2)}{2} \frac{1}{k!k} \\
&\approx \frac{1}{(k-1)!} \quad \text{für große } k
\end{aligned} \tag{3.25}$$

Im  $\mathbb{R}^3$  erhält man für  $C$  abhängig von  $k$  die vereinfachte Formel

$$C(0, 3, k, 2) = \frac{(k+1)(k+2)(k+3)}{6k!(k-\frac{1}{2})}. \tag{3.26}$$

Dann ergeben sich für  $C$  bei verschiedenen Element-Ordnungen  $k$  die Werte der Tabelle 3.1.

Ordnung $k$	Wert von $C$	
	im $\mathbb{R}^2$	im $\mathbb{R}^3$
1	3	8
2	1.5	3.333
3	0.5555	1.3333
4	0.15625	0.4167
5	0.035	0.1037
6	0.006809	0.02121
7	0.001020	0.003663
8	0.0001395	0.0005463
$\vdots$	$\vdots$	$\vdots$

Tabelle 3.1: Werte für  $C$

### 3.4 Fehler der Galerkin-Finite-Elemente-Methode

In diesem Abschnitt wird mit Hilfe des Interpolationsfehlers der Fehler der konformen Finite-Elemente-Methode abgeschätzt. Der Raum  $\mathbf{X}_h \subset \mathbf{X} = \mathbf{W}_0^{1,2}(\Omega) \subset \mathbf{W}^{1,2}(\Omega)$  sei so konstruiert, dass

$$\mathbf{X}_h = \mathcal{C}_{\mathcal{T}} \cap \mathbf{X}$$

gilt mit

$$\mathcal{C}_{\mathcal{T}} = \{\Pi_{\mathcal{T}} v : v \in \mathbf{C}^0(\bar{\Omega})\} \tag{3.27}$$

und stückweise polynomialen Funktionen vom Grad  $k$ . Man betrachtet die zum Problem (2.1), (2.2) gehörende Variationsgleichung:

Finde  $u \in \mathbf{X}$ , so dass

$$a(u, v) = f(v) \quad \text{für alle } v \in \mathbf{X} \quad (3.28)$$

gilt.

**Satz 3.17** Die Lösung  $u$  der Variationsgleichung (3.28) sei regulär nach Satz 2.5 mit  $m = 0$ , also gelte

$$u \in \mathbf{X} \cap \mathbf{W}^{k+1,2}(\Omega). \quad (3.29)$$

Die Voraussetzungen des Satzes 3.15 und des Lemmas 2.3 seien erfüllt.

(i) Dann ist die Variationsgleichung im diskreten Raum eindeutig lösbar:

Finde  $u_h \in \mathbf{X}_h$ , so dass

$$a(u_h, v) = f(v) \quad \text{für alle } v \in \mathbf{X}_h \quad (3.30)$$

gilt.

(ii) Für den Fehler gilt die Abschätzung

$$\|u - u_h\|_{\mathbf{X}} \leq Ch^k |u|_{k+1,\Omega}.$$

**Beweis :**

(i) Aus  $\mathbf{X}_h \subset \mathbf{X}$  folgen Stetigkeit und Elliptizität der Bilinearform  $a(\cdot, \cdot)$  in (3.30). Dann wendet man das Lemma 1.21 von Lax–Milgram an und erhält die Behauptung.

(ii) Das Lemma 1.23 liefert

$$\|u - u_h\|_{\mathbf{X}} \leq C \inf_{v \in \mathbf{X}_h} \|u - v\|_{\mathbf{X}} \leq C \|u - \Pi_{\mathcal{T}} u\|_{\mathbf{X}}.$$

Die Behauptung folgt aus Satz 3.15(iii) mit  $r = 1$ . □

Diese Fehlerabschätzung ist im Vergleich zur Interpolationsfehler–Abschätzung aus Satz 3.15 nicht optimal bzgl. der  $\mathbf{L}^2$ –Norm. Im folgenden Satz wird deshalb mit einem Dualitätsargument, dem sogenannten „Aubin–Nitsche–Trick“, die erwartete Ordnung des Fehlers bewiesen.

**Satz 3.18** Sei  $\mathbf{X} \hookrightarrow \mathbf{W}^{1,2}(\Omega)$  eine stetige Einbettung. Die Lösung  $u$  der Variationsgleichung (3.28) sei regulär gemäß (3.29) mit  $k = 1$ . Weiter seien die Voraussetzungen von Satz 3.15 erfüllt. Die zu (3.28) adjungierte Variationsgleichung lautet:

Finde  $w \in \mathbf{X} \cap \mathbf{W}^{2,2}(\Omega)$ , so dass

$$a^*(w, v) := a(v, w) = g(v) \quad \text{für alle } v \in \mathbf{X}^* \quad (3.31)$$

gilt. Sie besitzt für alle  $g \in \mathbf{L}^2(\Omega)$  genau eine Lösung  $w$  mit der oberen Schranke

$$|w|_{2,\Omega} \leq C \|g\|_{0,\Omega}. \quad (3.32)$$

Dann gilt für den Fehler des diskreten Problems (3.30) die Abschätzung

$$\|u - u_h\|_{0,\Omega} \leq Ch^{k+1} |u|_{k+1,\Omega}. \quad (3.33)$$

**Beweis :** Satz 7.11 aus [Lub02] liefert

$$\|u - u_h\|_{0,\Omega} \leq M \|u - u_h\|_{\mathbf{X}} \sup_{g \in \mathbf{L}^2(\Omega)} \frac{\inf_{v \in \mathbf{X}_h} \|w - v\|_{\mathbf{X}}}{\|g\|_{0,\Omega}} \quad (3.34)$$

mit der Konstanten  $M$  aus (1.32). Außerdem gilt

$$\inf_{v \in \mathbf{X}_h} \|w - v\|_{\mathbf{X}} \stackrel{\text{Satz 3.15}}{\leq} Ch |w|_{2,\Omega} \stackrel{(3.32)}{\leq} Ch \|g\|_{0,\Omega}. \quad (3.35)$$

Dann folgt mit Satz 3.17 und (3.34) die Behauptung.  $\square$

Erfüllt das Gebiet nicht die Regularitätsanforderungen aus Kapitel 2, wie im Beispiel 2.4 mit einer Singularität auf dem Rand, muss man eine Möglichkeit finden, eine gute Approximation zu bekommen. Eine Verfeinerung der Zerlegung auf dem gesamten Gebiet mit dem Ziel der Verbesserung der Approximation an Singularitäten (im Beispiel 2.4 die Singularität bei  $r = 0$ ) ist ineffektiv. Dagegen hilft aber ein **adaptiver Löser**. Dieser berechnet den Approximationsfehler auf jedem Element mit einem lokalen Fehlerschätzer. Ist der Fehler größer als eine bestimmte Schranke, wird die Zerlegung lokal verfeinert, um den Fehler zu verringern. Eine solche, adaptiv verfeinerte Zerlegung sieht man in Abbildung 3.11.

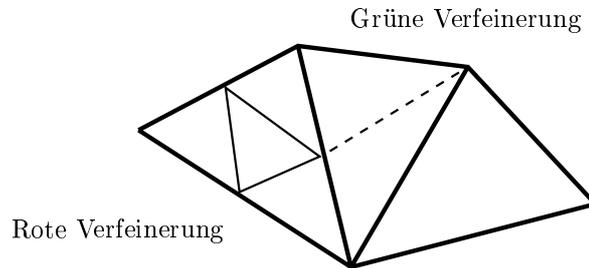


Abbildung 3.7: Rote und grüne Verfeinerung

Bei adaptiven Verfeinerungen benutzt man neben roten Verfeinerungen die **grüne Verfeinerung**. Dabei halbiert man ein Dreieck entlang einer Seitenhalbierenden. Die Klassifizierung der Zerlegung ändert sich, weil die Dreiecke spitzwinklicher werden. Es ist hierbei jedoch wichtiger, die Verfeinerung lokal zu halten und **hängende Knoten** wie in Abbildung 3.2 zu vermeiden.

Die Arbeitsweise eines adaptiven Löser ist folgende: Er verfeinert ein Dreieck mit großem Approximationsfehler rot. Dabei entstehen auf den Kanten der Nachbardreiecke freie Knoten, die in diesem Nachbardreieck bisher keine Bedeutung haben (hängende Knoten). Das macht die Zerlegung unzulässig. Verfeinert man dieses Nachbardreieck jetzt auch rot, entstehen wieder hängende Knoten auf den neuen Nachbardreiecken. Dabei entfernt man sich mit der Verfeinerung der Zerlegung immer weiter vom eigentlich kritischen Bereich. Um eine unnötige Ausbreitung der Verfeinerung zu verhindern, werden diese Nachbardreiecke grün verfeinert und der bisher hängende Knoten zum Eckpunkt der zwei neuen Dreiecke.

### 3.5 Numerische Experimente

Die Anwendung von Elementen höherer Ordnung ( $P_k$ -Elemente mit  $k \geq 2$ ) hat Vor- und Nachteile

- Bei einer glatten Lösung der linearen, elliptischen Differentialgleichung zweiter Ordnung ist eine grobe Zerlegung mit wenigen Teilgebieten  $K$  ausreichend, um einen kleineren Fehler der Approximations-Lösung zu erzielen. Es reicht deshalb,  $k$  zu erhöhen, um einen noch kleineren Fehler zu erhalten. Dabei steigt die Anzahl der Freiheitsgrade auf jedem Teilgebiet  $K$  bei einer Erhöhung der Ordnung  $k$  auf  $k + 1$  um den Faktor  $(k + n + 1)/(k + 1) \notin \mathbb{N}$  im  $\mathbb{R}^n$ .  
In Abbildung 3.8 sieht man das Ergebnis für das Problem aus Beispiel 3.19. Für  $P_k$ -Elemente mit  $k = 1$  bis  $k = 6$  wurden die erwarteten Konvergenzordnungen  $h^k$  bzgl. der  $\mathbf{H}^1$ -Norm und  $h^{k+1}$  bzgl. der  $\mathbf{L}^2$ -Norm und die in FEMLAB erreichten Konvergenzordnungen modulo einer Konstanten ermittelt. Dabei wurde  $h$  von  $\frac{1}{2}$  bis zu  $\frac{1}{32}$  verfeinert.  $H1_T$  bzw.  $L2_T$  sind die nach Satz 3.15 zu erwartenden Fehler,  $H1_{FEM}$  bzw.  $L2_{FEM}$  die Fehler der numerischen Approximation des Problems des FEMLAB-Programms.
- Ist die Lösung nicht glatt, empfiehlt es sich, die Zerlegung wenigstens in der Umgebung von Singularitäten zu verfeinern, da ein Element höherer Ordnung dort die Lösung nicht gut approximieren kann.

Der Effekt von Elementen höherer Ordnung auf einer groben Zerlegung gegenüber linearen Elementen auf einer feinen Zerlegung ist eine größere Genauigkeit der approximierten Lösung auf jedem Teilgebiet und damit ein kleinerer Gesamtfehler, wenn die Lösung ausreichend glatt ist wie in den Beispielen 3.19 und 6.3.

#### Beispiel 3.19 (Diffusionsdominanter Fall)

Man gibt die lineare, elliptische Differentialgleichung zweiter Ordnung

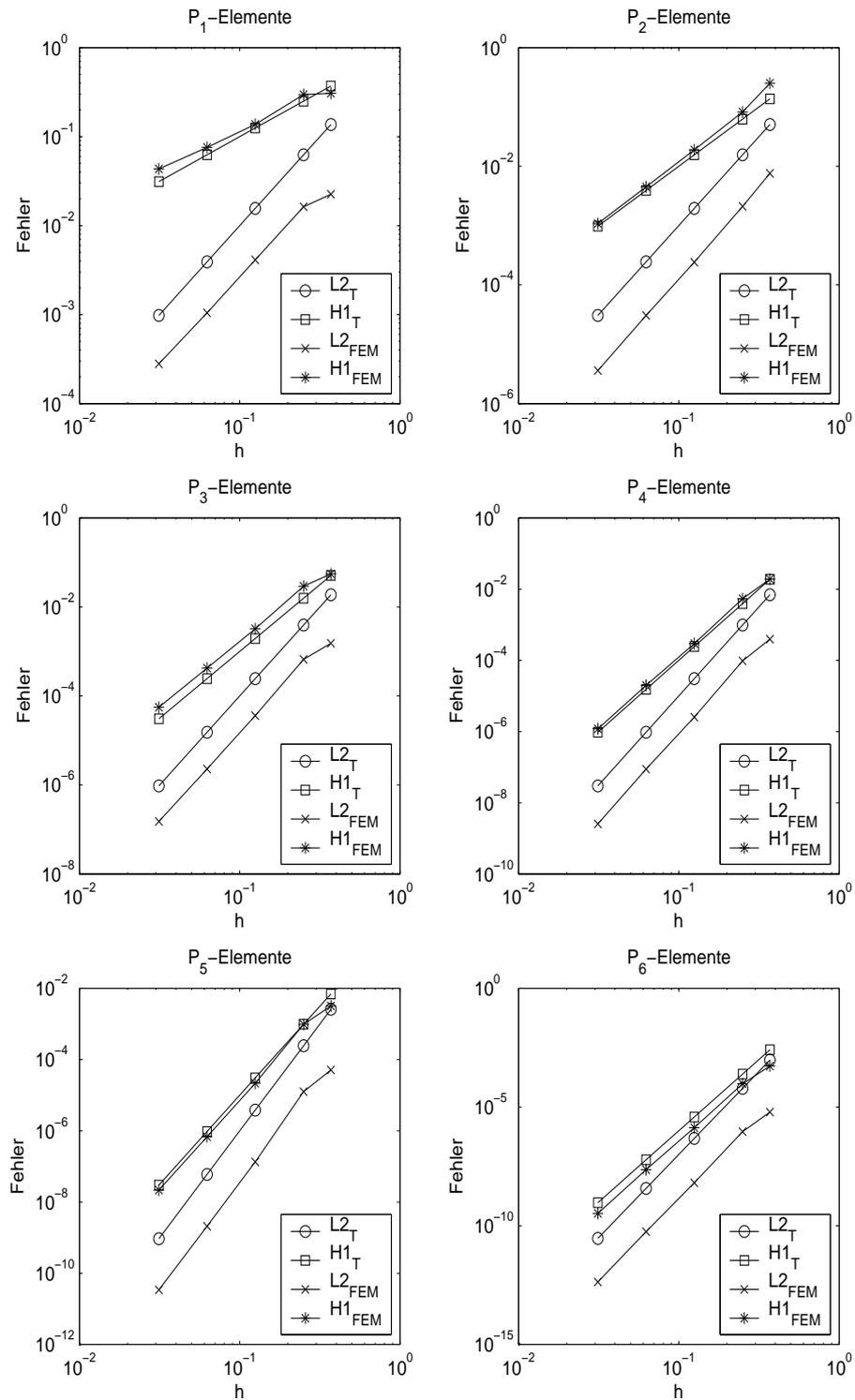
$$\begin{aligned} -\Delta u(x, y) + \vec{b}(x, y)\nabla u(x, y) + c(x, y)u(x, y) &= f(x, y) && \text{in } \Omega, \\ u(x, y) &= 0 && \text{auf } \Gamma \end{aligned}$$

mit

$$\begin{aligned} \vec{b}(x, y) &= (0, 0)^T, \\ c(x, y) &= 0, \\ f(x, y) &= e^{-xy} \{ (2\pi^2 - y^2 - x^2) \sin(\pi x) \sin(\pi y) + 2\pi[x \sin(\pi x) \cos(\pi y) \\ &\quad + y \cos(\pi x) \sin(\pi y)] \} \end{aligned}$$

und  $\Omega = [0, 1]^2$  in FEMLAB ein und vergleicht die errechnete Approximationslösung mit der Lösung des Problems  $u = \sin(\pi x) \sin(\pi y)e^{-xy}$ . Dabei erhält man die Resultate der Abbildung 3.9 für:

- lineare Elemente gerechnet auf einer Zerlegung, die sukzessiv rot verfeinert wird („Verfeinerung“), d.h. jeder Punkt auf der Kurve entspricht einer rot verfeinerten Zerlegung der Zerlegung im vorherigen Punkt. Der Punkt am linken Rand entspricht der Ausgangszerlegung.

Abbildung 3.8: Konvergenz von  $h = 1/3$  bis  $h = 1/32$  mit  $\vec{b} = \vec{0}$

- $P_k$ -Elemente mit  $k = 1, \dots, 8$  gerechnet auf der Ausgangszerlegung („ $k$ -Erhöhung“), d.h. für jeden Punkt auf der Kurve ist mit Elementen einer um 1 höheren Ordnung auf der Ausgangszerlegung das Problem gerechnet worden. Der Punkt am linken Rand entspricht der Ordnung  $k = 1$ .

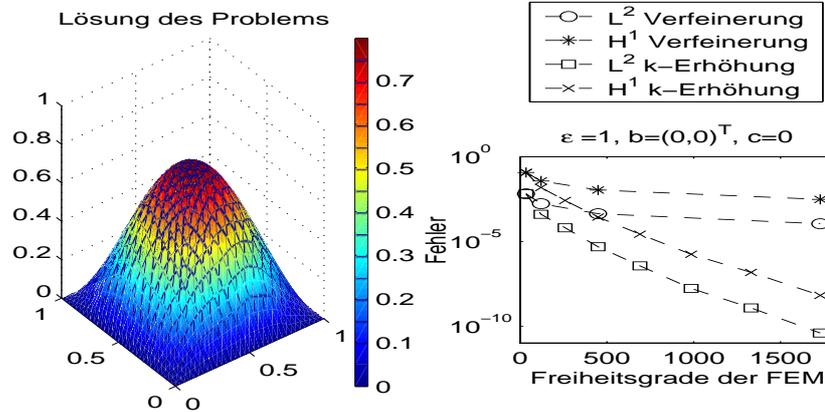


Abbildung 3.9: Echte Lösung und Approximationsfehler in der  $L^2$ - und  $H^1$ -Norm des diffusionsdominanten Beispiels 3.19

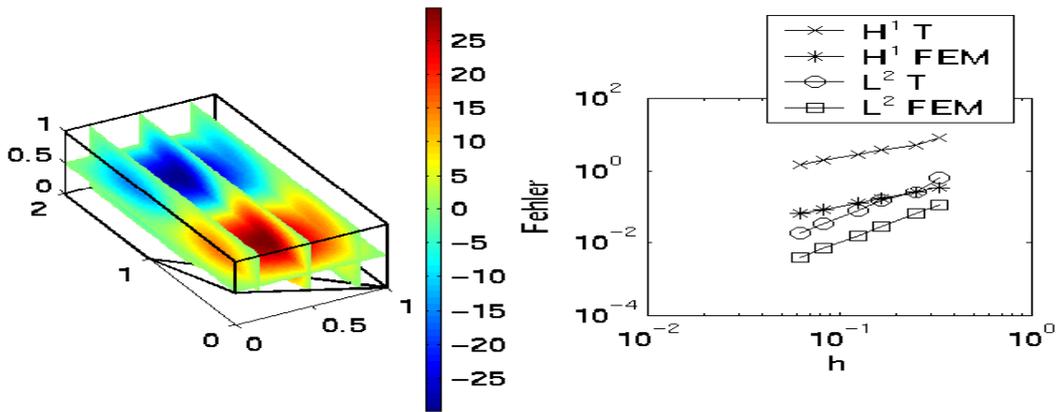


Abbildung 3.10: Lösung zu Beispiel 3.20 und Fehlerdiagramm

**Beispiel 3.20** (Dreidimensionales Beispiel zur Geometrie ‘fem3d’)

Das zu lösende Problem auf dem Gebiet  $\Omega := [0, 1] \times [0, 2] \times [0, 1] \setminus \{(x, y, z) \in \mathbb{R}^3 : x < 1 \text{ und } y < 1 \text{ und } z < \frac{1}{2}\}$  lautet

Finde  $u \in \mathbf{X}$ , so dass

$$\begin{aligned}
 -\Delta u(x, y, z) &= 90\pi^2 \sin(\pi x) \sin(\pi y) \sin(\pi z) && \text{in } \Omega, \\
 u(x, y, z) &= 0 && \text{auf den Randflächen 1, 2 und 4 – 7,} \\
 u(x, y, z) &= 30 \sin(\pi x) \sin(\pi y) \sin(\pi z) && \text{auf Randfläche 3}
 \end{aligned}$$

mit der Randflächennummerierung wie in Abbildung 3.1 gilt.

Die Lösung ist  $u(x, y, z) = 30 \sin(\pi x) \sin(\pi y) \sin(\pi z)$ . Links in Abbildung 3.10 ist ein Scheibendiagramm der Lösung mit 3 Scheiben in  $z$ -Richtung und einer Scheibe in  $x$ -Richtung zu sehen. Rechts sieht man ein Approximations-Fehlerdiagramm für  $\mathbf{L}^2$ - und  $\mathbf{H}^1$ -Norm mit einem  $h$  aus (3.1) zwischen  $\frac{1}{3}$  und  $\frac{1}{16}$ . Die Konvergenzrate für die hier verwendeten linearen Elemente liegt in der  $\mathbf{H}^1$ -Norm bei  $h$  (in der Legende der Abbildung 3.10 mit  $\mathbf{H}^1 T$  bezeichnet), in der  $\mathbf{L}^2$ -Norm bei  $h^2$  ( $\mathbf{L}^2 T$ ).  $\mathbf{H}^1 FEM$  und  $\mathbf{L}^2 FEM$  sind die entsprechenden Normen der FEMLAB-Lösung.

### Beispiel 3.21 (Adaptive Verfeinerung)

Auf Beispiel 2.4 wird ein adaptiver Löser angewandt mit den im FEMLAB-Code angegebenen Optionen. Dabei erhält man die in Abbildung 3.11 dargestellte Lösung und die dieser zugrunde liegende adaptive Verfeinerung.

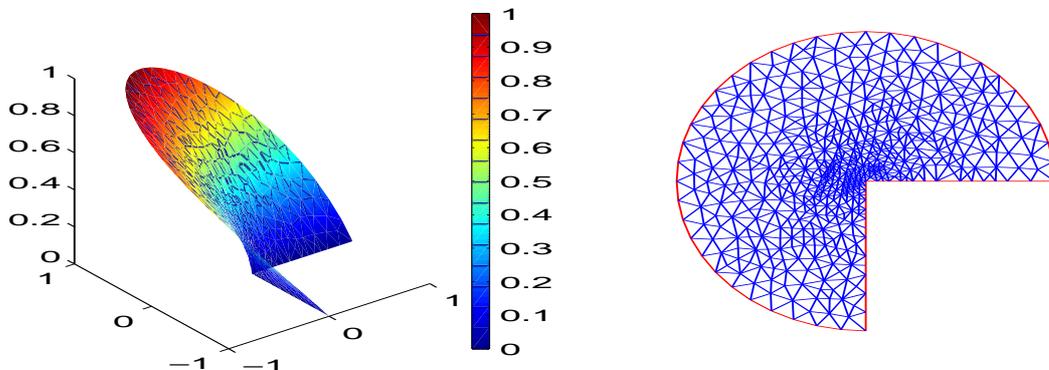


Abbildung 3.11: Die FEMLAB-Lösung und die Zerlegung des adaptiven Lözers zu Beispiel 2.4

Im folgenden FEMLAB-Code ist der Aufruf eines adaptiven Lözers implementiert, der aufgrund des mit Hilfe eines Fehlerschätzers errechneten Fehlers auf jedem Element die Zerlegung lokal verfeinert und zur Lösung auf der neuen Zerlegung den direkten oder den iterativen Löser aus IV verwendet.

```
sol3 = adaption(fem,'maxt',600,'eefun','fleel2','report','on');

% Optionen:
% 'maxt',600: Adaptive Verfeinerung der Zerlegung bis zur ersten Verfeine-
%             rung mit mehr als 600 finiten Elementen.
% 'eefun','fleel2': Fehlerschätzer in FEMLAB, der auf jedem Element den Feh-
%                   ler schätzt und dann ggf. eine adaptive Verfeinerung auf die-
%                   sem vornimmt (andere Fehlerschätzer sind: 'fleeceng' und
%                   'fleelfun'; man kann auch eigene schreiben).
% 'report','on': erzeugt eine Ausgabe der Adaptionsschritte auf den Bild-
%                schirm.
```

Es gibt noch eine zweite Möglichkeit, eine adaptive Zerlegung zu erzeugen. Dazu ist aber die Kenntnis der Position von Singularitäten, Unstetigkeiten oder Sprüngen der Lösung notwendig.

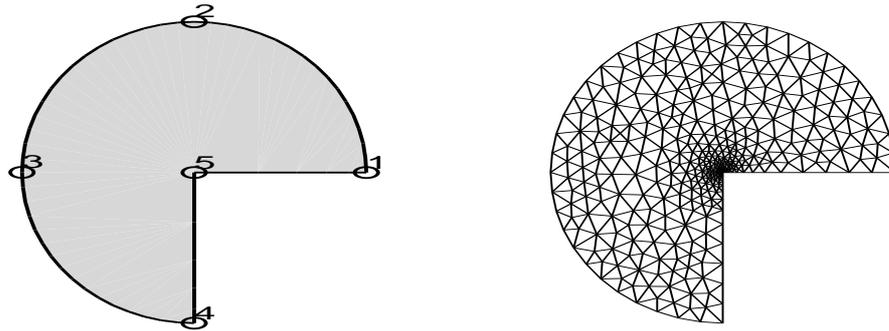


Abbildung 3.12: Gebiet mit nummerierten spezifischen Punkten in FEMLAB und die direkt erzeugte adaptive Zerlegung

Die Vorgehensweise ist folgende:

Man definiert sich seine Geometrie wie im Programmabschnitt I und definiert zusätzlich die Problemzonen als geometrische Objekte und addiert sie zur Gebietsgeometrie. Anschließend initialisiert man eine adaptive Zerlegung mit Hilfe des ‚meshinit‘-Befehls und löst das Problem auf dieser Zerlegung mit dem direkten Löser ‚femlin‘. Programmiert sieht dieses Verfahren („Verfahren mit gezielter Verfeinerung“) bezogen auf Beispiel 2.4 so aus:

```
fem.geom=arc2(0,0,1,0,3/2*pi);
subplot(1,2,1)
geomplot(fem,'pointlabels','on');
fem.mesh=meshinit(fem,'hmax',{[] [5;1/128]});

% Wie man im Geometrieplot links in Abbildung 3.12 erkennen kann, erhält
% die Ecke mit der Singularität aus Beispiel 2.4 die Nummer 5 im FEMLAB-
% Code. Deshalb wird im Aufruf von 'meshinit' die Nummer der Ecke und die
% Feinheit der Zerlegung in diesem Punkt angegeben. Die leere eckige Klamm-
% mer davor steht für ein nicht spezifiziertes maximales h auf dem ganzen
% Gebiet.
subplot(1,2,2)
meshplot(fem);

% Hier kommt jetzt Teil III des Programmcode, dann
fem.sol=femlin(fem);
```

Beim Verfahren mit gezielter Verfeinerung werden 696 Dreiecke erzeugt. Beim adaptiven Löser FEMLABs sind es 1084, denn die Abbruchbedingung „erste Verfeinerung mit mehr als 600 Dreiecken“ greift erst jetzt. Die Qualität der Lösung ist nicht gleich (vgl. Tabellen 3.2, 3.3 und die grafische Darstellung der Tabellenwerte in Abbildung 3.13 im Konvergenzdiagramm). Durch die Nutzung von mehr Informationen im Vorfeld erhält man eine bessere Approximation bei groben Zerlegungen. Wird die Zerlegung ausreichend fein in der Problemzone, ist eine stärkere globale Verfeinerung sinnvoll.

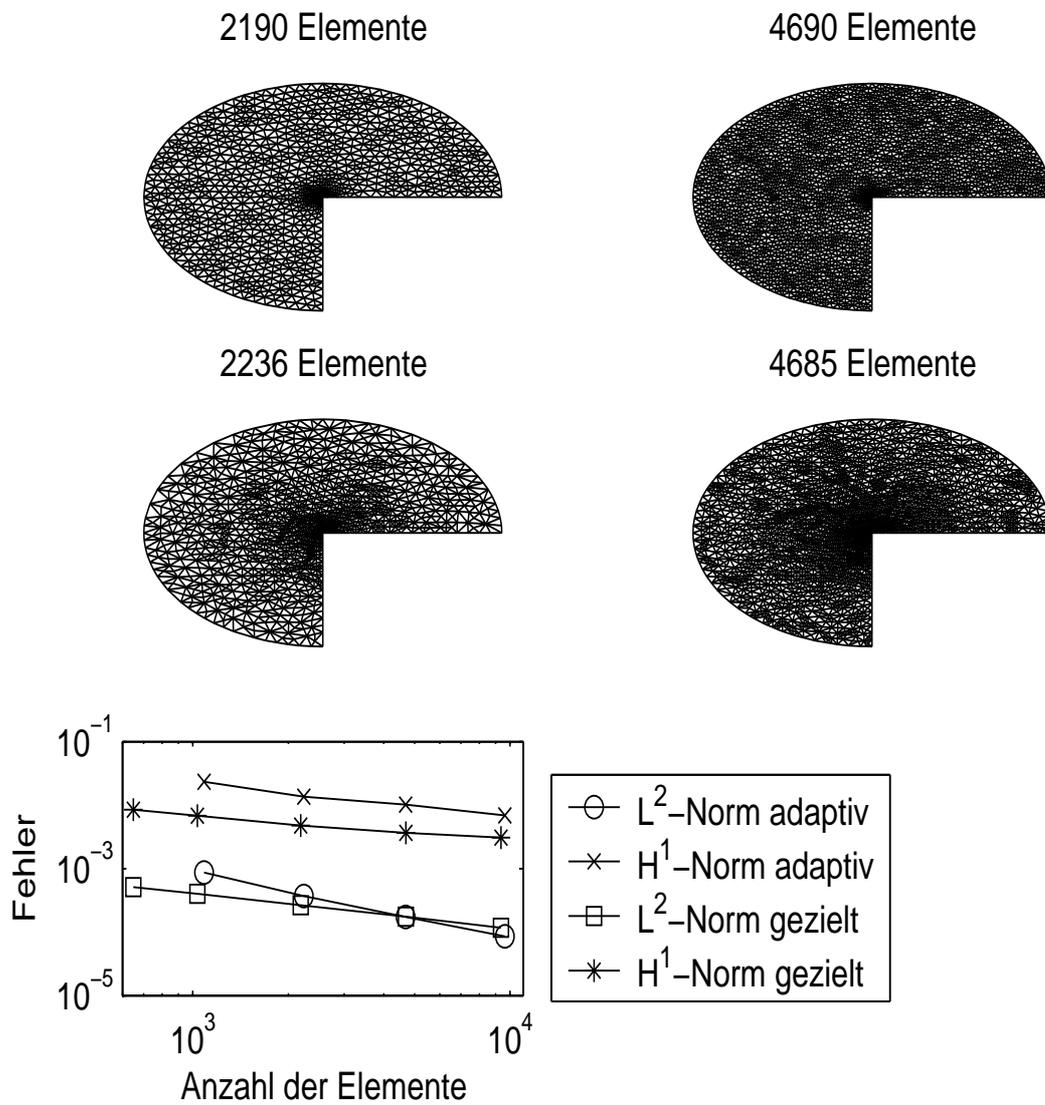


Abbildung 3.13: Feinere adaptive Zerlegungen und Fehlerdiagramm

Elemente	$L^2$ -Norm	$H^1$ -Norm
1084	$8.6539 \cdot 10^{-4}$	$2.3250 \cdot 10^{-2}$
2236	$3.6799 \cdot 10^{-4}$	$1.3580 \cdot 10^{-2}$
4685	$1.7265 \cdot 10^{-4}$	$1.0150 \cdot 10^{-2}$
9626	$8.5885 \cdot 10^{-5}$	$6.8567 \cdot 10^{-3}$

Tabelle 3.2: Normen der adaptiven Lösung

Elemente	$\mathbf{L}^2$ -Norm	$\mathbf{H}^1$ -Norm
1034	$4.0117 \cdot 10^{-4}$	$6.7743 \cdot 10^{-3}$
2190	$2.6552 \cdot 10^{-4}$	$4.7688 \cdot 10^{-3}$
4690	$1.7476 \cdot 10^{-4}$	$3.6684 \cdot 10^{-3}$
9362	$1.1712 \cdot 10^{-4}$	$3.0589 \cdot 10^{-3}$

Tabelle 3.3: Normen der Lösung bei gezielter Verfeinerung

Wir verfeinern bei beiden Verfahren weiter adaptiv. Danach stellen wir über der Elementanzahl der jeweiligen adaptiven Zerlegungen den Fehler in  $\mathbf{L}^2$ - und  $\mathbf{H}^1$ -Normen grafisch dar und erhalten das in Abbildung 3.13 dargestellte Konvergenzdiagramm nebst vier, zum Vergleich ausgewählten, adaptiven Zerlegungen. Die ersten beiden sind vom gezielt verfeinernden Algorithmus, die anderen beiden von FEMLABs adaptivem Algorithmus im Löser.



## Kapitel 4

# Einarbeitung von Randbedingungen in schwacher Form

In diesem Kapitel wird für das inhomogene Dirichlet–Randwertproblem der Poisson–Gleichung zunächst die Methode von Babuška aus [Bab73] in der Bearbeitung von Pitkäranta [Pit80] vorgestellt und anschließend in FEMLAB realisiert. Danach wird die stabilisierte Methode von Barbosa und Hughes erläutert und der Zusammenhang zur klassischen Methode von Nitsche hergestellt, wie bereits 1995 von Stenberg (vgl. [Ste95]). Abschließend wird noch allgemein auf andere Arten von Randbedingungen eingegangen. Dazu werden mittels FEMLAB einige Beispiele mit gemischten Randbedingungen gerechnet.

### 4.1 Das Dirichlet–Randwert–Problem

Die Grundlagen für die Theorie gemischter Finite–Elemente–Methoden sind von Babuška und Brezzi schon 1973 erarbeitet worden. Von ihnen wurden die Randbedingungen eines inhomogenen Dirichlet–Randwert–Problems durch Verwendung eines Lagrange–Multiplikators approximiert. Diese Methode konvergiert optimal, wenn die Kombination aus dem Funktionenraum auf dem Gebiet  $\Omega$  und dem Raum des Lagrange–Multiplikators auf dem Rand  $\Gamma$  einer ‚inf–sup–Bedingung‘ (vgl. (4.15)) genügt. Leider blieb die Frage nach der Konstruktion solcher Finite–Elemente–Unterräume unbeantwortet. Pitkäranta fand heraus, dass die Wahl dieser Räume sehr eingeschränkt ist.

Sei  $\Omega$  ein Gebiet im  $\mathbb{R}^n$ ,  $n = 2, 3$ , mit glattem Rand  $\Gamma$ . Betrachtet wird hier das **Dirichlet–Randwert–Problem der Poisson–Gleichung**:  
Finde  $u \in \mathbf{X}$ , so dass

$$-\Delta u = f \quad \text{in } \Omega, \tag{4.1}$$

$$u = g \quad \text{auf } \Gamma \tag{4.2}$$

gilt.

### Die schwache Formulierung

Gegeben seien  $f \in \mathbf{L}^2(\Omega)$  und  $g \in \mathbf{H}^{\frac{1}{2}}(\Gamma)$  (vgl. 1.23). Dann lautet die **schwache Formulierung** der Variationsgleichung:

Finde  $u \in \mathbf{H}^1(\Omega)$  und  $\lambda \in \mathbf{H}^{-\frac{1}{2}}(\Gamma)$ , so dass

$$\mathcal{B}(u, \lambda; v, \mu) = (f, v) + \langle g, \mu \rangle \quad \text{für alle } (v, \mu) \in \mathbf{H}^1(\Omega) \times \mathbf{H}^{-\frac{1}{2}}(\Gamma) \quad (4.3)$$

gilt.

Für die Bilinearform  $\mathcal{B}$  gilt

$$\mathcal{B}(u, \lambda; v, \mu) := (\nabla u, \nabla v) + \langle \lambda, v \rangle + \langle \mu, u \rangle. \quad (4.4)$$

Mit  $(\cdot, \cdot)$  wird das Skalarprodukt im  $\mathbf{L}^2(\Omega)$  aus (1.8) bezeichnet und mit  $\langle \cdot, \cdot \rangle$  das duale Produkt aus (1.26). Das Problem hat eine eindeutige Lösung (vgl. [Bab73]).

### Bedingung für den Lagrange-Multiplikator

Durch Umformen des Problems und Einsetzen der vom Dirichlet-Problem geforderten Bedingungen an  $u$  sowohl im Inneren von  $\Omega$  als auch auf dem Rand  $\Gamma$  erhält man eine Bedingung für den Lagrange-Multiplikator  $\lambda$ . Dazu nimmt man die schwache Formulierung her und formt sie um.

$$\begin{aligned} & \mathcal{B}(u, \lambda; v, \mu) - (f, v) - \langle g, \mu \rangle \\ &= (\nabla u, \nabla v) + \langle \lambda, v \rangle + \langle \mu, u \rangle - (f, v) - \langle g, \mu \rangle \\ &= 0 \quad \text{für alle } (v, \mu) \in \mathbf{H}^1(\Omega) \times \mathbf{H}^{-\frac{1}{2}}(\Gamma). \end{aligned}$$

Wenn  $u$  hinreichend glatt ist (d.h. z.B.  $u \in \mathbf{H}^2(\Omega)$ ), kann man den ersten Term partiell integrieren und erhält

$$\begin{aligned} & - \int_{\Omega} \Delta u(x) v(x) dx + \int_{\Gamma} \frac{\partial u(x)}{\partial n} v(x) ds + \langle \lambda, v \rangle + \langle u, \mu \rangle - \int_{\Omega} f(x) v(x) dx - \langle g, \mu \rangle \\ &= 0 \quad \text{für alle } (v, \mu) \in \mathbf{H}^1(\Omega) \times \mathbf{H}^{-\frac{1}{2}}(\Gamma). \end{aligned}$$

Schränkt man in der obigen Gleichung den Raum der Testfunktionen  $(v, \mu)$  aus (4.3) ein auf  $\mathbf{C}_0^\infty(\Omega) \times \mathbf{L}^2(\Gamma)$ , gelten (4.1) und (4.2) im distributionellen Sinn. Also folgt

$$\begin{aligned} & - \int_{\Omega} \Delta u(x) v(x) dx + \int_{\Gamma} \frac{\partial u(x)}{\partial n} v(x) ds + \langle \lambda, v \rangle + \langle g, \mu \rangle - \langle g, \mu \rangle + \int_{\Omega} \Delta u(x) v(x) dx \\ &= \int_{\Gamma} \frac{\partial u(x)}{\partial n} v(x) ds + \int_{\Gamma} \lambda v(x) ds \\ &= 0 \quad \text{für alle } (v, \mu) \in \mathbf{C}_0^\infty(\Omega) \times \mathbf{L}^2(\Gamma) \end{aligned}$$

und für  $\lambda$  erhält man

$$\lambda = - \frac{\partial u}{\partial n}. \quad (4.5)$$

### Finite–Elemente–Räume

Sei  $\mathcal{T}_h$  eine zulässige Zerlegung (vgl. Definition 3.1) von  $\overline{\Omega}$ . Wir betrachten den Finite–Elemente–Raum  $\mathcal{C}_{\mathcal{T}}$  aus (3.27). Dann ist

$$\mathbf{X}_h = \{v \in \mathbf{H}^1(\Omega) \mid v|_K \in \mathbb{P}_k(K) \text{ für alle } K \in \mathcal{T}_h\}$$

mit  $\mathbb{P}_k$  als Raum der Polynome vom Grad  $k \geq 1$  der Raum der Finite–Elemente–Testfunktionen  $v$  aus  $\mathbf{H}^1(\Omega)$ . Ferner sei

$$\mathbf{M}_h = \{\mu \in \mathbf{L}^2(\Gamma) \mid \mu(x)|_E = \hat{\mu}(F_E^{-1}(x)) \text{ für irgendein } \hat{\mu} \in \mathbb{P}_l(\hat{E}), \text{ für alle } E \in \mathcal{E}_h\},$$

mit  $l \geq 0$  der Raum des Lagrange–Multiplikators  $\lambda$ , wobei  $\mathcal{E}_h = \{E_j\}_{j=1}^{M_\Gamma}$  die Zerlegung des Randes  $\Gamma$  in Teilgebiete  $E_j$  ist mit Index  $h := \max_{E \in \mathcal{E}_h} h_E$  (im  $\mathbb{R}^2$  Kurvenstücke, im  $\mathbb{R}^3$  krummflächige Dreiecke). Außerdem ist  $F_E : \hat{E} \rightarrow E$  die Abbildung vom Referenzelement  $\hat{E}$  auf  $E$  (vgl. (3.8) für  $K \in \mathcal{T}_h$ ). An die Randzerlegung  $\mathcal{E}_h$  fordern wir die Regularität

$$\|F_E\|_2 \leq Ch_E \quad \text{und} \quad \|F_E^{-1}\|_2 \leq Ch_E^{-1} \quad \text{für alle } E \in \mathcal{E}_h, \quad (4.6)$$

wobei  $h_E$  der Durchmesser von  $E$  ist. Für isotrope Zerlegungen (vgl. Definition 3.2) gilt zusätzlich

$$C_1 h_K \leq h_E \leq C_2 h_K \quad \text{für alle } K \in \mathcal{T}_h \text{ und } E \in \mathcal{E}_h : K \cap E \neq \emptyset. \quad (4.7)$$

### Originalmethode von Babuška

Gesucht werden  $(u_h, \lambda_h) \in \mathbf{X}_h \times \mathbf{M}_h$ , so dass

$$\mathcal{B}(u_h, \lambda_h; v, \mu) = (f, v) + \langle \mu, g \rangle \quad \text{für alle } (v, \mu) \in \mathbf{X}_h \times \mathbf{M}_h \quad (4.8)$$

mit  $f \in \mathbf{L}^2(\Omega)$  und  $g \in \mathbf{H}^{\frac{1}{2}}(\Gamma)$  gilt.

Unter Verwendung der folgenden zerlegungsabhängigen Normen lässt sich die Stabilität und die Fehlerabschätzung des Problems (4.8) leichter zeigen als es in [Bab73] getan wurde.

$$\|v\|_{\frac{1}{2},h}^2 = \sum_{E \in \mathcal{E}_h} h_E^{-1} \|v\|_{0,E}^2 \quad \text{für } v \in \mathbf{H}^1(\Omega) \quad (4.9)$$

$$\text{und} \quad \|z\|_{-\frac{1}{2},h}^2 = \sum_{E \in \mathcal{E}_h} h_E \|z\|_{0,E}^2 \quad \text{für } z \in \mathbf{L}^2(\Gamma). \quad (4.10)$$

Für diese Normen gilt dann

$$\langle v, z \rangle \leq \|v\|_{\frac{1}{2},h} \|z\|_{-\frac{1}{2},h} \quad \text{für alle } (v, z) \in \mathbf{H}^1(\Omega) \times \mathbf{L}^2(\Gamma). \quad (4.11)$$

Außerdem definiert man

$$\|v\|_{1,h} = \|v\|_{1,\Omega} + \|v\|_{\frac{1}{2},h} \quad \text{für alle } v \in \mathbf{H}^1(\Omega). \quad (4.12)$$

Es folgen zwei Lemmata. Das erste beinhaltet eine Interpolations–Abschätzung auf  $\mathbf{M}_h$ , das zweite eine diskrete Spurungleichung für die Normalenableitung einer Funktion aus  $\mathbf{X}_h$ .

**Lemma 4.1** Für  $\lambda \in \mathbf{H}^{l+1}(\Gamma)$  gilt

$$\inf_{\mu \in \mathbf{M}_h} \|\lambda - \mu\|_{-\frac{1}{2},h} \leq Ch^{l+\frac{3}{2}} \|\lambda\|_{l+1,\Gamma}. \quad (4.13)$$

**Lemma 4.2** Es gibt eine Konstante  $C_I$ , so dass

$$C_I \left\| \frac{\partial v}{\partial n} \right\|_{-\frac{1}{2},h} \leq \|\nabla v\|_{0,\Omega} \quad \text{für alle } v \in \mathbf{X}_h \quad (4.14)$$

gilt.

Die Konvergenz der von Pitkäranta variierten Methode von Babuška ist gegeben durch den folgenden Satz.

**Satz 4.3** Der Finite-Elemente-Raum, Unterraum von  $\mathbf{H}^1(\Omega) \times \mathbf{L}^2(\Gamma)$ , erfülle die Bedingungen

$$\inf_{\mu \in \mathbf{M}_h} \sup_{v \in \mathbf{X}_h \setminus \{0\}} \frac{\langle \mu, v \rangle}{\|v\|_{1,h} \|\mu\|_{-\frac{1}{2},h}} \geq C \quad (4.15)$$

und

$$|v|_{1,\Omega}^2 \geq C \|v\|_{1,h}^2 \quad \text{für alle } v \in \{v \in \mathbf{X}_h \mid \langle \mu, v \rangle = 0 \text{ für alle } \mu \in \mathbf{M}_h\}. \quad (4.16)$$

Für die Lösung  $(u_h, \lambda_h) \in \mathbf{X}_h \times \mathbf{M}_h$  von (4.8) gilt dann

$$\|u - u_h\|_{1,h} + \|\lambda - \lambda_h\|_{-\frac{1}{2},h} \leq C \left( h^k \|u\|_{k+1,\Omega} + h^{l+\frac{3}{2}} \|\lambda\|_{l+1,\Gamma} \right) \quad (4.17)$$

für  $u \in \mathbf{H}^{k+1}(\Omega)$  und  $\lambda \in \mathbf{H}^{l+1}(\Gamma)$ .

## Realisierung in FEMLAB

In FEMLAB kann man auch schwache Randbedingungen mit Hilfe des Lagrange-Multiplikators  $\lambda$  programmieren. Das ist jedoch etwas aufwändiger. Man braucht Zerlegungen des Gebietes  $\Omega$  und des Randes  $\Gamma$ . Diese werden zusammengefügt und das neue Problem dann mit FEMLAB gelöst. Dazu benötigt man einige Funktionen, die nicht in FEMLAB implementiert sind. Für das folgende Beispiel wurden Programmroutinen aus [Rap03] benutzt. Am Ende des nächsten Abschnitts wird noch einmal dieses Beispiel aufgegriffen und die im Programm von Rapin vorhandene Stabilisierung benutzt.

**Beispiel 4.4** Zur Veranschaulichung betrachten wir das Problem auf dem Einheitsquadrat  $\Omega := [0, 1]^2$ :

Finde  $(u, \lambda) \in \mathbf{X}_h \times \mathbf{M}_h$ , so dass

$$\begin{aligned} -\Delta u(x, y) &= 2\pi^2 \sin(\pi x) \sin(\pi y) && \text{für alle } (x, y) \in \Omega, \\ u(x, y) &= 0 && \text{für alle } (x, y) \in \Gamma \end{aligned}$$

gilt.

Die Lösung ist  $u(x, y) = \sin(\pi x) \sin(\pi y)$ . Mit der Bedingung (4.5) gilt

$$\lambda(x, y) = -\frac{\partial u}{\partial n}(x, y) = \begin{cases} \pi \sin(\pi y) & , x = 0 \vee x = 1 \text{ für alle } y \in [0, 1] \\ \pi \sin(\pi x) & , y = 0 \vee y = 1 \text{ für alle } x \in [0, 1] \end{cases} \quad (4.18)$$

In Abbildung 4.1 sieht man acht Grafiken. Die erste zeigt die Zerlegung für  $u$  im Inneren des Gebietes bis zum Rand mit  $h = \frac{1}{18}$ . Die Grafiken zwei bis vier sind für  $h = \frac{1}{75}$  erstellt. Die ersten beiden davon zeigen das Gebiet  $\Omega$ . Zuerst die Randzerlegung für  $\lambda_h$  und schließlich die gemeinsamen Randpunkte von  $\lambda$  und  $u$  mit  $h_E = 1/60$  und  $h_K = 1/75$ . Das Ergebnis für  $\lambda$  nach (4.18) und für das approximierte  $\lambda_h$  sieht man in den Grafiken vier und fünf. Das sechste Bild zeigt den  $L^2$ -Fehler nach Definition (4.10) des Lagrange-Multiplikators der Ordnung  $h^{l+\frac{3}{2}}$  mit  $l = -\frac{1}{2}$  aus (4.17) mit  $h_E = 1/40$  bis  $h_E = 1/125$ . Die ‚inf-sup-Bedingung‘ (4.15) ist erfüllt. In Grafik sieben und acht sieht man für eine verletzte ‚inf-sup-Bedingung‘ die gemeinsamen Randpunkte von  $\lambda$  und  $u$  mit  $h_E = 1/18$  und  $h_K = 1/12$  und das approximierte  $\lambda_h$  für diesen Fall.

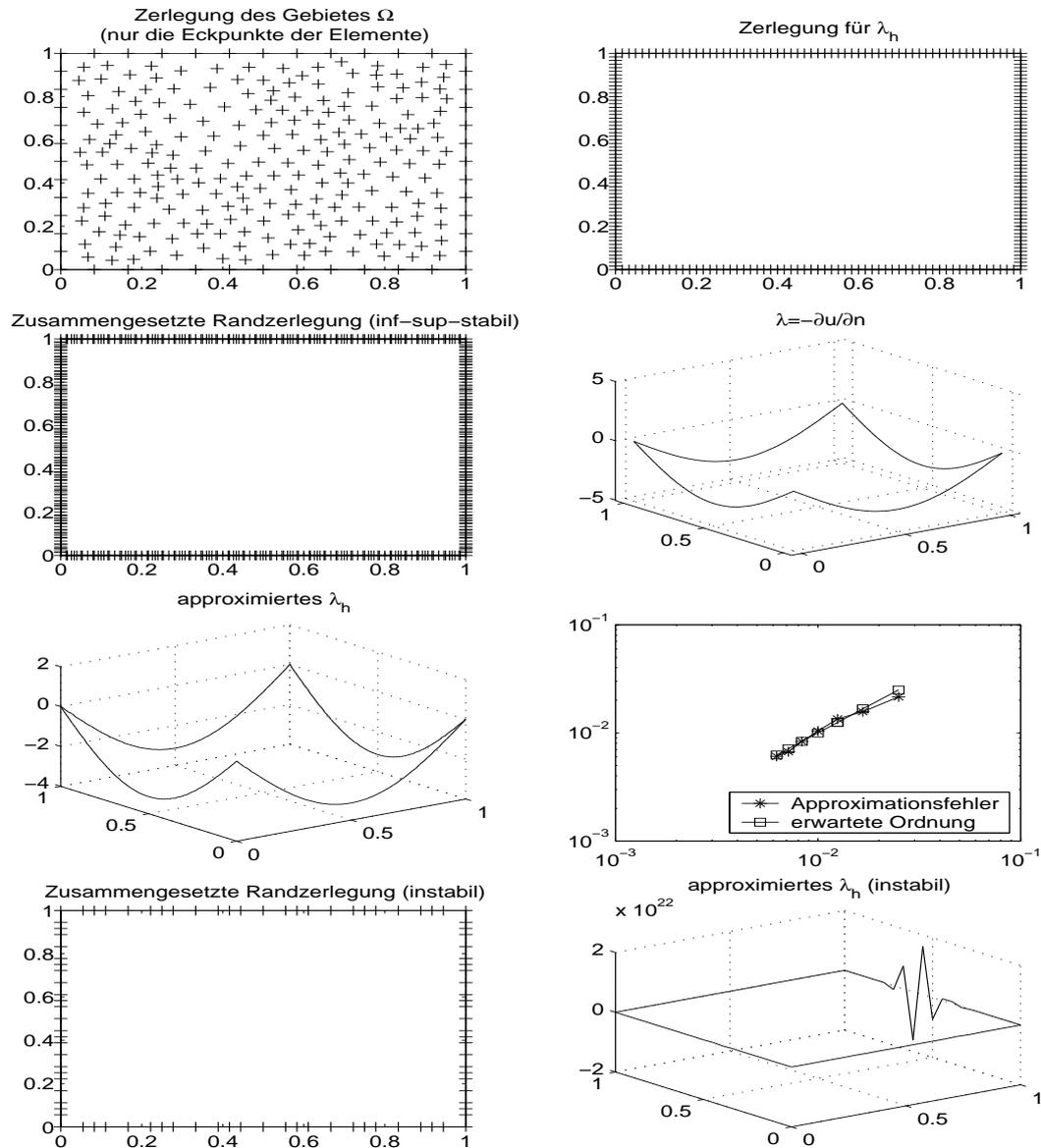


Abbildung 4.1: Grafische Ausgabe zu Beispiel 4.4

## 4.2 Stabilisierte Methoden

Barbosa und Hughes erhielten 1991 durch die Addition von Stabilisierungstermen in der schwachen Form des Dirichlet–Randwert–Problems Stabilität ohne die ‚inf-sup-Bedingung‘ (vgl. [Bar91]) zu erfüllen. Diese gemischte Finite–Elemente–Methode gewann dadurch an Potential als Lösungsmethode. Die Stabilisierung ist im folgenden für die schwache Formulierung (4.3) des Problems (4.1), (4.2) dargestellt.

### Methode von Barbosa und Hughes

Die **stabilisierte Methode** lautet:

Finde  $(u_h, \lambda_h) \in \mathbf{X}_h \times \mathbf{M}_h$ , so dass

$$\mathcal{B}_h(u_h, \lambda_h; v, \mu) = (f, v) + \langle g, \mu \rangle \quad \text{für alle } (v, \mu) \in \mathbf{X}_h \times \mathbf{M}_h \quad (4.19)$$

gilt. Dabei ist

$$\mathcal{B}_h(u_h, \lambda_h; v, \mu) := \mathcal{B}(u_h, \lambda_h; v, \mu) - \alpha \sum_{E \in \mathcal{E}_h} h_E \left\langle \lambda + \frac{\partial u}{\partial n}, \mu + \frac{\partial v}{\partial n} \right\rangle_E, \quad (4.20)$$

wobei  $\mathcal{B}$  die in (4.4) definierte Bilinearform ist und  $\langle \cdot, \cdot \rangle_E$  das Skalarprodukt im  $\mathbf{L}^2$  auf  $E \in \mathcal{E}_h$  ist.

Zuerst erhält man eine Konsistenzaussage.

**Lemma 4.5** *Für die exakte Lösung  $(u, \lambda) \in \mathbf{H}^2(\Omega) \cap \mathbf{H}_0^1(\Omega) \times \mathbf{L}^2(\Gamma)$  von (4.3) gilt*

$$\mathcal{B}_h(u, \lambda; v, \mu) = (f, v) + \langle g, \mu \rangle \quad \text{für alle } (v, \mu) \in \mathbf{X}_h \times \mathbf{M}_h. \quad (4.21)$$

**Beweis :** Wie schon gesehen, gilt in der schwachen Formulierung  $\lambda = -\frac{\partial u}{\partial n}$ . Setzt man diese Bedingung für  $\lambda$  in (4.19) ein, erhält man wieder (4.3).  $\square$

Als nächstes zeigen wir die Stabilität in Form einer diskreten ‚Babuška–Brezzi–Bedingung‘.

**Lemma 4.6** *Sei  $0 < \alpha < C_7^2$ . Dann existiert eine Konstante  $C > 0$ , so dass*

$$\inf_{(v, \mu) \in \mathbf{X}_h \times \mathbf{M}_h} \sup_{(z, \eta) \in \mathbf{X}_h \times \mathbf{M}_h} \frac{\mathcal{B}_h(v, \mu; z, \eta)}{\left( \|z\|_{1,h} + \|\eta\|_{-\frac{1}{2},h} \right) \left( \|v\|_{1,h} + \|\mu\|_{-\frac{1}{2},h} \right)} \geq C \quad (4.22)$$

gilt.

**Beweis :** Seien  $(v, \mu) \in \mathbf{X}_h \times \mathbf{M}_h$  beliebig. Dann gilt nach (4.14)

$$\begin{aligned}
\mathcal{B}_h(v, \mu; v, -\mu) &= \|\nabla v\|_{0,\Omega}^2 + \underbrace{\langle v, \mu \rangle + \langle v, -\mu \rangle}_{=0} + \alpha \sum_{E \in \mathcal{E}_h} h_E \left\langle \mu + \frac{\partial v}{\partial n}, \mu - \frac{\partial v}{\partial n} \right\rangle_E \\
&= \|\nabla v\|_{0,\Omega}^2 + \alpha \sum_{E \in \mathcal{E}_h} h_E \left( \|\mu\|_{0,E}^2 - \left\| \frac{\partial v}{\partial n} \right\|_{0,E}^2 \right) \\
&\stackrel{(4.10),(4.14)}{\geq} \left( 1 - \frac{\alpha}{C_I^2} \right) \|\nabla v\|_{0,\Omega}^2 + \alpha \sum_{E \in \mathcal{E}_h} h_E \|\mu\|_{0,E}^2 \\
&\geq C_1 \left( \|\nabla v\|_{0,\Omega}^2 + \|\mu\|_{-\frac{1}{2},h}^2 \right)
\end{aligned} \tag{4.23}$$

mit der Bedingung  $0 < \alpha < C_I^2$ . Sei nun  $\Pi_h : \mathbf{L}^2(\Gamma) \rightarrow \mathbf{M}_h$  die  $\mathbf{L}^2$ -Projektion mit

$$\langle g, \mu \rangle = \langle \Pi_h g, \mu \rangle \quad \text{für alle } \mu \in \mathbf{M}_h. \tag{4.24}$$

Da die Funktionen in  $\mathbf{M}_h$  unstetig sind, lässt sich  $\bar{\mu} \in \mathbf{M}_h$  definieren als  $\bar{\mu}|_E = h_E^{-1} \Pi_h v|_E$  für alle  $E \in \mathcal{E}_h$ . Dann gilt

$$\|\bar{\mu}\|_{-\frac{1}{2},h} = \|\Pi_h v\|_{\frac{1}{2},h}. \tag{4.25}$$

**Beweis :**(der Behauptung (4.25))

$$\begin{aligned}
\|\bar{\mu}\|_{-\frac{1}{2},h}^2|_E &\stackrel{(4.10)}{=} h_E \|\bar{\mu}\|_{0,E}^2 \\
&= h_E^2 h_E^{-1} \|\bar{\mu}\|_{0,E}^2 \\
&\stackrel{(4.9)}{=} h_E^2 \|\bar{\mu}\|_{\frac{1}{2},h}^2|_E
\end{aligned}$$

und durch Ziehen der Wurzel aus dieser Gleichung erhält man

$$\begin{aligned}
\|\bar{\mu}\|_{-\frac{1}{2},h} &= h_E \|\bar{\mu}\|_{\frac{1}{2},h} \\
&\stackrel{n.Def.}{=} h_E \|h_E^{-1} \Pi_h v\|_{\frac{1}{2},h} \\
&= \|\Pi_h v\|_{\frac{1}{2},h}.
\end{aligned} \quad \square$$

Mit Benutzung von (4.11), (4.14) und der Youngschen Ungleichung ergibt sich

$$\begin{aligned}
\mathcal{B}_h(v, \mu; 0, \bar{\mu}) &= \underbrace{\langle \nabla v, \nabla 0 \rangle}_{=0} + \underbrace{\langle v, \bar{\mu} \rangle + \langle \mu, 0 \rangle}_{=0} - \alpha \sum_{E \in \mathcal{E}_h} h_E \left\langle \mu + \frac{\partial v}{\partial n}, \bar{\mu} \right\rangle_E \\
&\stackrel{(4.25)}{=} \langle v, h_E^{-1} \Pi_h v \rangle - \alpha \sum_{E \in \mathcal{E}_h} h_E \left\langle \mu + \frac{\partial v}{\partial n}, h_E^{-1} \Pi_h v \right\rangle_E \\
&\stackrel{(4.24)}{=} \sum_{E \in \mathcal{E}_h} h_E^{-1} \|\Pi_h v\|_{0,E}^2 - \alpha \sum_{E \in \mathcal{E}_h} \left\langle \mu + \frac{\partial v}{\partial n}, \Pi_h v \right\rangle_E
\end{aligned}$$

$$\begin{aligned}
&= \sum_{E \in \mathcal{E}_h} h_E^{-1} \|\Pi_h v\|_{0,E}^2 - \alpha \sum_{E \in \mathcal{E}_h} \sqrt{h_E} \left( \left\| \frac{\partial v}{\partial n} \right\|_{0,E} + \|\mu\|_{0,E} \right) \\
&\quad \sqrt{h_E^{-1}} \|\Pi_h v\|_{0,E} \\
&\stackrel{(4.9)}{\geq} \|\Pi_h v\|_{\frac{1}{2},h}^2 - \alpha \|\Pi_h v\|_{\frac{1}{2},h} \left( \left\| \frac{\partial v}{\partial n} \right\|_{-\frac{1}{2},h} + \|\mu\|_{-\frac{1}{2},h} \right) \\
&\stackrel{(4.14)}{\geq} \|\Pi_h v\|_{\frac{1}{2},h}^2 - \|\Pi_h v\|_{\frac{1}{2},h} \left( \alpha \left( C_I^{-1} \|\nabla v\|_{0,\Omega} + \|\mu\|_{-\frac{1}{2},h} \right) \right) \\
&\stackrel{quadr. Erg.}{\geq} \|\Pi_h v\|_{\frac{1}{2},h}^2 - \frac{1}{2} \|\Pi_h v\|_{\frac{1}{2},h}^2 - \frac{1}{2} \alpha \left( C_I^{-1} \|\nabla v\|_{0,\Omega} + \|\mu\|_{-\frac{1}{2},h} \right)^2 \\
&\geq \frac{1}{2} \|\Pi_h v\|_{\frac{1}{2},h}^2 - C_3 \left( \|\nabla v\|_{0,\Omega}^2 + \|\mu\|_{-\frac{1}{2},h}^2 \right). \tag{4.26}
\end{aligned}$$

Sei  $(z, \eta) = (v, -\mu + \delta \bar{\mu})$  mit  $\delta > 0$ . Man benutzt die bisherigen Ergebnisse und bekommt folgende neue Aussagen:

$$\begin{aligned}
1) \quad \mathcal{B}_h(v, \mu; z, \eta) &= \mathcal{B}_h(v, \mu; v, -\mu) + \delta \mathcal{B}_h(v, \mu; 0, \bar{\mu}) \\
&\stackrel{(4.23), (4.26)}{\geq} (C_1 - \delta C_3) \left( \|\nabla v\|_{0,\Omega}^2 + \|\mu\|_{-\frac{1}{2},h}^2 \right) + \frac{1}{2} \delta \|\Pi_h v\|_{\frac{1}{2},h}^2 \\
&\stackrel{\delta < \frac{C_1}{C_3}}{\geq} C \left( \|\nabla v\|_{0,\Omega}^2 + \|\mu\|_{-\frac{1}{2},h}^2 + \|\Pi_h v\|_{\frac{1}{2},h}^2 \right), \tag{4.27}
\end{aligned}$$

$$\begin{aligned}
2) \quad \|\nabla v\|_{0,\Omega}^2 + \|\Pi_h v\|_{\frac{1}{2},h}^2 &\stackrel{(4.24)}{=} |v|_{1,\Omega}^2 + |v|_{\frac{1}{2},h}^2 \\
&\stackrel{(4.16)}{\geq} C |v|_{1,h}^2 + \underbrace{|v|_{\frac{1}{2},h}^2}_{\geq 0} \\
&\geq C |v|_{1,h}^2, \tag{4.28}
\end{aligned}$$

$$\begin{aligned}
3) \quad \|z\|_{1,h} + \|\eta\|_{-\frac{1}{2},h} &= \|v\|_{1,h} + \|-\mu + \delta \bar{\mu}\|_{-\frac{1}{2},h} \\
&\stackrel{\Delta\text{-Ungl.}}{\leq} \|v\|_{1,h} + \|\mu\|_{-\frac{1}{2},h} + \delta \|\bar{\mu}\|_{-\frac{1}{2},h} \\
&\stackrel{(4.25)}{=} \|v\|_{1,h} + \|\mu\|_{-\frac{1}{2},h} + \delta \|\Pi_h v\|_{\frac{1}{2},h} \\
&\stackrel{(4.12), (4.24)}{=} \|v\|_{1,\Omega} + \|v\|_{\frac{1}{2},h} + \|\mu\|_{-\frac{1}{2},h} + \delta \|v\|_{\frac{1}{2},h} \\
&\leq (1 + \delta) \left( \|v\|_{1,\Omega} + \|v\|_{\frac{1}{2},h} \right) + \|\mu\|_{-\frac{1}{2},h} \\
&\stackrel{(4.12)}{=} (1 + \delta) \|v\|_{1,h} + \|\mu\|_{-\frac{1}{2},h} \\
&\leq C \left( \|v\|_{1,h} + \|\mu\|_{-\frac{1}{2},h} \right). \tag{4.29}
\end{aligned}$$

Setzt man die Aussagen (4.27)–(4.29) zusammen, erhält man

$$\begin{aligned}
\sup_{(z,\eta) \in \mathbf{X}_h \times \mathbf{M}_h} \frac{\mathcal{B}_h(v, \mu; z, \eta)}{\left(\|z\|_{1,h} + \|\eta\|_{-\frac{1}{2},h}\right) \left(\|v\|_{1,h} + \|\mu\|_{-\frac{1}{2},h}\right)} &\stackrel{(4.29)}{\geq} \frac{\mathcal{B}_h(v, \mu; z, \eta)}{\tilde{C} \left(\|v\|_{1,h} + \|\mu\|_{-\frac{1}{2},h}\right)^2} \\
&\stackrel{(4.27),(4.28)}{\geq} \frac{\hat{C} \left(\|v\|_{1,h}^2 + \|\mu\|_{-\frac{1}{2},h}^2\right)}{\tilde{C} \left(\|v\|_{1,h} + \|\mu\|_{-\frac{1}{2},h}\right)^2} \\
&\geq C
\end{aligned}$$

Damit ist die Stabilität gezeigt.  $\square$

Es fehlt nun noch die Stetigkeit der Bilinearform (4.20). Man nimmt zwei Lemmata zu Hilfe. Das Lemma 4.7 stellt Fehlerabschätzungen für den Interpolationsfehler auf einem Randstück  $E$  bereit, das Lemma 4.8 gibt eine Abschätzung der Norm der Normalenableitung auf dem Gebietsrand  $\Gamma$ .

**Lemma 4.7** *Für  $u \in \mathbf{H}^{k+1}(\Omega)$ ,  $k \geq 1$ , existiert ein Interpolationsoperator  $\Pi_h$  und positive Konstanten  $C_J$  und  $C_K$ , so dass für alle  $E \subset \mathcal{E}_h$*

$$(i) \quad \|u - \Pi_h u\|_E \leq C_K h_E^{k+\frac{1}{2}} \|u\|_{k+1,K} \quad \text{und} \quad (4.30)$$

$$(ii) \quad \left\| \frac{\partial(u - \Pi_h u)}{\partial n} \right\|_E \leq C_J h_E^{k-\frac{1}{2}} \|u\|_{k+1,K} \quad \text{mit } E \subset K \quad (4.31)$$

gelten.

**Beweis :** (i) siehe [Pit80], Prop. 4.1, (4.11), (ii) folgt aus (i).  $\square$

**Lemma 4.8** *Für  $v := u - \Pi_h u$ ,  $u \in \mathbf{H}^{k+1}(\Omega)$ ,  $k \geq 1$ , gibt es eine Konstante  $C_1 > 0$ , so dass*

$$\left\| \frac{\partial v}{\partial n} \right\|_{-\frac{1}{2},h} \leq C_1 h^k \|u\|_{k+1,\Omega} \quad (4.32)$$

gilt.

**Beweis :**

$$\begin{aligned}
\left\| \frac{\partial v}{\partial n} \right\|_{-\frac{1}{2},h}^2 &= \sum_{E \in \mathcal{E}_h} h_E \left\langle \frac{\partial v}{\partial n}, \frac{\partial v}{\partial n} \right\rangle_E \\
&\stackrel{(4.31)}{\leq} \sum_{E \in \mathcal{E}_h} h_E C_J^2 h_E^{2(k-\frac{1}{2})} \|u\|_{k+1,K}^2 \\
&\leq C_1^2 h^{2k} \|u\|_{k+1,\Omega}^2.
\end{aligned}$$

Durch Ziehen der Wurzel folgt die Behauptung.  $\square$

Mit dem nächsten Lemma wird die Stetigkeit der Bilinearform (4.20) gezeigt:

**Lemma 4.9** Für  $v := u - \Pi_h u$ ,  $u \in \mathbf{H}^{k+1}(\Omega)$ ,  $\mu \in \mathbf{L}^2(\Gamma)$  und eine positive Konstante  $C$  gilt

$$|\mathcal{B}_h(v, \mu; z_h, \eta_h)| \leq C \left( \|z_h\|_{1,h} + \|\eta_h\|_{-\frac{1}{2},h} \right) \left( C_1 h^k \|u\|_{k+1,\Omega} + \|\mu\|_{-\frac{1}{2},h} \right) \quad \text{für alle } (z_h, \eta_h) \in \mathbf{X}_h \times \mathbf{M}_h. \quad (4.33)$$

**Beweis :**

$$\begin{aligned} |\mathcal{B}_h(v, \mu; z_h, \eta_h)| &\leq C \left( \|v\|_{1,h} + \|\mu\|_{-\frac{1}{2},h} \right) \left( \|z_h\|_{1,h} + \|\eta_h\|_{-\frac{1}{2},h} \right) \\ &\quad - \alpha \sum_{E \in \mathcal{E}_h} h_E \left\langle \mu + \frac{\partial v}{\partial n}, \eta_h + \frac{\partial z_h}{\partial n} \right\rangle_E \\ &\leq C \left( \underbrace{\|v\|_{1,h}}_{\stackrel{(4.30)}{\leq} C_1 h^k \|u\|_{k+1,\Omega}} + \|\mu\|_{-\frac{1}{2},h} \right) \left( \|z_h\|_{1,h} + \|\eta_h\|_{-\frac{1}{2},h} \right) \\ &\quad + \alpha \left( \|\mu\|_{-\frac{1}{2},h} + \underbrace{\left\| \frac{\partial v}{\partial n} \right\|_{-\frac{1}{2},h}}_{\stackrel{(4.32)}{\leq} C_1 h^k \|u\|_{k+1,\Omega}} \right) \left( \|\eta_h\|_{-\frac{1}{2},h} + \underbrace{\left\| \frac{\partial z_h}{\partial n} \right\|_{-\frac{1}{2},h}}_{\stackrel{(4.14)}{\leq} C_I \|z_h\|_{1,h}} \right) \\ &\leq C \left( \|z_h\|_{1,h} + \|\eta_h\|_{-\frac{1}{2},h} \right) \left( C_1 h^k \|u\|_{k+1,\Omega} + \|\mu\|_{-\frac{1}{2},h} \right). \quad \square \end{aligned}$$

Der nächste Satz liefert schließlich eine Fehlerabschätzung des Variationsproblems (4.19):

**Satz 4.10** Sei  $(u_h, \lambda_h) \in \mathbf{X}_h \times \mathbf{M}_h$  eine Lösung des Problems (4.19) und sei  $0 < \alpha < C_I^2$ . Mit  $u \in \mathbf{H}^{k+1}(\Omega)$ ,  $k \geq 1$  und  $\lambda \in \mathbf{H}^{l+1}(\Gamma)$ ,  $l \geq 0$  gilt dann

$$\|u - u_h\|_{1,h} + \|\lambda - \lambda_h\|_{-\frac{1}{2},h} \leq C \left( h^k \|u\|_{k+1,\Omega} + h^{l+\frac{3}{2}} \|\lambda\|_{l+1,\Gamma} \right). \quad (4.34)$$

**Beweis :** Sei  $\mu_h \in \mathbf{M}_h$  beliebig. Dann folgt

$$\begin{aligned} \|u - u_h\|_{1,h} + \|\lambda - \lambda_h\|_{-\frac{1}{2},h} &\stackrel{\Delta\text{-Ungl.}}{\leq} \|u - \Pi_h u\|_{1,h} + \|\Pi_h u - u_h\|_{1,h} \\ &\quad + \|\lambda - \mu_h\|_{-\frac{1}{2},h} + \|\mu_h - \lambda_h\|_{-\frac{1}{2},h} \\ &\stackrel{(4.30)}{\leq} C_K h^k \|u\|_{k+1,\Omega} + \|\lambda - \mu_h\|_{-\frac{1}{2},h} \\ &\quad + \|\Pi_h u - u_h\|_{1,h} + \|\mu_h - \lambda_h\|_{-\frac{1}{2},h} \\ &\stackrel{(4.22)}{\leq} C_K h^k \|u\|_{k+1,\Omega} + \|\lambda - \mu_h\|_{-\frac{1}{2},h} \\ &\quad + \sup_{(z_h, \eta_h) \in \mathbf{X}_h \times \mathbf{M}_h} \frac{\mathcal{B}_h(\Pi_h u - u_h, \mu_h - \lambda_h; z_h, \eta_h)}{\|z_h\|_{1,h} + \|\eta_h\|_{-\frac{1}{2},h}} \\ &\stackrel{(1.41)}{\leq} C_K h^k \|u\|_{k+1,\Omega} + \|\lambda - \mu_h\|_{-\frac{1}{2},h} \end{aligned}$$

$$\begin{aligned}
 & + \sup_{(z_h, \eta_h)} \frac{\mathcal{B}_h(\Pi_h u - u, \mu_h - \lambda; z_h, \eta_h)}{\|z_h\|_{1,h} + \|\eta_h\|_{-\frac{1}{2},h}} \\
 \stackrel{(4.33)}{\leq} & C_K h^k \|u\|_{k+1,\Omega} + \|\lambda - \mu_h\|_{-\frac{1}{2},h} \\
 & + \left( C_1 h^k \|u\|_{k+1,\Omega} + \|\mu\|_{-\frac{1}{2},h} \right) \\
 & \sup_{(z_h, \eta_h)} \frac{C \left( \|z_h\|_{1,h} + \|\eta_h\|_{-\frac{1}{2},h} \right)}{\|z_h\|_{1,h} + \|\eta_h\|_{-\frac{1}{2},h}} \\
 \stackrel{\mu := \lambda - \mu_h}{\leq} & C \left( h^k \|u\|_{k+1,\Omega} + \|\lambda - \mu_h\|_{-\frac{1}{2},h} \right).
 \end{aligned}$$

Jetzt schätzt man  $\|\lambda - \mu_h\|_{-\frac{1}{2},h}$  mit Lemma 4.1 und geeigneter Wahl von  $\mu_h$  nach oben ab und erhält die Behauptung.  $\square$

Die Finite-Elemente-Räume können nach Erreichen dieses Resultats unabhängig gewählt werden.

Die Abbildung 4.2 zeigt das approximierte  $\lambda_h$  und dessen Fehlerdiagramm für eine stabilisierte Methode mit verletzter ‚inf-sup-Bedingung‘ (4.15). Verletzt ist die Bedingung, wenn die Zerlegung für  $\lambda_h$  auf dem Rand feiner ist als für das Gebiet  $\Omega$ . Die erreichte Fehlerordnung für  $h_E$  von 1/100 bis 1/250 liegt bei der Stabilisierung von Rapin wie im ‚inf-sup-stabilen‘ Fall bei  $\mathcal{O}(h)$ . Es ist nicht die gleiche Stabilisierung wie von Barbosa und Hughes. [Rap03] beinhaltet außerdem die Erweiterung des Poisson-Problems auf Diffusions-Konvektions-Reaktions-Gleichungen.

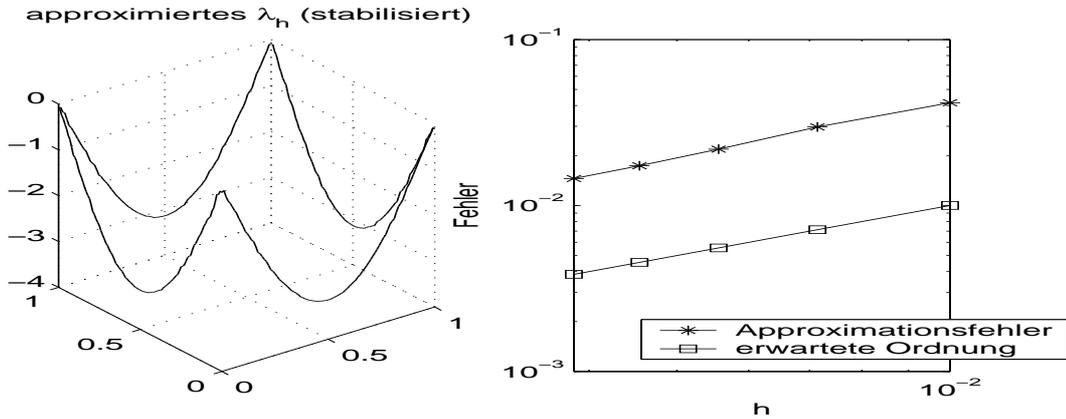


Abbildung 4.2:  $\lambda_h$  und Fehlerdiagramm für die stabilisierte Methode

Betrachtet man den Raum  $\mathbf{M}_h$  noch etwas genauer, so erkennt man, da die Funktionen in diesem Raum unstetig sind, dass man auf jedem Randelement  $E \in \mathcal{E}_h$  die Variable  $\lambda_h$  lokal eliminieren kann. Wir setzen eine Testfunktion  $\mu \in \mathbf{M}_h$  in das Variationsproblem (4.19) ein und

erhalten

$$\begin{aligned} \mathcal{B}_h(u_h, \lambda_h; v, \mu) &= (\nabla u_h, \nabla v) + \langle \lambda_h, v \rangle + \langle \mu, u_h \rangle - \alpha \sum_{E \in \mathcal{E}_h} h_E \left\langle \lambda_h + \frac{\partial u_h}{\partial n}, \mu + \frac{\partial v}{\partial n} \right\rangle_E \\ &= (f, v) + \langle g, \mu \rangle. \end{aligned}$$

Diese Formulierung kann man jetzt in zwei Gleichungen schreiben (**Sattelpunktform**):

$$(\nabla u_h, \nabla v) + \langle \lambda_h, v \rangle - \alpha \sum_{E \in \mathcal{E}_h} h_E \left\langle \lambda_h + \frac{\partial u_h}{\partial n}, \frac{\partial v}{\partial n} \right\rangle_E = (f, v), \quad (4.35)$$

$$\langle \mu, u_h \rangle - \alpha \sum_{E \in \mathcal{E}_h} h_E \left\langle \lambda_h + \frac{\partial u_h}{\partial n}, \mu \right\rangle_E = \langle g, \mu \rangle. \quad (4.36)$$

Man löst die Gleichung (4.36) mit Hilfe der  $\mathbf{L}^2$ -Projektion  $\Pi_h$  nach  $\lambda_h$  auf, indem man zuerst die Terme mit  $\lambda_h$  isoliert

$$\langle \mu, u_h \rangle - \langle \mu, g \rangle - \alpha \sum_{E \in \mathcal{E}_h} h_E \left\langle \frac{\partial u_h}{\partial n}, \mu \right\rangle_E = \alpha \sum_{E \in \mathcal{E}_h} h_E \left\langle \lambda_h, \mu \right\rangle_E$$

und dann die Projektion  $\Pi_h$  anwendet

$$\langle \mu, \Pi_h u_h - \Pi_h g \rangle - \alpha \sum_{E \in \mathcal{E}_h} h_E \left\langle \Pi_h \frac{\partial u_h}{\partial n}, \mu \right\rangle_E = \alpha \sum_{E \in \mathcal{E}_h} h_E \left\langle \lambda_h, \mu \right\rangle_E.$$

Weiterhin gilt diese Gleichung für jedes  $\mu \in \mathbf{M}_h$ , also gilt auch

$$(\Pi_h u_h - \Pi_h g)|_E - \alpha h_E \left( \Pi_h \frac{\partial u_h}{\partial n} \right) \Big|_E = \alpha h_E \lambda_h|_E,$$

wenn man das Ganze auf  $E \in \mathcal{E}_h$  einschränkt. Schließlich teilt man durch  $(\alpha h_E)$  und erhält einen Ausdruck für  $\lambda_h$  auf  $E \in \mathcal{E}_h$

$$\lambda_h|_E = (\alpha h_E)^{-1} (\Pi_h u_h - \Pi_h g)|_E - \left( \Pi_h \frac{\partial u_h}{\partial n} \right) \Big|_E. \quad (4.37)$$

Dies setzt man in die Gleichung (4.35) ein und benutzt, dass  $\Pi_h$  eine orthogonale Projektion ist. Dann erhält man für alle  $v \in \mathbf{X}_h$

$$\begin{aligned} & (\nabla u, \nabla v) - \left\langle \Pi_h \frac{\partial u_h}{\partial n}, \Pi_h v \right\rangle - \left\langle \Pi_h \frac{\partial v}{\partial n}, \Pi_h u_h \right\rangle \\ & + \sum_{E \in \mathcal{E}_h} (\alpha h_E)^{-1} \langle \Pi_h u_h, \Pi_h v \rangle_E + \alpha \sum_{E \in \mathcal{E}_h} h_E \left\langle (\Pi_h - I) \frac{\partial u_h}{\partial n}, (\Pi_h - I) \frac{\partial v}{\partial n} \right\rangle_E \\ & = (f, v) - \left\langle \Pi_h \frac{\partial v}{\partial n}, \Pi_h g \right\rangle + \sum_{E \in \mathcal{E}_h} (\alpha h_E)^{-1} \langle \Pi_h g, \Pi_h v \rangle_E. \end{aligned} \quad (4.38)$$

### Die klassische Methode von Nitsche

Da der Raum  $\mathbf{M}_h$  mit der Stabilisierung von Barbosa und Hughes beliebig gewählt werden kann, identifizieren wir  $\mathbf{M}_h$  mit  $\mathbf{L}^2(\Gamma)$ , d.h.  $\Pi_h = I$ . Man erhält dann die **klassische Formulierung**: Finde  $u_h \in \mathbf{X}_h$ , so dass

$$\begin{aligned} \mathcal{B}_h(u_h; v) &:= (\nabla u, \nabla v) - \left\langle \frac{\partial u_h}{\partial n}, v \right\rangle - \left\langle \frac{\partial v}{\partial n}, u_h \right\rangle + \gamma \sum_{E \in \mathcal{E}_h} h_E^{-1} \langle u_h, v \rangle_E \\ &= (f, v) - \left\langle \frac{\partial v}{\partial n}, g \right\rangle + \gamma \sum_{E \in \mathcal{E}_h} h_E^{-1} \langle g, v \rangle_E \\ &=: \mathcal{F}_h(v) \end{aligned} \tag{4.39}$$

gilt.

Sie geht auf einen Artikel von Nitsche aus dem Jahr 1971 zurück (vgl. [Nit71]). Für diese Methode bekommt man direkt eine optimale Fehlerabschätzung.

**Satz 4.11** *Sei  $u_h \in \mathbf{X}_h$  die Lösung vom Problem (4.39) und sei  $\gamma > C_I^{-2}$ . Für  $u \in \mathbf{H}^{k+1}(\Omega)$  gilt dann*

$$\|u - u_h\|_{1,h} \leq Ch^k \|u\|_{k+1,\Omega}. \tag{4.40}$$

**Beweis :** Die Konsistenz erkennt man direkt aus der Formulierung, die Stabilität zeigt man mit Cauchy–Schwarzscher und Youngscher Ungleichung sowie (4.14) für alle  $v \in \mathbf{X}_h$

$$\begin{aligned} \mathcal{B}_h(v; v) &\stackrel{(4.9)}{=} \|\nabla v\|_{0,\Omega}^2 - 2 \left\langle v, \frac{\partial v}{\partial n} \right\rangle + \gamma \|v\|_{\frac{1}{2},h}^2 \\ &\stackrel{(1.25)}{\geq} \|\nabla v\|_{0,\Omega}^2 - 2 \|v\|_{\frac{1}{2},h} \left\| \frac{\partial v}{\partial n} \right\|_{-\frac{1}{2},h} + \gamma \|v\|_{\frac{1}{2},h}^2 \\ &\stackrel{quadr.Erg.}{\geq} \|\nabla v\|_{0,\Omega}^2 - \frac{1}{\varepsilon} \left\| \frac{\partial v}{\partial n} \right\|_{-\frac{1}{2},h}^2 + (\gamma - \varepsilon) \|v\|_{\frac{1}{2},h}^2 \\ &\stackrel{(4.14)}{\geq} \left(1 - \frac{1}{\varepsilon C_I^2}\right) \|\nabla v\|_{0,\Omega}^2 + (\gamma - \varepsilon) \|v\|_{\frac{1}{2},h}^2 \\ &\stackrel{(4.12)}{\geq} C \|v\|_{1,h}^2. \end{aligned} \tag{4.41}$$

Mit Konsistenz und Stabilität folgt aus (4.30), der Galerkin–Orthogonalität (1.41) und der Beschränktheit von  $\mathcal{B}_h$  die Behauptung.  $\square$

## 4.3 Gemischte Randwert–Probleme

### Randbedingungen zweiter und dritter Art

In (0.2) wurden homogene Randbedingungen erster Art oder Dirichlet–Randbedingungen schon definiert. Im Programmcode in Kapitel 3 stehen inhomogene Randbedingungen erster Art. Hier werden jetzt noch zwei andere Arten von Randbedingungen vorgestellt:

Die **Randbedingung zweiter Art** oder **Neumann–Randbedingung** lautet

$$\frac{\partial u}{\partial n} = g \text{ auf } \Gamma. \quad (4.42)$$

Die **Randbedingung dritter Art** oder **Robin–Randbedingung** ist eine Kombination aus Randbedingungen erster und zweiter Art. Sie wird formuliert durch

$$\frac{\partial u}{\partial n} + \alpha u = g \text{ auf } \Gamma. \quad (4.43)$$

Werden auf verschiedenen Teilstücken des Randes ( $\Gamma_1, \Gamma_2, \dots$ ) verschiedene Randbedingungen (z.B. Dirichlet–Randbedingungen auf  $\Gamma_1$ , Neumann–Randbedingungen auf  $\Gamma_2$ , usw.) gefordert, spricht man von **gemischten Randbedingungen**. In FEMLAB realisiert man diese gemischten Randbedingungen über vier Randparameter.

```
% Angabe gemischter Randbedingungen:
% Eine Randbedingung mit q=0 ist eine Neumann-Randbedingung, sonst eine Ro-
% bin-Randbedingung, vorausgesetzt h=0 und r=0.
%   nu c nabla u + qu = g - h^t lambda      (NEUMANN/ROBIN-Randbedingung)
%   h u = r                                (DIRICHLET-Randbedingung)
fem2D.bnd.h={0 0 1}
fem2D.bnd.r={0 0 '0.5*(x.^2+y.^2)'}
fem2D.bnd.q={0 0 0}      % das alpha aus der Gleichung 4.44
fem2D.bnd.g={0 1 0}      % das g aus der Gleichung 4.43 bzw. 4.44
% Bei diesem Beispiel sind die drei Randbedingungen also:
% 1) homogene Neumann-Randbedingung
% 2) inhomogene Neumann-Randbedingung
% 3) inhomogene Dirichlet-Randbedingung
fem2D.bnd.ind=[3 2 3 1];
```

## Numerische Beispiele

**Beispiel 4.12** *Finde  $u \in \mathbf{X}$ , so dass*

$$\begin{aligned} -\Delta u(x, y) &= -2 && \text{für alle } (x, y) \in \Omega := [0, 1]^2, \\ \frac{\partial u}{\partial n}(1, y) &= 1 && \text{und} \\ \frac{\partial u}{\partial n}(0, y) &= 0 && \text{für alle } y \in [0, 1] \text{ und} \\ u(x, 0) &= u(x, 1) = \frac{1}{2}(x^2 + y^2) && \text{für alle } x \in [0, 1] \end{aligned}$$

*gilt. Die Lösung des Problems ist  $u(x, y) = \frac{1}{2}(x^2 + y^2)$ . Grafisch sind die Approximation der Lösung und das Fehlerdiagramm mit  $\mathbf{L}^2$ - und  $\mathbf{H}^1$ -Norm für  $P_1$ -Elemente in Abbildung 4.3 dargestellt. Dabei sind  $\mathcal{O}(h^2)$  für die  $\mathbf{L}^2$ -Norm und  $\mathcal{O}(h)$  für die  $\mathbf{H}^1$ -Norm die erreichten Fehlerordnungen.*

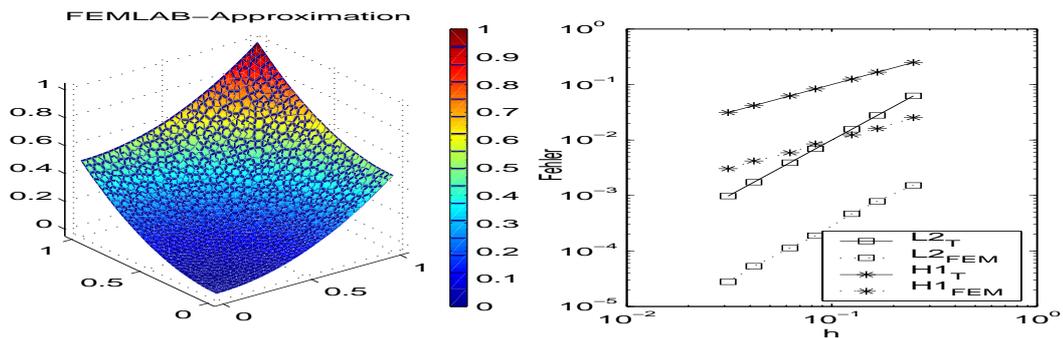


Abbildung 4.3: FEMLAB-Approximation und Fehlerdiagramm zu 4.12

Schließlich sieht man in Abbildung 4.4 die Approximation von FEMLAB parallel zum Programmcode aus Abschnitt 3.2, in dem bereits gemischte Randbedingungen benutzt wurden. Nimmt man den besten iterativen Löser für dieses Problem (der angegebene iterative Löser im Programmcode), sieht man in der dritten Grafik, dass dieser wesentlich mehr Zeit für das Lösen des Gleichungssystems benötigt. In der vierten Grafik ist der bei kleinem  $h$  größere Approximationsfehler dargestellt. Der Unterschied zwischen den beiden Lösungen ist sehr gering, vergrößert sich jedoch mit zunehmender Zahl an Vergleichspunkten. Die Qualität des iterativen Löser nimmt also ab.

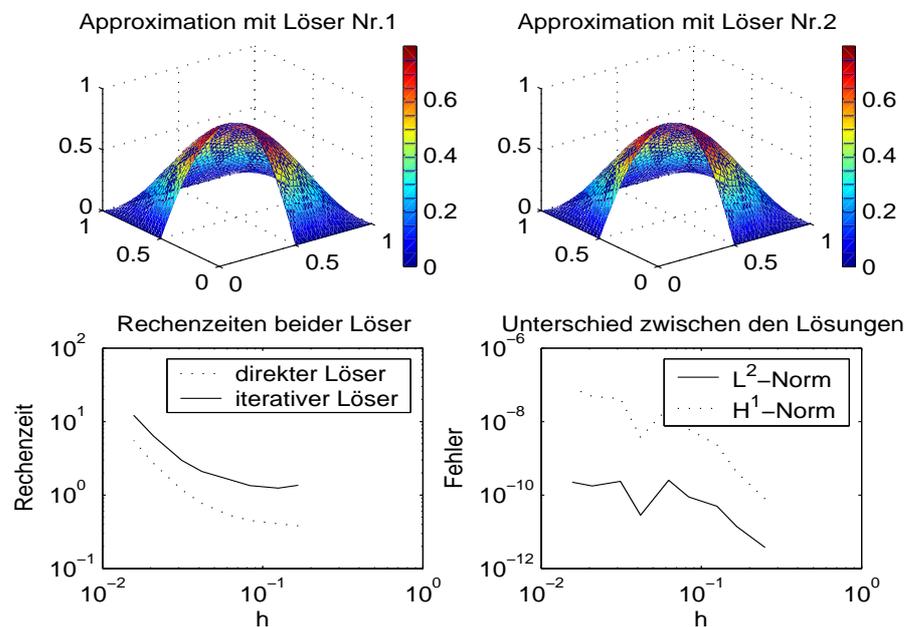


Abbildung 4.4: FEM-Approximation für gemischte Dirichlet-Randbedingungen mit direktem (1) und iterativem (2) Löser, Rechenzeiten der beiden Löser und Unterschied der Lösungen in  $\mathbf{L}^2$ - und  $\mathbf{H}^1$ -Norm



# Kapitel 5

## Isoparametrische Elemente

In diesem Kapitel wird die Voraussetzung eines stückweise polyedrischen Randes  $\Gamma$  des Gebietes  $\Omega$  durch die allgemeinere Vorschrift eines krummlinigen, Lipschitz-stetigen Randes ersetzt. Die Lösung mit Finite-Elemente-Methoden muss dem neuen Rand angepasst werden, da sich die Konvergenzordnung mit vorhergehenden Methoden durch ungenaue Randapproximation verschlechtert. Die Verwendung isoparametrischer Elemente wird einer Verfeinerung der Zerlegung in Randnähe gegenüber gestellt. Die Anwendbarkeit und Effektivität beider Methoden wird mit Hilfe von FEMLAB verglichen.

### 5.1 Theorie

Bei der Anwendung von Finite-Elemente-Methoden haben wir bisher nur stückweise polynomiale Elemente der Ordnung  $k$  auf dem Rand  $\Gamma$  des Gebietes  $\Omega$  benutzt. Besonders für  $k = 1$  ist die Approximation eines nicht polyedrischen Randes schlecht (vgl. Abbildung 5.1). Eine Fehlerabschätzung bestätigt diese Aussage:

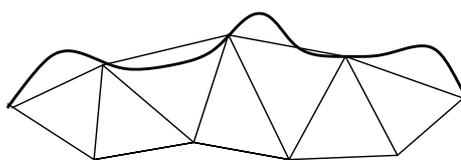


Abbildung 5.1: Krummliniger Rand und eine Zerlegung mit linearen Elementen

**Satz 5.1** *Sei  $\Omega$  ein Gebiet mit  $C^2$ -Rand. Für die Finite-Elemente-Approximation mit linearen, finiten Elementen gilt bei quasiuniformen Triangulierungen*

$$\|u - u_h\|_{1,\Omega} \leq Ch |u|_{2,\Omega} \quad (5.1)$$

und

$$\|u - u_h\|_{0,\Omega} \leq Ch^{\frac{3}{2}} |u|_{2,\Omega}. \quad (5.2)$$

**Beweis :** siehe [Bra97], III.1.7 und III.1.8. □

Man verliert bei der Approximation mit linearen Elementen in der  $\mathbf{L}^2$ -Norm im Vergleich mit dem Interpolationsfehler eine halbe  $h$ -Potenz. Um diesen Verlust auszugleichen, sind isoparametrische Elemente höherer Ordnung am Rand  $\Gamma$  nützlich. Zur Vereinfachung der Darstellung wird hier nur der wichtige Fall quadratischer, isoparametrischer Elemente betrachtet.

Eine Näherung mit höherer Genauigkeit erhält man demnach, wenn man den Gebietsrand stückweise mit Polynomen vom Grad  $l \geq 2$  approximiert. Dann geht der Fehler in der  $\mathbf{L}^2$ -Norm auf  $\mathcal{O}(h^l)$  zurück. Die verlorene halbe  $h$ -Potenz wird also zurück gewonnen. Die betroffenen Dreiecke bzw. Tetraeder

$$K_\Gamma := \{K \in \mathcal{T}_h : K \cap \Gamma \neq \emptyset\}$$

haben dann eine gebogene Kante (vgl. Abbildung 5.2) bzw. Fläche.

Sei  $(\tilde{K}, \mathcal{P}_{\tilde{K}}, \tilde{\Sigma})$  ein finites Element (vgl. Definition 3.4), wobei  $\tilde{K}$  das Referenzdreieck aus Abbil-

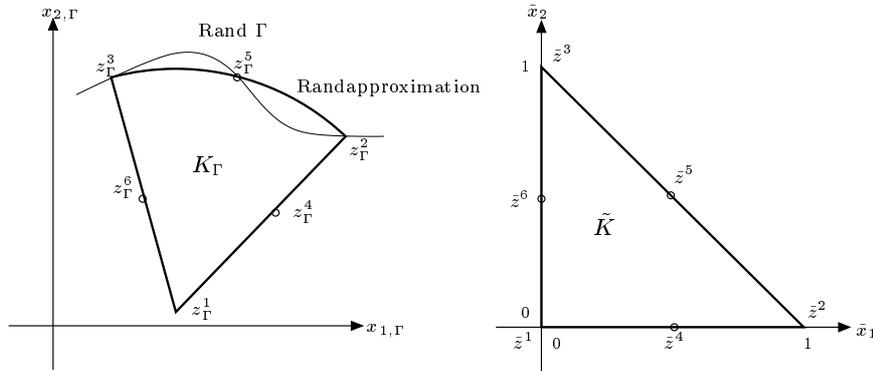


Abbildung 5.2: isoparametrisches Dreieck  $K_\Gamma$  und Referenzdreieck  $\tilde{K}$  mit  $l = 2$

dung 5.2 ist. Außerdem sei  $F$  die affin-lineare Abbildung aus (3.8) von  $\tilde{K}$  nach  $K_\Gamma$ . Man definiert dann

$$\mathcal{P}_K := \{p : p(x) = \tilde{p}(F^{-1}(x)), x \in K_\Gamma, \tilde{p} \in \mathcal{P}_{\tilde{K}}\} \quad (5.3)$$

und

$$\Sigma_K := \{p_i \in K_\Gamma \text{ aus (3.6)} : p_i = F(\tilde{p}_i)\}. \quad (5.4)$$

Damit ist auch  $(K, \mathcal{P}_K, \Sigma_K)$  ein finites Element mit den Funktionen  $p \in \mathcal{P}_K$ , definiert durch  $F^{-1} : K_\Gamma \rightarrow \tilde{K}$  und die polynomialen Funktionen  $\tilde{p} : \tilde{K} \rightarrow \mathbb{R}$ ,  $\tilde{p} \in \mathcal{P}_{\tilde{K}}$ .

**Definition 5.2** Wenn für jede Komponente  $F_i$  der Abbildung  $F$

$$F_i \in \mathcal{P}_{\tilde{K}}, \quad i = 1, \dots, n,$$

gilt, dann heißt  $(K, \mathcal{P}_K, \Sigma_K)$  **isoparametrisch**.

**Bemerkung 5.3** *Im allgemeinen ist die inverse Abbildung  $F^{-1}$  nur ganzrational, wenn  $K_\Gamma$  ein Dreieck ohne gebogene Kante ist und somit  $F$  und  $F^{-1}$  linear sind. Das Gleiche gilt für die Funktionen  $p \in \mathcal{P}_K$ .*

Um eine bessere Vorstellung über die Vorgehensweise bei isoparametrischen Elementen zu haben, folgt ein Beispiel.

**Beispiel 5.4** *Sei  $\tilde{K}$  das Referenzdreieck aus Abbildung 5.2 mit den Eckpunkten  $\tilde{z}^i$ ,  $i = 1, 2, 3$ , und den Seitenmittelpunkten  $\tilde{z}^i$ ,  $i = 4, 5, 6$ . Außerdem sei  $\mathcal{P}_{\tilde{K}} = \mathbb{P}_2(\tilde{K})$ ,  $\Sigma_{\tilde{K}}$  beinhalte die Funktionswerte der Punkte  $\tilde{z}^i$ ,  $i = 1, \dots, 6$ , und  $\tilde{\phi}_i \in \mathbb{P}_2(\tilde{K})$ ,  $i = 1, \dots, 6$ , seien die entsprechenden Basisfunktionen mit  $\tilde{\phi}_i(\tilde{z}^j) = \delta_{ij}$  (Kronecker-Symbol). Sei ein Dreieck  $K$  mit einer gebogenen und zwei geraden Kanten wie in Abbildung 5.2 gegeben, das in der  $(x_1, x_2)$ -Ebene liege, wobei  $z^4$  und  $z^6$  die Mittelpunkte der geraden Seiten  $\overline{z^1 z^2}$  und  $\overline{z^1 z^3}$  seien. Dann definiert man eine Transformationsabbildung  $F$  durch*

$$F(\tilde{x}) = \sum_{j=1}^6 z^j \tilde{\phi}_j(\tilde{x}) \quad \text{für alle } \tilde{x} \in \tilde{K} \quad (5.5)$$

mit  $K = F(\tilde{K}) = \{x \in \mathbb{R}^2 : x = F(\tilde{x}), \tilde{x} \in \tilde{K}\}$ . Also ist  $F$  eine Abbildung vom Referenzdreieck  $\tilde{K}$  auf das Dreieck  $K$  mit  $F(\tilde{z}^i) = z^i$ ,  $i = 1, \dots, 6$ .

Welche Eigenschaften hat die Transformationsabbildung im Beispiel 5.4, d.h. unter welchen Voraussetzungen gibt es eine inverse Abbildung  $F^{-1}$ ?

Die im Beispiel 5.4 vorliegende Abbildung  $F$  ist nur unter bestimmten Voraussetzungen bijektiv (und besitzt somit eine Inverse  $F^{-1}$ ). In einer kleinen Umgebung von jedem Punkt  $\tilde{x} \in \tilde{K}$  ist  $F$  lokal bijektiv, wenn

$$\det J(\tilde{x}) = \begin{vmatrix} \frac{\partial F_1}{\partial \tilde{x}_1} & \frac{\partial F_1}{\partial \tilde{x}_2} \\ \frac{\partial F_2}{\partial \tilde{x}_1} & \frac{\partial F_2}{\partial \tilde{x}_2} \end{vmatrix} \neq 0 \quad \text{für alle } \tilde{x} \in \tilde{K} \quad (5.6)$$

gilt, wobei  $J(\tilde{x})$  die Jakobi-Matrix von  $F$  an der Stelle  $\tilde{x}$  ist. Im allgemeinen reicht (5.6) nicht aus für eine globale Bijektivität von  $F$ . In diesem Beispiel werden die Seiten von  $\tilde{K}$  jedoch bijektiv auf die Seiten von  $K$  abgebildet. Damit kann man folgenden Satz zeigen.

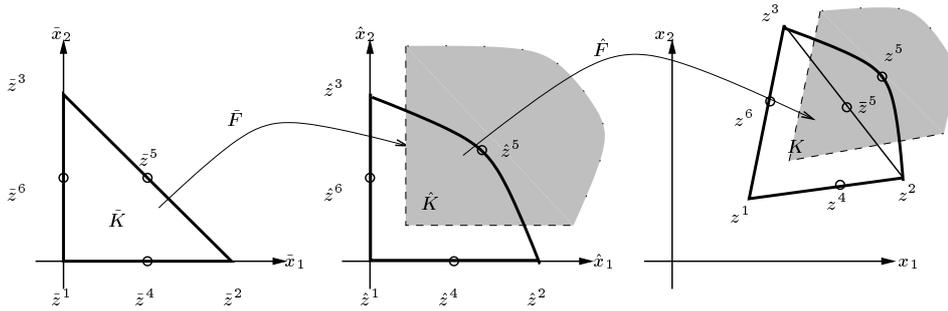
**Satz 5.5** *Erfüllt die Abbildung  $F$  aus Beispiel 5.4 die Bedingung (5.6) und werden die Seiten des Referenzdreiecks  $\tilde{K}$  bijektiv auf die Seiten eines beliebigen Dreiecks  $K$  einer Zerlegung  $\mathcal{T}_h$  abgebildet, dann ist  $F$  sogar global bijektiv.*

**Beweis :** Man splittet die Transformationsabbildung  $F$  in zwei Abbildungen  $\hat{F}$  und  $\tilde{F}$ , so dass  $F(\tilde{x}) = \hat{F}(\tilde{F}(\tilde{x}))$  gilt (vgl. Abbildung 5.3).  $\tilde{F}$  ist die affine Abbildung, die die Eckpunkte  $\tilde{z}^j$ ,  $j = 1, 2, 3$ , auf die Punkte  $z^j = (z_1^j, z_2^j)$ ,  $j = 1, 2, 3$ , abbildet.  $\hat{F}$  hat also die Form

$$\hat{F}(y) = By + b \quad \text{für alle } y \in \{\hat{x} \in \mathbb{R}^2 : x = \hat{F}(\hat{x}), x \in K\}$$

mit

$$B = \begin{pmatrix} z_1^2 - z_1^1 & z_1^3 - z_1^1 \\ z_2^2 - z_2^1 & z_2^3 - z_2^1 \end{pmatrix} \quad \text{und} \quad b = \begin{pmatrix} z_1^1 \\ z_2^1 \end{pmatrix}$$

Abbildung 5.3: Die gesplittete Transformationsabbildung  $F = \hat{F} \circ \tilde{F}$ 

Da  $z^1, z^2$  und  $z^3$  nicht auf einer Geraden liegen, ist  $\det B \neq 0$  und damit  $\hat{F}$  bijektiv. Für die Abbildung  $\tilde{F}$  gilt

$$\tilde{F}_i(\tilde{x}) = \tilde{x}_i + d_i \tilde{x}_1 \tilde{x}_2, \quad i = 1, 2$$

mit  $d_i := 4z_i^5 - 2$ . Die entsprechende Jacobi-Matrix  $\tilde{J}$  zu  $\tilde{F}$  ist gegeben durch

$$\tilde{J}(\tilde{x}) = \begin{pmatrix} 1 + d_1 \tilde{x}_2 & d_1 \tilde{x}_1 \\ d_2 \tilde{x}_2 & 1 + d_2 \tilde{x}_1 \end{pmatrix}.$$

Die Determinante von  $\tilde{J}(\tilde{x})$

$$\det \tilde{J}(\tilde{x}) = 1 + d_1 \tilde{x}_2 + d_2 \tilde{x}_1 \tag{5.7}$$

ist also linear in  $\tilde{x}$ . Wenn

$$\tilde{z}_i^5 > \frac{1}{4}, \quad i = 1, 2, \tag{5.8}$$

ist (das gilt im grau unterlegten Bereich im mittleren Koordinatensystem von Abbildung 5.3), so ist  $d_i > -1$  für  $i = 1, 2$  und mit (5.7) gilt

$$\begin{aligned} \det \tilde{J}(\tilde{z}^1) &= \det \tilde{J}(0, 0) = 1 > 0, \\ \det \tilde{J}(\tilde{z}^2) &= \det \tilde{J}(1, 0) = 1 + d_2 > 0 \quad \text{und} \\ \det \tilde{J}(\tilde{z}^3) &= \det \tilde{J}(0, 1) = 1 + d_1 > 0. \end{aligned}$$

Also ist  $\det \tilde{J} > 0$  im Dreieck  $\tilde{K}$  und also  $\tilde{F}$  bijektiv, wenn  $\tilde{z}^5$  und  $z^5$  im grau unterlegten Bereich der beiden rechten Koordinatensysteme von Abbildung 5.3 liegen (Bedingung (5.8)). Unter dieser Voraussetzung ist dann auch  $F$  bijektiv.  $\square$

**Bemerkung 5.6** Die Entfernung zwischen  $z^5$  und  $\tilde{z}^5$  auf der Strecke  $\overline{z^2 z^3}$  (vgl. Abbildung 5.3) muss also klein sein, nämlich  $|z^5 - \tilde{z}^5| = \mathcal{O}(h_K^2)$  mit  $h_K$  als Durchmesser von  $K$ . Für ausreichend feine Zerlegungen ist diese Bedingung erfüllt und  $F$  bijektiv.

Die folgende, durch quadratische, isoparametrische Elemente verbesserte Fehlerabschätzung gilt für quadratische Elemente auf dem gesamten Gebiet  $\Omega$  mit isoparametrischen Elementen auf dem Rand  $\Gamma$ .

**Satz 5.7** Sei  $\mathcal{T}_h$  eine quasiuniforme Triangulierung von  $\Omega$  mit quadratischen, isoparametrischen Elementen. Dann existiert eine von  $h$  aus (3.1) unabhängige Konstante  $C$ , so dass

$$\|u - \Pi_h u\|_{m,\Omega} \leq Ch^{3-m} \|u\|_{3,\Omega} \quad \text{für } m = 0, 1 \text{ und alle } u \in \mathbf{H}^3(\Omega)$$

gilt.

**Beweis :** siehe [Cia91], Kapitel VI, Theorem 42.1. □

## 5.2 Numerische Experimente

```
fem.sshape=2;                % Randapproximation mit P2-Elementen
% Achtung ! Ist der Befehl 'fem.shape' für Elemente höherer Ordnung im In-
% neren des Gebiets definiert, wird die Ordnung der Randelemente automa-
% tisch auf den höchsten Wert in 'fem.shape' gesetzt.
```

Isoparametrische Elemente am Gebietsrand  $\Gamma$  sind in FEMLAB also einfach realisierbar (s.o.). Es ist außerdem eine Routine vorhanden, die auch für  $P_1$ -Elemente bei krummlinigen Rändern eine gute Approximation ohne isoparametrische Elemente liefert. Sie ist mit dem Parameter ‚hcurve‘ des ‚meshinit()‘-Befehls gekoppelt. Diese Routine analysiert den Radius  $r_\Gamma$  der Krümmung des Randes. ‚hcurve‘ liefert einen zusätzlichen Parameter zur Berechnung des maximalen Randelement-Durchmessers  $h_{max,\Gamma}$ , der bei zweidimensionalen Geometrien auf 0.3 eingestellt ist. Daraus ergibt sich  $h_{max,\Gamma} = r_\Gamma * 0.3$ . Den Effekt dieses Vorgangs sieht man in Abbildung 5.4. Mit

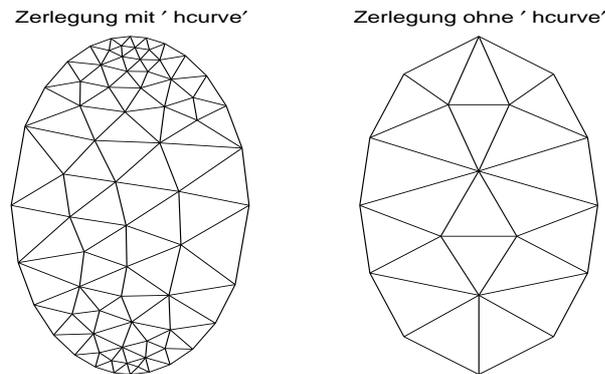


Abbildung 5.4: Effekt des Parameters ‚hcurve‘ bei krummen Rändern

folgendem Programmcode ist Abbildung 5.4 unter Vorgabe des Randkrümmungsradius  $r_\Gamma = \frac{1}{4}$  und dem daraus resultierenden Parameter  $h_{max,\Gamma} = \frac{3}{40}$  erstellt worden.

```
fem.geom=ellip2(0,0,1/4,1/2,0,'center');
% Zerlegung mit Beeinflussung von 'hcurve' (erfolgt automatisch):
```

```
fem.mesh=meshinit(fem,'hmax',1/4);

% Zerlegung ohne Beeinflussung von 'hcurve' durch Angabe eines wirkungslo-
% sen, hohen Wertes beim Parameter 'hcurve', der hier sonst auf 3/40 wäre:
fem.mesh=meshinit(fem,'hmax',1/4,'hcurve',4);
```

FEMLAB verfeinert demnach auf Kosten einer quasi-uniformen Zerlegung an gekrümmten Rändern.

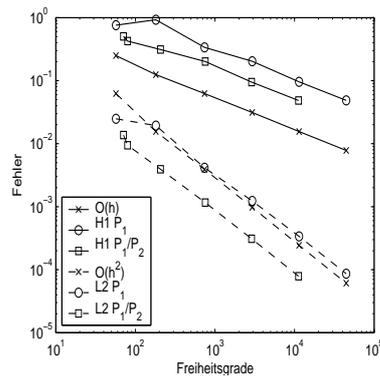


Abbildung 5.5: Fehlerdiagramm zu Beispiel 5.8

**Beispiel 5.8** Wir betrachten ein Problem auf der Ellipse  $\Omega := \{(x, y) \in \mathbb{R}^2 \mid 16x^2 + 4y^2 \leq 1\}$ :  
Finde  $u \in \mathbf{X}$ , so dass

$$\begin{aligned}
 -\Delta u(x, y) &= e^{x+y}[(32\pi^2 - 2) \sin(4\pi x) \sin(4\pi y) \\
 &\quad - 8\pi(\sin(4\pi x) \cos(4\pi y) + \cos(4\pi x) \sin(4\pi y))] && \text{in } \Omega \\
 u(x, y) &= \sin(4\pi x) \sin(4\pi y)e^{x+y} && \text{auf } \Gamma
 \end{aligned}$$

gilt.

Die Lösung des Problems ist  $u(x, y) = \sin(4\pi x) \sin(4\pi y)e^{x+y}$ . Wir rechnen mit FEMLAB mit den in Abbildung 5.4 gezeigten Zerlegungen (links mit  $P_1$ -Elementen und dem Parameter ‚hcurve‘, rechts mit  $P_1$ -Elementen im Gebiet  $\Omega$  und  $P_2$ -Elementen auf dem Rand  $\Gamma$ ) und vergleichen die Ergebnisse. Man stellt fest, dass die Randkrümmung der Ellipse durch die Verfeinerung in FEMLAB genauso gut approximiert wird wie mit isoparametrischen Randelementen. In Abbildung 5.5 sieht man die entsprechenden Fehlerkurven dazu.  $O(h)$  bzw.  $O(h^2)$  stellen die nach Interpolations-Fehlerabschätzungen zu erwartenden Fehler für  $\mathbf{H}^1$ - bzw.  $\mathbf{L}^2$ -Norm dar. ‚ $H^1 P_1$ ‘ ist der Fehler bei Benutzung des Parameters ‚hcurve‘, entsprechend ‚ $L^2 P_1$ ‘. ‚ $H^1 P_1/P_2$ ‘ ist der Fehler für isoparametrische Randelemente ohne Parameter ‚hcurve‘, entsprechend ‚ $L^2 P_1/P_2$ ‘. Der Unterschied ist:

- Mit Parameter ‚hcurve‘ erhält man eine lokal verfeinerte, isotrope Zerlegung, rechnet aber schneller bis zu  $h = \frac{1}{128}$ .

- Mit isoparametrischen Randelementen bleibt man quasi-uniform, rechnet länger, erhält aber ein Ergebnis mit einer kleineren, multiplikativen Konstante.

Beide erreichen die optimale Fehlerordnung bzgl. ihrer Interpolations-Fehlerabschätzung.

Als nächstes betrachten wir ein Beispiel mit sowohl nach außen als auch nach innen gekrümmten Rändern. Ziel ist es, die Aussage des Satzes 5.7 praktisch nachzuprüfen.

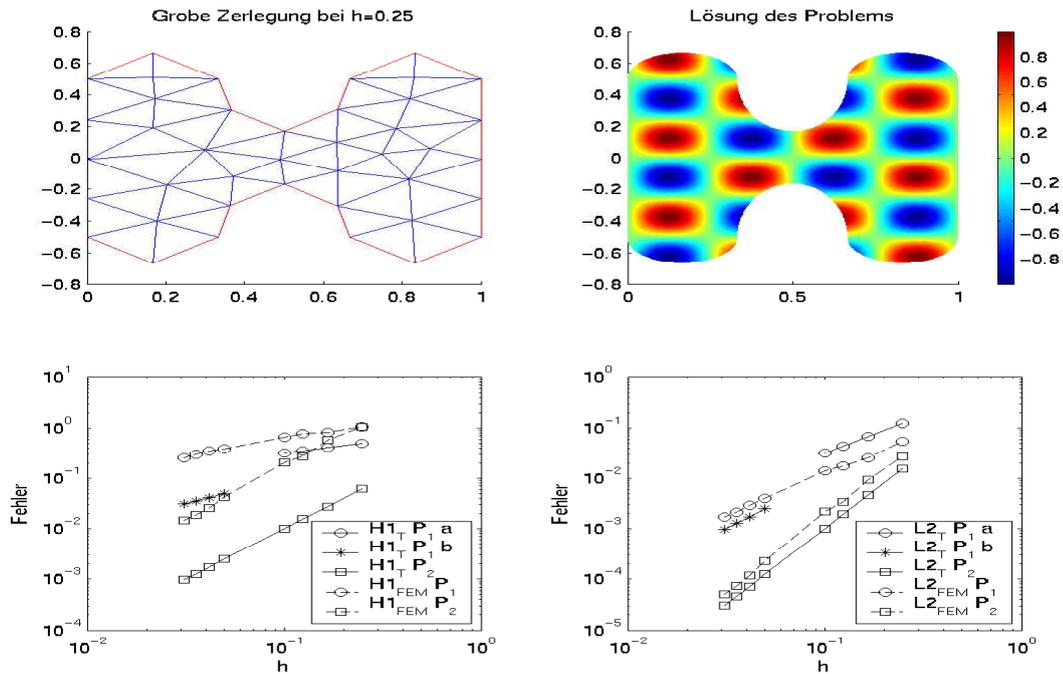


Abbildung 5.6: Zerlegung, Lösung und Fehler in  $\mathbf{H}^1$ - und  $\mathbf{L}^2$ -Norm zu Beispiel 5.9

**Beispiel 5.9** Ein Gebiet mit starken Randkrümmungen ist das  $H$ -Gebiet. Es hat die Form eines großen, gedruckten  $\mathbf{H}$  mit abgerundeten Ecken (vgl. Abbildung 5.6). Auf diesem Gebiet betrachten wir das Problem:

Finde  $u \in \mathbf{X}$ , so dass

$$\begin{aligned} -\Delta u(x, y) &= 32\pi^2 \sin(4\pi x) \sin(4\pi y) && \text{in } \Omega \\ u(x, y) &= \sin(4\pi x) \sin(4\pi y) && \text{auf } \Gamma \end{aligned}$$

gilt.

Die Lösung des Problems ist  $u(x, y) = \sin(4\pi x) \sin(4\pi y)$ .

Man sieht in Abbildung 5.6 eine grobe Zerlegung des Gebietes  $\Omega$ , die Lösung und zwei Fehlerdiagramme. Der maximale Elementdurchmesser  $h$  fällt von  $\frac{1}{4}$  auf  $\frac{1}{25}$ . Die erwartete ( $\mathbf{H}^1_T$  bzw.  $\mathbf{L}^2_T$ ) und die erreichte ( $\mathbf{H}^1_{FEM}$  bzw.  $\mathbf{L}^2_{FEM}$ ) Fehlerordnung  $\mathcal{O}(h^j)$  sind darüber geplottet ( $j = 1$  für  $P_1$ -Elemente bzw.  $j = 2$  für isoparametrische  $P_2$ -Elemente in der  $\mathbf{H}^1$ -Norm links und  $j = \frac{3}{2}$  bzw.  $j = 3$  in der  $\mathbf{L}^2$ -Norm rechts).

Die Besonderheit ist der Bruch in den Fehlerdiagrammen sowohl bei der  $\mathbf{H}^1$ - als auch der  $\mathbf{L}^2$ -Norm für  $P_1$ -Elemente (bezeichnet in der Legende mit  $P_{1a}$  und  $P_{1b}$ ). Hier fällt der erwartete Fehler von  $\mathcal{O}(h^{\frac{1}{2}})$  auf  $\mathcal{O}(h)$  bzw. von  $\mathcal{O}(h^{\frac{3}{2}})$  auf  $\mathcal{O}(h^2)$ , weil der oben bewiesene Effekt eintritt. Der Abstand zwischen dem Punkt auf dem Rand des Gebietes und dem imaginären Punkt auf der Kantenmitte des linearen Elementes wird so klein, dass die vernachlässigte Krümmung im Gesamtfehler keine Rolle mehr spielt. Die theoretischen Fehlerordnungen bzgl.  $h$  aus Satz 5.7 werden tatsächlich erreicht.

**Beispiel 5.10** Zum Abschluss folgt noch ein Problem mit gemischten Randformen. Finde  $u \in \mathbf{X}$ , so dass

$$\begin{aligned} -\Delta u(x, y) &= e^{xy}[(x^2 + y^2 - 2\pi^2) \sin(\pi x) \cos(\pi y) \\ &\quad + 2\pi y \cos(\pi x) \cos(\pi y) - 2\pi x \sin(\pi x) \sin(\pi y)] \quad \text{in } \Omega, \\ u(0, y) &= u(1, y) = 0 \quad \text{für alle } y \in [0, 1], \\ u(x, 0) &= \sin(\pi x) \quad \text{und} \\ u(x, 1) &= -\sin(\pi x)e^x \quad \text{für alle } x \in [0, 1] \text{ und} \\ u(x, y) &= \sin(\pi x) \cos(\pi y)e^{xy} \quad \text{auf } \Gamma_i \end{aligned}$$

mit  $\Omega := \{(x, y) \in [0, 1]^2 \mid x^2 + y^2 \geq \frac{1}{4}\}$  und  $\Gamma_i := \{(x, y) \in [0, 1]^2 \mid x^2 + y^2 = \frac{1}{4}\}$  gilt.

Die Lösung dieses Problems ist  $u(x, y) = \sin(\pi x) \cos(\pi y)e^{xy}$ . Das Fehlerdiagramm rechts in Abbildung 5.7 zeigt den zu erwartenden Interpolationsfehler ( $H_{1T}$ ,  $L_{2T}$ ) und den FEM-Approximationsfehler ( $H_{1FEM}$ ,  $L_{2FEM}$ ) in  $\mathbf{H}^1$ - und  $\mathbf{L}^2$ -Norm bei einer Rechnung mit isoparametrischen Elementen auf dem Rand  $\Gamma$  und  $P_2$ -Elementen im Gebiet  $\Omega$ .

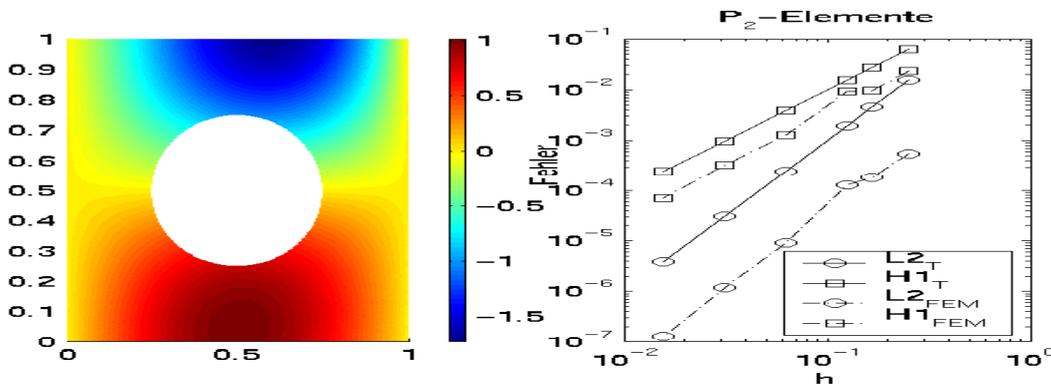


Abbildung 5.7: Lösung und Fehlerdiagramm zu Beispiel 5.10

# Kapitel 6

## Singulär gestörte Probleme

Wie in Kapitel 2 schon gesehen, tritt bei  $\varepsilon$  nahe 0 eine Veränderung des Problems

$$-\varepsilon\Delta u + \vec{b}\nabla u + cu = f \quad \text{in } \Omega$$

auf. Die lineare, elliptische Differentialgleichung zweiter Ordnung reduziert sich zu einer linearen, hyperbolischen Differentialgleichung erster Ordnung. Deshalb wird in diesem Kapitel allein der wichtige grenznahe Fall  $0 < \varepsilon \ll 1$  behandelt.

### 6.1 Stabilisierung in Strömungsrichtung („Streamline Diffusion“)

Man erhält für  $0 < \varepsilon \ll 1$  eine konvektionsdominante Diffusions–Konvektions–Reaktions–Gleichung, deren Lösung mit einer konformen Finite–Elemente–Methode wie in Kapitel 3 im Bereich der durch den starken Konvektionsterm neu auftretenden Grenzschichten oszilliert. Deshalb stabilisiert man diese Methode durch Addition zusätzlicher Terme, die mit einem Parameter  $\delta$  gewichtet sind.

Die neue Formulierung enthält den Term  $\sum_{K \in \mathcal{T}_h} \delta_K (\vec{b}\nabla u_h, \vec{b}\nabla v)$ . Das ist die mit  $\sqrt{\delta_K}$  gewichtete  $\mathbf{L}^2$ –Norm der Richtungsableitung von  $u_h$  bzw.  $v$  in Richtung des Vektorfeldes  $\vec{b}$ . Die Methode heißt „Streamline Diffusion“, weil durch den obigen Term künstliche Diffusion in Richtung des sogenannten Strömungsfeldes  $\vec{b}$  addiert wird. Die stabilisierte Fassung der elliptischen Variationsgleichung (1.39) heißt also:

Finde  $u_h \in \mathbf{X}_h$ , so dass

$$a_{Stab}(u_h, v) = f_{Stab}(v) \quad \text{für alle } v \in \mathbf{X}_h \quad (6.1)$$

gilt mit

$$\begin{aligned} a_{Stab}(u_h, v) &:= a(u_h, v) + \sum_{K \in \mathcal{T}_h} \delta_K (-\varepsilon\Delta u_h + \vec{b}\nabla u_h + cu_h, \vec{b}\nabla v)_K \\ &= \varepsilon(\nabla u_h, \nabla v) + (\vec{b}\nabla u_h + cu_h, v) \\ &\quad + \sum_{K \in \mathcal{T}_h} (-\varepsilon\Delta u_h + \vec{b}\nabla u_h + cu_h, \delta_K \vec{b}\nabla v)_K \end{aligned} \quad (6.2)$$

und

$$\begin{aligned} f_{Stab}(v) &:= (f, v) + \sum_{K \in \mathcal{T}_h} \delta_K (f, \vec{b} \nabla v)_K \\ &= \sum_{K \in \mathcal{T}_h} (f, v + \delta_K \vec{b} \nabla v)_K. \end{aligned} \quad (6.3)$$

Dabei ist  $K$  ein Dreieck der Zerlegung  $\mathcal{T}_h$  mit  $h$  aus (3.1). Die Stabilisierung ist konsistent, weil die Galerkin-Orthogonalität (1.41) durch die Addition von Stabilisierungstermen auf beiden Seiten der Variationsgleichung erhalten bleibt.

Im folgenden Lemma 6.1 wird der Effekt der Stabilisierungsmethode gezeigt. Mit dem Resultat des Lemmas 6.1 ist das Lemma 1.21 von Lax–Milgram anwendbar und die stabilisierte Variationsgleichung (6.1) hat genau eine Lösung.

**Lemma 6.1** *Sei der Parameter  $\delta_K$  in Abhängigkeit von den Problemdaten auf jedem Teildreieck so gewählt, dass*

$$\begin{aligned} \delta_K &= 0 & \text{für } \varepsilon \geq \frac{\|c\|_{\mathbf{L}^\infty(K)}^2 h_K^2}{c_{inv} c_0} \\ \delta_K &\leq \frac{c_0}{2 \|c\|_{\mathbf{L}^\infty(K)}^2} & \text{für } 0 < \varepsilon < \frac{\|c\|_{\mathbf{L}^\infty(K)}^2 h_K^2}{c_{inv} c_0} \end{aligned} \quad (6.4)$$

mit  $c_{inv}$  aus (3.3) und  $c_0$  aus (2.7) gilt.

Dann ist  $a_{Stab}(\cdot, \cdot)$   $\mathbf{X}_h$ -elliptisch, d.h. es gilt

$$a_{Stab}(v, v) \geq C_0 \|v\|^2 \quad (6.5)$$

mit  $C_0 = 1 - \frac{1}{\sqrt{2}}$  und

$$\|v\| := \left( \varepsilon |v|_{1,\Omega}^2 + \sum_{K \in \mathcal{T}_h} \delta_K \|\vec{b} \nabla v\|_{0,K}^2 + c_0 \|v\|_{0,\Omega}^2 \right)^{\frac{1}{2}}. \quad (6.6)$$

**Beweis :** Für  $v \in \mathbf{X}_h$  gilt mit den Ungleichungen

$$ab \leq \frac{1}{2} \left( \frac{1}{\sqrt{2}} a^2 + \sqrt{2} b^2 \right) \quad (6.7)$$

$$\varepsilon \delta_K < \frac{h_K^2}{2c_{inv}} \quad (6.8)$$

nach (6.4) und der inversen Ungleichung (3.3)

$$\begin{aligned} \left| \sum_{K \in \mathcal{T}_h} \delta_K (-\varepsilon \Delta v + cv, \vec{b} \nabla v)_K \right| &\leq \left( \sum_{K \in \mathcal{T}_h} \varepsilon^2 \delta_K \|\Delta v\|_{0,K}^2 \right)^{\frac{1}{2}} \left( \sum_{K \in \mathcal{T}_h} \delta_K \|\vec{b} \nabla v\|_{0,K}^2 \right)^{\frac{1}{2}} \\ &\quad + \left( \sum_{K \in \mathcal{T}_h} \varepsilon \delta_K \|cv\|_{0,K}^2 \right)^{\frac{1}{2}} \left( \sum_{K \in \mathcal{T}_h} \delta_K \|\vec{b} \nabla v\|_{0,K}^2 \right)^{\frac{1}{2}} \end{aligned}$$

$$\begin{aligned}
 &\stackrel{(6.7)}{\leq} \frac{1}{2} \left( \frac{2}{\sqrt{2}} \sum_{K \in \mathcal{T}_h} \delta_K \|\vec{b} \nabla v\|_{0,K}^2 + \sqrt{2} \varepsilon^2 \sum_{K \in \mathcal{T}_h} \delta_K \|\Delta v\|_{0,K}^2 \right. \\
 &\quad \left. + \sqrt{2} \sum_{K \in \mathcal{T}_h} \delta_K \|cv\|_{0,K}^2 \right) \\
 &\stackrel{(6.8)}{\leq} \frac{\sqrt{2}}{2} \left( \sum_{K \in \mathcal{T}_h} \delta_K \|\vec{b} \nabla v\|_{0,K}^2 + \frac{\varepsilon}{2c_{inv}} \sum_{K \in \mathcal{T}_h} h_K^2 \|\Delta v\|_{0,K}^2 \right. \\
 &\quad \left. + \sum_{K \in \mathcal{T}_h} \frac{c_0}{2 \|c\|_{\mathbf{L}^\infty(K)}^2} \|c\|_{\mathbf{L}^\infty(K)}^2 \|v\|_{0,K}^2 \right) \\
 &\stackrel{(3.3)}{\leq} \frac{1}{\sqrt{2}} \left( \sum_{K \in \mathcal{T}_h} \delta_K \|\vec{b} \nabla v\|_{0,K}^2 + \varepsilon |v|_{1,\Omega}^2 + c_0 \|v\|_{0,\Omega}^2 \right) \quad (6.9)
 \end{aligned}$$

und damit

$$\begin{aligned}
 a_{Stab}(v, v) &= a(v, v) + \sum_{K \in \mathcal{T}_h} \delta_K (-\varepsilon \Delta v + \vec{b} \nabla v + cv, \vec{b} \nabla v)_K \\
 &\stackrel{(2.8)}{\geq} \varepsilon |\nabla v|_{1,\Omega}^2 + c_0 \|v\|_{0,\Omega}^2 + \sum_{K \in \mathcal{T}_h} \delta_K \left[ (-\varepsilon \Delta v + cv, \vec{b} \nabla v)_K + \|\vec{b} \nabla v\|_{0,K}^2 \right] \\
 &\stackrel{(6.9)}{\geq} C_0 \|v\|^2.
 \end{aligned}$$

□

Bei der Approximation der Lösung mit (6.1) erhält man eine verbesserte Fehlerabschätzung. Das folgende Lemma schätzt den Fehler der stabilisierten Variationsgleichung gegen den Fehler der Approximation auf dem diskreten Raum  $\mathbf{X}_h$  ab. Dabei wird die Wahl des Parameters  $\delta_K$  für jedes Dreieck  $K \in \mathcal{T}_h$  spezifiziert.

**Lemma 6.2** *Seien  $u$  die Lösung der stabilisierten Variationsgleichung (6.1) und  $u_h$  die Lösung der Variationsgleichung (1.39). Dann gilt für  $u \in \mathbf{W}^{k+1,2}(\Omega)$  mit einer Konstanten  $C > 0$*

$$\| \|u - u_h\| \| \leq Ch^{k+\frac{1}{2}} |u|_{k+1,\Omega}.$$

**Beweis :** Im folgenden Beweis wird die Konstante  $C > 0$  in jedem Schritt entsprechend angepasst. Zuerst erhält man mit der Dreiecksungleichung

$$\| \|u - u_h\| \| \leq \| \|u - \Pi_{\mathcal{T}} u\| \| + \| \|\Pi_{\mathcal{T}} u - u_h\| \| . \quad (6.10)$$

Dabei ist  $\Pi_{\mathcal{T}}$  der Interpolationsoperator aus Satz 3.15.

Als erstes betrachtet man den zweiten Summanden von (6.10) genauer.

$$\begin{aligned} C_0 \|\Pi_{\mathcal{T}}u - u_h\|^2 &\leq a_{Stab}(\Pi_{\mathcal{T}}u - u_h, \Pi_{\mathcal{T}}u - u_h) \\ &\stackrel{(1.41)}{=} a_{Stab}(\Pi_{\mathcal{T}}u - u, \Pi_{\mathcal{T}}u - u_h) \\ &= \varepsilon \int_{\Omega} \nabla(\Pi_{\mathcal{T}}u - u) \nabla(\Pi_{\mathcal{T}}u - u_h) dx \end{aligned} \quad (6.11)$$

$$+ \int_{\Omega} [\vec{b} \nabla(\Pi_{\mathcal{T}}u - u) + c(\Pi_{\mathcal{T}}u - u)](\Pi_{\mathcal{T}}u - u_h) dx \quad (6.12)$$

$$+ \sum_{K \in \mathcal{T}_h} \delta_K (-\varepsilon \Delta(\Pi_{\mathcal{T}}u - u) + c(\Pi_{\mathcal{T}}u - u), \Pi_{\mathcal{T}}u - u_h)_K. \quad (6.13)$$

Im folgenden werden die drei Summanden (6.11), (6.12) und (6.13) getrennt abgeschätzt.

$$\begin{aligned} \varepsilon \int_{\Omega} \nabla(\Pi_{\mathcal{T}}u - u) \nabla(\Pi_{\mathcal{T}}u - u_h) dx &\leq \sqrt{\varepsilon} |\Pi_{\mathcal{T}}u - u|_{1,\Omega} \|\Pi_{\mathcal{T}}u - u_h\| \\ &\stackrel{(3.21)}{\leq} C \sqrt{\varepsilon} h^k |u|_{k+1,\Omega} \|\Pi_{\mathcal{T}}u - u_h\|, \end{aligned}$$

$$\begin{aligned} &\int_{\Omega} [\vec{b} \nabla(\Pi_{\mathcal{T}}u - u) + c(\Pi_{\mathcal{T}}u - u)](\Pi_{\mathcal{T}}u - u_h) dx \\ &= \int_{\Omega} (c - \nabla \cdot \vec{b})(\Pi_{\mathcal{T}}u - u)(\Pi_{\mathcal{T}}u - u_h) dx - \int_{\Omega} (\Pi_{\mathcal{T}}u - u) \vec{b} \nabla(\Pi_{\mathcal{T}}u - u_h) dx \\ &\leq \|c - \nabla \cdot \vec{b}\|_{\mathbf{L}^\infty(\Omega)} \|\Pi_{\mathcal{T}}u - u\|_{0,\Omega} \|\Pi_{\mathcal{T}}u - u_h\|_{0,\Omega} \\ &\quad + \|\Pi_{\mathcal{T}}u - u\|_{0,\Omega} \|\vec{b} \nabla(\Pi_{\mathcal{T}}u - u_h)\|_{0,\Omega} \\ &\leq \left[ \left( \sum_{K \in \mathcal{T}_h} \|\Pi_{\mathcal{T}}u - u\|_{0,K}^2 \right)^{\frac{1}{2}} + \left( \sum_{K \in \mathcal{T}_h} \delta_K^{-1} \|\Pi_{\mathcal{T}}u - u\|_{0,K}^2 \right)^{\frac{1}{2}} \right] \|\Pi_{\mathcal{T}}u - u_h\| \\ &\stackrel{(3.19)}{\leq} Ch^k \left( \sum_{K \in \mathcal{T}_h} (1 + \delta_K^{-1}) h_K^2 |u|_{k+1,K}^2 \right)^{\frac{1}{2}} \|\Pi_{\mathcal{T}}u - u_h\| \end{aligned}$$

$$\begin{aligned} &\left| \sum_{K \in \mathcal{T}_h} \delta_K (-\varepsilon \Delta(\Pi_{\mathcal{T}}u - u) + \vec{b} \nabla(\Pi_{\mathcal{T}}u - u) + c(\Pi_{\mathcal{T}}u - u), \vec{b} \nabla(\Pi_{\mathcal{T}}u - u_h))_K \right| \\ &\stackrel{(3.19)}{\leq} C \sum_{K \in \mathcal{T}_h} \sqrt{\delta_K} (\varepsilon h_K^{k-1} + \|\vec{b}\|_{\mathbf{L}^\infty(K)} h_K^k + \|c\|_{\mathbf{L}^\infty(K)} h_K^{k+1}) |u|_{k+1,K} \\ &\quad \cdot \sqrt{\delta_K} \|\vec{b} \nabla(\Pi_{\mathcal{T}}u - u_h)\|_{0,K} \\ &\stackrel{(3.1)}{\leq} Ch^k \left( \sum_{K \in \mathcal{T}_h} \delta_K (\varepsilon^2 h_K^{-2} + \|\vec{b}\|_{\mathbf{L}^\infty(K)}^2 + \|c\|_{\mathbf{L}^\infty(K)}^2 h_K^2) |u|_{k+1,K}^2 \right)^{\frac{1}{2}} \\ &\quad \cdot \left( \sum_{K \in \mathcal{T}_h} \delta_K \|\vec{b} \nabla(\Pi_{\mathcal{T}}u - u_h)\|_{0,K}^2 \right)^{\frac{1}{2}} \\ &\stackrel{(6.8)}{\leq} Ch^k \left( \sum_{K \in \mathcal{T}_h} (\varepsilon + \delta_K \|\vec{b}\|_{\mathbf{L}^\infty(K)}^2) |u|_{k+1,K}^2 \right)^{\frac{1}{2}} \|\Pi_{\mathcal{T}}u - u_h\|. \end{aligned}$$

Zusammengefasst ergibt sich daraus die Abschätzung:

$$\begin{aligned}
 \|\Pi_{\mathcal{T}}u - u_h\| &\leq Ch^k \left[ \sqrt{\varepsilon} |u|_{k+1, \Omega} \right. \\
 &\quad \left. + \left( \sum_{K \in \mathcal{T}_h} (1 + \delta_K^{-1}) h_K^2 |u|_{k+1, K}^2 \right)^{\frac{1}{2}} + \left( \sum_{K \in \mathcal{T}_h} (\varepsilon + \delta_K) |u|_{k+1, K}^2 \right)^{\frac{1}{2}} \right] \\
 &\stackrel{(3.22)}{\leq} Ch^k \left( \sum_{K \in \mathcal{T}_h} \left( \varepsilon + h_K^2 + \frac{h_K^2}{\delta_K} + \delta_K \|\vec{b}\|_{\mathbf{L}^\infty(K)}^2 \right) |u|_{k+1, K}^2 \right)^{\frac{1}{2}}.
 \end{aligned}$$

Anschließend berechnet man eine Abschätzung für den ersten Summanden aus (6.10).

$$\begin{aligned}
 \|u - \Pi_{\mathcal{T}}u\|^2 &= \varepsilon |u - \Pi_{\mathcal{T}}u|_{1, \Omega}^2 + \sum_{K \in \mathcal{T}_h} \delta_K \|\vec{b} \nabla (u - \Pi_{\mathcal{T}}u)\|_{0, K}^2 + c_0 \|u - \Pi_{\mathcal{T}}u\|_{0, \Omega}^2 \\
 &\stackrel{(3.19)}{\leq} C^2 \sum_{K \in \mathcal{T}_h} \left( \varepsilon h_K^{2k} + c_0 h_K^{2(k+1)} + \delta_K \|\vec{b}\|_{\mathbf{L}^\infty(K)}^2 h_K^{2k} \right) |u|_{k+1, K}^2 \\
 &\stackrel{(3.1)}{\leq} C^2 h^{2k} \sum_{K \in \mathcal{T}_h} \left( \varepsilon + h_K^2 + \delta_K \|\vec{b}\|_{\mathbf{L}^\infty(K)}^2 \right) |u|_{k+1, K}^2.
 \end{aligned}$$

Dann folgt für (6.10)

$$\|u - u_h\| \leq Ch^k \left( \sum_{K \in \mathcal{T}_h} \left( \varepsilon + h_K^2 + \frac{h_K^2}{\delta_K} + \delta_K \|\vec{b}\|_{\mathbf{L}^\infty(K)}^2 \right) |u|_{k+1, K}^2 \right)^{\frac{1}{2}}.$$

Jetzt benötigt man eine lokale Abschätzung für  $\delta_K$ . Dazu approximiert man

$$\frac{h_K^2}{\delta_K} \approx \delta_K \|\vec{b}\|_{\mathbf{L}^\infty(K)}^2. \tag{6.14}$$

Mit der lokalen **Peclet**-Zahl  $Pe_K := h_K \|\vec{b}\|_{\mathbf{L}^\infty(K)} / \varepsilon$  unterscheidet man zwei Fälle

$$\delta_K \sim \frac{h_K}{\|\vec{b}\|_{\mathbf{L}^\infty(K)}} \quad \text{falls } Pe_K \geq 1 \tag{6.15}$$

$$\delta_K \sim \frac{h_K^2}{\varepsilon} \quad \text{falls } Pe_K \leq 1. \tag{6.16}$$

Dann ergibt sich die Behauptung mit

$$\varepsilon + h_K^2 + \frac{h_K^2}{\delta_K} + \delta_K \sim \varepsilon + \|\vec{b}\|_{\mathbf{L}^\infty(K)} h_K + h_K^2 \leq Ch_K, \tag{6.17}$$

(3.1) und (3.22). □

## 6.2 Numerische Experimente

### Beispiel 6.3 (Konvektionsdominanter Fall)

In Beispiel (3.19) ändert man  $\varepsilon = 1$  zu  $\varepsilon = 10^{-10}$ , den Konvektionsterm  $\vec{b} = \vec{0}$  zu  $\vec{b} = (-y, x)^T$  und passt  $f$  entsprechend an die Lösung des Problems  $u = \sin(\pi x) \sin(\pi y) e^{-xy}$  an, die ihrerseits in  $C^\infty(\bar{\Omega})$  liegt und nicht von  $\varepsilon$  abhängt. Dann lautet der Term  $f$  der rechten Seite

$$f(x, y) = e^{-xy} \left\{ \varepsilon \left[ (2\pi^2 - y^2 - x^2) \sin(\pi x) \sin(\pi y) + 2\pi [x \sin(\pi x) \cos(\pi y) + y \cos(\pi x) \sin(\pi y)] \right] - y \sin(\pi y) [\pi \cos(\pi x) - y \sin(\pi x)] + x \sin(\pi x) [\pi \cos(\pi y) - x \sin(\pi y)] \right\}.$$

In Abbildung 6.1 sieht man links das Fehlerdiagramm der nicht stabilisierten, konformen Finite-Elemente-Methode. Rechts ist das Fehlerdiagramm der FEMLAB-Lösung mit der implementierten Stabilisierung „sd“ zu sehen. Beide Fehlerdiagramme sind analog zum Fehlerdiagramm rechts in Abbildung 3.9 aus Beispiel 3.19 erzeugt.

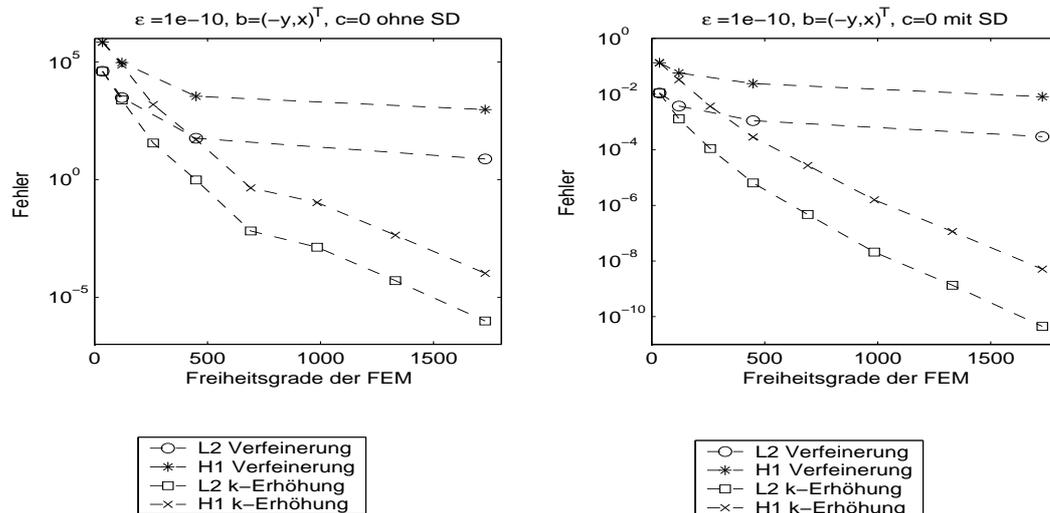


Abbildung 6.1: Approximationsfehler in der  $L^2$ - und  $H^1$ -Norm des konvektionsdominanten Beispiels 6.3 mit und ohne „Streamline Diffusion“ (SD)

Auf den folgenden Seiten sieht man in den Abbildungen 6.2 und 6.3 die erwarteten Konvergenzen in der  $\mathbf{L}^2$ -Norm  $h^{k+1}$  und der  $\mathbf{H}^1$ -Norm  $h^k$  für Elemente der Ordnungen  $k = 1, \dots, 6$  ( $P_1$ - $P_6$ -Elemente). Verglichen werden diese mit den in FEMLAB jeweils erreichten Konvergenzen zuzüglich einer multiplikativen Konstanten für das Problem aus Beispiel 6.3. Dabei wurde  $h$  von  $\frac{1}{2}$  bis zu  $\frac{1}{32}$  verfeinert. Abbildung 6.2 zeigt die nicht stabilisierte Finite-Elemente-Methode, in Abbildung 6.3 die mit „Streamline Diffusion“ stabilisierte Finite-Elemente-Methode aus FEMLAB. Diese Fehlerdiagramme sind vergleichbar mit der Abbildung 3.8 zu Beispiel 3.19. In FEMLAB ist die „Streamline Diffusion“-Methode implementiert mit freier Wahl des Parameters  $\delta$ . Man aktiviert die Methode im Aufruf des FEMLAB-Lösers als Parameter.

```
fem.sol=femlin(fem,'sd','on'); % Löser mit Stabilisierung mit delta=1
```

Dadurch wird die Standard-Testfunktion  $v$  zu  $v + \delta \vec{b} \nabla v$  verändert. Man kann auch statt des ‘on’ einen anderen Zahlenwert als 1 für  $\delta$  angeben.

Die Programmierung der Ausgabe von Diagrammen wie in Abbildung 6.3 erfolgt so:

```

h=[1/2 1/4 1/8 1/16 1/32];      % die maximalen Element-Durchmesser
k=[1 2 3 4 5 6];                % die Ordnung der Elemente
for j=1:length(k)
  for i=1:length(h)
%-----
% Hier werden jeweils die Abschnitte II,III und IV eingefügt
%-----
    gloses = fun2grid(fem2D,'fProbe(x,y)');% eigene Funktion,die die Lösung
                                           % an den von FEMLAB generierten Punkten des
                                           % Gebietes Omega mit der Lösungsfunktion
                                           % 'fProbe(x,y)' berechnet
    errL2(i)=L2Norm(fem2D,gloses);% Fehler in der L2-Norm (eigene Funktion)
    errH1(i)=H1Norm(fem2D,gloses);% Fehler in der H1-Norm (eigene Funktion)
  end;
  subplot(3,2,j),
  loglog(h,h.^(k(j)+1),'k-o'), % theoretische L2-Fehlerordnung
  hold on,
  loglog(h,h.^(k(j)),'k-s'),   % theoretische H1-Fehlerordnung
  loglog(h,errL2,'k-x'),       % erreichte Fehlerordnung in L2
  loglog(h1,errH1,'k-*'),      % erreichte Fehlerordnung in H1
  legend('L2_{T}','H1_{T}','L2_{FEM}','H1_{FEM}',4);
  title(strcat('P_{',num2str(k(j)),'}-Elemente')),
  hold off;
end
% Befehle:
% length(): liefert die Länge eines Vektors oder die größte Dimension
%           einer Matrix.
% Befehle für die Grafikausgabe:
% loglog:   doppelt logarithmischer Plot
% hold on:  bis zum Befehl 'hold off' wird im aktuellen Grafik-Fenster
%           weitergezeichnet.
% legend:   schreibt eine Legende zur Grafik, wobei die verschiedenen Li-
%           nien in Farbe und Form in der Reihenfolge ihrer Zeichnung au-
%           tomatisch eingefügt werden.

```

**Beispiel 6.4** Wir betrachten ein Problem auf dem Einheitsquadrat mit einem kreisrunden Loch  $\Omega := [0, 1]^2 \setminus \{(x, y) \in \mathbb{R}^2 : \|(x, y) - \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}\| < \frac{1}{4}\}$ :  
Finde  $u \in \mathbf{X}$ , so dass

$$\begin{aligned}
 -\varepsilon \Delta u(x, y) + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \nabla u(x, y) &= 0 && \text{in } \Omega, \\
 u(0, y) &= 1 && \text{für alle } y \in (0, 1] \text{ und}
 \end{aligned}$$

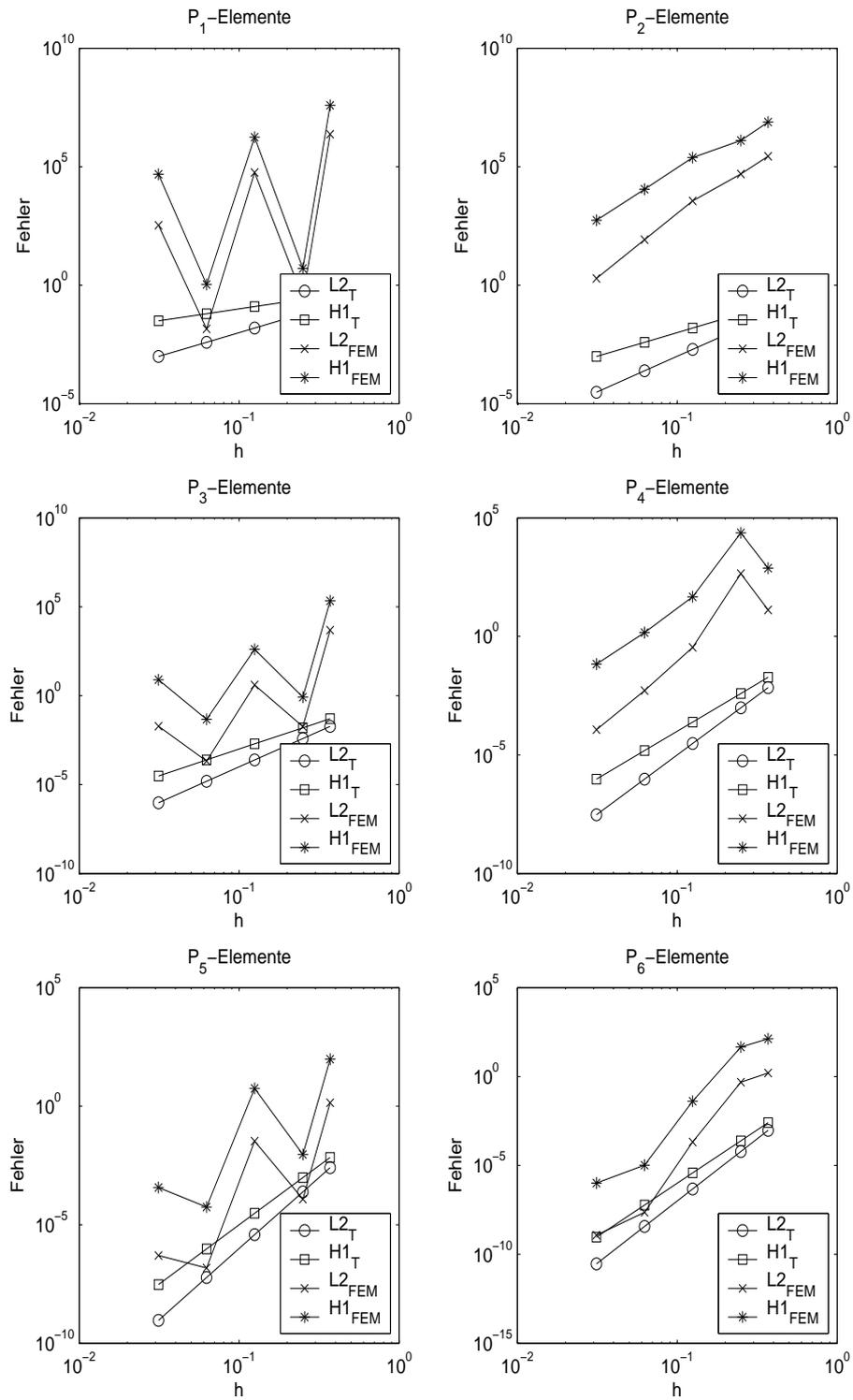


Abbildung 6.2: Konvergenz bis  $h = 1/32$  mit  $\vec{b} = \begin{pmatrix} -y \\ x \end{pmatrix}$  und  $\varepsilon = 10^{-10}$  ohne SD

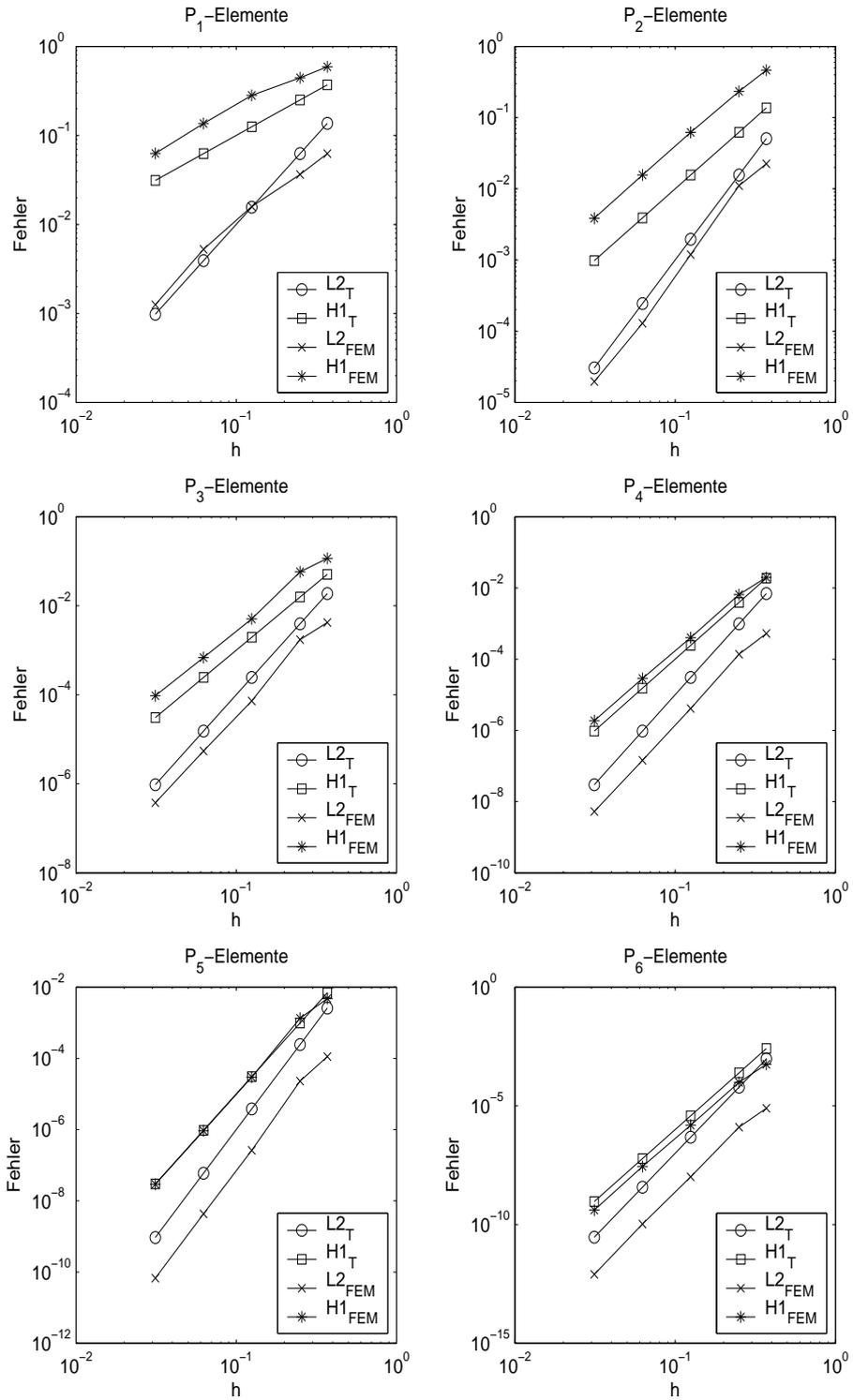


Abbildung 6.3: Konvergenz bis  $h = 1/32$  mit  $\vec{b} = \begin{pmatrix} -y \\ x \end{pmatrix}$  und  $\varepsilon = 10^{-10}$  mit SD

$$u(x, y) = 0 \quad \text{auf } \Gamma \setminus \{(x, y) \in \mathbb{R}^2 : x = 0\}$$

gilt.

Dabei ist  $\Gamma_- := \{(x, y) \in [0, 1]^2 : x = 0 \text{ oder } y = 0 \text{ oder } \|(x, y) - \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}\| = \frac{1}{4} \text{ für } x + y > 1\}$  der für den Grenzfall  $\varepsilon = 0$  wichtige einströmende Rand.

In Abbildung 6.4 stellen die oberen drei Abbildungen die Lösungen für ein zwischen 1 und  $10^{-6}$  variiertes  $\varepsilon$  dar, die unteren drei mit dem jeweils gleichen  $\varepsilon$  die Lösung mit der „Streamline Diffusion“-Stabilisierung aus FEMLAB.

In der ersten Spalte sieht man den für  $\varepsilon = 1$  diffusionsdominanten Fall, der eigentlich nicht stabilisiert werden müsste. Diese Stabilisierung stört aber nicht, da sie konsistent ist.

Die zweite Spalte zeigt deutlich die an Wichtigkeit zunehmende Rolle des Konvektionsterms, denn die unstetigen Randinformationen auf  $\Gamma_-$  werden in das Gebiet hineingetragen und nicht mehr geglättet. Dadurch entstehen Randgrenzschichten bei  $y = 1$  für  $0 \leq x \leq \frac{3}{8}$  und bei  $\|(x, y) - \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}\| = \frac{1}{4}$  für  $\frac{3}{8} \leq y \leq \frac{5}{8}$ . Außerdem bilden sich zwei innere Grenzschichten bei  $x = y$  für  $x < \frac{3}{8}$  und bei  $x + \frac{1}{2} = y$  für  $\frac{3}{8} < x \leq \frac{1}{2}$ . Durch die Bedingung  $u(x, 0) = 0$  vom Randstück  $y = 0$  entstehen keine Grenzschichten, da die Randbedingung gleich ist mit dem Wert von  $u$  im Inneren des Gebietes.

In der dritten Spalte ist das  $\varepsilon$  schon so klein, dass ohne Stabilisierung keine vernünftige Lösung mehr erreicht wird. Die Grenzschichten werden immer schärfer, d.h. der Gradient der Lösung wird wesentlich steiler.

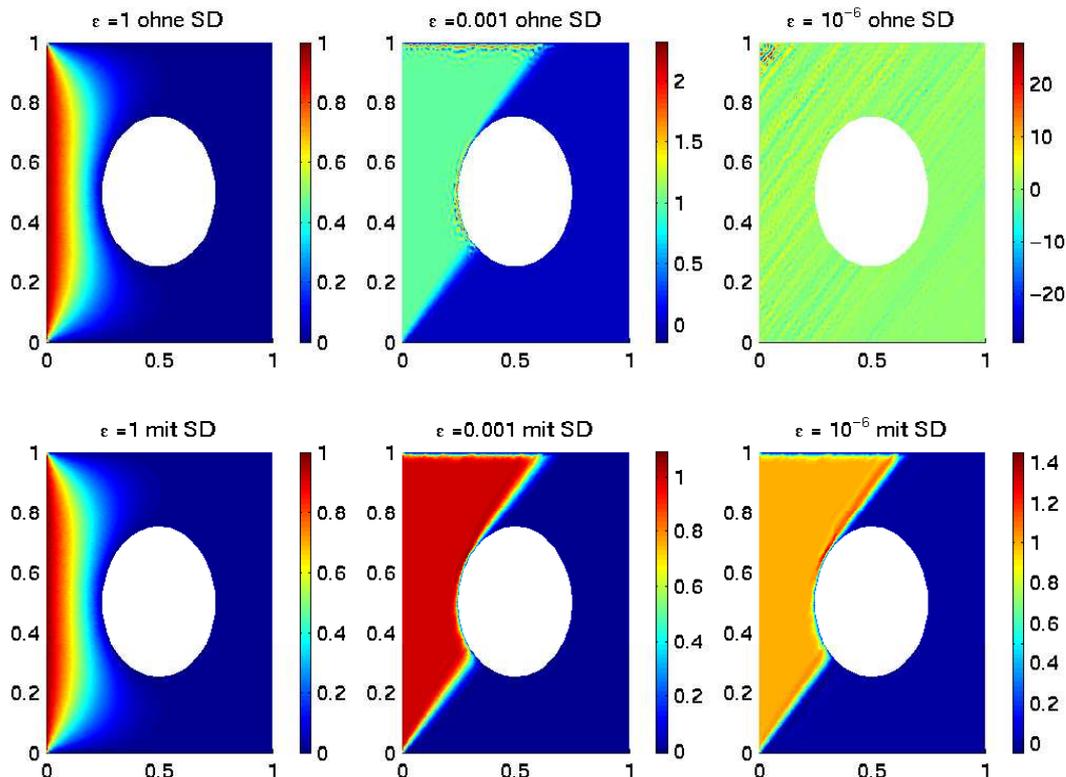


Abbildung 6.4: Beispiel 6.4

# Zusammenfassung und Ausblick

Ich werde in diesem Abschnitt ein Fazit aus meinen numerischen Ergebnissen ziehen und einen Blick in die Zukunft wagen, was die Entwicklung und Weiterentwicklung von Programmen des Typs FEMLAB zur Lösung von partiellen Differentialgleichungen betrifft.

Wenn man die dargestellten Konvergenzresultate für lineare, elliptische Differentialgleichungen zweiter Ordnung aus der ausgearbeiteten Theorie und der Praxis in FEMLAB miteinander vergleicht, kann man schnell erkennen, dass FEMLAB bereits mit modernen, der aktuellen Theorie entsprechenden Routinen ausgestattet ist. Denn alle theoretischen Konvergenzraten werden auch in FEMLAB erreicht. Sehr überzeugend stellt dies z.B. die Abbildung 3.8 dar.

Man kann den Nutzen bzw. die Funktionalität von Elementen höherer Ordnung in ein paar Punkten festhalten:

- + Der Konvergenzfehler nimmt mit steigender Ordnung der Elemente stärker ab als bei Verfeinerungen mit gleichem Speicheraufwand.
- + Die Rechenzeit erhöht sich wie bei einer Verfeinerung. Bei Elementen höherer Ordnung steigt sie abhängig von der Ordnung der Elemente, die die Steifigkeitsmatrizen immer dichter besetzt. Die Dimension der Steifigkeitsmatrizen bleibt erhalten.
- + Der Speicherbedarf bleibt relativ gering. Verfeinert man hingegen die Zerlegung, werden die Steifigkeitsmatrizen größer und man benötigt mehr Speicher.
- + Gekrümmte Ränder werden von Elementen höherer Ordnung besser approximiert.
- Man erhält für Lösungen mit Sprüngen bessere Resultate in der Nähe der Sprünge durch eine Verfeinerung der Zerlegung. Elemente höherer Ordnung eignen sich in den glatten Bereichen der Lösung.

Zusammengefasst bedeutet das, dass für speichersparende Verfahren grobe Zerlegungen mit Elementen hoher Ordnung besser geeignet sind. In FEMLAB merkt man das besonders bei dreidimensionalen Geometrien. Denn die Zerlegungen dieser Geometrien benötigen ein Vielfaches an Speicher gegenüber noch so fein zerlegten zweidimensionalen Geometrien. Der Speicherbedarf wächst noch stärker, wenn man die dreidimensionalen Zerlegungen auch nur geringfügig verfeinert. Die Laufzeit erhöht sich überproportional, weil diese großen Matrizen nicht mehr so leicht vom Hauptspeicher gehandhabt werden können. Deshalb konnte ich in Kapitel 3 auch nur ein kleines, dreidimensionales Beispiel rechnen (vgl. Beispiel 3.20).

In Kapitel 5 wird neben isoparametrischen Elementen zur Approximation krummliniger Ränder ein raffiniertes, adaptives Verfahren zur Verfeinerung der Gebietszerlegung von FEMLAB eingesetzt, um die dabei verloren gegangene Konvergenzordnung zurück zu gewinnen. Dies Verfahren

erfüllt alle numerischen Anforderungen, die auch isoparametrische Elemente leisten, nämlich die Rückgewinnung der alten Konvergenzordnung.

Auch mit der „Streamline–Diffusion“-Stabilisierung in FEMLAB kann man zufrieden sein. Die Abbildungen 6.3 und 6.4 in Kapitel 6 zeigen ihren Erfolg deutlich.

Zu bemängeln sind unübersichtliche und vor allem nicht klar verständliche Beschreibungen zu numerisch wichtigen Methoden im Handbuch [COM01] von FEMLAB. Die Themen sind häufig nicht an einer Stelle übersichtlich abgehandelt, sondern auf andere Themen im Handbuch verteilt. So auch die laut Handbuch mögliche Einarbeitung von Randbedingungen in schwacher Form (vgl. Kapitel 4). Hat man die Erläuterungen an Beispielen ausprobiert, kommen häufig mathematisch unverständliche Ergebnisse heraus. Deshalb habe ich mich im speziellen Fall des Kapitels 4 auch der in Matlab programmierten Routinen Rapins (vgl.[Rap03]) bedient. Bei anderen Schwierigkeiten habe ich es durch aufwändiges, systematisches Ausprobieren dann geschafft, dass FEMLAB seine Routinen richtig benutzt. Auch ist die Befehlssyntax häufig besser der Hilfe unter ‚help‘ zu entnehmen.

Abschließend möchte ich zusammenfassen: FEMLAB ist als Computersoftware für Finite–Elemente–Methoden recht gut geeignet. Sowohl mit linearen Elementen, als auch Elementen höherer Ordnung liefert es gute Ergebnisse für die, in dieser Arbeit behandelten, linearen, elliptischen Differentialgleichungen zweiter Ordnung. Auch komplizierte und dreidimensionale Gebiete lassen sich konstruieren. Der Rechenaufwand ist stellenweise noch etwas zeitintensiv, aber das ändert sich hoffentlich mit zunehmend besser entwickelten Rechnern und neuen Versionen des Programms FEMLAB. Ein Schritt in nächster Zukunft wird es sein, in FEMLAB die Kopplung von linearen Elementen in Bereichen von Sprüngen der Lösung mit Elementen höherer Ordnung auf dem glatten Lösungsgebiet zu erreichen (gemischte Methoden).

# Literaturverzeichnis

- [Ada75] Adams, Robert A.: *Sobolev spaces*. Academic Press, New York, 1975.
- [Alt99] Alt, Hans W.: *Lineare Funktionalanalysis*. Springer Verlag, Berlin-Heidelberg-New York, 3. Auflage, 1999.
- [Arc76] Arcangeli, R. und Jean-Louis Gout: *Sur l'évaluation de l'erreur d'interpolation de Lagrange dans un ouvert de  $\mathbb{R}^n$* . In: Rev. fr. d'Aut., Inf. et Rech. Op., Série rouge Anal. Numer., 10, S. 5–27, 1976.
- [Bab73] Babuška, Ivo: *The finite element method with Lagrangian multipliers*. In: Numer. Math., 20, S. 179–192, 1973.
- [Bar91] Barbosa, Helio J. C. und Thomas J. R. Hughes: *The finite element method with Lagrange multipliers on the boundary: circumventing the Babuska-Brezzi condition*. In: Comput. Methods Appl. Mech. Eng., 85, S. 109–128, 1991.
- [Bra97] Braess, Dietrich: *Finite Elemente*. Springer Verlag, Berlin-Heidelberg-New York, 2. Auflage, 1997.
- [Bre91] Brezzi, Franco und Michel Fortin: *Mixed and hybrid finite element methods*. Springer Verlag, Berlin-Heidelberg-New York, 1991.
- [Bre02] Brenner, Susanne C. und L. Ridgway Scott: *The mathematical theory of finite element methods*. Springer Verlag, Berlin-Heidelberg-New York, 2. Auflage, 2002.
- [Cia91] Ciarlet, Philippe G.: *Basic error estimates for elliptic problems*. In: Ciarlet, Philippe G. und Jacques L. Lions (Hg.), *Handbook of numerical analysis II (finite element methods, part 1)*. North-Holland, Amsterdam, New York, Oxford, Tokyo, 1991.
- [COM01] COMSOL Inc., Burlington, MA: *FEMLAB 2.2, Reference Manual*, 2001.
- [Gro94] Großmann, Christian und Hans-Görg Roos: *Numerik partieller Differentialgleichungen*. B.G. Teubner Stuttgart, 2. Auflage, 1994.
- [Joh87] Johnson, Claes: *Numerical solution of partial differential equations by the finite element method*. Cambridge University Press, 1987.
- [Kli01] Klimanis, Nils: *Lösungsverfahren und Vorkonditionierer in objekt-orientierter Implementierung für Konvektions-Diffusions-Reaktions-Probleme*. Diplomarbeit, Institut für Numerische und Angewandte Mathematik, Georg-August-Universität Göttingen, 2001.

- [Kna00] Knabner, Peter und Lutz Angermann: *Numerik partieller Differentialgleichungen: Eine anwendungsorientierte Einführung*. Springer Verlag, Berlin-Heidelberg-New York, 2000.
- [Lad68] Ladyzhenskaja, Olga A. und Nina N. Ural'ceva: *Linear and quasilinear elliptic equations*. Academic Press, New York, 1968.
- [Lio68] Lions, Jacques L. und Enrico Magenes: *Problèmes aux limites non-homogènes et applications*, Band 1. Dunod, Paris, 1968.
- [Lub02] Lube, Gert: *Numerik elliptischer Randwertprobleme*. Vorlesungsskript, Institut für Numerische und Angewandte Mathematik, Georg-August-Universität Göttingen, WS 2001/02.  
URL <http://www.num.math.uni-goettingen.de/lube/lehre.html>
- [Lub03] Lube, Gert: *Numerik instationärer partieller Differentialgleichungen*. Vorlesungsskript, Institut für Numerische und Angewandte Mathematik, Georg-August-Universität Göttingen, WS 2002/03.  
URL <http://www.num.math.uni-goettingen.de/lube/lehre.html>
- [Lub94] Lube, Gert: *Stabilized Galerkin finite element methods for convection dominated and incompressible flow problems*. In: Numerical analysis and mathematical modelling Banach center publications, 29, S. 85–104, 1994.
- [Nit71] Nitsche, Joachim: *Über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind*. In: Abh. Math. Univ. Hamburg, 36, S. 9–15, 1970/71.
- [Pit80] Pitkäranta, Juhani: *Local stability conditions for the Babuska method of Lagrange multipliers*. In: Math. Comput., 35, S. 1113–1129, 1980.
- [Qua94] Quarteroni, Alfio und Alberto Valli: *Numerical approximation of partial differential equations*. Springer Verlag, Berlin-Heidelberg-New York, 1994.
- [Rap03] Rapin, Gerd: *The Three-field Formulation for Elliptic Equations: Stabilization and Decoupling Strategies*. Doktorarbeit, Institut für Numerische und Angewandte Mathematik, Georg-August-Universität Göttingen, 2003.
- [Roo96] Roos, Hans-Görg, Martin Stynes und Lutz Tobiska: *Numerical methods for singularly perturbed differential equations*. Springer Verlag, Berlin-Heidelberg-New York, 1996.
- [Ste95] Stenberg, Rolf: *On some techniques for approximating boundary conditions in the finite element method*. In: J. Comput. Appl. Math., 63, S. 139–148, 1995.

# Danksagung

Ich danke allen, die mich in der Zeit während der Erstellung meiner Diplomarbeit unterstützt haben.

Ganz besonderen Dank schulde ich Herrn Prof. Dr. Gert Lube, der jederzeit ein offenes Ohr für mich und meine Probleme hatte. Er motivierte mich immer wieder mit seinem sachlichen und konsequenten Auftreten. Ich bedanke mich auch bei der Arbeitsgruppe mit Andreas, Gerd, Tobias K., Nils, Benjamin, Stephan und Tobias G. und dem Operating-Team Herrn Wasmann und Herrn Perske für die angenehme Zeit, die ich bei Euch/Ihnen verbracht habe.

Danken möchte ich auch Katharina und Hendrik für das Korrekturlesen meiner Arbeit.

Herzlichen Dank sage ich Melanie, die mir auch in stressigen Zeiten liebevoll zur Seite gestanden hat.

Besonderen Dank erhalten auch meine Eltern dafür, dass sie mir mein Studium ermöglicht haben, nicht nur in finanzieller Hinsicht.



# Versicherung

Hiermit versichere ich, dass ich die vorliegende Diplomarbeit selbständig verfasst habe und keine anderen als die angegebenen Hilfsmittel verwendet habe.

Göttingen, den 28.07.2003