

**Numerische Untersuchungen eines
unstetigen Galerkin–Verfahrens
zur Lösung der
Konvektions–Diffusions–Gleichung**

Diplomarbeit

zur Erlangung des akademischen Grades
eines Diplommathematikers
der Naturwissenschaftlichen-Technischen-Fakultät I
– Mathematik und Informatik –
der Universität des Saarlandes

vorgelegt von

Matthias Albert Augustin

Saarbrücken, 2009

Eidesstattliche Erklärung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Saarbrücken, den 17.12.2009

Matthias Augustin

Danksagung

Ich möchte meinen Dank aussprechen

- Herrn Prof. Dr. Volker John für die Vergabe des Themas und die gute Betreuung.
- Herrn Rudolf Umla für die freundschaftliche Zusammenarbeit.
- Simone Müller, Karsten Schwarz und Rudolf Umla fürs Korrekturlesen.
- meiner Familie und meinen Freunden, die mich immer unterstützt haben.
- der Studienstiftung des deutschen Volkes, ohne deren Stipendium mein Studium zumindest um einiges härter gewesen wäre.

Inhaltsverzeichnis

1	Einleitung	1
2	Grundlagen	3
2.1	Partielle Differentialgleichungen und klassische Funktionenräume	3
2.2	Schwache Formulierung von Differentialgleichungen	6
2.2.1	Allgemeinere Funktionenräume	6
2.2.2	Einige Aussagen aus der Funktionalanalysis	13
2.2.3	Formulierung als Variationsproblem	16
3	Galerkin–Verfahren und Finite–Elemente–Methoden	19
3.1	Galerkin–Verfahren	19
3.2	Finite–Elemente–Methoden	21
3.2.1	Grundideen	21
3.2.2	Polynomielle Finite–Elemente–Räume	23
3.2.3	Unstetige Finite–Elemente	28
3.2.4	Abschätzung des Interpolationsfehlers	34
3.2.5	Spur- und inverse Ungleichungen	37
3.2.6	Allgemeine Fehlerabschätzungen	38
4	Die Konvektions–Diffusions–Gleichung	41
4.1	Starke Formulierung und analytische Aussagen	41
4.2	Schwache Formulierung	44
4.3	Das Standard–Galerkin–Verfahren	47
4.4	Die Streamline–Diffusion–Methode	48
4.5	Discontinuous–Galerkin–Verfahren	53
4.5.1	Formulierung als Interior–Penalty–Methode	53
4.5.2	Flussformulierung	66
5	Numerische Ergebnisse	69
5.1	Einige technische Details	69
5.1.1	Implementierungsaspekte	69
5.1.2	Wahl des Parameters der Streamline–Diffusion–Methode	73
5.1.3	Wahl der Gebietszerlegung	73
5.2	Rechnungen für verschiedene Beispiele	74
5.2.1	Test der Implementierung	75
5.2.2	Beispiel ohne Grenzschichten	80
5.2.3	Beispiel mit inneren Grenzschichten	88
5.2.4	Beispiel mit Randgrenzschichten	99
5.2.5	Qualität der Randbedingungen	105
6	Zusammenfassung und Ausblick	109

Symbolverzeichnis

(Ω, V_h, Σ)	Finites-Element, Seite 23
(\cdot, \cdot)	Skalarprodukt im Raum $L^2(\Omega)$, Seite 8
$(\cdot, \cdot)_X$	Skalarprodukt im Raum X , Seite 8
$\langle \cdot, \cdot \rangle$	duales Produkt, Seite 13
T'	dualer Operator zum Operator T , Seite 14
T^{-1}	inverser Operator zum Operator T , Seite 14
U^\perp	orthogonales Komplement zum Unterraum U , Seite 15
V'	Dualraum zum Vektorraum V , Seite 13
v^+	$:= \lim_{\substack{\epsilon \rightarrow 0 \\ \epsilon > 0}} v(x + \epsilon b)$, Seite 30
v^-	$:= \lim_{\substack{\epsilon \rightarrow 0 \\ \epsilon > 0}} v(x - \epsilon b)$, Seite 30
$X \hookrightarrow Y$	Einbettung von X in Y , Seite 9
$\ T\ $	Operatornorm des Operators T , Seite 13
$\ \cdot\ _{1,h}$	Norm im Sobolev-Raum $H^1(\mathcal{T})$, Seite 30
$\ \cdot\ _X$	Norm im Raum X
$\ \cdot\ _{\text{disc}}$	Energienorm zu a_{disc} , Seite 65
$\ \cdot\ _{\text{konv}}$	Energienorm zu a_{konv} , Seite 62
$\ \cdot\ _{SD}$	Energienorm des Streamline-Diffusion-Verfahrens, Seite 49
$\ \cdot\ _a$	Energienorm zur Bilinearform a , Seite 20
$\ a\ $	Norm der Bilinearform a , Seite 13
$[\cdot]_E$	Sprung einer Funktion an der Kante E , Seite 29
$\langle \cdot \rangle_E$	Mittelwert einer Funktion an der Kante E , Seite 30
$[\cdot]_K$	Sprung einer Funktion am Rand einer Zelle K in eine vorgegebene Richtung, Seite 30
$ \alpha $	Ordnung des Multiindex α , Seite 3

$ \cdot $	Betrag in \mathbb{R} , Seite 5
$ \cdot $	euklidische Norm auf \mathbb{R}^d , Seite 5
$ \cdot _X$	Halbnorm auf dem Raum X , Seite 9
$\nabla \cdot b$	Divergenz von b , Seite 45
∇v	Gradient von v , Seite 5
Δ	Laplace-Operator, Seite 41
$\partial_n v$	Normalenableitung von v , Seite 37
$\partial\Omega$	Rand von Ω , Seite 3
$\partial_+ K$	Ausströmrand von K , Seite 30
$\partial_- K$	Einströmrand von K , Seite 30
\mathbb{N}	Menge der natürlichen Zahlen, Seite 3
\mathbb{N}_0	Menge der natürlichen Zahlen inklusive der Null, Seite 3
\mathbb{N}_0^d	Menge der d -dimensionalen Multiindizes, Seite 3
\mathbb{R}	Menge der reellen Zahlen, Seite 3
\mathbb{R}^d	d -dimensionaler Raum, Seite 3
A	Darstellungsoperator der Bilinearform a , Seite 13
C_0^∞	Raum der beliebig oft differenzierbaren Funktionen mit kompaktem Träger, Seite 8
$C^{k,s}(\bar{\Omega})$	Raum der k -mal stetig differenzierbaren Funktionen mit Hölder-stetigen k -ten Ableitungen auf $\bar{\Omega}$ zum Exponenten s , Seite 5
$C^k(\Omega)$	Raum der k -mal stetig differenzierbaren Funktionen auf Ω , Seite 3
$C_0^k(\Omega)$	Raum der k -mal stetig differenzierbaren Funktionen mit kompaktem Träger in Ω , Seite 10
C_F	Friedrichskonstante, Seite 12
$D(T)$	Definitionsbereich des Operators T , Seite 14
$D^\alpha v$	partielle Ableitung von v zum Multiindex α , Seite 3
$D^k v$	Menge aller partieller Ableitungen von v zu Multiindizes mit Ordnung k , Seite 3
E	Kante zwischen zwei Zellen K und K' einer Zerlegung \mathcal{T} , Seite 28
F_K	Referenztransformation von \hat{K} auf K , Seite 24
$H^{k+s}(\Omega)$	Sobolev-Slobodecki-Raum, Seite 12

$H^k(\Omega)$	Sobolev–Raum $W^{k,2}$, Seite 9
$H_0^k(\Omega)$	Sobolev–Raum $W_0^{k,2}$, Seite 10
$H_1(u)$	$:= u - u_h _{H^1(\mathcal{T})}$, Seite 75
I	Interpolationsoperator, Seite 34
I_h^p	Lagrange–Interpolationsoperator, Seite 34
K	Zelle der Zerlegung \mathcal{T} , Seite 22
\hat{K}	Referenzzelle eines parametrischen Finiten–Elementes, Seite 24
$L(\alpha)$	Langevin–Funktion, Seite 73
$L_{\text{loc}}^1(\Omega)$	Raum der lokal integrierbaren Funktionen, Seite 8
$L^p(\Omega)$	Lebesgue–Raum, $1 \leq p \leq \infty$, Seite 7
$L_2(u)$	$:= \ u - u_h\ _{L^2(\mathcal{T})}$, Seite 75
$\mathbb{P}_p(K)$	Raum der Polynome vom Grad $\leq p$ auf K , Seite 24
$P_p(\Omega)$	Raum der (stetigen) stückweisen Polynome vom Grad $\leq p$ auf Ω , Seite 24
$P_p^{\text{disc}}(\Omega)$	Raum der unstetigen stückweisen Polynome vom Grad $\leq p$ auf Ω , Seite 28
Pe_K	lokale Pécletzahl, Seite 51
$\mathbb{Q}_p(K)$	Tensorraum der Polynome vom Grad $\leq p$ in jeder Variablen auf K , Seite 24
$Q_p(\Omega)$	Tensorraum der (stetigen) stückweisen Polynome vom Grad $\leq p$ in jeder Variablen auf Ω , Seite 24
$Q_p^{\text{disc}}(\Omega)$	Tensorraum der unstetigen stückweisen Polynome vom Grad $\leq p$ in jeder Variablen auf Ω , Seite 28
$R(T)$	Bild des Operators T , Seite 14
SDFEM	Streamline–Diffusion–Finite–Element–Methode, Seite 48
SIPG	symmetrisches Interior–Penalty–Galerkin–Verfahren, Seite 57
T	Operator, Seite 13
V_h	endlich–dimensionaler Teilraum des Vektorraumes V , Seite 19
$W^{-k,q}(\Omega)$	Dualraum von $W_0^{k,p}(\Omega)$, Seite 11
$W^{k,p}(\Omega)$	Sobolev–Raum der Funktionen, deren Ableitungen mit Ordnungen $\leq k$ in L^p liegen, Seite 9
$W_0^{k,p}(\Omega)$	Abschluss von C_0^∞ bezüglich $\ \cdot\ _{W^{k,p}}$, Seite 10
$a(\cdot, \cdot)$	Bilinearform, Seite 13

$a_{SD}(\cdot, \cdot)$	Bilinearform des Streamline–Diffusion–Verfahrens, Seite 49
$a_{\text{diff}}(\cdot, \cdot)$	Anteil der diffusiven Terme an a_{disc} , Seite 57
$a_{\text{disc}}(\cdot, \cdot)$	Bilinearform des SIPG–Verfahrens, Seite 56
$a_{\text{flux}}(\zeta_h, \tau_h)$	Bilinearform der Flussformulierung, Seite 67
$a_{\text{konv}}(\cdot, \cdot)$	Anteil der konvektiven Terme an a_{disc} , Seite 61
$b(\cdot)$	Advektionsvektor, Seite 41
$b_{\text{flux}}(u_h, \tau_h)$	Bilinearform der Flussformulierung, Seite 67
$c(\cdot)$	Reaktionsfunktion in der Konvektions–Diffusions–Gleichung, Seite 41
$c_{\text{flux}}(u_h, v_h)$	Bilinearform der Flussformulierung, Seite 67
$\text{diam}(K)$	Durchmesser von K , Seite 22
$f(\cdot)$	rechte Seite der Konvektions–Diffusions–Gleichung, Seite 41
$f_{SD}(\cdot)$	Linearform des Streamline–Diffusion–Verfahrens, Seite 49
$f_{\text{disc}}(\cdot)$	Linearform des SIPG–Verfahrens, Seite 57
$f_{\text{flux}}(v_h)$	Linearform der Flussformulierung, Seite 67
$g_{\text{flux}}(\tau_h)$	Linearform der Flussformulierung, Seite 67
g_i	stetige Funktion von Γ_i nach \mathbb{R} , Seite 5
h	charakteristische Größe der Zerlegung \mathcal{T} , meist das Maximum der Zellen–durchmesser, Seite 22
h_E	Länge der Kante E , Seite 55
h_K	charakteristische Größe der Zelle K , meist der Durchmesser, Seite 22
$\ker(T)$	Kern des Operators T , Seite 14
$\ker(T)^\circ$	Annihilator zu $\ker(T)$, Seite 14
$n(\cdot)$	äußerer Normalenvektor, Seite 5
n_E	Normalenvektor zur Kante E , Seite 54
ord_i	Konvergenzordnung auf Gitterlevel i , Seite 74
u	exakte Lösung eines Variationsproblem im gesamten Raum V , Seite 19
u^D	vorgegebene Dirichlet–Randwerte, Seite 41
\tilde{u}^D	Fortsetzung von u^D , Seite 44
$u_{h,i}$	Lösung eines Finite–Elemente Verfahrens auf Gitterlevel i , Seite 74

u_h	Näherungslösung eines Galerkin- oder FE-Verfahrens, Seite 19
\hat{u}_h^c	konvektiver Fluss zu u_h , Seite 67
\hat{u}_h^d	diffusiver Fluss zu u_h , Seite 67
\hat{x}, \hat{y}	Koordinaten auf der Referenzzelle, Seite 25
Γ	Rand von Ω , Seite 5
Γ_i	Teilstück des Randes von Ω , Seite 5
Γ_{int}	Menge aller Punkte, welche zu inneren Kanten der Zerlegung \mathcal{T} gehören, Seite 28
Π_h	Projektionsoperator, Seite 38
Σ	Satz linear unabhängiger, linearer, stetiger und bezüglich V_h unisolventer Funktionale, Seite 23
$\Phi_i(\cdot)$	Funktional, Seite 22
Ω	Teilmenge von \mathbb{R}^d , meist ein Gebiet, Seite 3
$\alpha = (\alpha_1, \dots, \alpha_d)$	d -dimensionaler Multiindex, Seite 3
γ	Spuroperator, Seite 11
γ	Konstante, Seite 62
δ	Skalierungsparameter der Youngschen Ungleichung, Seite 15
$\delta_1, \delta_2, \delta_3$	Funktionen zur Festlegung numerischer Flüsse, Seite 67
ϵ	Diffusionskonstante, Seite 41
ζ	$:= \epsilon \nabla u$ im Rahmen der Flussformulierung, Seite 66
$\hat{\zeta}_h^d$	diffusiver Fluss zu ζ_h , Seite 67
κ_E	Parameter der Norm $\ \cdot\ _{1,h}$, Seite 30
μ	Lebesgue-Maß, Seite 6
μ_{inv}	Konstante der inversen Ungleichung, Seite 36
$\xi_i(\hat{x})$	Basisfunktion in einer Dimension auf dem Intervall $[-1, 1]$, Seite 27
π	Projektionsoperator, Seite 15
π_h	L^2 -Projektionsoperator, Seite 36
ρ	Reaktionsfunktion in der Konvektions-Diffusions-Gleichung; zeitweise an Stelle von c benutzt um Verwechslungen mit Konstanten vorzubeugen, Seite 65
ρ_K	Umkreisradius, Seite 34

σ	(konstanter Anteil des) Stabilisierungsparameter eines Discontinuous–Galerkin–Verfahrens, Seite 55
σ_E	Stabilisierungsparameter eines Discontinuous–Galerkin–Verfahrens für eine Kante E , Seite 55
σ_h	(allgemeiner) Stabilisierungsparameter eines Discontinuous–Galerkin–Verfahrens, Seite 55
σ_K	Innkreisradius, Seite 34
τ_0	Parameter des Streamline–Diffusion–Verfahrens im konvektions–dominanten Fall, Seite 51
τ_1	Parameter des Streamline–Diffusion–Verfahrens im diffusions–dominanten Fall, Seite 51
τ_K	Parameter des Streamline–Diffusion–Verfahrens, Seite 49
ϕ_i	Basisfunktion in V_h , Seite 19
$\hat{\phi}$	Basisfunktion auf der Referenzzelle, Seite 24
\mathcal{E}	Menge aller Kanten der Zerlegung \mathcal{T} , Seite 28
\mathcal{L}	linearer Differentialoperator zweiter Ordnung, Seite 4
\mathcal{E}^D	Menge aller Kanten der Zerlegung \mathcal{T} , auf welchen eine Dirichlet–Randbedingung vorgegeben ist, Seite 28
\mathcal{E}_{int}	Menge aller innerer Kanten der Zerlegung \mathcal{T} , Seite 28
$\mathcal{L}(V, Q)$	Raum der stetigen linearen Operatoren von V nach Q , Seite 13
\mathcal{L}^p	\mathcal{L}^p –Raum, $1 \leq p \leq \infty$, Seite 6
$\mathcal{T}, \mathcal{T}_h$	Zerlegung von Ω , Seite 22

Kapitel 1

Einleitung

Zahlreiche Anwendungen, vor allem aus der Physik, führen in ihrer mathematischen Formulierung auf (partielle) Differentialgleichungen. Diese sind oftmals nicht auf rein analytischem Wege lösbar, weshalb numerische Verfahren wie die Finite-Differenzen-Methode, die Finite-Volumen-Methode oder die Finite-Elemente-Methode zum Einsatz kommen. In vielen Fällen muss man jedoch feststellen, dass die Standardvarianten dieser Methoden keine brauchbaren Lösungen liefern.

Eine noch recht einfache Gleichung, bei der dies schon der Fall ist, ist die so genannte Konvektions-Diffusions-Gleichung

$$-\epsilon \Delta u + b(x) \cdot \nabla u + c(x) u = f(x) \text{ in } \Omega.$$

Hierbei ist u eine skalare Funktion auf einem Gebiet $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, welche die Verteilung einer physikalischen Größe, beispielsweise der Temperatur oder einer Konzentration, beschreibt. Die Größen ϵ , b , c und f sind hinreichend reguläre Funktionen. Bei geeignet vorgegebenen Randbedingungen besitzt die Konvektions-Diffusions-Gleichung eine eindeutige Lösung auf Ω . Jedoch führt der Fall $\|b\|_{L^\infty(\Omega)} \gg \epsilon$ zu Problemen bei der numerischen Bestimmung einer Näherungslösung. Hier liefert das Standardverfahren der Finiten-Elemente-Methode häufig eine Lösung, welche unphysikalische Oszillationen zeigt. Um dies zu verhindern, muss das Verfahren modifiziert werden.

Eine Möglichkeit dazu, welche im Rahmen dieser Arbeit behandelt werden soll, ist die Verwendung von unstetigen Funktionen. Dies führt auf die so genannten Discontinuous-Galerkin-Verfahren. Ziel ist es, ein Discontinuous-Galerkin-Verfahren im Programmpaket MooNMD („Mathematics and object oriented Numerics in Magdeburg“) zu implementieren und an Hand geeigneter Beispiele zu untersuchen. Dabei soll auch ein Vergleich mit dem so genannten Streamline-Diffusion-Verfahren stattfinden, welches eine in der Praxis oftmals benutzte und bewährte Stabilisierungsmethode ist.

Discontinuous-Galerkin-Verfahren wurden erstmals von Reed und Hill ([Ree73](#)) benutzt um das Neutronentransportproblem zu behandeln. Ihr Verfahren wurde von LeSaint und Raviart ([LeS74](#)) analysiert, welche analytische Abschätzungen für den Verfahrensfehler angaben. Eine der ersten Arbeiten zur Verwendung unstetiger Finiten-Elemente für elliptische Gleichungen wurde von Arnold veröffentlicht ([Arn82](#)). Seitdem wurden zahlreiche weitere Methoden vorgeschlagen (für einen Überblick, siehe ([Arn02](#)) und ([Coc00](#))). Für diese Arbeit wurde das Symmetric-Interior-Penalty-Verfahren (SIPG) verwendet, welches beispielsweise in ([Kan07](#)) analysiert wird.

Die vorliegende Arbeit gliedert sich wie folgt:

- Zuerst werden in **Kapitel 2** die Grundlagen aus der Funktionalanalysis kurz bereit gestellt. Dabei werden wir sehen, wie man prinzipiell zu einer gegebenen Differentialgleichung ein Variationsproblem aufstellt, die Theorie der Lebesgue- und Sobolev-Räume

kurz anschneiden und einige grundlegende Sätze dazu angeben.

- In **Kapitel 3** werden Galerkin–Verfahren und Finite–Elemente–Methoden vorgestellt. Dabei werden die Ideen hinter den Verfahren kurz erläutert und einige Aussagen über die Lösbarkeit und die mit ihrer Hilfe bestimmten Lösungen formuliert. Wir werden in diesem Kapitel sowohl Finite–Elemente mit stetigen als auch mit unstetigen Basisfunktionen behandeln.
- **Kapitel 4** stellt zunächst die Konvektions–Diffusions–Gleichung und einige analytische Aussagen dazu vor. Anschließend werden wir die schwache Formulierung und das Standard–Galerkin–Verfahren für diese Gleichung präsentieren. Da das Standard–Galerkin–Verfahren, wie schon erwähnt, oftmals nicht in der Lage ist, eine brauchbare Lösung zu bestimmen, werden wir als stabilisierte Methode das Streamline–Diffusion–Verfahren kurz erläutern und einige Aussagen bezüglich der Güte der mittels dieses Verfahrens bestimmten Lösung treffen. Als Alternative zum Streamline–Diffusion–Verfahren stellen wir das SIPG–Verfahren als Beispiel eines Discontinuous–Galerkin–Verfahren vor. Auch für dieses werden wir einige Aussagen bezüglich der Qualität der numerisch bestimmten Lösung angeben.
- Eine numerische Untersuchung des SIPG–Verfahrens ist Gegenstand von **Kapitel 5**. In diesem Kapitel werden wir zunächst kurz die Implementierung der einzelnen Terme des SIPG–Verfahrens testen. Anschließend werden wir an Hand verschiedener Beispiele untersuchen, in welchen Fällen das SIPG–Verfahren geeignet ist, die Konvektions–Diffusions–Gleichung zu lösen. Als Vergleich dienen uns jeweils die Ergebnisse eines Streamline–Diffusions–Verfahrens, welches bereits in MooNMD implementiert ist. Dabei werden wir auch die Laufzeiten der beiden Methoden vergleichen. Zudem werden wir kurz untersuchen, inwiefern die Randbedingungen erfüllt sind, welche beim SIPG–Verfahren nur schwach implementiert sind.
- Zum Abschluss fasst **Kapitel 6** die Ergebnisse dieser Arbeit zusammen und gibt einen Ausblick auf weitere mögliche Untersuchungen zu Discontinuous–Galerkin–Verfahren für die Konvektions–Diffusions–Gleichung.

Kapitel 2

Grundlagen

In diesem Kapitel sollen zunächst einige Grundlagen zu Differentialgleichungen besprochen werden. Ziel ist es dabei, eine Formulierung zu erhalten, welche für die numerische Behandlung geeignet ist.

2.1 Partielle Differentialgleichungen und klassische Funktionenräume

Wir wollen zunächst definieren, was eine partielle Differentialgleichung überhaupt ist. Dazu führen wir folgende Notationen ein:

Definition und Lemma 2.1. (Partielle Ableitungen und klassische Funktionenräume)

Sei Ω eine beschränkte offene Menge aus \mathbb{R}^d , $d \in \mathbb{N}$, $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}_0^d$ ein Multiindex und $k \in \mathbb{N}_0$.

Die partielle Ableitung $D^\alpha v$ ist definiert als

$$(D^\alpha v)(x) := \left(\frac{\partial^{|\alpha|} v}{\partial x_1^{\alpha_1} \cdot \dots \cdot \partial x_d^{\alpha_d}} \right)(x).$$

Dabei ist $|\alpha| := \sum_i \alpha_i$ die Ordnung der Ableitung. Die Menge aller partiellen Ableitungen der Ordnung k von v bezeichnen wir mit $(D^k v)(x) := \{(D^\alpha v)(x) : |\alpha| = k\}$.

Die Menge der k -mal stetig differenzierbaren Funktionen auf dem Abschluss von Ω sei definiert durch:

$$C^k(\bar{\Omega}) := \{v : \bar{\Omega} \rightarrow \mathbb{R} : v \text{ ist } k\text{-mal stetig differenzierbar auf } \Omega \\ \text{und für } |\alpha| \leq k \text{ ist } D^\alpha v \text{ stetig auf } \bar{\Omega} \text{ fortsetzbar}\}.$$

$C^k(\bar{\Omega})$ ist vollständig bezüglich folgender Norm

$$\|v\|_{C^k(\Omega)} := \max_{0 \leq |\alpha| \leq k} \sup_{x \in \Omega} |(D^\alpha v)(x)|.$$

Beweis.

Für den eindimensionalen Fall wird beispielsweise in (Wer95) auf Seite 6 gezeigt, dass $(C^k(\bar{\Omega}), \|v\|_{C^k(\Omega)})$ ein Banachraum ist. Für höhere Dimensionen kann man dies analog nachrechnen. \square

Eine partielle Differentialgleichung lässt sich nun wie folgt definieren:

Definition 2.2. (Partielle Differentialgleichung und klassische Lösung)

Sei Ω ein beschränktes Gebiet aus \mathbb{R}^d , $d \in \mathbb{N}$, $d > 1$ und sei eine Abbildung $\mathcal{F} : \mathbb{R}^{dk} \times \mathbb{R}^{d(k-1)} \times \dots \times \mathbb{R}^d \times \mathbb{R} \times \Omega \rightarrow \mathbb{R}$, $k \in \mathbb{N}_0$ gegeben. Die Gleichung

$$\mathcal{F}((D^k v)(x), (D^{k-1}v)(x), \dots, v(x), x) = 0 \quad \forall x \in \Omega$$

heißt partielle Differentialgleichung (PDGl) der Ordnung k , falls mindestens eine Ableitung der Ordnung k tatsächlich auftritt und keine Ableitungen mit Ordnung größer als k auftreten. Eine Funktion $v \in C^k(\bar{\Omega})$, welche die obige Differentialgleichung erfüllt, heißt klassische Lösung.

Bemerkung 2.3.

Obige Definition lässt sich direkt auf Systeme von Differentialgleichungen übertragen. Für ein System aus m Gleichungen ist dann $\mathcal{F} : \mathbb{R}^{mdk} \times \mathbb{R}^{md(k-1)} \times \dots \times \mathbb{R}^{md} \times \mathbb{R}^{\geq} \times \Omega \rightarrow \mathbb{R}^m$, $k \in \mathbb{N}_0$ und $v \in C^k(\bar{\Omega}, \mathbb{R}^m)$, $v = (v_1, \dots, v_m)$.

Die Eigenschaften einer PDGl hängen von der Struktur von \mathcal{F} ab. Man kann dabei verschiedene Arten von Differentialgleichungen unterscheiden. In dieser Arbeit treten nur lineare Differentialgleichungen auf.

Definition 2.4.

Eine PDGl heißt linear, wenn sie die Form

$$\sum_{|\alpha| \leq k} a_\alpha(x) (D^\alpha v)(x) = f(x)$$

mit gegebenen Funktionen $a_\alpha(x)$ und $f(x)$ besitzt. Gilt zusätzlich $f \equiv 0$, so heißt die PDGl homogen.

Die Differentialgleichungen, welche wir in dieser Arbeit betrachten wollen, werden höchstens von zweiter Ordnung sein. Für diese unterscheidet man üblicherweise drei Typen.

Definition 2.5.

Sei Ω ein beschränktes Gebiet aus \mathbb{R}^d , $d \in \mathbb{N}$, $v : \Omega \rightarrow \mathbb{R}$. Weiter sei \mathcal{L} ein linearer Differentialoperator zweiter Ordnung, so dass gilt

$$v(x) \mapsto \mathcal{L}v(x) = \sum_{i,j=1}^d a_{ij}(x) \frac{\partial^2 v(x)}{\partial x_i \partial x_j} + \sum_{i=1}^d b_i(x) \frac{\partial v(x)}{\partial x_i} + c(x) v(x) + f(x).$$

Die PDGl $\mathcal{L}v(x) = 0$ heißt

- (i) elliptisch, falls die Matrix $(a_{ij})_{i,j=1}^d$ positiv (negativ) definit ist,
- (ii) hyperbolisch, falls die Matrix $(a_{ij})_{i,j=1}^d$ indefinit ist,
- (iii) parabolisch, falls die Matrix $(a_{ij})_{i,j=1}^d$ semidefinit ist.

Üblicherweise sucht man eine Lösung der PDGl, die zusätzlich noch bestimmte Eigenschaften am Rand $\partial\Omega$ des Gebiets Ω erfüllt; man stellt so genannte Randbedingungen an die Lösung. Ebenso wird bei zeitabhängigen Problemen meist eine Bedingung an die Lösung zum Zeitpunkt $t = 0$ gestellt. Zeitabhängige Probleme werden in dieser Arbeit jedoch nicht behandelt.

Die am häufigsten vorkommenden Randbedingungen sind:

$$\begin{aligned} \text{Dirichlet-Randbedingung:} & \quad v = g_1 \quad \text{auf } \Gamma_1. \\ \text{Neumann-Randbedingung:} & \quad n \cdot \nabla v = g_2 \quad \text{auf } \Gamma_2. \\ \text{Robin-Randbedingung:} & \quad \beta v + n \cdot \nabla v = g_3 \quad \text{auf } \Gamma_3. \end{aligned}$$

Dabei sind Γ_i , $i = 1, 2, 3$, disjunkte Teilstücke des Gebietsrandes $\partial\Omega$, so dass $\Gamma_1 \cup \Gamma_2 \cup \Gamma_3 = \partial\Omega$ gilt, $n = n(x)$ der äußere Normalenvektor von $\partial\Omega$ an der Stelle x , $(\nabla v)(x)$ der Gradient der Funktion v , $\beta(x)$ eine stetige Funktion von Γ_3 nach \mathbb{R} und g_i stetige Funktionen von Γ_i nach \mathbb{R} . Ein häufiger Spezialfall sind homogene Dirichlet-Randbedingungen, das heißt

$$v = 0 \text{ auf } \partial\Omega.$$

Neben der Struktur der PDGl an sich und den vorgegebenen Randbedingungen spielen auch die Eigenschaften des Gebiets Ω eine Rolle bei der Frage nach Existenz und Eindeutigkeit der Lösung einer PDGl. Oftmals wird dabei verlangt, dass Ω einen so genannten Lipschitz-Rand $\partial\Omega$ besitzt. Das hierbei auftretende Konzept der Lipschitz-Stetigkeit ist ein Spezialfall der Hölder-Stetigkeit, welche wir später auch noch benötigen werden.

Lemma 2.6. (Hölder-Stetigkeit)

Seien die Voraussetzungen von Lemma 2.1 gegeben und sei $0 \leq |\alpha| \leq k$. Der Raum der k -mal stetig differenzierbaren Funktionen mit Hölder-stetigen k -ten Ableitungen $C^{k,s}(\bar{\Omega})$, $0 < s \leq 1$, sei der Teilraum von $C^k(\bar{\Omega})$ bestehend aus den Funktionen v , für welche $D^\alpha v$ Hölder-stetig mit Exponenten s ist, das heißt es gibt eine Konstante $C > 0$ mit:

$$|(D^\alpha v)(x) - (D^\alpha v)(y)| \leq C |x - y|^s \quad \forall x, y \in \Omega. \tag{2.1.1}$$

Im Fall $s = 1$ spricht man von Lipschitz-Stetigkeit.

Für $r \geq s$ gilt die Inklusion $C^{k,r} \subset C^{k,s}$.

$C^{k,s}(\bar{\Omega})$ ist ein Banachraum bezüglich der Norm

$$\|v\|_{C^{k,s}(\Omega)} = \|v\|_{C^k(\Omega)} + \max_{0 \leq |\alpha| \leq k} \sup_{\substack{x,y \in \Omega \\ x \neq y}} \frac{|(D^\alpha v)(x) - (D^\alpha v)(y)|}{|x - y|^s}. \tag{2.1.2}$$

Beweis.

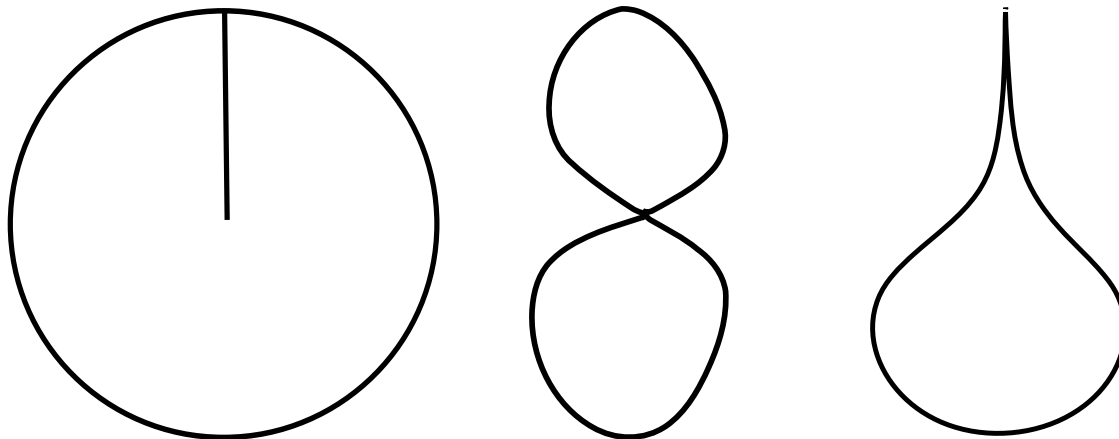
Dass $C^{k,s}(\bar{\Omega})$ ein Banachraum ist, zeigt man analog zum Beweis von Lemma 2.1. Für den Beweis der Inklusion siehe (Ada75) Theorem 1.31. \square

Definition 2.7. (Lipschitz-Rand)

Sei $\Omega \subset \mathbb{R}^d$, $d \in \mathbb{N}$, ein beschränktes Gebiet. Dann heißt $\partial\Omega$ ein Lipschitz-Rand beziehungsweise Ω ein Lipschitz-Gebiet, wenn es endlich viele Mengen $\omega_1, \dots, \omega_n \subset \mathbb{R}^d$ gibt, für welche gilt:

- (1.) Ω wird durch $\omega_1, \dots, \omega_n$ überdeckt.
- (2.) $\partial\Omega \cap \omega_i$ ist Graph einer Lipschitz-stetigen Funktion $\forall i = 1, \dots, n$.
- (3.) $\Omega \cap \omega_i$ liegt auf einer Seite dieses Graphen.

Beispiele für Lipschitzgebiete sind polygonal oder polyhedral berandete Gebiete. Gebiete mit nicht Lipschitz-stetigem Rand zeigt Abbildung 2.1.



Beispiele für Gebiete mit *nicht* Lipschitz–stetigem Rand. Der Grund für die fehlende Regularität ist links der Schlitz im Gebietsrand, in der Mitte die Überschneidung und rechts der Scheitel.

(Abbildung 2.1)

2.2 Schwache Formulierung von Differentialgleichungen

Betrachtet man Differentialgleichungen in der bisher eingeführten Form, so sind die Ansprüche an die Lösung bezüglich Stetigkeit und Differenzierbarkeit recht stark. In den Anwendungen sind oftmals die Voraussetzungen für Existenz und Eindeutigkeit der Lösung nicht erfüllt. Trotzdem laufen die beschriebenen Prozesse ab, was bedeutet, dass es Lösungen gibt. Um diese zu beschreiben, benötigt man allgemeinere mathematische Konzepte wie das Konzept der schwachen Differenzierbarkeit. Mit diesem lassen sich Lösungen von Differentialgleichungen in einem allgemeineren Sinne definieren. Dazu ist es zunächst notwendig, allgemeinere Funktionenräume einzuführen.

2.2.1 Allgemeinere Funktionenräume

Den Anfang machen hier die so genannten Lebesgue–Räume.

Definition 2.8.

Seien μ das Lebesgue–Maß auf dem \mathbb{R}^n , $\Omega \subset \mathbb{R}^n$ eine beschränkte Menge und $f : \Omega \rightarrow \mathbb{R}$ eine messbare Abbildung. Für $1 \leq p \leq \infty$ definieren wir die Halbnormen $\|\cdot\|_{L^p(\Omega)}$ durch

$$\|f\|_{L^p(\Omega)} = \left(\int_{\Omega} |f|^p d\mu \right)^{1/p}, \quad 1 \leq p < \infty,$$

$$\|f\|_{L^\infty(\Omega)} = \inf_{\mu(\mathcal{N})=0} \sup_{x \in \Omega \setminus \mathcal{N}} |f(x)|$$

und die Räume $\mathcal{L}^p(\Omega)$ durch

$$\mathcal{L}^p(\Omega) = \{f : \Omega \rightarrow \mathbb{R} : f \text{ messbar und } \|f\|_{L^p(\Omega)} < \infty\}.$$

Dass es sich bei $\|\cdot\|_{L^p(\Omega)}$ tatsächlich nur um Halbnormen handelt, zeigt folgendes Beispiel: Gegeben sei die Funktion $f : [0, 1] \rightarrow \mathbb{R}$ mit

$$f(x) = \begin{cases} 1, & \text{falls } x = 1, \\ 0, & \text{sonst.} \end{cases}$$

Diese ist offenbar nicht identisch verschwindend. Trotzdem ergibt sich $\|f\|_{L^p([0,1])} = 0$. Demnach können auch Funktionen, welche nicht identisch Null sind, durch $\|\cdot\|_{L^p(\Omega)}$ auf Null abgebildet werden.

Um aus den \mathcal{L}^p -Räumen die Lebesgue-Räume $L^p(\Omega)$ zu erhalten, unterteilt man die \mathcal{L}^p -Räume in Äquivalenzklassen.

Definition 2.9. Lebesgue-Räume

Sei

$$\mathcal{N}^p = \{f \in \mathcal{L}^p(\Omega) : f = 0 \text{ fast überall bezüglich } \mu\}.$$

Dann sind die Lebesgue-Räume definiert durch den Faktorraum

$$L^p(\Omega) = \mathcal{L}^p(\Omega) / \mathcal{N}^p.$$

Die Lebesgue-Räume sind entgegen den \mathcal{L}^p -Räumen normierte Räume. Die zugehörige Norm bildet ein Element von L^p nach \mathbb{R} ab, indem man in $\mathcal{L}^p(\Omega)$ einen beliebigen Repräsentanten f der jeweiligen Äquivalenzklasse wählt und dann $\|f\|_{L^p(\Omega)}$ auswertet. Gemäß der Konstruktion der Lebesgue-Räume ist die so definierte Norm wohldefiniert: Für zwei beliebige Funktionen f, g aus der gleichen Äquivalenzklasse gilt $\|f\|_{L^p(\Omega)} = \|g\|_{L^p(\Omega)}$. Weiterhin ist die Unterteilung in Äquivalenzklassen nicht zu grob in dem Sinne, dass zwei unterschiedliche stetige Funktionen nie der gleichen Äquivalenzklasse angehören. Diese Aussage kann noch verstärkt werden, da $L^p(\Omega) \cap C^\infty(\Omega)$ dicht in $C^\infty(\Omega)$ liegt (für $p = \infty$ gilt diese Aussage nicht).

Das folgende Lemma fasst einige Eigenschaften der Lebesgue-Räume zusammen:

Lemma 2.10. Sei $1 \leq p, q \leq \infty$.

(i) Für $f, g \in L^p(\Omega)$ gelten $f + g \in L^p(\Omega)$ und die *Minkowski-Ungleichung*:

$$\|f + g\|_{L^p(\Omega)} \leq \|f\|_{L^p(\Omega)} + \|g\|_{L^p(\Omega)}.$$

(ii) Für $f \in L^p(\Omega)$ und $g \in L^q(\Omega)$ mit $1/p + 1/q = 1$, ist $fg \in L^1(\Omega)$ und die *Hölder-Ungleichung* gilt:

$$\|fg\|_{L^1(\Omega)} \leq \|f\|_{L^p(\Omega)} \|g\|_{L^q(\Omega)}.$$

(iii) Es gilt $L^p(\Omega) \subset L^q(\Omega)$ für $p \geq q$, falls Ω beschränkt ist.

(iv) Die Lebesgue-Räume sind vollständig.

(v) Für die Dualräume der Lebesgue-Räume gilt:

$$\begin{aligned} (L^p(\Omega))' &= L^q(\Omega) \quad \text{für } 1 < p, q < \infty, \frac{1}{p} + \frac{1}{q} = 1, \\ (L^1(\Omega))' &= L^\infty(\Omega), \\ (L^\infty(\Omega))' &\neq L^1(\Omega). \end{aligned}$$

Insbesondere sind also die Lebesgue-Räume mit $1 < p < \infty$ reflexiv, die Räume $L^1(\Omega)$ und $L^\infty(\Omega)$ hingegen nicht.

Beweis.

siehe (Ada75) Kapitel 2 \square

Eine besondere Rolle unter den Lebesgue-Räumen spielt der Raum $L^2(\Omega)$. Auf diesem kann durch

$$(f, g)_{L^2(\Omega)} = \int_{\Omega} fg \, d\mu$$

ein Skalarprodukt definiert werden. Damit ist $L^2(\Omega)$ ein Hilbertraum. Das L^2 -Skalarprodukt kürzen wir im Folgenden mit (\cdot, \cdot) ab.

Um nun das Konzept der schwachen Ableitung einzuführen, benötigen wir noch einen weiteren Funktionenraum:

Definition 2.11.

Der Raum der lokal integrierbaren Funktionen ist definiert durch

$$L^1_{\text{loc}}(\Omega) = \left\{ f : \Omega \rightarrow \mathbb{R} \mid f \text{ messbar und } \forall K \subset \Omega \text{ kompakt: } \int_K |f| \, d\mu < \infty \right\}.$$

Damit können wir nun definieren, was wir unter der schwachen Ableitung einer Funktion verstehen:

Definition 2.12. (Schwache Ableitung)

Sei Ω ein beschränktes Gebiet in \mathbb{R}^d , $d \in \mathbb{N}$, $f \in L^1_{\text{loc}}(\Omega)$ und α ein Multiindex. Falls eine Funktion $g \in L^1_{\text{loc}}(\Omega)$ existiert, so dass

$$\int_{\Omega} f D^\alpha v \, d\mu = (-1)^{|\alpha|} \int_{\Omega} gv \, d\mu$$

für alle v aus $C_0^\infty(\Omega)$ gilt, so heißt g die schwache Ableitung von f zum Multiindex α . Hierbei bezeichnet D die klassische Ableitung und $C_0^\infty(\Omega)$ den Raum aller beliebig oft differenzierbaren Funktionen mit kompaktem Träger in Ω .

Analog zur klassischen Ableitung schreiben wir $D^\alpha f$ für die schwache Ableitung von f zum Multiindex α . Dies stellt kein Problem dar, da die Definition so gewählt ist, dass eine im klassischen Sinne differenzierbare Funktion auch schwach ableitbar ist und dass ihre schwache Ableitung fast überall mit ihrer klassischen übereinstimmt. Natürlich gibt es aber durchaus Funktionen, welche zwar schwach, jedoch nicht klassisch differenzierbar sind. Dies führt uns auf die Definition der Sobolev-Räume.

Lemma 2.13. (Sobolev–Räume)

Seien $k \in \mathbb{N}_0$, $1 \leq p \leq \infty$ und Ω ein beschränktes Gebiet in \mathbb{R}^d , $d \in \mathbb{N}$. Die Sobolev–Räume sind definiert durch

$$W^{k,p}(\Omega) = \{f : \Omega \rightarrow \mathbb{R} : \text{für alle } \alpha \text{ mit } |\alpha| \leq k \text{ ist } D^\alpha f \in L^p(\Omega)\}.$$

Die Sobolev–Räume werden vermittels folgender Normen zu Banachräumen:

$$\|f\|_{W^{k,p}(\Omega)} = \left(\sum_{|\alpha| \leq k} (\|D^\alpha f\|_{L^p(\Omega)})^p \right)^{1/p} \quad \text{für } 1 \leq p < \infty,$$

$$\|f\|_{W^{k,\infty}(\Omega)} = \sum_{|\alpha| \leq k} \|D^\alpha f\|_{L^\infty(\Omega)}.$$

Die Sobolev–Räume mit $p = 2$ sind Hilberträume bezüglich des Skalarprodukts

$$(\cdot, \cdot)_{W^{k,2}(\Omega)} : W^{k,2}(\Omega) \times W^{k,2}(\Omega) \rightarrow \mathbb{R}$$

$$(u, v)_{W^{k,2}(\Omega)} = \sum_{|\alpha| \leq k} (D^\alpha u, D^\alpha v).$$

Den Raum $W^{k,2}(\Omega)$ kürzen wir im Folgenden durch $H^k(\Omega)$ ab.

Beweis.

Dass es sich bei $\|f\|_{W^{k,p}(\Omega)}$ um eine Norm handelt, rechnet man nach. Dass der Raum $W^{k,p}(\Omega)$ versehen mit dieser Norm vollständig ist, findet sich beispielsweise in (Ada75). \square

Oft wird $W^{k,p}(\Omega)$ auch als Abschluss von $C^\infty(\Omega)$ bezüglich der $\|\cdot\|_{W^{k,p}(\Omega)}$ -Norm definiert. Beide Definitionen sind unabhängig vom Gebiet Ω äquivalent zueinander.

Neben den $\|\cdot\|_{W^{k,p}(\Omega)}$ -Normen treten des öfteren folgende Halbnormen auf:

$$|f|_{W^{k,p}(\Omega)} = \left(\sum_{|\alpha|=k} (\|D^\alpha f\|_{L^p(\Omega)})^p \right)^{1/p} \quad \text{für } 1 \leq p < \infty,$$

$$|f|_{W^{k,\infty}(\Omega)} = \sum_{|\alpha|=k} \|D^\alpha f\|_{L^\infty(\Omega)}.$$

Wir wollen noch einige Eigenschaften der Sobolev–Räume angeben. Dazu brauchen wir zunächst den Begriff der Einbettung.

Definition 2.14. (Einbettung)

Seien X, Y zwei normierte Vektorräume. Wir sagen, X ist in Y eingebettet, falls gilt:

- (i) X ist ein Untervektorraum von Y .
- (ii) Der Identitätsoperator $I : X \rightarrow Y$, $Iv = v$ für alle $v \in X$ ist stetig.

Eine Einbettung zeigen wir durch $X \hookrightarrow Y$ an.

Für die Sobolev–Räume ergibt sich daraus sofort, dass $W^{k,p}(\Omega) \hookrightarrow W^{l,q}(\Omega)$ äquivalent ist zu $W^{k,p}(\Omega) \subset W^{l,q}(\Omega)$. Betrachtet man hingegen Einbettungen von Sobolev–Räumen

in klassische Funktionenräume, so muss man berücksichtigen, dass Elemente der Sobolev-Räume keine Funktionen sondern Äquivalenzklassen von Funktionen sind. Wir schreiben $W^{k,p}(\Omega) \hookrightarrow C^l(\Omega)$, falls es zu jeder Äquivalenzklasse in $W^{k,p}(\Omega)$ einen Repräsentanten aus $C^l(\Omega)$ gibt und falls der Identitätsoperator die Menge dieser Repräsentanten stetig nach $C^l(\Omega)$ abbildet. Einige Aussagen zu solchen Einbettungen liefert der folgende Satz:

Satz 2.15. (Sobolevsche Einbettungstheoreme und Dichtheit)

Sei $\Omega \subset \mathbb{R}^d$, $d \in \mathbb{N}$, ein beschränktes Gebiet mit Lipschitz-Rand. Seien ferner $j, m \in \mathbb{N}_0$ und $1 \leq p, q < \infty$. Dann gelten:

(i) Für $mp < d$ und $q \leq \frac{dp}{d-mp}$:

$$W^{j+m,p}(\Omega) \hookrightarrow W^{j,q}(\Omega).$$

(ii) Für $mp = d$:

$$W^{j+m,p}(\Omega) \hookrightarrow W^{j,q}(\Omega).$$

Ist speziell $p = 1$ und damit $m = d$, so gilt die Einbettung auch für $q = \infty$.

(iii) Für $mp > d$:

$$W^{j+m,p}(\Omega) \hookrightarrow C^j(\Omega).$$

Weiterhin sei $C_0^k(\Omega)$ der Raum der k -mal stetig differenzierbaren Funktionen mit kompaktem Träger in Ω . Für $k \geq l$ liegt $C_0^k(\Omega)$ dicht in $H^l(\Omega)$.

Beweis.

Die Einbettungstheoreme werden beispielsweise in (Ada75) bewiesen. Hier werden auch Verallgemeinerungen bezüglich des Gebiets besprochen. Der Beweis der Dichtheitsaussage findet sich in (Wlo82) Satz 3.6. \square

Wichtige Spezialfälle der Einbettungstheoreme sind $H^2(\Omega) \subset C(\Omega)$ für $d = 2, 3$ und $W^{1,1}(\Omega) \subset L^2(\Omega)$ für $d = 2$.

Die Tatsache, dass die Elemente der Sobolev-Räume Äquivalenzklassen sind, bereitet noch ein Problem. Da der Rand $\partial\Omega$ eines Gebietes $\Omega \subset \mathbb{R}^d$ bezüglich des d -dimensionalen Lebesgue-Maßes eine Nullmenge darstellt, gehören Funktionen mit verschiedenen Randwerten derselben Äquivalenzklasse an, so dass es keinen Sinn mehr macht, Bedingungen an die Randwerte im klassischen Sinne zu stellen. Um trotzdem Randbedingungen stellen zu können, bieten sich zwei Möglichkeiten an:

Definition 2.16.

Der Sobolev-Raum $W_0^{k,p}(\Omega)$ ist definiert als der Abschluss von $C_0^\infty(\Omega)$ bezüglich der Norm $\|\cdot\|_{W^{k,p}}$. Für $p = 2$ schreiben wir wieder $H_0^k(\Omega)$.

Die Funktionen $u \in W_0^{k,p}(\Omega)$ erfüllen die homogenen Dirichlet-Randbedingungen

$$u = 0 \quad \text{auf } \partial\Omega$$

in dem Sinne, dass es eine Folge $u_n \in C_0^\infty(\Omega)$ gibt, welche bezüglich $\|\cdot\|_{W^{k,p}(\Omega)}$ gegen u konvergiert und für deren Elemente insbesondere $u_n|_{\partial\Omega} = 0$ gilt.

Wie für die Sobolev-Räume gilt auch hier, dass $W_0^{k,p}(\Omega)$ ein Banachraum und H_0^k ein Hilbertraum ist.

Zudem führen die Sobolev-Räume $W_0^{k,p}(\Omega)$ auf die Definition von Sobolev-Räumen mit negativem Exponenten k , was sich in folgendem Lemma zeigt:

Lemma 2.17.

Seien Ω ein beschränktes Gebiet und $1 < p, q < \infty$ mit $1/p + 1/q = 1$. Der Raum $W^{-k,q}(\Omega)$, $k \in \mathbb{N}_0$, ist definiert durch

$$W^{-k,q}(\Omega) := \{u' \in (C_0^\infty(\Omega))' : \|u'\|_{W^{-k,q}(\Omega)} < \infty\},$$

wobei $(C_0^\infty(\Omega))'$ der Dualraum von $C_0^\infty(\Omega)$ ist und

$$\|u'\|_{W^{-k,q}(\Omega)} := \sup_{0 \neq u \in C_0^\infty(\Omega)} \frac{u'(u)}{\|u\|_{W^{k,p}(\Omega)}}.$$

Dann ist $W^{-k,q}(\Omega)$ der Dualraum von $W_0^{k,p}(\Omega)$.

Um auch inhomogene Dirichlet-Randbedingungen

$$u = g \quad \text{auf } \partial\Omega$$

betrachten zu können, ist es notwendig zu definieren, was man unter den Randwerten einer Funktion $u \in W^{k,p}(\Omega)$ versteht. Dazu führen wir den Spuroperator ein:

Satz 2.18. (Spuroperator und Spur einer Sobolev-Funktion)

Sei $\Omega \subset \mathbb{R}^d$ ein Gebiet mit Lipschitz-Rand $\partial\Omega$ und $2 \leq p \leq \infty$. Dann gibt es genau einen linearen, stetigen Operator $\gamma : W^{k,p}(\Omega) \rightarrow L^2(\partial\Omega)$, der die Randwerte von Funktionen u aus $W^{1,p} \cap C(\bar{\Omega})$ unverändert lässt:

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

γ wird Spuroperator genannt.

Die Spur einer Funktion u aus $W^{k,p}(\Omega)$, $k \geq 1$ ist dann definiert als

$$\lim_{n \rightarrow \infty} (\gamma u_n) = \gamma u,$$

wobei $\{u_n\}_{n=1}^\infty$ eine Folge aus $C(\bar{\Omega})$ mit Grenzwert u sei. Die Zuordnung $u \mapsto \gamma u$ ist eindeutig.

Beweis.

Existenz und Eindeutigkeit des Spuroperators γ werden in (Wlo82) Satz 8.7. bewiesen. Die Existenz einer Folge $\{u_n\}_{n=1}^\infty$ aus $C(\bar{\Omega})$ mit Grenzwert u folgt aus der Dichtheitsaussage in Satz 2.15. Die Existenz des Spuroperators liefert dann die Existenz der Spur. Die Eindeutigkeit der Spur folgt aus der Eindeutigkeit und Stetigkeit des Spuroperators. \square

Bemerkung 2.19.

Definition 2.16 und Satz 2.18 sind miteinander vereinbar, da sich

$$\gamma u = 0 \quad \text{auf } \partial\Omega \quad \forall u \in W_0^{k,p}(\Omega)$$

ergibt. Für $k \geq 2$ kann man zusätzlich zeigen, dass auch

$$\gamma D^\alpha u = 0 \quad \text{auf } \partial\Omega \quad \forall u \in W_0^{k,p}(\Omega) \quad \text{mit } |\alpha| \leq k - 1$$

gilt.

Bemerkung 2.20.

Neben Sobolev-Räumen mit ganzzahligen positiven oder negativen Exponenten können auch Sobolev-Räume mit nichtganzzahligen Exponenten, die so genannten Sobolev-Slobodecki-Räume definiert werden:

$$H^{k+s}(\Omega) = \{f \in L^2(\Omega) : D^\alpha f \in L^2(\Omega) \text{ für } |\alpha| \leq k \text{ und } |D^\alpha f|_s < \infty\},$$

mit $k \in \mathbb{N}$, $0 < s < 1$, $\Omega \subset \mathbb{R}^d$ und

$$|f|_s = \int_{\Omega} \int_{\Omega} \frac{|f(x) - f(y)|^2}{|x - y|^{2s+d}} dx dy.$$

Diese treten zum Beispiel im Beweis zu Satz 2.18 in (Wlo82) auf, der zeigt, dass der Spuroperator eine Funktion aus $H^k(\Omega)$ auf den Funktionenraum $H^{k-1/2}(\partial\Omega)$ abbildet. Somit besitzt die Spur eigentlich eine höhere Regularität als in Satz 2.18 angegeben.

Gemäß (Wlo82) Satz 3.1 ist $H^{k+s}(\Omega)$ ein Hilbertraum. Für weitere Resultate zu den Sobolev-Slobodecki-Räumen siehe (Wlo82) oder auch (Tre75).

Weitere Aussagen über Sobolev-Räume liefern die folgenden Ungleichungen:

Lemma 2.21. (Friedrichs-Ungleichungen)

Seien $\Omega \subset \mathbb{R}^d$, $d \in \mathbb{N}$, ein beschränktes Gebiet und $1 \leq p \leq \infty$. Dann gilt

(i) für alle $u \in W_0^{k,p}(\Omega)$

$$\|u\|_{W^{k,p}(\Omega)} \leq C_F |u|_{W^{k,p}(\Omega)}, \tag{2.2.1}$$

(ii) für alle $u \in H^1(\Omega)$

$$\|u\|_{L^2(\Omega)}^2 \leq C^2 (\|\nabla u\|_{L^2(\Omega)} + \|u\|_{L^2(\Gamma)}), \tag{2.2.2}$$

$$\|u\|_{L^2(\Omega)}^2 \leq \tilde{C}^2 \left(\|\nabla u\|_{L^2(\Omega)} + \left(\int_{\Gamma} |u| dx \right)^2 \right), \tag{2.2.3}$$

mit einer Konstanten $C_F > 0$, die nur von p , d und dem Durchmesser von Ω abhängt und Konstanten C, \tilde{C} die von Ω abhängen. C_F, C und \tilde{C} werden auch als Friedrichs-Konstanten bezeichnet.

Beweis.

Siehe (Ada75). \square

Die Friedrichs-Ungleichungen gewährleisten insbesondere, dass für $u \in W_0^{k,p}(\Omega)$ mit $u \neq 0$ die Abschätzung $|u|_{W^{k,p}(\Omega)} > 0$ gilt. Damit ist $|\cdot|_{W^{k,p}(\Omega)}$ sogar eine Norm auf $W_0^{k,p}(\Omega)$.

Bevor wir nun von der starken zur schwachen Formulierung einer Differentialgleichung übergehen, benötigen wir noch einige Aussagen aus der Funktionalanalysis.

2.2.2 Einige Aussagen aus der Funktionalanalysis

Wir wollen hier zunächst einige Begriffe und Sätze aus der Funktionalanalysis in abstrakter Form rekapitulieren.

Lemma 2.22. (Operatornorm)

Seien $(V, \|\cdot\|_V)$ und $(Q, \|\cdot\|_Q)$ normierte Vektorräume und sei weiter $\mathcal{L}(V, Q)$ der Raum aller linearen, stetigen Operatoren, die von V nach Q abbilden. Für $T \in \mathcal{L}(V, Q)$ wird durch

$$\|T\| := \sup_{0 \neq v \in V} \frac{\|Tv\|_Q}{\|v\|_V}$$

eine Norm auf $\mathcal{L}(V, Q)$ definiert. Diese Norm heißt Operatornorm.

Mit der Operatornorm wird $\mathcal{L}(V, Q)$ zum normierten Raum. Ist Q zusätzlich ein Banachraum, so auch $\mathcal{L}(V, Q)$.

Sei weiterhin $S \in \mathcal{L}(Q, V)$. Dann gilt

$$\|TS\| \leq \|T\|\|S\|,$$

das heißt die Operatornorm ist submultiplikativ.

Beweis.

Siehe (Wer95) Seite 47f. \square

Auch für Bilinearformen lässt sich eine Norm definieren:

Lemma 2.23.

Seien $(V, \|\cdot\|_V)$ und $(Q, \|\cdot\|_Q)$ normierte Vektorräume und $a(\cdot, \cdot) : V \times Q \rightarrow \mathbb{R}$ eine stetige Bilinearform. Dann ist durch

$$\|a\| := \sup_{0 \neq v \in V} \sup_{0 \neq q \in Q} \frac{a(v, q)}{\|v\|_V \|q\|_Q}$$

eine Norm auf $\mathcal{L}(V \times Q, \mathbb{R})$ definiert.

Beweis.

Der Beweis folgt schnell unter Verwendung der Stetigkeit von a und der Normen sowie der Dreiecksungleichung für $\|\cdot\|_V$ beziehungsweise $\|\cdot\|_Q$. \square

Die Beschränktheit von a ist offenbar äquivalent zur Aussage $\|a\| = C < \infty$. Um eine Verknüpfung zwischen stetigen Bilinearformen und stetigen linearen Operatoren zu schaffen, betrachten wir den Dualraum V' von V . Das duale Produkt zwischen $v \in V$ und $v' \in V'$ ist gegeben durch

$$\langle v', v \rangle := v'(v).$$

Damit können wir das folgende Lemma formulieren:

Lemma 2.24.

Seien $(V, \|\cdot\|_V)$ und $(Q, \|\cdot\|_Q)$ normierte Vektorräume und $a(\cdot, \cdot) : V \times Q \rightarrow \mathbb{R}$ eine stetige Bilinearform. Dann gibt es genau einen stetigen linearen Operator $A : V \rightarrow Q'$, so dass für alle $v \in V$ und alle $q \in Q$ gilt:

$$a(v, q) = \langle Av, q \rangle_Q. \tag{2.2.4}$$

Den Operator A nennen wir den Darstellungsoperator von a . Die Normen von A und a sind gleich, sprich $\|A\| = \|a\|$.

Beweis.

Man definiert $A : V \rightarrow Q'$, indem man für das Bild Av fordert, dass $(Av)(q) = a(v, q)$ gilt. Damit ist A wohldefiniert. Aus der Stetigkeit und Linearität von a folgt, dass auch A stetig und linear sein muss. Die Eindeutigkeit von A folgt dann aus (2.2.4) mittels eines Widerspruchsbeweises. \square

Mit dem Konzept des Dualraumes lässt sich auch der duale Operator T' zum Operator T definieren:

Lemma 2.25. (Dualer Operator)

Seien $(V, \|\cdot\|_V)$ und $(Q, \|\cdot\|_Q)$ normierte Vektorräume. Zu einem Operator $T \in \mathcal{L}(V, Q)$ gibt es genau einen Operator $T' : Q' \rightarrow V'$ mit

$$\langle T'q', v \rangle_V = \langle q', Tv \rangle_Q \quad \forall v \in V \text{ und } \forall q' \in Q'.$$

Den Operator $T' : Q' \rightarrow V'$ nennen wir den zu T dualen Operator (auch adjungierter Operator oder Adjungierte). Der duale Operator ist stetig und linear.

Beweis.

Siehe (Yos65) Seite 193 bis 195 oder (Alt85) Seite 262. \square

Ist der Operator T nur auf einer Teilmenge $D(T) \subsetneq V$ definiert, so gilt obiges Lemma immer noch, wenn der Definitionsbereich des dualen Operators eingeschränkt wird auf

$$D(T') := \{q' \in Q' : v \mapsto \langle q', Tv \rangle \text{ ist stetig auf } V\}.$$

Wir wollen noch einige weitere Eigenschaften des dualen Operators angeben.

Lemma 2.26.

Seien $(V, \|\cdot\|_V)$ und $(Q, \|\cdot\|_Q)$ normierte Vektorräume. Der Operator $T \in \mathcal{L}(V, Q)$ und sein dualer Operator T' besitzen die gleiche Operatornorm.

Sind $(V, \|\cdot\|_V)$ und $(Q, \|\cdot\|_Q)$ Banachräume, so existiert die Inverse $T^{-1} \in \mathcal{L}(V, Q)$ genau dann, wenn der duale Operator eine Inverse $(T')^{-1} \in \mathcal{L}(V', Q')$ besitzt. Zudem gilt

$$(T^{-1})' = (T')^{-1}.$$

Beweis.

Siehe (Alt85) Seite 264 und (Yos65) Seite 195. \square

Für zwei Vektorräume V und Q und einen Operator $T : V \rightarrow Q$ bezeichne $\ker(T)$ den Kern des Operators T und $R(T)$ seinen Bildbereich.

Satz 2.27. (Satz vom abgeschlossenen Bild)

Seien $(V, \|\cdot\|_V)$ und $(Q, \|\cdot\|_Q)$ Banachräume und T ein stetiger Operator von V nach Q mit einer Definitionsmenge, welche dicht in V liegt. Dann sind die folgenden Aussagen äquivalent:

- (i) $R(T)$ ist abgeschlossen in Q .
- (ii) $R(T')$ ist abgeschlossen in V' .
- (iii) $R(T) = \ker(T')^\circ = \{q \in Q : \langle q', q \rangle = 0 \text{ für alle } q' \in \ker(T')\}$.
- (iv) $R(T') = \ker(T)^\circ = \{v \in V' : \langle v', v \rangle = 0 \text{ für alle } v \in \ker(T)\}$.

$\ker(T')^\circ$ und $\ker(T)^\circ$ bezeichnet man als Annihilatoren von $\ker(T')$ und $\ker(T)$.

Beweis.

Siehe (Yos65), Seite 205 bis 207. \square

Satz 2.28. (Projektionssatz)

Seien V ein Hilbertraum mit Skalarprodukt a , U ein abgeschlossener Unterraum von V und U^\perp das orthogonale Komplement von U , also

$$U^\perp = \{v \in V : a(v, u) = 0 \quad \forall u \in U\}.$$

Dann ist auch U^\perp ein abgeschlossener Unterraum von V und jedes $v \in V$ kann in eindeutiger Weise zerlegt werden in

$$v = u + u^\perp \quad \text{mit } u \in U \text{ und } u^\perp \in U^\perp.$$

Sei über diese Zerlegung der Operator $\pi : V \rightarrow U$ definiert durch $\pi(v) = u$. Dann ist π ein linearer stetiger Operator mit den Eigenschaften:

$$\begin{aligned} \pi^2 &= \pi && (\pi \text{ ist idempotent}), \\ a(\pi v, u) &= a(v, \pi u) \quad \forall u, v \in V && (\pi \text{ ist symmetrisch}). \end{aligned}$$

Man bezeichnet π als den Projektionsoperator oder die Projektion auf U .

Beweis.

Siehe (Yos65) Seite 82 und 83. \square

Zum Abschluss dieses Abschnittes führen wir noch eine Ungleichung an, welche wir im Laufe dieser Arbeit verwenden werden.

Lemma 2.29. (Young–Ungleichungen)

Seien $a, b \in \mathbb{R}_0^+$. Dann gelten die Youngschen Ungleichungen:

$$\begin{aligned} ab &\leq \frac{1}{2\delta} a^2 + \frac{\delta}{2} b^2 \text{ für alle } \delta > 0, \\ ab &\leq \frac{1}{p} a^p + \frac{1}{q} b^q \text{ für alle } 1 < p, q < \infty \text{ mit } \frac{1}{p} + \frac{1}{q} = 1. \end{aligned}$$

Auch die zweite Ungleichung kann dabei ähnlich wie die erste mit einem $\delta > 0$ skaliert werden.

Beweis.

Die erste Ungleichung rechnet man elementar nach.

Zum Beweis der zweiten Ungleichung unterscheiden wir zwei Fälle:

(a) $a^p = b^q$:

In diesem Fall ist

$$ab = ab^{\frac{a}{b}} = a(b^q)^{\frac{1}{q}} = a(a^p)^{\frac{1}{q}} = a^{\frac{p}{q}} a^{\frac{p}{q}} = a^{p(\frac{1}{q} + \frac{1}{q})} = a^p = \left(\frac{1}{p} + \frac{1}{q}\right) a^p = \frac{a^p}{p} + \frac{a^p}{q} = \frac{a^p}{p} + \frac{b^q}{q}.$$

(b) $a^p \neq b^q$:

Es gilt:

$$\ln(ab) = \ln(a) + \ln(b) = \frac{\ln(a^p)}{p} + \frac{\ln(b^q)}{q}.$$

Auf Grund der strengen Monotonie des natürlichen Logarithmus ist $\ln(a^p) \neq \ln(b^q)$. Mit der strikten Konvexität der Exponentialfunktion erhält man dann

$$ab = \exp(\ln(ab)) = \exp\left(\frac{\ln(a^p)}{p} + \frac{\ln(b^q)}{q}\right) < \frac{\exp(\ln(a^p))}{p} + \frac{\exp(\ln(b^q))}{q} = \frac{a^p}{p} + \frac{b^q}{q}. \quad \square$$

2.2.3 Formulierung als Variationsproblem

An dieser Stelle soll erläutert werden, wie man prinzipiell von der starken zur schwachen Formulierung einer Differentialgleichung gelangt. Die konkrete Umsetzung ist dabei immer von der Art der Differentialgleichung abhängig (siehe zum Beispiel Kapitel 4).

Als Ausgangspunkt wählen wir eine Differentialgleichung der Form

$$\mathcal{L}u(x) = 0, \tag{2.2.5}$$

wobei \mathcal{L} ein Differentialoperator wie in Definition 2.5 sein soll. Als Randbedingungen wählen wir hier homogene Dirichlet-Randbedingungen auf dem gesamten Gebietsrand $\partial\Omega$. Gesucht ist eine Lösung $u \in H^1(\Omega)$ von (2.2.5). Um eine geeignete Gleichung zur Bestimmung einer solchen Lösung aufzustellen, multiplizieren wir die PDGl (2.2.5) mit einer beliebigen Testfunktion $v \in H_0^1(\Omega)$ und integrieren über das Gebiet Ω . Unter der Annahme, dass $\mathcal{L}u \in H^{-1}(\Omega)$ ergibt sich

$$\langle \mathcal{L}u, v \rangle = 0. \tag{2.2.6}$$

Üblicherweise wird hier noch partielle integriert, um Ableitungen von der Lösung zu entfernen und auf die Testfunktion zu übertragen. Dadurch sind die Ansprüche an die Differenzierbarkeit von u geringer als an die klassische Lösung. Des Weiteren wird diese Gleichung noch umsortiert, indem alle Ausdrücke, die sowohl u als auch v enthalten auf der linken Seite in einer Bilinearform $a(u, v)$ und alle Ausdrücke, welche nur v und nicht u enthalten in einer Linearform $\tilde{f}(v)$ auf der rechten Seite zusammengefasst werden. Als schwache Formulierung der Differentialgleichung ergibt sich dann:

Finde $u \in H^1(\Omega)$ mit $u = 0$ auf $\partial\Omega$ und

$$a(u, v) = \tilde{f}(v) \quad \forall v \in H^1(\Omega). \tag{2.2.7}$$

Gemäß dieser Herleitung ist jede klassische Lösung von (2.2.5) auch Lösung der schwachen Formulierung (2.2.7). Die Umkehrung gilt im Allgemeinen nicht. Aussagen über die Lösbarkeit von (2.2.7) hängen von den Eigenschaften der Bilinearform $a(\cdot, \cdot)$ ab.

Definition 2.30. (Stetigkeit und Koerzitivität von Bilinearformen)

Seien V ein reeller Hilbertraum mit Skalarprodukt $(\cdot, \cdot)_V$ sowie induzierter Norm $\|\cdot\|_V = (\cdot, \cdot)_V^{1/2}$ und $a : V \times V \rightarrow \mathbb{R}$ eine Bilinearform. a heißt

- (i) stetig, falls es eine Konstante $C > 0$ gibt, so dass $a(u, v) \leq C\|u\|_V\|v\|_V \quad \forall u, v \in V$,
- (ii) strikt koerzitiv, falls es eine Konstante $\tilde{C} > 0$ gibt, so dass $a(u, u) \geq \tilde{C}\|u\|_V^2 \quad \forall u \in V$.

Sei $a(\cdot, \cdot)$ eine stetige, strikt koerzitive Bilinearform auf V und sei weiter $f \in V'$. Unter einem Variationsproblem versteht man dann die Aufgabe:

Finde ein $u \in V$, so dass für alle $v \in V$ gilt:

$$a(u, v) = f(v). \tag{2.2.8}$$

Ist a eine symmetrische, positiv (negativ) definite Bilinearform, also ein Skalarprodukt, so folgt die Lösbarkeit des Variationsproblems aus dem Rieszschen Darstellungssatz:

Satz 2.31. (Rieszscher Darstellungssatz)

Sei V ein Hilbertraum und a ein Skalarprodukt auf V . Zu jedem beschränkten linearen Funktional $f \in V'$ gibt es ein eindeutig bestimmtes $u \in V$ mit

$$a(u, v) = \langle f, v \rangle \quad \forall v \in V.$$

Die Abbildung $R_V : V' \rightarrow V, f \mapsto u$ ist ein isometrischer Isomorphismus. R_V wird Rieszscher Darstellungsoperator genannt.

Beweis.

Siehe (Yos65) Seite 90. \square

Bemerkung 2.32.

Oftmals wird statt des obigen Problems auch folgende Aufgabenstellung als Variationsproblem definiert:

Gegeben sei das lineare Funktional

$$J(v) := \frac{1}{2}\|v\|_V^2 - f(v) \quad \text{mit } \|v\|_V^2 = a(v, v).$$

Finde u aus V , so dass gilt:

$$J(u) = \min_{0 \neq v \in V} J(v).$$

Den Beweis, dass diese Fragestellung zu (2.2.8) äquivalent ist, findet man beispielsweise in (Cia78) Theorem 1.1.2.

Eine Verallgemeinerung für nicht-symmetrische Bilinearformen liefert der folgende Satz:

Satz 2.33. (Satz von Lax–Milgram)

Sei $a(\cdot, \cdot) : V \times V \rightarrow \mathbb{R}$ eine beschränkte und bezüglich $\|\cdot\|_V$ strikt koerzitive Bilinearform. Zu jedem beschränkten linearen Funktional $f \in V'$ gibt es ein eindeutig bestimmtes $u \in V$ mit

$$a(u, v) = \langle f, v \rangle \quad \forall v \in V.$$

Beweis.

Der Beweis erfolgt mit Hilfe des Rieszchen Darstellungssatzes, siehe (Yos65) Seite 92. \square

Kapitel 3

Galerkin–Verfahren und Finite–Elemente–Methoden

Eine Lösung der im vorherigen Kapitel vorgestellten Variationsprobleme lässt sich analytisch oft nicht oder nur mit erheblichem Aufwand bestimmen. Stattdessen werden häufig numerische Verfahren benutzt, um eine Näherungslösung zu bestimmen. Wir wollen in dieser Arbeit eine Variante der so genannten Finite–Elemente–Methoden (FEM) behandeln. Die FEM beruhen auf dem Galerkin–Verfahren. Beide sollen in diesem Kapitel vorgestellt werden.

3.1 Galerkin–Verfahren

Wir betrachten folgendes Variationsproblem:

Sei V ein Hilbertraum mit Norm $\|\cdot\|_V$ und sei $f \in V'$. Finde $u \in V$, so dass für alle $v \in V$ gilt:

$$a(u, v) = f(v). \quad (3.1.1)$$

Falls $a : V \times V \rightarrow \mathbb{R}$ stetig und bezüglich der Norm $\|\cdot\|_V$ strikt koerzitiv ist, so besitzt dieses Variationsproblem, wie bereits erwähnt, nach dem Satz von Lax–Milgram 2.33 eine eindeutige Lösung.

Für die numerische Behandlung besteht das Problem darin, dass der Raum V meist unendlich-dimensional ist. Die Idee der Galerkin–Verfahren ist, eine Näherungslösung u_h zu bestimmen, welche in einem endlich-dimensionalen Teilraum $V_h \subset V$ liegt. Das Variationsproblem wird dann zu:

Finde $u_h \in V_h$, so dass für alle $v_h \in V_h$ gilt:

$$a(u_h, v_h) = f(v_h). \quad (3.1.2)$$

Auch dieses Problem besitzt nach dem Satz von Lax–Milgram eine eindeutige Lösung, da a auch im Unterraum V_h koerzitiv ist und endlich-dimensionale Unterräume von Hilberträumen wieder Hilberträume sind.

Um aus (3.1.2) ein endlich-dimensionales lineares Gleichungssystem zu bestimmen, wählen wir eine beliebige Basis $\{\phi_i\}_{i=1}^N$ von V_h . Da sich alle Funktionen $v_h \in V_h$ als Linearkombination der Basisfunktionen darstellen lassen, ist auf Grund der Linearität von a und f in (3.1.2) unmittelbar klar, dass es schon genügt, wenn (3.1.2) für alle Basisfunktionen $\{\phi_i\}_{i=1}^N$ erfüllt ist. Des Weiteren lässt sich auch die gesuchte Lösung u_h als Linearkombination der Basisfunktionen schreiben:

$$u_h = \sum_{j=1}^N u_j \phi_j.$$

Setzen wir dies alles in (3.1.2) ein, so erhalten wir

$$a\left(\sum_{j=1}^N u_j \phi_j, \phi_i\right) = f(\phi_i) \quad \text{für } i = 1, \dots, N$$

und unter Ausnutzung der Linearität von a

$$\sum_{j=1}^N u_j a(\phi_j, \phi_i) = f(\phi_i) \quad \text{für } i = 1, \dots, N.$$

Mit der Notation

$$\mathbf{A} := \{a(\phi_j, \phi_i)\}_{i,j=1}^N, \quad \mathbf{u} := (u_i)_{i=1}^N \quad \text{und} \quad \mathbf{f} := (f(\phi_i))_{i=1}^N, \quad (3.1.3)$$

ergibt sich die äquivalente Formulierung

$$\mathbf{A}\mathbf{u} = \mathbf{f}. \quad (3.1.4)$$

Damit haben wir ein lineares Gleichungssystem erhalten, aus dem sich die unbekanntenen Koeffizienten u_i und damit die gesuchte Lösung u_h berechnen lassen. Die hier auftretende Matrix \mathbf{A} ist positiv definit, was eine Folge der strikten Koerzitivität von a ist.

Wie bei jedem Näherungsverfahren stellt sich auch hier die Frage, wie gut die tatsächliche Lösung u durch die Näherungslösung u_h approximiert werden kann. Dies ist natürlich von den Räumen V und V_h abhängig. Allerdings lassen sich auch ohne eine nähere Spezifizierung der Räume schon einige Ergebnisse dazu formulieren:

Lemma 3.1.

Seien u die Lösung von (3.1.1) und u_h die zugehörige Näherungslösung eines Galerkin–Verfahrens (3.1.2). Ist die Bilinearform a stetig, symmetrisch und strikt koerzitiv bezüglich der von a induzierten Energienorm $\|\cdot\|_a = a(\cdot, \cdot)^{1/2}$, so gilt:

$$\|u - u_h\|_a = \min_{v_h \in V_h} \|u - v_h\|_a. \quad (3.1.5)$$

Beweis.

Sei v_h eine beliebige Testfunktion aus V_h . Setzt man diese in (3.1.1) und (3.1.2) ein und subtrahiert dann (3.1.2) von (3.1.1), so ergibt sich:

$$a(u - u_h, v_h) = 0 \quad \forall v_h \in V_h. \quad (3.1.6)$$

Daraus folgt für alle $v_h \in V_h$

$$\|u - u_h\|_a^2 = a(u - u_h, u - u_h) = a(u - u_h, u - v_h) \leq \|u - u_h\|_a \|u - v_h\|_a. \quad (3.1.7)$$

Die zweite Identität ergibt sich aus (3.1.6) und die Abschätzung nach oben aus der Cauchy–Schwarz–Ungleichung. Division durch $\|u - u_h\|_a$ und Übergang zum Infimum liefern die Behauptung. \square

Bemerkung 3.2.

Der im Beweis auftretende Zusammenhang

$$a(u - u_h, v_h) = 0 \quad \forall v_h \in V_h \quad (3.1.8)$$

wird als Galerkin-Orthogonalität bezeichnet und ist auch für weitere Approximationsätze nützlich. Galerkin-Orthogonalität ist auch dann gegeben, wenn auf die Symmetrie von a verzichtet wird.

Nicht immer ist die in Lemma 3.1 geforderte Symmetrie von a gegeben. In diesem Fall liefert das folgende Lemma eine Aussage:

Lemma 3.3. (Lemma von Cea)

Seien u die Lösung von (3.1.1) und u_h die zugehörige Näherungslösung eines Galerkin-Verfahrens (3.1.2). Ist die Bilinearform a stetig und strikt koerzitiv bezüglich $\|\cdot\|_V$, so gilt:

$$\|u - u_h\|_V \leq C \min_{v_h \in V_h} \|u - v_h\|_V \quad (3.1.9)$$

mit $C > 0$.

Beweis.

Ähnlich zum Beweis von Lemma 3.1 können wir für alle v_h aus V_h schreiben:

$$\delta \|u - u_h\|_V^2 \leq a(u - u_h, u - u_h) = a(u - u_h, u - v_h) \leq c \|u - u_h\|_V \|u - v_h\|_V.$$

Die erste Ungleichung folgt aus der strikten Koerzitivität von a mit der Koerzitivitäts-Konstanten δ , die zweite aus der Galerkin-Orthogonalität von a . Die Stetigkeit der Bilinearform impliziert die Existenz von $c > 0$, so dass $a(u, v) \leq c \|u\|_V \|v\|_V$, was die dritte Ungleichung liefert. Division durch $\|u - u_h\|_V$ und Übergang zum Infimum liefern die Behauptung, wobei $C := c/\delta$. \square

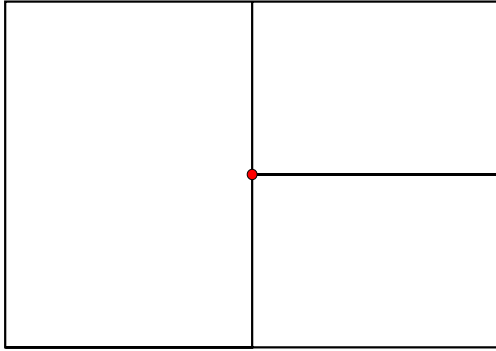
Die Aussage des Lemmas von Cea lässt sich auch so formulieren: In jeder zu $\|\cdot\|_V$ äquivalenten Norm wird die Lösung u durch die Näherungslösung u_h bestmöglich bis auf Multiplikation mit einer Konstanten approximiert. Diese Eigenschaft nennt man quasi-optimale Approximation.

3.2 Finite-Elemente-Methoden

3.2.1 Grundideen

Obwohl das Galerkin-Verfahren bereits eine Möglichkeit liefert, numerische Näherungslösungen von Variationsproblemen zu berechnen, kann es in der Praxis Probleme bereiten. Beim Galerkin-Verfahren wird versucht, eine Näherungslösung auf dem ganzen Gebiet Ω zu bestimmen. Je nachdem, wie Ω aussieht, ist schon das Finden einer geeigneten Basis nicht einfach. Zudem können sich, je nach Art der Basisfunktionen, Schwierigkeiten bei der Berechnung der Integrale der Bilinearform und der Linearform auf der rechten Seite ergeben.

Die Idee der Finite-Elemente-Methoden liegt nun darin, das Gebiet Ω in kleinere Teilgebiete zu zerlegen und das Variationsproblem auf jedem Teilgebiet zu lösen. Dazu wird auf jedem Teilgebiet ein passender Raum gesucht, so dass die jeweiligen Basisfunktionen auf möglichst vielen der anderen Teilgebiete verschwinden. Die Anzahl der Basisfunktionen ist



Unzulässige Zerlegung mit einem hängendem Knoten (roter Punkt). Die Schnittmenge der Zellen K und K' besteht aus mehr als einem Punkt. Aufgrund des hängenden Knotens teilen sich beide Zellen dennoch keine ganze Kante.

(Abbildung 3.1)

dadurch insgesamt recht groß. Da diese aber auf den meisten Teilgebieten verschwinden, ist die sich ergebende Matrix \mathbf{A} dünn besetzt, da sich die entsprechenden Skalarprodukte zu Null ergeben. Da die Teilgebiete meist klein sind, genügt es außerdem oft kleine Räume mit wenigen Basisfunktionen zu betrachten, um bereits gute Näherungslösungen zu erhalten.

Der Einfachheit halber beschränken wir uns auf polygonal berandete Gebiete $\Omega \subset \mathbb{R}^d$, $d = 2, 3$ und auf eine gewisse Sorte von Zerlegungen:

Definition 3.4. (Zulässige Zerlegung)

Eine Zerlegung $\mathcal{T} = \{K_1, \dots, K_M\}$ von $\Omega \subset \mathbb{R}^d$ mit abgeschlossenen Gitterzellen K_i heißt zulässig genau dann, wenn gilt:

- (i) Die Zerlegung überdeckt das ganze Gebiet, das heißt $\bar{\Omega} = \cup_{i=1}^M K_i$.
- (ii) Besteht der Durchschnitt $K_i \cap K_j$ zweier Zellen aus genau einem Punkt p , so ist p Eckpunkt von K_i und K_j .
- (iii) Besteht der Durchschnitt $K_i \cap K_j$, $i \neq j$, zweier Zellen aus mehr als einem Punkt, so entspricht die Menge $K_i \cap K_j$ einer Kante von K_i und einer Kante von K_j oder einer Seitenfläche von K_i und einer Seitenfläche von K_j .

Weiterhin definieren wir $h_K := \text{diam}(K)$ und die Gitterweite h der Zerlegung als $h := \max_{K \in \mathcal{T}} (h_K)$.

Insbesondere dürfen zulässige Zerlegungen keine hängenden Knoten enthalten, da Bedingung (iii) dann verletzt ist. Die in Abbildung 3.1 gezeigte Zerlegung ist also ein Beispiel für eine unzulässige Zerlegung.

Zur Festlegung der Basisfunktionen auf den einzelnen Zellen werden meist Funktionale $\Phi : V_h \rightarrow \mathbb{R}$ benutzt. Es wird vorausgesetzt, dass die Funktionale linear unabhängig voneinander, stetig und linear sind. Des Weiteren verlangen wir, dass sie unisolvent bezüglich V_h sind:

Definition 3.5. (Unisolvenz)

Sei N die Dimension des Raums V_h . Die Funktionale Φ_1, \dots, Φ_N werden unisolvent bezüglich V_h genannt, falls es zu jedem $\mathbf{a} = (a_1, \dots, a_N)^T \in \mathbb{R}^N$ genau ein $u \in V_h$ gibt, so dass

$$\Phi_i(u) = a_i \quad \text{für } i = 1, \dots, N \tag{3.2.1}$$

erfüllt ist.

Häufig werden als Funktionale Punktauswertungen $\Phi(u) = u(\mathbf{x})$, gewichtete Integralmittelwerte über einzelne Zellen der Gebietszerlegung $\Phi(u) = \int_K w(\mathbf{x}) u(\mathbf{x}) d\mathbf{x}$ oder Auswertungen der ersten Ableitung $\Phi(u) = \partial_i u(\mathbf{x})$ verwendet. Wir werden in dieser Arbeit die ersten beiden hier erwähnten Arten von Funktionalen benutzen.

Um mit den Funktionalen Φ_i eine Basis zu konstruieren, kann man sich für die Vektoren \mathbf{a} eine Basis des \mathbb{R}^N wählen. Mit den Standard-Einheitsvektoren ergibt sich zum Beispiel aus der Unisolvenz der Funktionalen, dass es einen eindeutigen Satz von Basisfunktionen $\phi_j \in V_h$ gibt, so dass gilt:

$$\Phi_i(\phi_j) = \delta_{ij} \quad \forall 0 \leq i, j \leq N. \quad (3.2.2)$$

Wir können nun ein Finites-Element wie folgt definieren:

Definition 3.6. (Finites-Element)

Es seien

- Ω ein nichtleeres, abgeschlossenes Gebiet des \mathbb{R}^d ($d \geq 1$).
- V_h ein N -dimensionaler Raum von Funktionen mit Definitionsbereich Ω und Bildbereich in \mathbb{R} .
- $\Sigma = \{\Phi_i\}_{i=1}^N$ ein Satz linear unabhängiger, linearer, stetiger und bezüglich V_h unisolventer Funktionalen.

Dann bezeichnen wir das Tripel (Ω, V_h, Σ) als Finites-Element. V_h wird in diesem Zusammenhang Finite-Elemente-Raum genannt.

Diese Definition eines Finiten-Elements ist sehr allgemein. Beispielsweise macht sie keine Aussage über die Zerlegung des Gebiets Ω oder über die Art der Funktionen in V_h . In der Praxis wird meist für jedes Teilgebiet K_i ein Polynomraum gewählt. V_h besteht dann aus dem Raum all jener Funktionen, die sich auf Ω gemäß der Zerlegung als stückweise Polynome darstellen lassen. Der Vorteil von Polynomen liegt darin, dass sie sich auch mit numerischen Methoden leicht und exakt integrieren und differenzieren lassen. Ebenso lässt sich für Polynomräume meist ohne großen Aufwand eine Basis zur gewählten Familie von Funktionalen berechnen. Andere mögliche Räume V_h sind Räume stückweise rationaler (Sev07) oder stückweise trigonometrischer Funktionen (Bab02), worauf wir aber hier nicht weiter eingehen wollen.

3.2.2 Polynomielle Finite-Elemente-Räume

Wir gehen im Folgenden immer davon aus, dass eine zulässige Zerlegung \mathcal{T} des Gebiets Ω in Zellen K_i , $i = 1 \dots N$ gegeben ist. Den Raum $V_h(\Omega)$ setzen wir dann aus den Räumen zusammen, die auf den einzelnen Zellen gegeben sind durch:

$$V_h(\Omega) := \{v \in C(\bar{\Omega}) : v|_K \in V_h(K) \quad \forall K \in \mathcal{T}\}. \quad (3.2.3)$$

Möchte man für $V_h(K)$ nun Polynomräume wählen, so werden in der Literatur zwei Arten von Räumen unterschieden:

$$\mathbb{P}_p(K) := \left\{ p : K \rightarrow \mathbb{R} : p(\mathbf{x}) = \sum_{|\alpha| \leq p} \mu_\alpha \mathbf{x}^\alpha \right\}, \quad \mu_\alpha \in \mathbb{R}, \alpha = \text{Multiindex}, \quad (3.2.4a)$$

$$\mathbb{Q}_p(K) := \left\{ p : K \rightarrow \mathbb{R} : p(\mathbf{x}) = \sum_{\{\alpha: \alpha_1, \alpha_2 \leq p\}} \mu_\alpha \mathbf{x}^\alpha \right\}, \quad \mu_\alpha \in \mathbb{R}, \alpha = \text{Multiindex}. \quad (3.2.4b)$$

\mathbb{P}_p umfasst also alle Funktionen, welche sich auf K als Polynom mit Grad kleiner gleich p darstellen lassen, wohingegen \mathbb{Q}_p der Raum aller Polynome auf K mit Grad kleiner gleich p in jeder Variablen ist. Die zugehörigen Räume auf ganz Ω definieren wir als

$$P_p(\Omega) := \{v \in C(\bar{\Omega}) : v|_K \in \mathbb{P}_p(K) \forall K \in \mathcal{T}\}, \quad p \geq 1, \quad (3.2.5a)$$

$$Q_p(\Omega) := \{v \in C(\bar{\Omega}) : v|_K \in \mathbb{Q}_p(K) \forall K \in \mathcal{T}\}, \quad p \geq 1. \quad (3.2.5b)$$

Prinzipiell ist es auch möglich, \mathbb{P}_p - und \mathbb{Q}_p -Räume auf verschiedenen Zellen zu mischen oder auf verschiedenen Zellen Polynome unterschiedlichen Grades zu benutzen. Wir wollen darauf hier jedoch nicht näher eingehen.

Als nächstes müssen wir für die Räume die entsprechenden Basisfunktionen bestimmen. Naiv betrachtet, würde man nun für jeden Raum auf jeder Zelle einen Satz unisolventer Funktionale vorgeben, damit auf jeder Zelle eine lokale Basis bestimmen und aus diesen lokalen Basen dann eine globale Basis für den entsprechenden Raum auf ganz Ω konstruieren. Dies ist jedoch vor allem bei feinen Zerlegungen mit vielen Zellen mit einem großen Speicheraufwand verbunden. Um diesen zu vermindern, benutzt man so genannte parametrische Finite–Elemente, wie sie auch im Programmpaket MooNMD verwendet werden.

Die Grundlage parametrischer Elemente ist, dass die Zellen der Zerlegung \mathcal{T} alle eine ähnliche geometrische Struktur besitzen, zum Beispiel, dass alle Zellen Dreiecke oder Vierecke sind. In diesem Fall ist es möglich, eine Referenzzelle \hat{K} zu definieren und für jede Zelle eine Referenztransformation $F_K : \hat{K} \rightarrow K$ zu finden, welche die Referenzzelle \hat{K} auf die Zelle K abbildet. Die lokalen Basisfunktionen ϕ auf der Zelle K ergeben sich damit aus den Basisfunktionen $\hat{\phi}$ auf der Referenzzelle \hat{K} durch

$$\phi(\mathbf{x}) = \hat{\phi} \circ F_K^{-1}(\mathbf{x}). \quad (3.2.6)$$

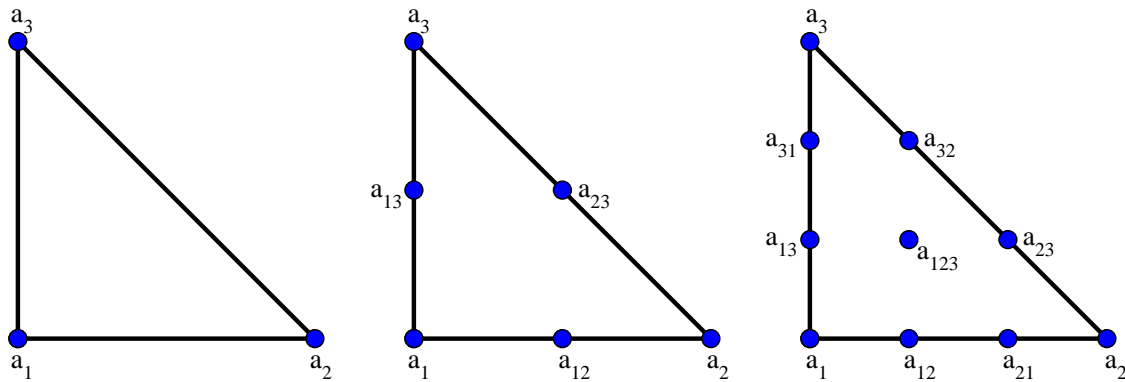
Es genügt nun, die Basisfunktionen auf der Referenzzelle \hat{K} und die Referenztransformation für jede Zelle K zu speichern. Auch für die Analysis ergeben sich Vorteile, da gewisse Aussagen nur für die Referenzzelle gezeigt werden müssen und sich dann durch die Transformation auf die anderen Zellen übertragen lassen.

Als Referenzzelle für Dreiecke wird üblicherweise das Dreieck mit den Eckpunkten $(0, 0)$, $(0, 1)$ und $(1, 0)$ gewählt. Für Vierecke wird das Einheitsquadrat $[0, 1]^2$ oder wie auch in MooNMD das große Einheitsquadrat $[-1, 1]^2$ benutzt. Bei Dreiecks- und Rechteckszerlegungen sind die Referenztransformationen affin linear. Bei Verwendung allgemeiner Vierecke ist dies nicht mehr der Fall. Zudem macht die Verwendung allgemeiner Vierecke es meist notwendig, statt stückweise polynomieller Funktionen stückweise rationale zuzulassen. Wir

beschränken uns im Folgenden auf die Verwendung von Rechtecken.

Für Gebiete $\Omega \subset \mathbb{R}^3$ werden an Stelle von Dreiecken Simplexes (Cia78) und an Stelle von Vierecken Hexaeder benutzt. Als Referenzzellen dienen der Einheitssimplex mit den Eckpunkten $(0, 0, 0)$, $(1, 0, 0)$, $(0, 1, 0)$ und $(0, 0, 1)$, der Einheitswürfel $[0, 1]^3$ oder der große Einheitswürfel $[-1, 1]^3$.

Zum Schluss dieses Abschnittes wollen wir noch einige Finite-Elemente und die dazugehörigen Basisfunktionen angeben. Wir beschränken uns dabei auf P -Elemente auf Dreiecken und Q -Elemente auf Rechtecken. Zur Konstruktion der Finiten-Elemente fehlen uns noch die zugehörigen unisolventen Funktionale $\{\Phi_i\}_{i=1}^N$. Wir wählen als Funktionale Punktauswertungen in gewissen Punkten entsprechend Abbildung 3.2 für P -Elemente auf Dreiecken und entsprechend Abbildung 3.3 für Q -Elemente auf Rechtecken. Die so erhaltenen Finiten-Elemente werden Lagrange-Elemente genannt.



Verteilung der Knoten bei den P_p -Elementen auf dem Referenzdreieck. Links für $p = 1$, in der Mitte für $p = 2$ und rechts für $p = 3$.

(Abbildung 3.2)

Für das Referenzdreieck \hat{K} sind die entsprechenden Basisfunktionen

- für die stetigen, stückweise linearen Finiten-Elemente P_1 durch

$$\hat{\phi}_1(\hat{x}, \hat{y}) = 1 - \hat{x} - \hat{y}, \quad \hat{\phi}_2(\hat{x}, \hat{y}) = \hat{x} \quad \text{und} \quad \hat{\phi}_3(\hat{x}, \hat{y}) = \hat{y},$$

wobei ϕ_i den Wert 1 im Eckpunkt \mathbf{a}_i annimmt.

- für die stetigen, stückweise quadratischen Finiten-Elemente P_2 durch

$$\begin{aligned} \hat{\phi}_1(\hat{x}, \hat{y}) &= (\hat{x} + \hat{y} - 1)(2\hat{x} + 2\hat{y} - 1), \\ \hat{\phi}_2(\hat{x}, \hat{y}) &= \hat{x}(2\hat{x} - 1) \quad \text{und} \\ \hat{\phi}_3(\hat{x}, \hat{y}) &= \hat{y}(2\hat{y} - 1), \end{aligned}$$

wobei $\hat{\phi}_i$ den Wert 1 im Eckpunkt \mathbf{a}_i annimmt sowie

$$\begin{aligned} \hat{\phi}_{12}(\hat{x}, \hat{y}) &= -4\hat{x}(\hat{x} + \hat{y} - 1), \\ \hat{\phi}_{13}(\hat{x}, \hat{y}) &= -4\hat{y}(\hat{x} + \hat{y} - 1) \quad \text{und} \\ \hat{\phi}_{23}(\hat{x}, \hat{y}) &= -4\hat{x}\hat{y}, \end{aligned}$$

wobei $\hat{\phi}_{ij}$ den Wert 1 im Mittelpunkt \mathbf{a}_{ij} der Kante mit den Eckpunkten \mathbf{a}_i und \mathbf{a}_j annimmt.

- für die stetigen, stückweise kubischen Finiten-Elemente P_3 durch

$$\begin{aligned}\hat{\phi}_1(\hat{x}, \hat{y}) &= -\frac{9}{2} \left(\hat{x} + \hat{y} - \frac{1}{3} \right) \left(\hat{x} + \hat{y} - \frac{2}{3} \right) (\hat{x} + \hat{y} - 1), \\ \hat{\phi}_2(\hat{x}, \hat{y}) &= \frac{9}{2} \hat{x} \left(\hat{x} - \frac{1}{3} \right) \left(\hat{x} - \frac{2}{3} \right) \text{ und} \\ \hat{\phi}_3(\hat{x}, \hat{y}) &= \frac{9}{2} \hat{y} \left(\hat{y} - \frac{1}{3} \right) \left(\hat{y} - \frac{2}{3} \right),\end{aligned}$$

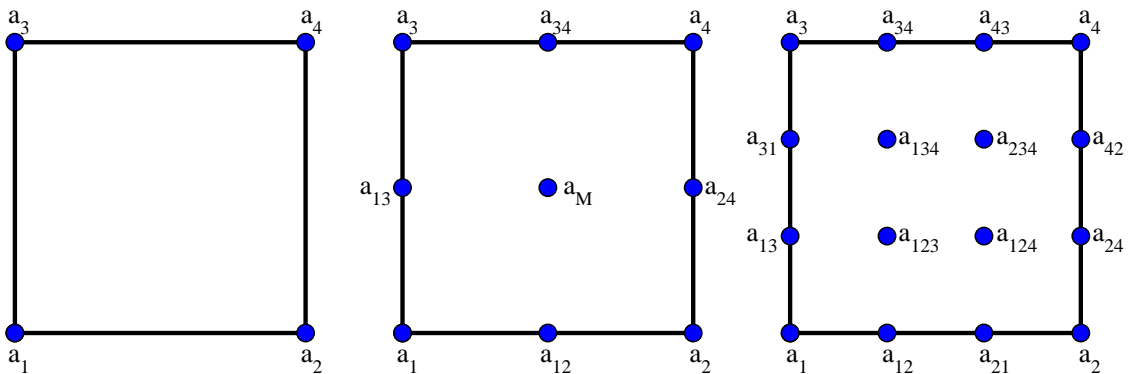
wobei $\hat{\phi}_i$ den Wert 1 im Eckpunkt \mathbf{a}_i annimmt,

$$\begin{aligned}\hat{\phi}_{12}(\hat{x}, \hat{y}) &= \frac{27}{2} \hat{x} \left(\hat{x} + \hat{y} - \frac{2}{3} \right) (\hat{x} + \hat{y} - 1), \\ \hat{\phi}_{21}(\hat{x}, \hat{y}) &= -\frac{27}{2} \hat{x} \left(\hat{x} - \frac{1}{3} \right) (\hat{x} + \hat{y} - 1), \\ \hat{\phi}_{13}(\hat{x}, \hat{y}) &= \frac{27}{2} \hat{y} \left(\hat{x} + \hat{y} - \frac{2}{3} \right) (\hat{x} + \hat{y} - 1), \\ \hat{\phi}_{31}(\hat{x}, \hat{y}) &= -\frac{27}{2} \hat{x} \hat{y} \left(\hat{y} - \frac{1}{3} \right), \\ \hat{\phi}_{23}(\hat{x}, \hat{y}) &= \frac{27}{2} \hat{x} \hat{y} \left(\hat{x} - \frac{1}{3} \right) \text{ und} \\ \hat{\phi}_{32}(\hat{x}, \hat{y}) &= \frac{27}{2} \hat{x} \hat{y} \left(\hat{y} - \frac{1}{3} \right),\end{aligned}$$

wobei $\hat{\phi}_{ij}$ den Wert 1 im Mittelpunkt \mathbf{a}_{ij} der Kante mit den Eckpunkten \mathbf{a}_i und \mathbf{a}_j annimmt (siehe Abbildung 3.2) sowie

$$\hat{\phi}_{123}(\hat{x}, \hat{y}) = -27\hat{x}\hat{y}(\hat{x} + \hat{y} - 1),$$

wobei $\hat{\phi}_{ij}$ den Wert 1 im Schwerpunkt \mathbf{a}_{123} des Dreiecks annimmt.



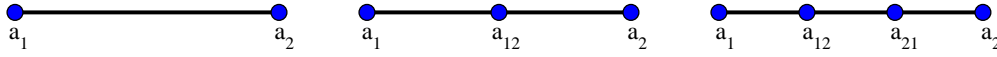
Verteilung der Knoten bei den Q_p -Elementen auf dem Referenzquadrat. Links für $p = 1$, in der Mitte für $p = 2$ und rechts für $p = 3$.

(Abbildung 3.3)

Zur Konstruktion der Basisfunktionen der Q -Elemente auf Rechtecken beachten wir, dass diese sich als Tensorprodukte eindimensionaler Finiten-Elemente schreiben lassen. Dies über-

trägt sich auf die Basisfunktionen, die daher direkt aus den Basisfunktionen der entsprechenden eindimensionalen Finiten-Elementen als Tensorprodukte hervorgehen.

In einer Dimension stellt Ω ein Intervall dar, welches in Teilintervalle zerlegt wird. Als Referenzzellen dienen hier das Einheitsintervall $[0, 1]$ oder das große Einheitsintervall $[-1, 1]$. Die Funktionale sind die Punktauswertungen in den entsprechend Abbildung 3.4 verteilten Punkten.



Verteilung der Knoten bei stetigen eindimensionalen Finiten-Elementen. Links lineare Elemente, in der Mitte quadratische, rechts kubische.

(Abbildung 3.4)

Damit ergeben sich

- die Basisfunktionen für stetige, stückweise lineare eindimensionale Finite-Elemente auf dem großen Referenzintervall zu:

$$\xi_1(\hat{x}) = \frac{1}{2}(1 - \hat{x}) \quad \text{und} \quad \xi_2(\hat{x}) = \frac{1}{2}(1 + \hat{x}). \quad (3.2.7)$$

Die Basisfunktionen $\hat{\phi}_i = \hat{\phi}_i(\hat{x}, \hat{y})$, $i = 1, \dots, 4$, von Q_1 auf dem großen Referenzquadrat sind dann:

$$\hat{\phi}_1 = \xi_1(\hat{x}) \xi_1(\hat{y}), \quad \hat{\phi}_2 = \xi_2(\hat{x}) \xi_1(\hat{y}), \quad \hat{\phi}_3 = \xi_1(\hat{x}) \xi_2(\hat{y}) \quad \text{und} \quad \hat{\phi}_4 = \xi_2(\hat{x}) \xi_2(\hat{y}). \quad (3.2.8)$$

Die Basisfunktion $\hat{\phi}_i$ nimmt dabei im Punkt \mathbf{a}_i den Wert 1 und in den anderen Punkten den Wert 0 an.

- die Basisfunktionen für stetige, stückweise quadratische eindimensionale Finite-Elemente auf dem großen Referenzintervall zu:

$$\xi_1(\hat{x}) = -\frac{1}{2}\hat{x}(1 - \hat{x}), \quad \xi_2(\hat{x}) = (1 + \hat{x})(1 - \hat{x}) \quad \text{und} \quad \xi_3(\hat{x}) = \frac{1}{2}\hat{x}(1 + \hat{x}). \quad (3.2.9)$$

Die 9 Basisfunktionen $\hat{\phi} = \hat{\phi}(\hat{x}, \hat{y})$ von Q_2 auf dem großen Referenzquadrat sind dann:

$$\begin{aligned} \hat{\phi}_1 &= \xi_1(\hat{x}) \xi_1(\hat{y}), & \hat{\phi}_2 &= \xi_3(\hat{x}) \xi_1(\hat{y}), & \hat{\phi}_3 &= \xi_1(\hat{x}) \xi_3(\hat{y}), & \hat{\phi}_4 &= \xi_3(\hat{x}) \xi_3(\hat{y}), \\ \hat{\phi}_{12} &= \xi_2(\hat{x}) \xi_1(\hat{y}), & \hat{\phi}_{13} &= \xi_1(\hat{x}) \xi_2(\hat{y}), & \hat{\phi}_{24} &= \xi_3(\hat{x}) \xi_2(\hat{y}), & \hat{\phi}_{34} &= \xi_2(\hat{x}) \xi_3(\hat{y}) \quad \text{und} \\ \hat{\phi}_M &= \xi_2(\hat{x}) \xi_2(\hat{y}). \end{aligned}$$

Die Basisfunktion $\hat{\phi}_i$ nimmt den Wert 1 im Eckpunkt \mathbf{a}_i , die Basisfunktion $\hat{\phi}_{ij}$ im Mittelpunkt \mathbf{a}_{ij} der Kante mit den Eckpunkten \mathbf{a}_i sowie \mathbf{a}_j und die Basisfunktion $\hat{\phi}_M$ im Mittelpunkt \mathbf{a}_M des Rechtecks.

- die Basisfunktionen für stetige, stückweise kubische eindimensionale Finite–Elemente auf dem großen Referenzintervall zu:

$$\begin{aligned}\xi_1(\hat{x}) &= -\frac{1}{16}(3\hat{x} + 1)(3\hat{x} - 1)(\hat{x} - 1), \\ \xi_2(\hat{x}) &= \frac{9}{16}(3\hat{x} - 1)(\hat{x} + 1)(\hat{x} - 1), \\ \xi_3(\hat{x}) &= -\frac{9}{16}(3\hat{x} + 1)(\hat{x} + 1)(\hat{x} - 1) \text{ und} \\ \xi_4(\hat{x}) &= \frac{1}{16}(3\hat{x} + 1)(3\hat{x} - 1)(\hat{x} + 1).\end{aligned}$$

Die 16 Basisfunktionen $\hat{\phi} = \hat{\phi}(\hat{x}, \hat{y})$ von Q_3 auf dem großen Referenzquadrat sind dann:

$$\begin{aligned}\hat{\phi}_1 &= \xi_1(\hat{x}) \xi_1(\hat{y}), & \hat{\phi}_2 &= \xi_4(\hat{x}) \xi_1(\hat{y}), & \hat{\phi}_3 &= \xi_1(\hat{x}) \xi_4(\hat{y}), & \hat{\phi}_4 &= \xi_4(\hat{x}) \xi_4(\hat{y}), \\ \hat{\phi}_{12} &= \xi_3(\hat{x}) \xi_1(\hat{y}), & \hat{\phi}_{21} &= \xi_2(\hat{x}) \xi_1(\hat{y}), & \hat{\phi}_{13} &= \xi_1(\hat{x}) \xi_3(\hat{y}), & \hat{\phi}_{31} &= \xi_1(\hat{x}) \xi_2(\hat{y}), \\ \hat{\phi}_{34} &= \xi_3(\hat{x}) \xi_4(\hat{y}), & \hat{\phi}_{43} &= \xi_2(\hat{x}) \xi_4(\hat{y}), & \hat{\phi}_{24} &= \xi_4(\hat{x}) \xi_3(\hat{y}), & \hat{\phi}_{42} &= \xi_4(\hat{x}) \xi_2(\hat{y}), \\ \hat{\phi}_{123} &= \xi_3(\hat{x}) \xi_3(\hat{y}), & \hat{\phi}_{124} &= \xi_2(\hat{x}) \xi_3(\hat{y}), & \hat{\phi}_{134} &= \xi_3(\hat{x}) \xi_2(\hat{y}), & \hat{\phi}_{234} &= \xi_2(\hat{x}) \xi_2(\hat{y}).\end{aligned}$$

Die Basisfunktionen $\hat{\phi}$ haben ihren 1-Knoten in dem gleich-indizierten Punkt **a** von Abbildung 3.3.

3.2.3 Unstetige Finite–Elemente

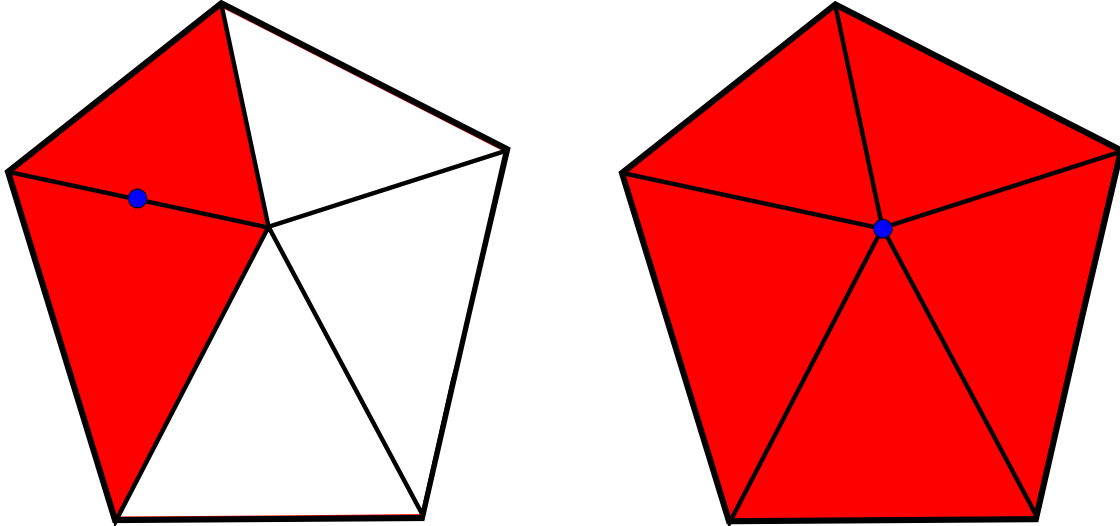
Wir wollen in dieser Arbeit so genannte Unstetige–Galerkin–Verfahren (englisch: **Discontinuous Galerkin Finite Element Methods**; DGFEM) untersuchen. Für diese können auf den einzelnen Zellen die gleichen Räume verwendet werden, wie sie im vorherigen Abschnitt eingeführt wurden. Man verzichtet jedoch bei der Konstruktion des Gesamttraumes auf die Bedingung, dass die Funktionen auf dem Gebiet Ω stetig sein sollen. Somit ergibt sich:

$$P_p^{disc}(\Omega) := \{v : \Omega \rightarrow \mathbb{R} : v|_K \in \mathbb{P}_p(K) \ \forall K \in \mathcal{T}\}, \quad p \geq 1, \quad (3.2.10a)$$

$$Q_p^{disc}(\Omega) := \{v : \Omega \rightarrow \mathbb{R} : v|_K \in \mathbb{Q}_p(K) \ \forall K \in \mathcal{T}\}, \quad p \geq 1. \quad (3.2.10b)$$

Man sieht sofort, dass die unstetigen Räume die stetigen als Teilmenge enthalten. Ebenso ist auch ein Unterschied zwischen stetigen und unstetigen Finiten–Elemente–Räumen sofort klar: Bei stetigen Finiten–Elementen teilen sich benachbarte Zellen einen Teil ihrer Basisfunktionen, da hier mehrere lokale Basisfunktionen zu einer globalen Basisfunktion zusammen genommen werden können. Das Einzugsgebiet einer globalen Basisfunktion besteht daher im Allgemeinen aus mehreren Zellen. Unter dem Einzugsgebiet einer globalen Basisfunktion versteht man dabei die Menge aller Zellen, auf denen sie nicht identisch verschwindet (siehe Abb. 3.5). Bei unstetigen Basisfunktionen hingegen ist das Einzugsgebiet einer Basisfunktion genau eine Zelle, da jede lokale Basisfunktion auf Grund der fehlenden Stetigkeitsbedingung am Rand der Zelle unstetig auf Null fallen kann und damit außerhalb “ihrer” Zelle identisch verschwindet. Daher können nicht mehrere lokale Basisfunktionen zu einer globalen Basisfunktion zusammengefasst werden, sondern jede lokale Basisfunktion ist zugleich eine globale Basisfunktion. Dadurch erhöht sich die Anzahl der globalen Basisfunktionen.

Zur Zerlegung \mathcal{T} definieren wir die Menge \mathcal{E} aller offenen eindimensionalen Grenzflächen



Beispiele für Einzugsgebiete globaler Basisfunktionen bei Verwendung stetiger Finiten-Elemente.

(Abbildung 3.5)

E (Kanten) zwischen je zwei Zellen K und K' . Zudem definieren wir

$$\begin{aligned}\mathcal{E}_{\text{int}} &:= \{E \in \mathcal{E} : E \subset \Omega\} \\ \Gamma_{\text{int}} &:= \{x \in \Omega : x \in E \text{ für ein } E \in \mathcal{E}_{\text{int}}\} \\ \mathcal{E}^D &:= \{E \subset \Gamma : \text{Auf } E \text{ ist eine Dirichlet-Randbedingung vorgegeben.}\}\end{aligned}$$

Hierbei und im Weiteren sei $\Gamma := \partial\Omega$ der Rand von Ω .

Da bei DG-Verfahren die Ansatzräume auf allen Zellen praktisch unabhängig voneinander sind, müssen wir auch einen allgemeineren Raum von Sobolev-Funktionen betrachten:

Definition 3.7 (Gebrochene Sobolevräume).

Sei Ω ein beschränktes Gebiet und \mathcal{T} eine zulässige Zerlegung von Ω . Der gebrochene Sobolevraum $W^{k,p}(\mathcal{T})$ ist dann definiert durch

$$W^{k,p}(\mathcal{T}) := \{v \in L^p(\Omega) \mid \forall K \in \mathcal{T} : v|_K \in W^{k,p}(K)\}$$

mit der Norm

$$\|\cdot\|_{W^{k,p}(\mathcal{T})} = \left(\sum_{K \in \mathcal{T}} \|\cdot\|_{W^{k,p}(K)}^p \right)^{\frac{1}{p}}.$$

Auch hier sieht man sofort, dass $W^{k,p}(\Omega) \subset W^{k,p}(\mathcal{T})$ gilt.

Die Funktionen aus $W^{k,p}(\mathcal{T})$ haben im Allgemeinen keine wohldefinierten Werte auf den Kanten der Zerlegung. Dies trifft auch auf die Funktionen aus $F_p^{\text{disc}}(\Omega)$ und $Q_p^{\text{disc}}(\Omega)$ zu, da diese beim Übergang von einer Zelle zur Nachbarzelle unstetig sind. Je nachdem, von welcher Zelle aus man sich einem Punkt der Kante nähert, erhält man also unterschiedliche Funktionswerte auf der Kante. Statt der Funktionswerte werden daher Sprünge und Mittelwerte der Funktion auf einer Kante betrachtet.

Für eine Funktion $v \in H^1(\mathcal{T})$ definieren wir den Sprung $[v]_E$ der Funktion über eine Kante

$E \in \mathcal{E}_{\text{int}}$ zwischen den Elementen K und K' durch

$$[v]_E := v|_{\partial K \cap E} - v|_{\partial K' \cap E}. \quad (3.2.11)$$

Hierbei bezeichnet ∂K den Rand der Zelle K , das heißt die Vereinigung über alle Kanten der Zelle K und $v|_{\partial K \cap E}$ den Grenzwert

$$\lim_{\substack{y \rightarrow x \\ y \in K}} v(y).$$

Das Vorzeichen des Sprungs ändert dabei offensichtlich sein Vorzeichen, wenn wir die Rollen von K und K' vertauschen.

Der Mittelwert $\langle v \rangle_E$ von v auf der Kante E wird definiert durch

$$\langle v \rangle_E := \frac{1}{2} (v|_{\partial K \cap E} + v|_{\partial K' \cap E}). \quad (3.2.12)$$

Im Gegensatz zum Sprung ändert dieser sein Vorzeichen nicht, wenn K und K' vertauscht werden.

Als drittes benötigen wir noch einen Sprung in Richtung eines vorgegebenen Vektors b . Dazu definieren wir zuerst Einströmanteil $\partial_- K$ und Ausströmanteil $\partial_+ K$ von ∂K für alle $K \in \mathcal{T}$ durch

$$\begin{aligned} \partial_- K &:= \{x \in \partial K : b(x) \cdot n_K(x) < 0\}, \\ \partial_+ K &:= \{x \in \partial K : b(x) \cdot n_K(x) \geq 0\}. \end{aligned}$$

Hierbei bezeichnet $n_K(x)$ den äußeren Einheitsnormalenvektor zu ∂K im Punkt $x \in \partial K$. Die Bezeichnungen Ein- und Ausströmanteil kommen aus der Fluidmechanik. Ist b der Richtungsvektor eines Strömungsfeldes, so ist $\partial_- K$ gerade derjenige Anteil von ∂K für den Fluid in die Zelle K hinein strömt. Der zu b gehörige Sprung ist definiert durch

$$[v]_K = v^+ - v^- := \lim_{\substack{\epsilon \rightarrow 0 \\ \epsilon > 0}} v(x + \epsilon b) - \lim_{\substack{\epsilon \rightarrow 0 \\ \epsilon > 0}} v(x - \epsilon b). \quad (3.2.13)$$

Der Sprung $[\cdot]_K$ ist nicht von der Nummerierung der Zellen, sondern von der Richtung von b abhängig. Sein Vorzeichen kann daher entlang einer Kante E wechseln, wenn sich das Vorzeichen von $b \cdot n_K$ ändert. Da er somit eher mit der Zelle K als mit der Kante E zusammen hängt, haben wir den Sprung $[\cdot]_K$ mit K statt mit E indiziert.

Es sei darauf hingewiesen, dass die Indizierung der Übersichtlichkeit halber in den folgenden Abschnitten und Kapiteln teilweise unterdrückt wird.

Definition 3.8.

Auf dem Raum $H^1(\mathcal{T})$ definieren wir folgende Norm

$$\|u\|_{1,h}^2 := \|u\|_W^{k,p}(\mathcal{T}) + \sum_{E \in \mathcal{E}_{\text{int}}} \kappa_E \| [u] \|_{L^2(E)}^2 + \sum_{E \in \mathcal{E}^D} \kappa_E \| u \|_{L^2(E)}^2.$$

Der hier auftretende Parameter κ_E wird dabei so gewählt, dass sich für das zu behandelnde Problem optimale Fehlerabschätzungen ergeben. Ein Beispiel werden wir in Kapitel 4 kennen lernen.

Bemerkung 3.9.

- (i) Die Norm in Definition 3.8 ist das unstetige Äquivalent der H^1 -Seminorm.
- (ii) Die Abschätzung

$$\|u\|_{L^2(\Omega)} \leq C \|u\|_{1,h}$$

ergibt sich aus der entsprechenden Abschätzung mit Hilfe der Friedrichs-Ungleichung auf jeder Zelle K .

Im Programmpaket MoonMD werden als Funktionale für unstetige Finite-Elemente nicht die Punktauswertungen benutzt, sondern bestimmte gewichtete Integralmittelwerte. Die Gewichtsfunktionen und Basisfunktionen für die einzelnen in dieser Arbeit verwendeten Räume sind im Folgenden angegeben. Als Referenzzellen werden hier wieder das Dreieck $(0, 0)$, $(0, 1)$, $(1, 0)$ für Dreiecksgitter und das große Einheitsquadrat für Rechtecksgitter benutzt.

Element	Gewichtsfunktionen	Basisfunktionen
P_1^{disc}	$w_1(\hat{x}, \hat{y}) = 1$ $w_2(\hat{x}, \hat{y}) = 12\hat{x} - 4$ $w_3(\hat{x}, \hat{y}) = 12\hat{y} - 4$	$\hat{\phi}_1(\hat{x}, \hat{y}) = 1$ $\hat{\phi}_2(\hat{x}, \hat{y}) = 2\hat{x} + \hat{y} - 1$ $\hat{\phi}_3(\hat{x}, \hat{y}) = 2\hat{y} + \hat{x} - 1$
P_2^{disc}	$w_1(\hat{x}, \hat{y}) = 1$ $w_2(\hat{x}, \hat{y}) = \hat{x} + \frac{1}{2}\hat{y} - \frac{1}{2}$ $w_3(\hat{x}, \hat{y}) = \hat{y} + \frac{1}{2}\hat{x} - \frac{1}{2}$ $w_4(\hat{x}, \hat{y}) = 90\hat{x}^2 - 72\hat{x} + 9$ $w_5(\hat{x}, \hat{y}) = 180\hat{x}\hat{y} - 36\hat{x} - 36\hat{y} + 9$ $w_6(\hat{x}, \hat{y}) = 90\hat{y}^2 - 72\hat{y} + 9$	$\hat{\phi}_1(\hat{x}, \hat{y}) = 1$ $\hat{\phi}_2(\hat{x}, \hat{y}) = 24\hat{x} - 8$ $\hat{\phi}_3(\hat{x}, \hat{y}) = 24\hat{y} - 8$ $\hat{\phi}_4(\hat{x}, \hat{y}) = 1 - 6\hat{x} - 2\hat{y} + 6\hat{x}^2 + 6\hat{x}\hat{y} + \hat{y}^2$ $\hat{\phi}_5(\hat{x}, \hat{y}) = 1 - 4\hat{x} - 4\hat{y} + 3\hat{x}^2 + 8\hat{x}\hat{y} + 3\hat{y}^2$ $\hat{\phi}_6(\hat{x}, \hat{y}) = 1 - 2\hat{x} - 6\hat{y} + \hat{x}^2 + 6\hat{x}\hat{y} + 6\hat{y}^2$
P_3^{disc}	wie P_2^{disc} und zusätzlich $w_7(\hat{x}, \hat{y}) = \frac{35}{2}\hat{x}^3 - \frac{45}{2}\hat{x}^2 + \frac{15}{2}\hat{x} - \frac{1}{2}$ $w_8(\hat{x}, \hat{y}) = \frac{105}{2}\hat{x}^2\hat{y} - 30\hat{x}\hat{y} - \frac{15}{2}\hat{x}^2 + 5\hat{x} + \frac{5}{2}\hat{y} - \frac{1}{2}$ $w_9(\hat{x}, \hat{y}) = \frac{105}{2}\hat{x}\hat{y}^2 - 30\hat{x}\hat{y} - \frac{15}{2}\hat{y}^2 + 5\hat{y} + \frac{5}{2}\hat{x} - \frac{1}{2}$ $w_{10}(\hat{x}, \hat{y}) = \frac{35}{2}\hat{y}^3 - \frac{45}{2}\hat{y}^2 + \frac{15}{2}\hat{y} - \frac{1}{2}$	wie P_2^{disc} und zusätzlich $\hat{\phi}_7(\hat{x}, \hat{y}) = -32 + 384\hat{x} + 96\hat{y} - 960\hat{x}^2 - 768\hat{x}\hat{y} + 960\hat{x}^2\hat{y} + 384\hat{x}\hat{y}^2 - 96\hat{y}^2 + 640\hat{x}^3 + 32\hat{y}^3$ $\hat{\phi}_8(\hat{x}, \hat{y}) = -32 + 288\hat{x} + 192\hat{y} - 576\hat{x}^2 - 1152\hat{x}\hat{y} + 1152\hat{x}^2\hat{y} + 864\hat{x}\hat{y}^2 - 288\hat{y}^2 + 320\hat{x}^3 + 128\hat{y}^3$ $\hat{\phi}_9(\hat{x}, \hat{y}) = -32 + 288\hat{y} + 192\hat{x} - 576\hat{y}^2 - 1152\hat{x}\hat{y} + 1152\hat{x}\hat{y}^2 + 864\hat{x}^2\hat{y} - 288\hat{x}^2 + 320\hat{y}^3 + 128\hat{x}^3$ $\hat{\phi}_{10}(\hat{x}, \hat{y}) = -32 + 384\hat{y} + 96\hat{x} - 960\hat{y}^2 - 768\hat{x}\hat{y} + 960\hat{x}\hat{y}^2 + 384\hat{x}^2\hat{y} - 96\hat{x}^2 + 640\hat{y}^3 + 32\hat{x}^3$

Basisfunktionen und Gewichtsfunktionen der Funktionale für P_p^{disc} -Elemente auf Dreiecken ($p = 1, 2, 3$).

(Tabelle 3.1)

P_1^{disc}	$w_1(\hat{x}, \hat{y}) = 1$ $w_2(\hat{x}, \hat{y}) = \hat{x}$ $w_3(\hat{x}, \hat{y}) = \hat{y}$	$\hat{\phi}_1(\hat{x}, \hat{y}) = 1$ $\hat{\phi}_2(\hat{x}, \hat{y}) = 3\hat{x}$ $\hat{\phi}_3(\hat{x}, \hat{y}) = 3\hat{y}$
P_2^{disc}	wie P_1^{disc} und zusätzlich $w_4(\hat{x}, \hat{y}) = 3\hat{x}^2 - 1$ $w_5(\hat{x}, \hat{y}) = \hat{x}\hat{y}$ $w_6(\hat{x}, \hat{y}) = 3\hat{y}^2 - 1$	wie P_1^{disc} und zusätzlich $\hat{\phi}_4(\hat{x}, \hat{y}) = \frac{5}{4}(3\hat{x}^2 - 1)$ $\hat{\phi}_5(\hat{x}, \hat{y}) = 9\hat{x}\hat{y}$ $\hat{\phi}_6(\hat{x}, \hat{y}) = \frac{5}{4}(3\hat{y}^2 - 1)$
P_3^{disc}	wie P_2^{disc} und zusätzlich $w_7(\hat{x}, \hat{y}) = 5\hat{x}^3 - 3\hat{x}$ $w_8(\hat{x}, \hat{y}) = 3\hat{x}^2\hat{y} - \hat{y}$ $w_9(\hat{x}, \hat{y}) = 3\hat{y}^2\hat{x} - \hat{x}$ $w_{10}(\hat{x}, \hat{y}) = 5\hat{y}^3 - 3\hat{y}$	wie P_2^{disc} und zusätzlich $\hat{\phi}_7(\hat{x}, \hat{y}) = \frac{7}{4}\hat{x}(5\hat{x}^2 - 3)$ $\hat{\phi}_8(\hat{x}, \hat{y}) = \frac{15}{4}\hat{y}(3\hat{x}^2 - 1)$ $\hat{\phi}_9(\hat{x}, \hat{y}) = \frac{15}{4}\hat{x}(3\hat{y}^2 - 1)$ $\hat{\phi}_{10}(\hat{x}, \hat{y}) = \frac{7}{4}\hat{y}(5\hat{y}^2 - 3)$
Q_1^{disc}	$w_1(\hat{x}, \hat{y}) = 1$ $w_2(\hat{x}, \hat{y}) = \hat{x}$ $w_3(\hat{x}, \hat{y}) = \hat{y}$ $w_4(\hat{x}, \hat{y}) = \hat{x}\hat{y}$	$\hat{\phi}_1(\hat{x}, \hat{y}) = 1$ $\hat{\phi}_2(\hat{x}, \hat{y}) = \hat{x}$ $\hat{\phi}_3(\hat{x}, \hat{y}) = \hat{y}$ $\hat{\phi}_4(\hat{x}, \hat{y}) = \hat{x}\hat{y}$
Q_2^{disc}	wie Q_1^{disc} und zusätzlich $w_5(\hat{x}, \hat{y}) = \frac{15}{4}\hat{x}^2 - \frac{5}{4}$ $w_6(\hat{x}, \hat{y}) = \frac{15}{4}\hat{y}^2 - \frac{5}{4}$ $w_7(\hat{x}, \hat{y}) = \hat{x}\left(\frac{15}{4}\hat{y}^2 - \frac{5}{4}\right)$ $w_8(\hat{x}, \hat{y}) = \hat{y}\left(\frac{15}{4}\hat{x}^2 - \frac{5}{4}\right)$ $w_9(\hat{x}, \hat{y}) = \left(\frac{15}{4}\hat{x}^2 - \frac{5}{4}\right)\left(\frac{15}{4}\hat{y}^2 - \frac{5}{4}\right)$	wie Q_1^{disc} und zusätzlich $\hat{\phi}_5(\hat{x}, \hat{y}) = -\frac{1}{2} + \frac{3}{2}\hat{x}^2$ $\hat{\phi}_6(\hat{x}, \hat{y}) = -\frac{1}{2} + \frac{3}{2}\hat{y}^2$ $\hat{\phi}_7(\hat{x}, \hat{y}) = \frac{1}{2}\hat{x}(3\hat{y}^2 - 1)$ $\hat{\phi}_8(\hat{x}, \hat{y}) = \frac{1}{2}\hat{y}(3\hat{x}^2 - 1)$ $\hat{\phi}_9(\hat{x}, \hat{y}) = \frac{1}{4}(3\hat{x}^2 - 1)(3\hat{y}^2 - 1)$
Q_3^{disc}	wie Q_2^{disc} und zusätzlich $w_{10}(\hat{x}, \hat{y}) = \frac{35}{4}\hat{x}^3 - \frac{21}{4}\hat{x}$ $w_{11}(\hat{x}, \hat{y}) = \frac{35}{4}\hat{y}^3 - \frac{21}{4}\hat{y}$ $w_{12}(\hat{x}, \hat{y}) = \hat{y}\left(\frac{35}{4}\hat{x}^3 - \frac{21}{4}\hat{x}\right)$ $w_{13}(\hat{x}, \hat{y}) = \hat{x}\left(\frac{35}{4}\hat{y}^3 - \frac{21}{4}\hat{y}\right)$ $w_{14}(\hat{x}, \hat{y}) = \left(\frac{15}{4}\hat{y}^2 - \frac{5}{4}\right)\left(\frac{35}{4}\hat{x}^3 - \frac{21}{4}\hat{x}\right)$ $w_{15}(\hat{x}, \hat{y}) = \left(\frac{15}{4}\hat{x}^2 - \frac{5}{4}\right)\left(\frac{35}{4}\hat{y}^3 - \frac{21}{4}\hat{y}\right)$ $w_{16}(\hat{x}, \hat{y}) = \left(\frac{35}{4}\hat{x}^3 - \frac{21}{4}\hat{x}\right)\left(\frac{35}{4}\hat{y}^3 - \frac{21}{4}\hat{y}\right)$	wie Q_2^{disc} und zusätzlich $\hat{\phi}_{10}(\hat{x}, \hat{y}) = \frac{5}{2}\hat{x}^3 - \frac{3}{2}\hat{x}$ $\hat{\phi}_{11}(\hat{x}, \hat{y}) = \frac{5}{2}\hat{y}^3 - \frac{3}{2}\hat{y}$ $\hat{\phi}_{12}(\hat{x}, \hat{y}) = \frac{1}{2}\hat{x}\hat{y}(5\hat{x}^2 - 3)$ $\hat{\phi}_{13}(\hat{x}, \hat{y}) = \frac{1}{2}\hat{x}\hat{y}(5\hat{y}^2 - 3)$ $\hat{\phi}_{14}(\hat{x}, \hat{y}) = \frac{1}{4}\hat{x}(5\hat{x}^2 - 3)(3\hat{y}^2 - 1)$ $\hat{\phi}_{15}(\hat{x}, \hat{y}) = \frac{1}{4}\hat{y}(5\hat{y}^2 - 3)(3\hat{x}^2 - 1)$ $\hat{\phi}_{16}(\hat{x}, \hat{y}) = \frac{1}{4}\hat{x}\hat{y}(5\hat{x}^2 - 3)(5\hat{y}^2 - 3)$

Basisfunktionen und Gewichtsfunktionen der Funktionale für P_p^{disc} - und Q_p^{disc} -Elemente auf Rechtecken ($p = 1, 2, 3$).

(Tabelle 3.2)

3.2.4 Abschätzung des Interpolationsfehlers

Bei der Verwendung Finiter–Elemente stellt sich die Frage, wie gut eine Funktion prinzipiell durch Funktionen des gewählten Finite–Elemente–Raumes approximiert werden kann. Um diese Frage zu beantworten, definiert man Interpolationsoperatoren auf den Räumen und schätzt dann den Interpolationsfehler ab. Die Abschätzungen haben die Form

$$\|u - Iu\|_{X(\mathcal{A})} \leq Ch^s \|u\|_{Y(\mathcal{B})}. \quad (3.2.14)$$

Dabei ist u die exakte Lösung, Iu ihre Interpolation in einem Finite–Elemente–Raum, I der zugehörige Interpolationsoperator, $\mathcal{A}, \mathcal{B} \subset \Omega$, $X(\mathcal{A})$ sowie $Y(\mathcal{B})$ geeignete Räume mit Normen $\|\cdot\|_{X(\mathcal{A})}$, $\|\cdot\|_{Y(\mathcal{B})}$, $s \in \mathbb{R}^+$, $C > 0$ gitterunabhängige Konstanten und h eine gitterabhängige Größe wie zum Beispiel das Maximum der Durchmesser der Gitterzellen oder das Maximum der Längen der Kante.

Die Aussage der Interpolationsfehlerabschätzung ist dabei, dass der Fehler $\|u - Iu\|_{X(\mathcal{A})}$ mit der Potenz s gegen Null geht, wenn h gegen Null geht, das heißt, mit zunehmend feineren Gittern. Die Größe s wird die Konvergenzordnung genannt. Die folgenden Abschätzungen sind in Bezug auf die Konvergenzordnung optimal, das heißt, andere Interpolationsoperatoren liefern im Allgemeinen keine bessere Konvergenzordnung $\tilde{s} > s$. Hingegen ist die Konstante C oftmals suboptimal, da sie bei geschickter Wahl des Interpolationsoperators verkleinert werden kann.

Die Güte der Approximation hängt auch von der Zerlegung des Gebiets ab. Gute Fehlerabschätzungen lassen sich nur für hinreichend reguläre Zerlegungen beweisen. Für die hier angegebenen Fehlerabschätzungen genügt es, quasi–uniforme Zerlegungen zu fordern. Die in dieser Arbeit verwendeten Gitter sind quasi–uniform.

Definition 3.10. (Quasi–uniforme Zerlegung)

Eine zulässige Zerlegung \mathcal{T}_h von $\Omega \subset \mathbb{R}^2$ heißt quasi–uniform, wenn

- (i) es eine Konstante $C > 0$ gibt, so dass für den Umkreisradius ρ_K und Inkreisradius σ_K jeder Zelle K gilt:

$$\max_{K \in \mathcal{T}_h} \frac{\sigma_K}{\rho_K} \leq C \quad (\text{Formregularität}), \quad (3.2.15)$$

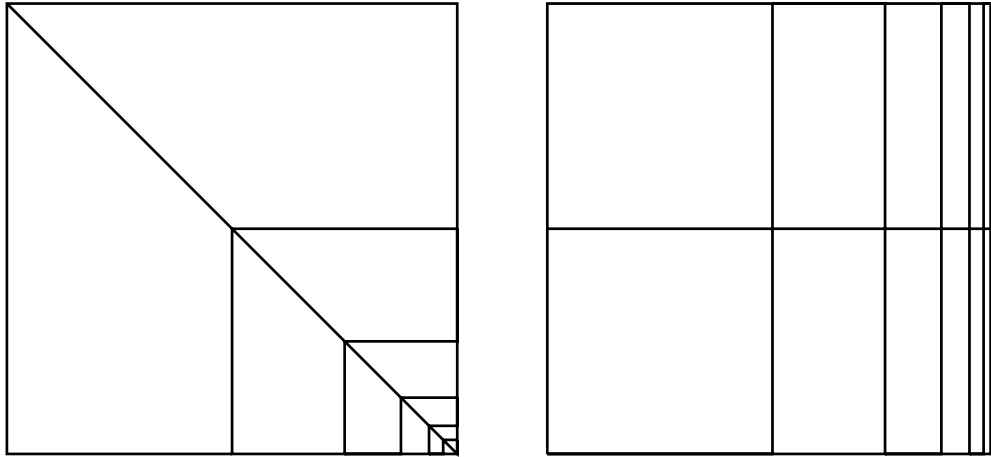
- (ii) wenn alle Zellen von der gleichen Größenordnung sind, das heißt, wenn es eine Konstante \tilde{c} gibt, so dass

$$\min_{K \in \mathcal{T}_h} h_K \leq \tilde{c} \max_{K \in \mathcal{T}_h} h_K \quad \text{mit } h_K := \text{diam}(K) \quad (\text{Größenregularität}). \quad (3.2.16)$$

Für dreidimensionale Gebiete Ω gilt obige Definition auch, wenn man den Umkreisradius ρ_K durch den Radius der kleinsten Kugel, welche die Gitterzelle K enthält ersetzt und ebenso den Inkreisradius σ_K durch den Radius der größten Kugel, welchen noch ganz in K enthalten ist. Abbildung 3.6 zeigt Gebietszerlegungen, welche nicht quasi–uniform sind.

Die Abschätzungen der Interpolationsfehler, welche wir nun angeben werden, gelten sowohl in zwei Dimensionen für Zerlegungen in Dreiecke oder Rechtecke als auch in drei Dimensionen für Zerlegungen in Tetraeder oder Hexaeder. Wir schreiben daher wieder V_h für den Finite–Elemente–Raum.

Als erstes betrachten wir den Lagrange–Interpolationsoperator I_h^p . Dieser ist auf K folgendermaßen definiert: Betrachte die lokalen Funktionale $\{\Phi_j\}_{j=1}^N$ des Lagrange–Elemente–



Beispiele nicht-quasi-uniformerer Zerlegungen:

Denkt man sich die Gebietszerlegung wie in der Abbildung angedeutet fortgesetzt, so ist die linke Zerlegung zwar form- aber nicht größenregulär. Die rechte Zerlegung ist weder form- noch größenregulär.

(Abbildung 3.6)

Raumes $V_h(\Omega)$ auf der Zelle K . Nun definieren wir

$$I_h^p(u)(x) = \sum_{j=1}^N a_j \phi_j(x), \quad (3.2.17)$$

wobei ϕ_j die zugehörigen Basisfunktionen auf K sind und die Koeffizienten a_j durch $a_j := \Phi_j(u)$ bestimmt sind. Nach Konstruktion erhält der Lagrange-Interpolationsoperator insbesondere homogene Randbedingungen.

Lemma 3.11. (Lagrange-Interpolation)

Seien \mathcal{T}_h eine quasi-uniforme Zerlegung von $\Omega \subset \mathbb{R}^d$, $d \geq 2$, und $I_h^p : H^k(K) \rightarrow V_h(K)$, $k > d/2$, der Lagrange-Interpolationsoperator auf $K \in \mathcal{T}_h$. Dann gilt für $u \in H^k(K)$ die lokale Interpolationsfehlerabschätzung:

$$\|u - I_h^p u\|_{H^m(K)} \leq C \frac{h_K^{l-m}}{p^{k-m}} \|u\|_{H^k(K)}, \quad (3.2.18)$$

mit $0 \leq m \leq l := \min(p+1, k)$ und der Konstanten $C > 0$.

Beweis.

Siehe (Hou01) Abschnitt 4. Die Abhängigkeit von der Elementordnung p folgt in Kombination mit (Sch98) Korollar 4.68. \square

Statt einer Abschätzung bezüglich der Sobolevnormen kann man auch eine Abschätzung bezüglich der Halbnormen treffen:

Lemma 3.12.

Sei \mathcal{T}_h eine quasi–uniforme Zerlegung von $\Omega \subset \mathbb{R}^d$, $d \geq 2$, und sei $I_h^{k-1} : H^k(K) \rightarrow V_h(K)$, $k > d/2$, der Lagrange–Interpolationsoperator auf $K \in \mathcal{T}_h$. Dann gilt für $u \in H^k(K)$ die lokale Interpolationsfehlerabschätzung:

$$|u - I_h^{k-1}u|_{H^m(K)} \leq Ch_K^{k-m} |u|_{H^k(K)}, \quad (3.2.19)$$

mit $0 \leq m \leq k$ und der Konstanten $C > 0$.

Beweis.

Siehe (Cia72) Theorem 2. \square

Man beachte, dass in Lemma 3.12 der Polynomgrad nicht mehr beliebig sein darf wie in Lemma 3.11, sondern auf $k - 1$ fixiert wird.

Eine weitere Fehlerabschätzung, die wir benötigen werden, ist die des Fehlers der globalen L^2 -Projektion π_h . Diese ist auf V_h definiert durch $\pi_h(u) = u_h$, wobei u_h definiert ist über

$$(u - u_h, v_h)_{L^2(\Omega)} = 0 \quad \forall v_h \in V_h(\Omega). \quad (3.2.20)$$

Lemma 3.13. (L^2 -Projektion)

Sei \mathcal{T}_h eine quasi–uniforme Zerlegung von $\Omega \subset \mathbb{R}^d$, $d \geq 2$, und sei $\pi_h : L^2(\Omega) \rightarrow V_h(\Omega)$ die globale L^2 -Projektion. Dann gelten für $u \in H^k(\Omega)$, $k \geq 2$, die globalen Interpolationsfehlerabschätzungen

$$\|u - \pi_h u\|_{H^m(\Omega)}^2 \leq C \sum_{K \in \mathcal{T}_h} h_K^{2(l-m)} |u|_{H^k(K)}, \quad (3.2.21)$$

$$\|u - \pi_h u\|_{L^2(\partial\Omega)}^2 \leq C \sum_{K \in \mathcal{T}_h} h_K^{2l} |u|_{H^k(\partial\Omega)} \quad (3.2.22)$$

mit $0 \leq m \leq l := \min(p + 1, k)$ und der Konstanten $C > 0$, beziehungsweise für $u \in W^{k+1,p}(K)$ mit $W^{k+1,p}(K) \hookrightarrow W^{m,q}(K)$

$$\|u - \pi_h u\|_{W^{m,q}(K)} \leq C |K|^{\frac{1}{q} - \frac{1}{p}} h_K^{k-m+1} |u|_{W^{k+1,p}(K)}. \quad (3.2.23)$$

Beweis.

Siehe (Roo08) Lemma 3.83 und Bemerkung 3.84. Der Beweis benutzt die Interpolationsfehler-Abschätzung aus Lemma 3.12. Für die zweite Abschätzung siehe (Cia78), Theorem 3.1.5. \square

Bemerkung 3.14.

Im Gegensatz zur Lagrange–Interpolation erhält die L^2 -Projektion im Allgemeinen keine homogenen Randbedingungen.

3.2.5 Spur- und inverse Ungleichungen

Zur Abschätzung der Fehler von Finite-Elemente-Methoden ist es teilweise notwendig, Ableitungen von Finite-Elemente-Funktionen nach oben gegen niedrigere Ableitungen oder gegen die Funktionswerte selbst abzuschätzen. Solche Abschätzungen werden als inverse Ungleichungen bezeichnet.

Lemma 3.15. (Inverse Ungleichung)

Sei \mathcal{T}_h eine quasi-uniforme Zerlegung von $\Omega \subset \mathbb{R}^d$, $d \geq 2$. Dann gibt es eine Konstante $\mu_{\text{inv}} > 0$, welche nur von der Formregularitäts-Konstanten aus Definition 3.10 abhängt, so dass für alle $u_h \in V_h(\Omega)$ und alle Gitterzellen $K \in \mathcal{T}_h$ gilt:

$$\|\nabla u_h\|_{L^2(K)} \leq \mu_{\text{inv}} h_K^{-1} p^2 \|u_h\|_{L^2(K)} \quad (3.2.24)$$

Beweis.

Siehe (Bra07) Kapitel 2, Satz 6.8. Die Abhängigkeit von der Elementordnung p folgt in Kombination mit (Sch98) Theorem 4.76. \square

Für die von uns verwendeten Methoden ist es in manchen Fällen nützlich, die Norm der Spur einer Sobolev-Funktion auf dem Rand einer Gitterzelle nach oben gegen die Norm im Inneren der Zelle abzuschätzen zu können. Eine solche Abschätzung nennt man Spurgleichung.

Lemma 3.16. (Spurgleichungen)

Sei \mathcal{T}_h eine quasi-uniforme Zerlegung von $\Omega \subset \mathbb{R}^d$, $d \geq 2$. Dann gelten für alle $K \in \mathcal{T}_h$ und $v \in W^{1,p}(K)$, $1 \leq p \leq \infty$, $1/p + 1/q = 1$, die Spurgleichungen

$$\|v\|_{L^p(\partial K)} \leq C \|v\|_{L^p(K)}^{\frac{1}{q}} \|v\|_{W^{1,p}(K)}^{\frac{1}{p}} \quad (3.2.25)$$

$$\leq C \frac{h_K^{\frac{1}{q}}}{p} \|v\|_{W^{1,p}(K)} + C \frac{h_K^{-\frac{1}{p}}}{q} \|v\|_{L^p(K)} \quad (3.2.26)$$

mit einer Konstanten $C > 0$. Ist $v \in H^1(K)$, so gilt

$$\|v\|_{L^2(\partial K)}^2 \leq C \left(h_K^{-1} \|v\|_{L^2(K)}^2 + h_K \|v\|_{H^1(K)}^2 \right). \quad (3.2.27)$$

Beweis.

Die Ungleichung 3.2.25 wird in (Bre94), Theorem 1.6.6 bewiesen. Die zweite Ungleichung folgt mit Hilfe der Youngschen Ungleichungen. Den Beweis zur Ungleichung (3.2.27) findet man in (Tho97), Seite 26. \square

Als Folge dieser beiden Lemmata ergibt sich

Korollar 3.17.

Für jede Funktion $v \in V_h$ gilt für $1 < p < \infty$ die Spurgleichung

$$\|\partial_n v\|_{L^p(\partial K)} \leq C h^{-\frac{1}{p}} \|\nabla v\|_{L^p(K)}. \quad (3.2.28)$$

Mit Hilfe der Spurgleichungen lässt sich auch eine Abschätzung für den Fehler der L^2 -Projektion auf dem Rand einer Zelle angeben:

Korollar 3.18.

Sei \mathcal{T} eine quasi–uniforme Zerlegung und sei $u \in W^{k+1,p}(\mathcal{T})$. Dann gilt für die L^2 -Projektion die Abschätzung

$$|u - \pi_h u|_{W^{m,p}(\partial K)} \leq C h_K^{k-m+\frac{1}{q}} |u|_{W^{k+1,p}(K)}. \quad (3.2.29)$$

Beweis.

Die Abschätzung folgt unmittelbar aus (3.2.23) und (3.2.25). \square

3.2.6 Allgemeine Fehlerabschätzungen

In diesem Abschnitt soll allgemein das Konzept erläutert werden, mit dessen Hilfe der Fehler $u - u_h$ zwischen der exakten Lösung u und der mittels eines Galerkin–Verfahrens gewonnenen Lösung u_h abgeschätzt wird.

Bei der Behandlung von Variationsproblemen mittels Finiter–Element–Methoden zeigt sich oftmals, dass bei der unmittelbar aus der starken Formulierung einer Differentialgleichung gewonnene schwache Formulierung Probleme auftreten. Diese können sich in mangelnder Stabilität oder auch schlechter Konvergenz der Verfahren zeigen. Um dies zu verbessern, wird die zu Grunde liegende Bilinearform a und die Linearform f modifiziert, beispielsweise durch Addition von Termen, welche sich für die exakte Lösung zu Null ergeben. Wir bezeichnen die neue Bilinearform mit a_h und die neue Linearform mit f_h .

Definition 3.19. Konsistenz

Ein Galerkin–Verfahren

$$a_h(u, v) = f_h(v), \quad (3.2.30)$$

welches aus (3.1.1) durch Modifikation der Bilinearform a und der Linearform f hervorgeht, heißt konsistent, falls für die Lösung u von (3.1.1) gilt:

$$a_h(u, v) = f_h(v) \quad \forall v \in V_h \oplus V. \quad (3.2.31)$$

Wir wollen nun Abschätzungen für den Fehler $u - u_h$ angeben.

Satz 3.20.

Sei $u \in H^s(\mathcal{T})$ eine Lösung von (3.2.31) und sei $u_h \in V_h$ die zugehörige Näherungslösung. Weiter seien folgende Voraussetzungen erfüllt:

- (i) Es existiert eine Norm $\|\cdot\|_A$ auf V_h , so dass gilt

$$\|v\|_A^2 \leq C a_h(v, v) \quad \forall v \in V_h \quad (3.2.32)$$

mit einer Konstanten C , d.h. $a_h(\cdot, \cdot)$ ist V_h -elliptisch.

- (ii) Es existiert ein Projektionsoperator $\Pi_h : V \rightarrow V_h$, für welchen die Fehlerabschätzung

$$\|v - \Pi_h v\|_A \leq c h^{s-1} |v|_{H^s(\mathcal{T})} \quad (3.2.33)$$

für alle $v \in H^s(\mathcal{T})$ für eine Konstante c erfüllt ist.

(iii) Der Projektionsfehler $u - \Pi_h u$ genügt der Abschätzung

$$a_h(u - \Pi_h u, v) \leq \tilde{c} h^{s-1} |u|_{H^s(\mathcal{T})} \|v\|_A \quad \forall v \in V_h \quad (3.2.34)$$

mit einer Konstanten \tilde{c} .

Dann gilt für den Fehler $u - u_h$ die Abschätzung

$$\|u - u_h\|_A \leq \hat{c} h^{s-1} |u|_{H^s(\mathcal{T})}, \quad (3.2.35)$$

wobei \hat{c} konstant ist.

Alle hierbei auftretenden Konstanten sind unabhängig von h .

Beweis.

Nach der Dreiecksungleichung gilt

$$\|u - u_h\|_A \leq \|u - \Pi_h u\|_A + \|\Pi_h u - u_h\|_A$$

Den ersten Term schätzen wir mit Hilfe von Voraussetzung (ii) ab. Für den zweiten beachten wir, dass $\Pi_h u - u_h \in V_h$ gilt. Somit ist Voraussetzung (i) anwendbar. Zusammen mit Voraussetzung (iii) und der Galerkin-Orthogonalität folgt dann

$$\begin{aligned} \frac{1}{c} \|\Pi_h u - u_h\|_A^2 &\leq a_h(u - \Pi_h u, u - \Pi_h u) \\ &\leq \tilde{c} h^{s-1} |u|_{H^s(\mathcal{T})} \|\Pi_h u - u_h\|_A. \end{aligned}$$

Setzen wir dies nun in die Dreiecksungleichung ein, so folgt die Behauptung. \square

Bemerkung 3.21.

- (i) Ist die Bilinearform a_h symmetrisch, so kann $\|v\|_A = (a_h(v, v))^{1/2}$ gewählt werden.
- (ii) Für elliptische Probleme und bei Verwendung von Polynomen vom Grad p als Ansatzfunktionen ist Abschätzung (ii) typischerweise für $s \leq p + 1$ erfüllt.

Um auch Abschätzungen in der L^2 -Norm oder schwächeren Normen zu erhalten, kann ein Dualitätsargument benutzt werden, welches von Aubin und Nitsche aufgestellt wurde.

Satz 3.22.

Seien die Voraussetzungen von Satz 3.20 erfüllt. Seien $t \in \mathbb{N}_0$, $\tilde{\Omega} \subset \Omega$ und seien weiterhin folgende Voraussetzungen erfüllt:

- (i) Für beliebige $f \in H^t(\Omega)$ sei z die Lösung des dualen Problems

$$a(\phi, z) = (f, \phi) \quad \forall \phi \in V$$

und erfülle die Abschätzung

$$|z|_{H^{t+2}(\Omega)} \leq c \|f\|_{H^t(\Omega)}.$$

- (ii) Die Bilinearform a_h erfülle für z und $v \in V_h$ die Abschätzung

$$a_h(v, z - \Pi_h z) \leq \tilde{c} \|v\|_A h^{t+1} |z|_{H^{t+2}(\mathcal{T})}.$$

(iii) Für $w \in H^s(\mathcal{T})$ und $z \in H^{t+2}(\mathcal{T})$ gelte die Abschätzung

$$a_h(w - \Pi_h w, z - \Pi_h z) \leq \hat{c} h^{s-1} |w|_{H^s(\mathcal{T})} h^{t+1} |z|_{H^{t+2}(\mathcal{T})}.$$

Dann gilt für die $H^{-t}(\tilde{\Omega})$ -Norm des Fehler $u - u_h$ die Abschätzung

$$\|u - u_h\|_{H^{-t}(\tilde{\Omega})} \leq Ch^{s+t} |u|_{H^s(\mathcal{T})}.$$

Beweis.

siehe (Kan07), Theorem 1.3.14 \square

Bemerkung 3.23.

- (i) Im Fall $t = 0$ liefert der Satz eine Abschätzung des Fehlers in der L^2 -Norm.
- (ii) Im Fall $t = 0$ wird Voraussetzung (i) als elliptische Regularität bezeichnet. Für weitere Details zu dieser Voraussetzung siehe (Gri85).

Kapitel 4

Die Konvektions–Diffusions–Gleichung

Nachdem in den vorherigen Kapiteln die Grundlagen aus Funktionalanalysis, Variationsrechnung und der allgemeinen Theorie der Finiten–Element–Methoden gelegt worden sind, sollen diese nun zur Behandlung der Konvektions–Diffusions–Gleichung benutzt werden. Dazu betrachten wir zunächst die Konvektions–Diffusions–Gleichung in ihrer starken Formulierung um daraus dann die schwache Formulierung und das Standard–Galerkin–Verfahren herzuleiten. In der Praxis stellt sich leider heraus, dass die mit Hilfe des Standard–Galerkin–Verfahrens berechneten Lösungen oft unphysikalische Oszillationen zeigen. Um dies zu beheben, muss das Verfahren durch Hinzunahme weiterer Terme stabilisiert werden. Eine häufig verwendete Methode ist dabei das Streamline–Diffusion–Verfahren, welches wir hier kurz vorstellen werden, da es uns im nächsten Kapitel zum Vergleich dient. Das Hauptaugenmerk dieser Arbeit liegt jedoch auf den Discontinuous–Galerkin–Verfahren. Diese stellen eine weitere Möglichkeit zur Behandlung der Konvektions–Diffusions–Gleichung mit Hilfe stabilisierender Terme dar. Wir werden ein Discontinuous–Galerkin–Verfahren in Form einer Interior–Penalty–Methode vorstellen und einige Aussagen zu Fehlerabschätzungen und Konvergenzordnungen zeigen. Abschließend stellen wir als Alternative die Flussformulierung der Discontinuous–Galerkin–Verfahren vor. Die numerische Umsetzung des Verfahrens ist Gegenstand des nächsten Kapitels.

4.1 Starke Formulierung und analytische Aussagen

Sei Ω ein beschränktes offenes Gebiet in \mathbb{R}^d , $d = 2, 3$ mit Rand Γ . Wir betrachten folgendes Randwertproblem:

Finde eine Funktion $u : \bar{\Omega} \rightarrow \mathbb{R}$ mit:

$$-\epsilon \Delta u + b(x) \cdot \nabla u + c(x) u = f(x) \quad \text{in } \Omega, \quad (4.1.1a)$$

$$u = u^D \quad \text{auf } \Gamma, \quad (4.1.1b)$$

mit $\epsilon > 0$ und ausreichend glatten Funktionen $b : \Omega \rightarrow \mathbb{R}^2$ sowie $c, f : \Omega \rightarrow \mathbb{R}$, $c(x) > 0 \forall x \in \Omega$. Gleichung (4.1.1a) heißt skalare Konvektions–Diffusions–Gleichung. Wir haben sie hier mit nicht–homogenen Dirichlet–Randbedingungen versehen.

Die Bezeichnung skalare Konvektions–Diffusions–Gleichung sagt schon aus, was durch diese Gleichung beschrieben wird: Die Verteilung einer skalaren Größe, wie beispielsweise der Temperatur, welche sich durch Diffusion (erster Term in (4.1.1a)), Konvektion (zweiter Term in (4.1.1a)) und Erzeugung oder Vernichtung (dritter Term in (4.1.1a)), beispielsweise durch

chemische Reaktionen ändern kann. ϵ ist dann die Diffusionskonstante, b ein Geschwindigkeitsfeld, durch welches u advektiv transportiert wird, und c beschreibt Verteilung und Stärke von Quellen und Senken. Die rechte Seite von (4.1.1a) gibt den Einfluss von äußeren Kräften wie beispielsweise der Schwerkraft durch die Kraftdichte f wieder.

Man kann drei Fälle der Konvektions–Diffusions–Gleichung unterscheiden:

(i) Diffusions–dominanter Fall:

$$\epsilon > \|b\|_{L^\infty(\Omega)} \text{ und } \epsilon \gtrsim \|c\|_{L^\infty(\Omega)}. \quad (4.1.2)$$

(ii) Konvektions–dominanter Fall:

$$\epsilon \ll \|b\|_{L^\infty(\Omega)} \text{ und } \|b\|_{L^\infty(\Omega)} \gtrsim \|c\|_{L^\infty(\Omega)}. \quad (4.1.3)$$

(iii) Reaktions–dominanter Fall:

$$\|c\|_{L^\infty(\Omega)} \gtrsim \epsilon \text{ und } \|c\|_{L^\infty(\Omega)} > \|b\|_{L^\infty(\Omega)} \quad (4.1.4)$$

Typisch für das Verhalten der Lösung u im konvektions–dominanten Fall sind so genannte Grenzschichten. Grenzschichten sind dabei kleine Teilgebiete von Ω , innerhalb derer sich die Lösung rasch ändert, also starke Gradienten besitzt. Im diffusions–dominanten Fall treten meist keine Grenzschichten auf; im reaktions–dominanten Fall kann es Grenzschichten geben. Man kann drei Typen von Grenzschichten unterscheiden:

(i) Parabolische Grenzschichten:

Parabolische Grenzschichten treten an Dirichlet–Rändern auf, welche parallel zur Strömung liegen, das heißt, wo $n(x) \cdot b(x) = 0$ gilt. $n(x)$ ist hierbei der äußere Normalenvektor zu Ω an Γ . Ihre Breite ist von der Ordnung $\mathcal{O}(\sqrt{\epsilon} \ln(1/\epsilon))$. Ein Beispiel für parabolische Grenzschichten gibt der Graph auf der linken Seite von Abbildung 4.1 an. Gezeigt wird die Lösung der Konvektions–Diffusionsgleichung zu den Vorgaben $\epsilon = 10^{-8}$, $b = (1, 0)^T$, $c = 0$, $f = 1$ auf $\Omega = [0, 1]^2$ und $u = 0$ auf Γ . Gemäß dieser Vorgaben erhalten wir parabolische Grenzschichten bei $y = 0$ und $y = 1$.

(ii) Exponentielle Grenzschichten

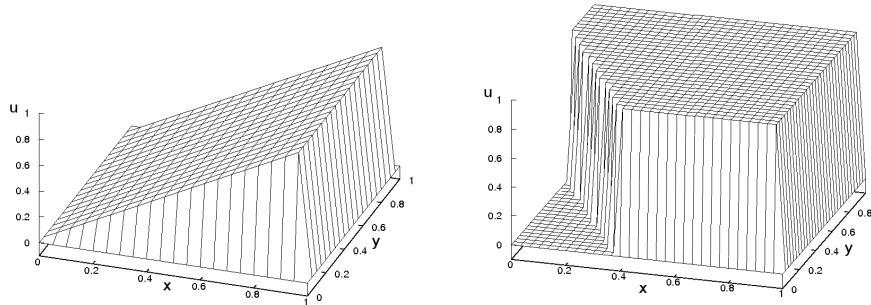
Exponentielle Grenzschichten treten an Dirichlet–Rändern im Ausflussbereich der Strömung auf, das heißt, wo $n(x) \cdot b(x) > 0$ gilt. Ihre Breite ist von der Ordnung $\mathcal{O}(\epsilon \ln(1/\epsilon))$. Damit sind sie noch schmaler als parabolische Grenzschichten, wodurch ihre numerische Behandlung weiter erschwert wird. Ein Beispiel für eine exponentielle Grenzschicht ist ebenfalls auf der linken Seite von Abbildung 4.1 zu sehen. Die exponentielle Grenzschicht liegt hier bei $x = 1$. Auf der rechten Seite der Abbildung ist ein weiteres Beispiel für eine exponentielle Grenzschicht gezeigt. Betrachtet wird dort die Konvektions–Diffusions–Gleichung mit $\epsilon = 10^{-8}$, $b = (\cos(-\pi/3), \sin(-\pi/3))^T$, $c = 0$, $f = 0$ auf $\Omega = [0, 1]^2$ und den Randwerten

$$u = \begin{cases} 0 & \text{für } (x, y) \in \Gamma \text{ und } x = 1 \text{ oder } y \leq 0.7, \\ 1 & \text{für andere Punkte auf } \Gamma. \end{cases}$$

Die exponentiellen Grenzschichten liegen hier bei $y = 1$ oder ungefähr ab $x \approx 0.4$ bei $y = 0$.

(iii) Innere Grenzschichten

Auch im Inneren des Gebiets Ω können Grenzschichten auftreten. Diese werden zum Beispiel durch Unstetigkeiten in den Dirichlet-Randwerten verursacht. Ihre Struktur ähnelt der von parabolischen Grenzschichten, weswegen sie manchmal auch als innere parabolische Grenzschichten bezeichnet werden. Der Graph auf der rechten Seite von Abbildung 4.1 zeigt eine innere Grenzschicht. Diese startet bei $(x, y) = (0, 0.7)$ und verläuft in Richtung des Advektionsvektors b . Die Lage der inneren Grenzschichten kann durch die Problemdata plausibel gemacht werden (Sty05).



Lösung u der Konvektions–Diffusions–Gleichung mit Grenzschichten (übernommen von (Joh07)). Für weitere Details siehe Text.

(Abbildung 4.1)

An Dirichlet-Rändern im Einströmbereich ($n(x) \cdot b(x) < 0$) treten üblicherweise keine Grenzschichten auf. Für weitere Details zu Grenzschichten verweisen wir auf (Heg08).

Wir wollen noch einen Satz zur Existenz einer klassischen Lösung zu (4.1.1) angeben.

Satz 4.1.

Sei Ω ein beschränktes Gebiet mit Lipschitz-stetigem Rand Γ und seien b , c und f Hölder-stetig auf $\bar{\Omega}$ mit $c(x) \geq 0$ für alle $x \in \bar{\Omega}$. Dann besitzt das Problem (4.1.1) für den Fall $u^D \equiv 0$ eine eindeutige Lösung u aus $C(\bar{\Omega}) \cap C^2(\Omega)$.

Beweis.

Siehe (Mic77). \square

Bemerkung 4.2.

- (i) Satz 4.1 kann auf den Fall inhomogener Dirichlet-Randbedingungen verallgemeinert werden, solange u^D hinreichend regulär ist. Dazu setzt man u^D vom Gebietsrand Γ ausgehend auf eine Funktion $\tilde{u}^D \in C(\bar{\Omega}) \cap C^2(\Omega)$ fort. Damit wird die Hilfsfunktion $\tilde{u} = u - \tilde{u}^D$ definiert. Setzt man $u = \tilde{u} + \tilde{u}^D$ in (4.1.1a) ein, so erhält man zusätzliche Terme, welche auf die rechte Seite der Gleichung gebracht werden. Die so erhaltene Konvektions–Diffusions–Gleichung für \tilde{u} ist mit homogenen Dirichlet-Randbedingungen versehen, wodurch sie nach Satz 4.1 eindeutig lösbar ist. Dieses Vorgehen bezeichnet man auch als Homogenisierung. Ob die Homogenisierung gelingt, hängt davon ab, ob eine Fortsetzung \tilde{u}^D existiert. Die Existenz der Fortsetzung ist dabei offenbar nur von dem Gebietsrand und der Randbedingung u^D abhängig. Eine ausführliche Diskussion zur Homogenisierung von Differentialgleichungen findet sich beispielsweise in (Cio99).

- (ii) Weitere Randbedingungen wie Neumann- oder Robin–Randbedingungen werden in (Wig70) untersucht. Auch Situationen, in denen der Typ der Randbedingung auf Γ variiert werden hier besprochen.
- (iii) Eine höhere Regularität der Lösung u (beispielsweise $u \in C^{2,s}(\bar{\Omega})$) lässt sich erreichen, wenn die Ecken des Gebiets zusätzliche Kompatibilitätsbedingungen erfüllen (siehe hierzu (Azz80) oder (Gri85)).

Weitere analytische Eigenschaften der Konvektions–Diffusions–Gleichung lassen sich beispielsweise in (Roo08) oder (Goe83) nachschlagen.

4.2 Schwache Formulierung

In Kapitel 2.2 haben wir bereits gesagt, dass Existenz- und Eindeigkeitssätze wie Satz 4.1 in der Praxis oft nicht anwendbar sind, da die Voraussetzungen nicht erfüllt sind. Wir wollen daher die Konvektions–Diffusions–Gleichung (4.1.1) nun, wie in Kapitel 2.2.3 bereits angedeutet, in ihre schwache Formulierung als Variationsproblem überführen. Die Lösung des Variationsproblems ist dann eine Sobolev–Funktion, die im Allgemeinen nicht mehr in $C(\bar{\Omega}) \cap C^2(\Omega)$ liegt. Dafür lassen sich nun allerdings die allgemeineren Konzepte zur Lösbarkeitstheorie von Variationsproblemen anwenden.

Wir gehen aus von (4.1.1), das heißt von der Konvektions–Diffusions–Gleichung in ihrer starken Form und der Vorgabe von Dirichlet–Randwerten auf dem gesamten Gebietsrand Γ . Die Daten des Problems seien dabei als hinreichend regulär vorausgesetzt, das heißt $b, c \in L^\infty(\Omega)$, $f \in L^2(\Omega)$ und $u^D \in L^2(\Gamma)$. Wir multiplizieren nun Gleichung (4.1.1a) mit einer Testfunktion $v \in H_0^1(\Omega)$, integrieren über das Gebiet Ω und wenden auf das Integral $(-\epsilon \Delta u, v)_{L^2(\Omega)}$ partielle Integration an. Damit erhalten wir

$$\epsilon (\nabla u, \nabla v)_{L^2(\Omega)} - \epsilon (n \cdot \nabla u, v)_{L^2(\Gamma)} + (b \cdot \nabla u, v)_{L^2(\Omega)} + (cu, v)_{L^2(\Omega)} = (f, v)_{L^2(\Omega)}. \quad (4.2.1)$$

Da die Testfunktion v auf dem Rand Γ identisch verschwindet, ergibt sich der zweite Term zu Null. Fassen wir die linke Seite nun in der Bilinearform $a(u, v)$ und die rechte Seite in der Bilinearform $f(v)$ zusammen, so erhalten wir als Variationsproblem:

Finde eine Funktion $u \in H^1(\Omega)$ mit $u = u^D$ auf Γ und

$$a(u, v) = f(v) \quad \forall v \in H_0^1(\Omega). \quad (4.2.2)$$

Aus der Herleitung ist ersichtlich, dass jede klassische Lösung von (4.1.1) auch Lösung des schwachen Problems ist. Die Umkehrung gilt im Allgemeinen nicht. Folgender Satz gibt jedoch ein hinreichendes Kriterium:

Satz 4.3.

Seien c, f und u^D stetige Funktionen, $b \in C^1(\bar{\Omega})$, Ω ein Lipschitzgebiet und sei eine Lösung $u \in C^2(\bar{\Omega})$ von (4.2.2) mit ausreichender Regularität bestimmt worden, welche zusätzlich der inhomogenen Dirichlet–Randbedingung $u = u^D$ auf Γ genügt. Dann löst u auch das klassische Randwertproblem (4.1.1).

Beweis.

Der Beweis findet sich in (Gro05), Satz 3.3. \square

Bemerkung 4.4.

Der angegebene Beweis zu Satz 4.3 zeigt, dass dieser auch auf Situationen verallgemeinert werden kann, in denen weitere Randbedingungen auftreten. Dazu müssen dann die in den Randbedingungen vorkommenden Funktionen ebenfalls gewisse Regularitätsbedingungen erfüllen. Zudem ergeben sich in der schwachen Formulierung bei Vorgabe von Neumann- oder Robin-Randbedingungen zusätzliche Terme, welche mit in die Linearform auf der rechten Seite eingehen. Die Verallgemeinerung zu Satz 4.3 sagt dann aus, dass diese Randbedingungen auch von u erfüllt werden. Neumann- und Robin-Randbedingungen zählen daher zu den so genannten natürlichen Randbedingungen.

Um sicherzustellen, dass Dirichlet-Randbedingungen in der schwachen Formulierung erfüllt werden, muss der Ansatzraum auf Funktionen aus H^1 eingeschränkt werden, welche diesen Randbedingungen genügen. Randbedingungen, welche im Ansatz- beziehungsweise im Testraum berücksichtigt werden, bezeichnet man auch als wesentliche Randbedingungen.

Wir wollen nun zeigen, dass das Variationsproblem (4.2.2) eine eindeutige Lösung hat. Der Einfachheit halber beschränken wir uns hier auf homogene Dirichlet-Randbedingungen. Wir zeigen, dass (4.2.2) unter passenden Bedingungen die Voraussetzungen des Satzes von Lax–Milgram erfüllt:

Satz 4.5.

Sei Ω ein Lipschitzgebiet aus \mathbb{R}^d , $d \in \mathbb{N}$, $b, c \in L^\infty(\Omega)$, $f \in L^2(\Omega)$ und

$$c - \frac{1}{2} \nabla \cdot b \geq c_0 > 0 \quad \text{fast überall auf } \Omega. \tag{4.2.3}$$

Dann besitzt das Variationsproblem (4.2.2) versehen mit homogenen Dirichlet-Randbedingungen eine eindeutige Lösung.

Beweis.

Wir zeigen, dass die Voraussetzungen des Satzes von Lax–Milgram (Satz 2.33) erfüllt sind:

1. H_0^1 ist ein Hilbertraum.
2. Das Funktional f ist nach Voraussetzung beschränkt und wegen der Linearität des L^2 -Skalarproduktes linear.
3. Die Bilinearform $a(\cdot, \cdot)$ ist beschränkt. Um dies zu zeigen, schätzen wir die Terme in $a(\cdot, \cdot)$ nach oben ab:

$$\epsilon \left| (\nabla u, \nabla v)_{L^2(\Omega)} \right| \leq \epsilon \|\nabla u\|_{L^2(\Omega)} \|\nabla v\|_{L^2(\Omega)} \leq \|u\|_{H^1(\Omega)} \|v\|_{H^1(\Omega)}, \tag{4.2.4}$$

wobei die Cauchy–Schwarz–Ungleichung benutzt wurde.

Für den zweiten und dritten Term in $a(\cdot, \cdot)$ ergibt sich

$$\begin{aligned} \left| (b \cdot \nabla u + cu, v)_{L^2(\Omega)} \right| &\leq \|b \cdot \nabla u + cu\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)} \\ &\leq (\|b \cdot \nabla u\|_{L^2(\Omega)} + \|cu\|_{L^2(\Omega)}) \|v\|_{L^2(\Omega)} \\ &\leq (\|b\|_{L^\infty(\Omega)} \|u\|_{H^1(\Omega)} + \|c\|_{L^\infty(\Omega)} \|u\|_{L^2(\Omega)}) \|v\|_{L^2(\Omega)} \\ &\leq (C \|u\|_{H^1(\Omega)} + C \|u\|_{L^2(\Omega)}) \|v\|_{L^2(\Omega)} \\ &\leq C \|u\|_{H^1(\Omega)} \|v\|_{H^1(\Omega)} \end{aligned} \tag{4.2.5}$$

aus der Cauchy–Schwarz– und der Minkowski–Ungleichung.

4. Die Bilinearform $a(\cdot, \cdot)$ ist strikt koerzitiv bzgl. $\|\cdot\|_{H^1(\Omega)}$.
Es gilt

$$\epsilon (\nabla u, \nabla u)_{L^2(\Omega)} \geq \epsilon |u|_{H^1(\Omega)}^2, \quad (4.2.6)$$

und unter Verwendung der partiellen Integration im dritten Schritt rechnet man nach

$$\begin{aligned} (b \cdot \nabla u + cu, u)_{L^2(\Omega)} &= \int_{\Omega} u b \cdot \nabla u \, dx + \int_{\Omega} c u^2 \, dx \\ &= \int_{\Omega} \frac{1}{2} b \cdot \nabla (u^2) \, dx + \int_{\Omega} c u^2 \, dx \\ &= -\frac{1}{2} \int_{\Omega} (\nabla \cdot b) u^2 \, dx + \int_{\Omega} c u^2 \, dx \\ &= \int_{\Omega} \underbrace{\left(c - \frac{1}{2} \nabla \cdot b \right)}_{\geq c_0} u^2 \, dx \\ &\geq c_0 \int_{\Omega} u^2 \, dx \\ &= c_0 \|u\|_{L^2(\Omega)}^2. \end{aligned} \quad (4.2.7)$$

Dabei sind die Randintegrale gleich Null, da die Funktion u auf $\partial\Omega$ verschwindet.
Aus (4.2.6) und (4.2.7) ergibt sich

$$a(u, u) \geq C \left(\epsilon |u|_{H^1(\Omega)}^2 + c_0 \|u\|_{L^2(\Omega)}^2 \right) = C \min\{\epsilon, c_0\} \|u\|_{H^1(\Omega)}^2. \quad (4.2.8)$$

Damit sind alle Voraussetzungen erfüllt und (4.2.2) besitzt nach dem Satz von Lax–Milgram eine eindeutige Lösung. \square

Bemerkung 4.6.

Satz 4.5 kann bezüglich der Randbedingungen noch verallgemeinert werden. Beispielsweise für den Fall inhomogener Dirichlet–Randbedingungen $u^D \in H^{1/2}(\partial\Omega)$ kann mit Hilfe von Satz 8.8 aus (Wlo82) u^D auf $\tilde{u}^D \in H^1(\Omega)$ fortgesetzt werden. Mit dieser Fortsetzung definiert man die Funktion $\tilde{u} = u - \tilde{u}^D$, setzt $u = \tilde{u} + \tilde{u}^D$ in die schwache Formulierung ein und bringt den Term mit \tilde{u}^D auf die rechte Seite. Man erhält somit eine Gleichung der Form (4.2.2) mit homogenen Randbedingungen für \tilde{u} , welche nach Satz 4.5 eindeutig lösbar ist. Daraus ergibt sich die eindeutige Lösbarkeit des Problems mit inhomogenen Dirichlet–Randbedingungen.

4.3 Das Standard–Galerkin–Verfahren

Mit Hilfe der im vorherigen Abschnitt eingeführten schwachen Formulierung können wir ein Galerkin–Verfahren für die Konvektions–Diffusions–Gleichung aufstellen. Der Einfachheit halber wollen wir von homogenen Dirichlet–Randbedingungen ausgehen. Wir wählen als Ansatz- und Testraum $V_h(\Omega)$ den Schnitt von $H_0^1(\Omega)$ mit einem stetigen polynomiellen Finite–Elemente–Raum auf Ω . Die Formulierung für Räume unstetiger Funktionen erfolgt in Abschnitt 4.5.

Das Standard–Galerkin–Verfahren lautet:

Finde $u_h \in V_h(\Omega)$ mit

$$a(u_h, v_h) = f(v_h) \quad \forall v_h \in V_h, \quad (4.3.1)$$

wobei die Bilinearform a und die Linearform f dieselben wie in der schwachen Formulierung sind. Die Größe h bezeichnet hier das Maximum der Zellendurchmesser. Wir nehmen an, dass Ω polyhedral berandet ist, so dass wir eine zulässige quasi–uniforme Zerlegung \mathcal{T}_h wählen können.

Da die Bilinearform a koerzitiv ist, gibt es auch in $V_h(\Omega) \subset H_0^1(\Omega)$ eine eindeutige Lösung u_h . Eine Aussage darüber, wie gut u_h die exakte Lösung u approximiert, gibt der folgende Satz:

Satz 4.7.

Seien Ω ein beschränktes Gebiet aus \mathbb{R}^d , $d \in \mathbb{N}$, $b, c \in L^\infty(\Omega)$, $f \in L^2(\Omega)$ und

$$c - \frac{1}{2} \nabla \cdot b \geq c_0 > 0 \quad \text{fast überall auf } \Omega. \quad (4.3.2)$$

Sei ferner die Lösung u der schwachen Formulierung (4.2.2) ausreichend regulär, das heißt, $u \in H^{p+1}(\Omega) \cap H_0^1(\Omega)$. Dann gilt für die Lösung u_h des Standard–Galerkin–Verfahrens im Raum $V_h(\Omega)$, welcher aus Polynomen vom Grad $\leq p$ bestehe, die folgende Fehlerabschätzung:

$$\|u - u_h\|_{H^1(\Omega)} \leq \frac{C}{\min\{c_0, \epsilon\}} h^p |u|_{H^{p+1}(\Omega)} \quad (4.3.3)$$

mit einer Konstanten $C > 0$ und dem maximalen Zellendurchmesser h .

Beweis.

Mit Hilfe der Dreiecksungleichung zerlegen wir den Fehler in einen Interpolationsfehler- und Diskretisierungsfehler-Anteil:

$$\|u - u_h\|_{H^1(\Omega)} \leq \|u - u_I\|_{H^1(\Omega)} + \|u_I - u_h\|_{H^1(\Omega)}, \quad (4.3.4)$$

wobei u_I die Lagrange–Interpolierte von u ist.

Für den Interpolationsfehler gilt nach Lemma 3.12 die Abschätzung:

$$\|u - u_I\|_{H^1(\Omega)} \leq Ch^p |u|_{H^{p+1}(\Omega)}. \quad (4.3.5)$$

Um eine Abschätzung für den Diskretisierungsfehler zu erhalten, gehen wir aus von

$$\|u_I - u_h\|_{H^1(\Omega)}^2 \leq \frac{C}{\min\{c_0, \epsilon\}} a(u_I - u_h, u_I - u_h) = \frac{C}{\min\{c_0, \epsilon\}} a(u_I - u, u_I - u_h). \quad (4.3.6)$$

Dabei haben wir zunächst die Koerzitivität von a und dann die Galerkin–Orthogonalität

ausgenutzt. Der Vorfaktor ergibt sich aus (4.2.6) und (4.2.7). Die Terme der Bilinearform a schätzen wir einzeln nach oben ab:

Unter Berücksichtigung der Cauchy–Schwarz–Ungleichung und der Interpolationsfehlerabschätzung (4.3.5) erhält man

$$\begin{aligned} \epsilon (\nabla(u_I - u), \nabla(u_I - u))_{L^2(\Omega)} &= \epsilon |u_I - u|_{H^1(\Omega)} |u_I - u_h|_{H^1(\Omega)} \\ &\leq \epsilon |u_I - u|_{H^1(\Omega)} \|u_I - u_h\|_{H^1(\Omega)} \\ &\leq \epsilon Ch^p |u|_{H^{p+1}(\Omega)} \|u_I - u_h\|_{H^1(\Omega)} \end{aligned} \quad (4.3.7)$$

und

$$\begin{aligned} (b \cdot \nabla(u_I - u), u_I - u_h)_{L^2(\Omega)} &\leq \|b \cdot \nabla(u_I - u)\|_{L^2(\Omega)} \|u_I - u_h\|_{H^1(\Omega)} \\ &\leq C |u_I - u|_{H^1(\Omega)} \|u_I - u_h\|_{H^1(\Omega)} \\ &\leq Ch^p |u|_{H^{p+1}(\Omega)} \|u_I - u_h\|_{H^1(\Omega)} \end{aligned} \quad (4.3.8)$$

sowie

$$\begin{aligned} (c(u_I - u), u_I - u_h)_{L^2(\Omega)} &\leq C |u_I - u|_{H^0(\Omega)} \|u_I - u_h\|_{H^1(\Omega)} \\ &\leq Ch^{p+1} |u|_{H^{p+1}(\Omega)} \|u_I - u_h\|_{H^1(\Omega)}. \end{aligned} \quad (4.3.9)$$

Dividiert man in (4.3.9) durch $\|u_I - u_h\|_{H^1(\Omega)}$ und fasst alle Abschätzungen zusammen, folgt die Behauptung. \square

Satz 4.7 sagt aus, dass der Fehler des Standard–Galerkin–Verfahrens beim Übergang $h \rightarrow 0$ mit optimaler Konvergenzrate verschwindet. Der Begriff optimal bezieht sich dabei auf die Konvergenzrate des Interpolationsfehlers, die vom Standard–Galerkin–Verfahren erreicht wird. Jedoch bereitet der Vorfaktor auf der rechten Seite von (4.3.3) Probleme, da er proportional zu $1/\epsilon$ ist. Im konvektions–dominanten Fall, wenn also $0 < \epsilon \ll 1$ gilt, wird der Vorfaktor sehr groß, so dass die Gitterweite h sehr klein gewählt werden muss, um dies aufzuwiegen und einen kleinen Fehler zu garantieren. Dadurch vergrößert sich der Rechenaufwand enorm, was nicht praktikabel ist. Bei größeren Gitterweiten ergeben sich jedoch größere Fehler, welche von unphysikalischen Oszillationen der Lösung auf dem ganzen Gebiet herrühren. Diese treten insbesondere dann auf, wenn Grenzschichten vorhanden sind. Um brauchbare Ergebnisse bei vernünftigen Gitterweiten h zu erhalten, muss das Standard–Galerkin–Verfahren daher modifiziert werden. Eine recht häufig benutzte Möglichkeit ist die so genannte Streamline–Diffusion–Methode, welche wir im nächsten Abschnitt vorstellen wollen.

4.4 Die Streamline–Diffusion–Methode

Es existieren verschiedene Ansätze zur Modifikation des Standard–Galerkin–Verfahrens, um auch im konvektions–dominanten Fall brauchbare Lösungen zu erhalten. Eine Idee ist dabei, Terme zu addieren, welche eine künstliche Diffusion hinzufügen. Erste Beispiele dazu sind die so genannten Upwind–Verfahren (Chr76; Tab77; Hei77). Nachteile dieser Methoden sind, dass die künstliche Diffusion oft zu stark ausfällt und ihr Mangel an Konsistenz, so dass Lösungen des ursprünglichen Variationsproblems keine Lösungen der Upwind–Verfahren mehr sind. Um diese Nachteile zu beheben, können die Upwind–Verfahren weiter modifiziert werden. Eine Möglichkeit bietet hier die Streamline–Diffusion–Methode (SDFEM), welche von

Brooks und Hughes (Hug79) entwickelt wurde. Bei dieser wird nur entlang der Stromlinien, also in Richtung von b künstliche Diffusion hinzuaddiert. Die künstliche Diffusion fällt daher geringer aus. Zudem ist das Verfahren konsistent.

Sei $\Omega \subset \mathbb{R}^2$ ein polygonal berandetes Gebiet (oder polyhedral in drei Dimensionen) und \mathcal{T}_h eine zugehörige quasi-uniforme Zerlegung. Als Approximationsraum wählen wir $V_h = P_p \cap H_0^1(\Omega)$ oder $V_h = Q_p \cap H_0^1(\Omega)$. Weiterhin seien die Daten des Problems hinreichend regulär. Die Streamline–Diffusion–Methode lautet dann:

Finde $u_h \in V_h$, so dass für alle $v_h \in V_h$ gilt:

$$a_{SD}(u_h, v_h) = f_{SD}(v_h), \quad \text{wobei} \quad (4.4.1a)$$

$$a_{SD}(u_h, v_h) = a(u_h, v_h) + \sum_{K \in \mathcal{T}_h} \tau_K (-\epsilon \Delta u_h + b \cdot \nabla u_h + c u_h, b \cdot \nabla v_h)_{L^2(K)}, \quad (4.4.1b)$$

$$f_{SD}(v_h) = f(v_h) + \sum_{K \in \mathcal{T}_h} \tau_K (f, b \cdot \nabla v_h)_{L^2(K)} \quad (4.4.1c)$$

mit $\tau_K \geq 0$. Zum Standard–Galerkin–Verfahren tritt hier der Term

$$\sum_{K \in \mathcal{T}_h} (\text{Res}, b \cdot \nabla v_h)_{L^2(K)}$$

hinzu, wobei Res das Residuum der Konvektions–Diffusions–Gleichung bezeichnet. Für eine ausreichend glatte Lösung $u \in H^2(\Omega) \cap H_0^1(\Omega)$ der schwachen Formulierung verschwindet das Residuum, so dass u ebenfalls (4.4.1a) erfüllt. Also ist die Streamline–Diffusion–Methode wie schon erwähnt konsistent und erfüllt daher die Galerkin–Orthogonalität

$$a_{SD}(u - u_h, v_h) = 0 \quad \forall v_h \in V_h. \quad (4.4.2)$$

Wir zeigen nun, dass die Streamline–Diffusion–Methode unter gewissen Einschränkungen die Voraussetzungen des Satzes von Lax–Milgram erfüllt. Damit besitzt sie eine eindeutige Lösung.

Dass wir auf einem Hilbertraum arbeiten, ist unmittelbar klar, ebenso wie die Linearität von f_{SD} . Die Beschränktheit von f_{SD} folgt aus der Beschränktheit von b und der Cauchy–Schwarz–Ungleichung. Die Beschränktheit von a_{SD} ergibt sich analog zur Beschränktheit von a im Beweis zu Satz 4.5. Zu zeigen bleibt noch die Koerzitivität von a_{SD} . Dazu führen wir die folgende Norm ein:

$$\|u\|_{SD} := \left(\epsilon |u|_{H^1(\Omega)}^2 + \sum_{K \in \mathcal{T}_h} \tau_K \|b \cdot \nabla u\|_{L^2(K)}^2 + c_0 \|u\|_{L^2(K)}^2 \right)^{1/2}. \quad (4.4.3)$$

Damit können wir folgenden Satz formulieren:

Satz 4.8. Seien $c_K = \max_{x \in K} |c(x)|$ und

$$c - \frac{1}{2} \nabla \cdot b \geq c_0 > 0 \quad \text{auf } \Omega. \quad (4.4.4)$$

Der SD-Parameter τ_K erfülle

$$0 < \tau_K < \frac{1}{2} \min \left\{ \frac{c_0}{c_K^2}, \frac{h_K^2}{\epsilon \mu_{inv}^2} \right\} \quad (4.4.5)$$

für alle $K \in \mathcal{T}_h$. Dann ist die Bilinearform $a_{SD}(\cdot, \cdot)$ koerzitiv bezüglich $\|\cdot\|_{SD}$, das heißt es gilt:

$$a_{SD}(v_h, v_h) \geq \frac{1}{2} \|v_h\|_{SD}^2 \quad \forall v_h \in V_h. \quad (4.4.6)$$

Beweis.

Wie im Beweis zu Satz 4.5 zur Koerzitivität von a ergibt sich zunächst

$$\begin{aligned} a_{SD}(v_h, v_h) &= \epsilon (\nabla v_h, \nabla v_h)_{L^2(\Omega)} + (b \cdot \nabla v_h, v_h)_{L^2(\Omega)} + (c v_h, v_h)_{L^2(\Omega)} \\ &\quad + \sum_{K \in \mathcal{T}_h} \tau_K (-\epsilon \Delta v_h + b \cdot \nabla v_h + c v_h, b \cdot \nabla v_h)_{L^2(K)} \\ &= \epsilon |v_h|_{H^1(\Omega)}^2 + (b \cdot \nabla v_h + c v_h, v_h)_{L^2(\Omega)} + \sum_{K \in \mathcal{T}_h} \tau_K \|b \cdot \nabla v_h\|_{L^2(K)}^2 \\ &\quad + \sum_{K \in \mathcal{T}_h} \tau_K (-\epsilon \Delta v_h + c v_h, b \cdot \nabla v_h)_{L^2(K)} \\ &\geq \epsilon |v_h|_{H^1(\Omega)}^2 + c_0 \|v_h\|_{L^2(\Omega)}^2 + \sum_{K \in \mathcal{T}_h} \tau_K \|b \cdot \nabla v_h\|_{L^2(K)}^2 \\ &\quad + \sum_{K \in \mathcal{T}_h} \tau_K (-\epsilon \Delta v_h + c v_h, b \cdot \nabla v_h)_{L^2(K)}. \end{aligned}$$

Um nun weiter nach unten abzuschätzen, machen wir folgende Überlegung: Der letzte Term in unserer Abschätzung kann als einziger negativ sein. Wir können daher weiter nach unten abschätzen, in dem wir diesen Term betragsmäßig nach oben abschätzen und dann mit negativem Vorzeichen eingehen lassen. Es ergibt sich

$$\begin{aligned} &\left| \sum_{K \in \mathcal{T}_h} \tau_K (-\epsilon \Delta v_h + c v_h, b \cdot \nabla v_h)_{L^2(K)} \right| \\ &\leq \sum_{K \in \mathcal{T}_h} \epsilon^2 \tau_K \|\Delta v_h\|_{L^2(K)}^2 + \sum_{K \in \mathcal{T}_h} c^2 \tau_K \|v_h\|_{L^2(K)}^2 + \frac{1}{2} \sum_{K \in \mathcal{T}_h} \|b \cdot \nabla v_h\|_{L^2(K)}^2 \\ &\leq \frac{\epsilon}{2} |v_h|_{H^1(\Omega)}^2 + \frac{c_0}{2} \|v_h\|_{L^2(\Omega)}^2 + \frac{1}{2} \sum_{K \in \mathcal{T}_h} \|b \cdot \nabla v_h\|_{L^2(K)}^2. \end{aligned} \quad (4.4.7)$$

Die erste Abschätzung folgt, da für jedes Skalarprodukt (\cdot, \cdot) und dadurch induzierte Norm

$\|u\|^2 = (u, u)$ wegen der Cauchy–Schwarz- und Young–Ungleichung gilt:

$$\begin{aligned} (a + b, c) &\leq (\|a\| + \|b\|) \|c\| \\ &\leq \frac{1}{2} (\|a\| + \|b\|)^2 + \frac{1}{2} \|c\|^2 \\ &\leq \|a\|^2 + \|b\|^2 + \frac{1}{2} \|c\|^2. \end{aligned}$$

Die zweite Abschätzung ergibt sich aus der Annahme (4.4.5) an τ_K und der lokal inversen Ungleichung (3.2.24).

Schließlich erhalten wir aus (4.4.7) die Behauptung:

$$\begin{aligned} a_{SD}(v_h, v_h) &\geq \epsilon |v_h|_{H^1(\Omega)}^2 + c_0 \|v_h\|_{L^2(\Omega)}^2 + \sum_{K \in \mathcal{T}_h} \tau_K \|b \cdot \nabla v_h\|_{L^2(K)}^2 \\ &\quad - \left| \sum_{K \in \mathcal{T}_h} \tau_K (-\epsilon \Delta v_h + c v_h, b \cdot \nabla v_h)_{L^2(K)} \right| \\ &\geq \frac{1}{2} \left(\epsilon |v_h|_{H^1(\Omega)}^2 + c_0 \|v_h\|_{L^2(\Omega)}^2 + \sum_{K \in \mathcal{T}_h} \|b \cdot \nabla v_h\|_{L^2(K)}^2 \right). \quad \square \end{aligned}$$

Nachdem wir nun wissen, dass eine eindeutige Lösung existiert, stellt sich noch die Frage nach der Güte dieser Lösung. Das folgende Lemma gibt eine Abschätzung des Diskretisierungsfehlers der Streamline–Diffusion–Methode:

Lemma 4.9.

Seien die Voraussetzungen von Satz 4.8 erfüllt und sei $u \in H^{p+1}(\Omega) \cap H_0^1(\bar{\Omega})$ mit $p \geq 1$. Dann lässt sich der Diskretisierungsfehler der Streamline–Diffusion–Methode abschätzen durch

$$\|u^I - u_h\|_{SD} \leq C \left[\sum_{K \in \mathcal{T}_h} (\epsilon + \tau_K + \tau_K^{-1} h_K^2 + h_K^2 + h_K (1 + \tau_K) + h_K^2 (1 + \tau_K)) h_K^{2p} |u|_{H^{p+1}(K)}^2 \right]^{\frac{1}{2}} \quad (4.4.8)$$

mit einer Konstanten $C > 0$, welche weder vom Gitter noch von ϵ abhängt, und der Lagrange–Interpolierten u^I von u .

Beweis.

Eine Beweisskizze findet man in (Roo08) Seite 305. \square

Um aus Lemma 4.9 die bestmögliche Konvergenzordnung zu erhalten, werden die Terme ϵ , τ_K , $\tau_K^{-1} h_K^2$, h_K^2 , $h_K(1 + \tau_K)$, $h_K^2(1 + \tau_K)$ gegeneinander balanciert. Dazu setzen wir

$$\tau_K = \begin{cases} \tau_0 h_K, & \text{falls } \text{Pe}_K > 1 \quad (\text{konvektions–dominanter Fall}), \\ \tau_1 h_K^2 / \epsilon, & \text{falls } \text{Pe}_K \leq 1 \quad (\text{diffusions–dominanter Fall}), \end{cases} \quad (4.4.9)$$

wobei τ_0, τ_1 zwei positive Konstanten sind und Pe_K die lokale Pécletzahl

$$\text{Pe}_K = \frac{\|b\|_{L^\infty(K)} h_K}{2\epsilon}. \quad (4.4.10)$$

ist. Mit dieser Wahl des Streamline–Diffusion–Parameters gelangen wir zu folgender Fehlerabschätzung:

Satz 4.10 (Globale Fehlerabschätzung der Streamline–Diffusion–Methode).

Sei τ_K gemäß (4.4.9) gewählt, die Bedingung von Satz 4.8 erfüllt und die Gitterweite h_K ausreichend klein, das heißt $h_K \leq 1$. Ist $u \in H^{p+1}(\Omega) \cap H_0^1(\bar{\Omega})$ ausreichend regulär, dann gilt für die Lösung u_h der Streamline–Diffusion–Methode die globale Fehlerabschätzung

$$\|u - u_h\|_{SD} \leq C (\epsilon^{1/2} + h^{1/2}) h^p |u|_{H^{p+1}(\Omega)} \quad (4.4.11)$$

mit einer Konstanten $C > 0$, welche weder vom Gitter noch von ϵ abhängt.

Beweis.

Mit Hilfe der Dreiecksungleichung zerlegen wir den Fehler in einen Interpolationsfehler- und einen Diskretisierungsfehler-Anteil:

$$\|u - u_h\|_{SD} \leq \|u^I - u_h\|_{SD} + \|u - u^I\|_{SD} \quad (4.4.12)$$

mit der Lagrange–Interpolierenden u^I von u .

Den Diskretisierungsfehler haben wir in Lemma 4.9 abgeschätzt. Setzt man in dieser Abschätzung die Parameter wie in (4.4.9) angegeben und fasst alle Terme der Ordnung $\mathcal{O}(h^{p+\frac{1}{2}})$ zusammen, so erhält man

$$\|u^I - u_h\|_{SD} \leq C (\epsilon^{1/2} + h^{1/2}) h^p |u|_{H^{p+1}(\Omega)}. \quad (4.4.13)$$

Für den Interpolationsfehler haben wir die Abschätzung (3.2.19):

$$\begin{aligned} \|u - u^I\|_{SD} &= \left(\epsilon |u - u^I|_{H^1(\Omega)}^2 + \sum_{K \in \mathcal{T}_h} \tau_K \|b \cdot \nabla (u - u^I)\|_{L^2(K)}^2 + c_0 \|u - u^I\|_{L^2(K)}^2 \right)^{1/2} \\ &\leq C \left(\epsilon h^{2p} |u|_{H^{p+1}(\Omega)}^2 + \sum_{K \in \mathcal{T}_h} \underbrace{\tau_K}_{\leq Ch_K} h^{2p} \|b\|_{L^\infty(K)}^2 |u|_{H^{p+1}(K)}^2 + c_0 h^{2(p+1)} |u|_{H^{p+1}(K)}^2 \right)^{1/2} \\ &\leq C (\epsilon^{1/2} + h^{1/2}) h^p |u|_{H^{p+1}(\Omega)}. \end{aligned}$$

Setzt man diese Abschätzungen in (4.4.12) ein, so folgt die Behauptung. \square

Bemerkung 4.11.

- (i) Im Gegensatz zur Fehlerabschätzung für das Standard–Galerkin–Verfahren enthält die Fehlerabschätzung für die Streamline–Diffusion–Methode auf der rechten Seite keine Terme negativer Potenz in ϵ . Daher ist die Abschätzung auch im konvektions–dominanten Fall bei vernünftigen Gitterweiten h brauchbar.
- (ii) Die Konvergenzordnung der Fehlerabschätzung für die Streamline–Diffusion–Methode ist suboptimal. Gemessen an der Konvergenzordnung der Interpolationsfehler verfehlt sie die optimalen Rate in der L^2 -Norm um den Faktor $h^{1/2}$. Im diffusions–dominanten Fall wird die optimale Konvergenzordnung h^p bezüglich der H^1 -Norm erreicht, während sie im konvektions–dominanten Fall um einen Faktor $h^{1/2}$ verfehlt wird. Bei hinreichender Regularität der exakten Lösung und Verwendung spezieller Gitter kann die Streamline–Diffusion–Methode so erweitert werden, dass die optimalen Konvergenzraten erreicht werden (Joh87; Zho96; Näv92; Zho97).

- (iii) Obwohl die Streamline–Diffusion–Methode das Standard–Galerkin–Verfahren verbessert, ergeben sich immer noch Probleme im Bereich von Grenzschichten. Um die Näherung in den Grenzschichten zu verbessern, kann die Streamline–Diffusion–Methode um so genannte shock–capturing Terme, auch SOLD–Terme (spurious oscillations at layers diminsing) genannt, ergänzt werden. Einen Überblick über SOLD–Methoden liefert zum Beispiel (Joh07; Joh08). Im Rahmen dieser Arbeit wollen wir aber nicht näher darauf eingehen.

4.5 Discontinuous–Galerkin–Verfahren

Wie wir gesehen haben, ist das Standard–Galerkin–Verfahren oft nicht geeignet, um im konvektions–dominanten Fall Lösungen der Konvektions–Diffusions–Gleichung numerisch zu berechnen. Alternativ können Räume unstetiger Funktionen für V_h gewählt werden, wodurch sich eine zusätzliche Stabilisierung ergibt.

4.5.1 Formulierung als Interior–Penalty–Methode

Die folgende Herleitung eines Discontinuous–Galerkin–Verfahrens ist an die entsprechenden Herleitungen in (Roo08) und (Kan07) angelehnt.

Wir gehen wieder von der starken Formulierung (4.1.1) aus und der Vorgabe von Dirichlet–Randwerten auf dem gesamten Gebietsrand Γ . Die Daten des Problems seien dabei hinreichend regulär, das heißt $b, c \in L^\infty(\Omega)$, $f \in L^2(\Omega)$ und $u^D \in H^{1/2}(\Gamma)$. Weiterhin nehmen wir an, dass eine quasi–uniforme Zerlegung \mathcal{T}_h des Gebietes Ω gegeben ist.

Bevor wir ein Discontinuous–Galerkin–Verfahren für die Konvektions–Diffusions–Gleichung herleiten, vergewissern wir uns schnell, dass für die exakte Lösung u der starken Formulierung gilt:

$$\begin{aligned} [u]_E &= 0 \quad \forall E \in \mathcal{E}_{\text{int}}, \\ [u]_K &= 0 \quad \forall K \in \mathcal{T}_h, \\ \langle u \rangle_E &= u \quad \forall E \in \mathcal{E}_{\text{int}}, \\ u^+ &= u^D \quad \forall E \in \mathcal{E}^D. \end{aligned}$$

Gleiches gilt für ∇u .

Des Weiteren brauchen wir noch folgende Rechenregeln für Sprünge und Mittelwerte:

$$\begin{aligned} [uv] &= [u] \langle v \rangle + [v] \langle u \rangle, \\ \langle uv \rangle &= \langle u \rangle \langle v \rangle + \frac{1}{4} [uv]. \end{aligned}$$

Diese rechnet man elementar nach.

Um zur schwachen Formulierung zu gelangen, gehen wir zunächst von homogenen Dirichlet–Randbedingungen aus, also $u^D = 0$. Wir multiplizieren (4.1.1) mit einer Testfunktion $v \in H^1(\mathcal{T}_h)$ und integrieren über das Gebiet Ω . Es ergibt sich

$$\int_{\Omega} (-\epsilon \Delta u + b \cdot \nabla u + cu) v \, dx = \int_{\Omega} f v \, dx, \quad (4.5.1a)$$

beziehungsweise wenn wir statt über Ω über jede Zelle K der Zerlegung \mathcal{T}_h integrieren und

diese Integrale dann aufsummieren

$$\sum_{K \in \mathcal{T}_h} \int_K (-\epsilon \Delta u + b \cdot \nabla u + cu) v \, dx = \sum_{K \in \mathcal{T}} \int_K f v \, dx. \quad (4.5.1b)$$

Für das Folgende teilen wir das Integral auf der linken Seite zunächst in drei Integrale auf und betrachten sie getrennt. Das Integral über den Reaktionsterm $c u v$ wird nicht weiter umgeformt, ebensowenig wie das Integral über den Term $f v$.

Wir betrachten zunächst das Integral über den Diffusionsteil, $\sum_{K \in \mathcal{T}_h} \int_K (-\epsilon \Delta u) v \, dx$. Dieses wandeln wir mittels partieller Integration und der ersten Greenschen Formel um:

$$\begin{aligned} \sum_{K \in \mathcal{T}_h} \int_K (-\epsilon \Delta u) v \, dx &= \sum_{K \in \mathcal{T}_h} \epsilon \int_K \nabla u \cdot \nabla v \, dx - \sum_{K \in \mathcal{T}_h} \epsilon \int_{\partial K} (\nabla u \cdot n_K) v \, ds \\ &= \sum_{K \in \mathcal{T}_h} \epsilon \int_K \nabla u \cdot \nabla v \, dx - \sum_{E \in \mathcal{E} \cap \Gamma} \epsilon \int_E (\nabla u \cdot n) v \, ds \\ &\quad - \sum_{E \in \mathcal{E}_{\text{int}}} \epsilon \int_E ((\nabla u \cdot n_K) v) |_{\partial K \cap E} + ((\nabla u \cdot n_{K'}) v) |_{\partial K' \cap E} \, ds, \end{aligned}$$

wobei K und K' die beiden Zellen sind, für welche $E \in \partial K$ sowie $E \in \partial K'$ gilt und n_K sowie $n_{K'}$ äußere Normalenvektoren zu den Zellen K beziehungsweise K' sind. Wir wollen die Summe der Integrale über $E \in \mathcal{E}_{\text{int}}$ noch etwas umformen:

$$\begin{aligned} \sum_{E \in \mathcal{E}_{\text{int}}} \epsilon \int_E ((\nabla u \cdot n_K) v) |_{\partial K \cap E} + ((\nabla u \cdot n_{K'}) v) |_{\partial K' \cap E} \, ds \\ &= \sum_{E \in \mathcal{E}_{\text{int}}} \epsilon \int_E ((\nabla u \cdot n_E) v) |_{\partial K \cap E} - ((\nabla u \cdot n_E) v) |_{\partial K' \cap E} \, ds \\ &= \sum_{E \in \mathcal{E}_{\text{int}}} \epsilon \int_E \langle \nabla u \cdot n_E \rangle_E [v]_E + \underbrace{[\nabla u \cdot n_E]_E}_{=0} \langle v \rangle_E \, ds \\ &= \sum_{E \in \mathcal{E}_{\text{int}}} \epsilon \int_E \langle \nabla u \cdot n_E \rangle_E [v]_E \, ds. \end{aligned}$$

Wir haben hier n_E als neuen Einheitnormalenvektor zu E eingeführt. Bei obiger Umformung gilt $n_E = n_K = -n_{K'}$.

Um die Notation zu vereinfachen, führen wir folgende Abkürzungen ein:

$$\begin{aligned} \sum_{E \in \mathcal{E}_{\text{int}}} \epsilon \int_E \langle \nabla u \cdot n_E \rangle_E [v]_E \, ds &= \epsilon \int_{\Gamma_{\text{int}}} \langle \nabla u \cdot n_E \rangle [v] \, ds, \\ \sum_{E \in \mathcal{E} \cap \Gamma} \epsilon \int_E (\nabla u \cdot n) v \, ds &= \epsilon \int_{\Gamma} (\nabla u \cdot n) v \, ds. \end{aligned}$$

Damit erhalten wir

$$\begin{aligned} \sum_{K \in \mathcal{T}_h} \int_K (-\epsilon \Delta u) v \, dx &= \sum_{K \in \mathcal{T}_h} \epsilon \int_K \nabla u \cdot \nabla v \, dx \\ &\quad - \epsilon \int_{\Gamma_{\text{int}}} \langle \nabla u \cdot n_E \rangle [v] \, ds - \epsilon \int_{\Gamma} (\nabla u \cdot n) v \, ds. \end{aligned}$$

Um dies zu symmetrisieren ziehen wir auf der rechten Seite die Terme

$$\epsilon \int_{\Gamma_{\text{int}}} [u] \langle \nabla v \cdot n_E \rangle \, ds \text{ und } \epsilon \int_{\Gamma} u (\nabla v \cdot n) \, ds \quad (4.5.2)$$

ab. Dabei gilt es zu beachten, dass sich diese Terme für die exakte Lösung zu Null ergeben. Mit den zusätzlichen Termen erhalten wir

$$\begin{aligned} \sum_{K \in \mathcal{T}_h} \int_K (-\epsilon \Delta u) v \, dx &= \sum_{K \in \mathcal{T}_h} \epsilon \int_K \nabla u \cdot \nabla v \, dx \\ &\quad - \epsilon \int_{\Gamma_{\text{int}}} ([u] \langle \nabla v \cdot n_E \rangle + \langle \nabla u \cdot n_E \rangle [v]) \, ds \\ &\quad - \epsilon \int_{\Gamma} (u (\nabla v \cdot n) + (\nabla u \cdot n) v) \, ds. \end{aligned}$$

In den Randintegralen, welche vom konvektiven Term herrühren, wählen wir für u den Standard–Upwind–Fluss

$$u^- = \lim_{\substack{\epsilon \rightarrow 0 \\ \epsilon > 0}} v(x - \epsilon b).$$

Dieser wurde so erstmals von Reed und Hill ([Ree73](#)) eingeführt und von LeSaint und Raviart ([LeS74](#)) analysiert. Die Verwendung des Standard–Upwind–Flusses führt hier bei Anwendung ähnlicher Umformungen wie oben für den diffusiven Anteil auf die Terme

$$- \sum_{K \in \mathcal{T}_h} \int_{\partial_- K \setminus \Gamma} (b \cdot n_K) [u] v^+ \, ds \text{ und } - \sum_{K \in \mathcal{T}_h} \int_{\partial_- K \cap \Gamma} (b \cdot n_K) u^+ v^+ \, ds.$$

Beide ergeben sich für die exakte Lösung u zu Null.

Koerzitivität der Bilinearform lässt sich erreichen, indem man noch die Strafterme

$$\sigma_h \int_{\Gamma_{\text{int}}} [u] [v] \, ds \text{ und } \sigma_h \int_{\Gamma} uv \, ds$$

addiert, welche ebenfalls für die exakte Lösung verschwinden. Hierbei ist σ_h der so genannte Unstetigkeitsstrafparameter oder Stabilisierungsparameter. Wir wollen zulassen, dass σ_h für jede Kante verschieden sein kann, das heißt, dass

$$\sigma_h|_E = \sigma_E \text{ für } E \in \mathcal{E}$$

gilt, wobei σ_E passend gewählt wird. Wir werden für unsere Untersuchungen

$$\sigma_E = \frac{\sigma}{h_E}$$

wählen, wobei σ eine Konstante und h_E die Länge der Kante E ist.

Summieren wir nun die einzelnen Integralterme auf, so erhalten wir die Bilinearform

$$\begin{aligned}
 a_{\text{disc}}(u, v) &:= \sum_{K \in \mathcal{T}_h} \left(\int_K \epsilon \nabla u \cdot \nabla v + (b \cdot \nabla u + cu) v \, dx \right) \\
 &\quad - \epsilon \int_{\Gamma_{\text{int}}} ([u] \langle \nabla v \cdot n_E \rangle + \langle \nabla u \cdot n_E \rangle [v]) \, ds - \sum_{K \in \mathcal{T}_h} \int_{\partial_- K \setminus \Gamma} (b \cdot n_K) [u] v^+ \, ds \\
 &\quad + \sigma_h \int_{\Gamma_{\text{int}}} [u] [v] \, ds \\
 &\quad - \epsilon \int_{\Gamma} (u (\nabla v \cdot n) + (\nabla u \cdot n) v) \, ds - \sum_{K \in \mathcal{T}_h} \int_{\partial_- K \cap \Gamma} (b \cdot n_K) u^+ v^+ \, ds \\
 &\quad + 2\sigma_h \int_{\Gamma} uv \, ds.
 \end{aligned}$$

In Anlehnung an (Kan07), Bemerkung 2.2.2 und 2.2.9 haben wir hier den Strafparameter σ_h im Integral über den Rand noch mit einem Faktor 2 versehen.

Um das Discontinuous–Galerkin–Verfahren auch für nicht–homogene Dirichlet-Randbedingungen zu formulieren, betrachten wir nochmals die Integrale, welche sich über den Rand Γ des Gebietes Ω erstrecken und welche wir mit dem Argumente, dass sie sich zu Null ergeben hinzu addiert haben. Diese sind:

$$-\epsilon \int_{\Gamma} u (\nabla v \cdot n) \, ds, \quad - \sum_{K \in \mathcal{T}_h} \int_{\partial_- K \cap \Gamma} (b \cdot n_K) u^+ v^+ \, ds, \quad + \sigma_h \int_{\Gamma} uv \, ds.$$

Für nicht-homogene Dirichlet-Randbedingungen ergeben sich diese nicht zu Null, sondern zu:

$$-\epsilon \int_{\Gamma} u^D (\nabla v \cdot n) \, ds, \quad - \sum_{K \in \mathcal{T}_h} \int_{\partial_- K \cap \Gamma} (b \cdot n_K) u^D v^+ \, ds, \quad + \sigma_h \int_{\Gamma} u^D v \, ds.$$

Um dies zu berücksichtigen, werden obige Terme auf die rechte Seite der schwachen Formulierung der Konvektions–Diffusions–Gleichung addiert. Anders als bei den bisher vorgestellten Verfahren werden damit die Randbedingungen nicht mehr stark durch die Wahl der Ansatz- und Testräume berücksichtigt, sondern schwach durch Hinzunahme entsprechender Terme in der Formulierung des Variationsproblems. Diese schwache Implementierung von Randbedingungen ist prinzipiell auch bei anderen Verfahren möglich.

Wir definieren daher als Linearform des Discontinuous–Galerkin–Verfahrens

$$\begin{aligned}
 f_{\text{disc}}(v) &:= \sum_{K \in \mathcal{T}_h} \int_K f v \, dx - \epsilon \int_{\Gamma} u^D (\nabla v \cdot n) \, ds \\
 &\quad - \sum_{K \in \mathcal{T}_h} \int_{\partial_- K \cap \Gamma} (b \cdot n_K) u^D v^+ \, ds + 2\sigma_h \int_{\Gamma} u^D v \, ds.
 \end{aligned}$$

Auch hier haben wir wieder einen Faktor 2 vor dem Strafparameter σ_h im Integral über den Rand hinzugefügt.

Als neue Formulierung erhalten wir:
 Finde $u \in H^1(\mathcal{T}_h)$, so dass für alle $v \in H^1(\mathcal{T}_h)$ gilt:

$$a_{\text{disc}}(u, v) = f_{\text{disc}}(v). \quad (4.5.3)$$

Wenn wir nun noch $u, v \in V_h$ mit $V_h = P_p^{\text{disc}}$ oder $V_h = Q_p^{\text{disc}}$ annehmen, erhalten wir das so genannte symmetrischen Interior–Penalty–Verfahren (SIPG) aus der Klasse der Discontinuous–Galerkin–Verfahren.

Bemerkung 4.12.

- (i) Statt die in (4.5.2) angegebenen Terme zu subtrahieren, können diese auch addiert oder ganz weggelassen werden. Die entsprechenden Verfahren sind dann nicht–symmetrisch (NIPG) oder unvollständig (englisch: incomplete; IIPG). Näheres hierzu findet sich beispielsweise in (Riv08) und der dort angegebenen Literatur.
- (ii) Statt der hier gezeigten Herleitung für den diffusiven Anteil lässt sich auch die Methode von Nitsche zur schwachen Behandlung von Dirichlet–Randbedingungen (Nit72) zellenweise anwenden, um die entsprechenden Terme herzuleiten (siehe (Kan07)).

Um den Fehler $u - u_h$ des SIPG–Verfahrens zu bestimmen, behandeln wir die einzelnen Terme wieder getrennt.

Wir betrachten zunächst den diffusiven Term, fassen dessen Beiträge zu a_{disc} und die Strafterme in der Bilinearform a_{diff} zusammen und zeigen zunächst, dass diese koerzitiv ist. Zur Vereinfachung der Darstellung setzen wir hier $\epsilon = 1$

Lemma 4.13.

Sei für jede Kante E zwischen den benachbarten Zellen K_1 und K_2 der Parameter σ_h so festgelegt, dass

$$\sigma_h = \frac{1}{2} (\sigma_E^+ + \sigma_E^-), \quad \sigma_E^\pm > \frac{1}{2} c_\pm^2$$

gilt. Hierbei ist c_\pm die jeweilige Konstante aus der Spurgleichung (3.2.28) für die Zellen K_1 und K_2 . Dann erfüllt die Bilinearform a_{diff} die Abschätzung

$$\|v\|_{1;h}^2 \leq C a_{\text{diff}}(v, v) \quad (4.5.4)$$

für alle $v \in V_h$ mit der Norm

$$\|v\|_{1;h}^2 := |v|_{H^1(\mathcal{T}_h)}^2 + \|\sqrt{\sigma_h} [v]\|_{L^2(\mathcal{E}_{\text{int}})}^2 + \|\sqrt{2\sigma_h} v\|_{L^2(\mathcal{E}^D)}^2 \quad (4.5.5)$$

und einer Konstanten C , welche unabhängig von σ_h oder der Gitterweite ist.

Beweis.

Dass $\|v\|_{1;h}$ eine Norm ist, rechnet man nach. Die Definitheit ergibt sich dabei mit Hilfe der Friedrichs–Ungleichungen aus Lemma 2.21.

Für die Ungleichung betrachten wir

$$a_{\text{diff}}(v, v) = \|v\|_{1;h}^2 - 2 \int_{\Gamma_{\text{int}}} \langle \nabla v \cdot n_E \rangle [v] ds - 2 \int_{\Gamma} (\nabla v \cdot n_E) v ds$$

Den zweiten und dritten Term schätzen wir weiter ab. Dazu benutzen wir die Spurgleichung (3.2.26) und die Youngsche Ungleichung aus Lemma 2.29 mit $\delta = c_*^2/2\sigma_h$, wobei c_*

die Konstante aus der Spurgleichung für die jeweilige Zelle K ist:

$$\begin{aligned}
 & 2 \int_{\Gamma_{\text{int}} \cup \Gamma} \langle \nabla v \cdot n_E \rangle [v] ds + 2 \int_{\Gamma} (\nabla v \cdot n_E) v ds \\
 & \leq \sum_{K \in \mathcal{T}_h} \left\{ \frac{\delta h_K}{c_*^2} \|\nabla v \cdot n_E\|_{L^2(\partial K_i)}^2 + \frac{c_*^2}{4\delta h_K} \| [v] \|_{L^2(\partial K_i)}^2 + \right. \\
 & \quad \left. \frac{\delta h_K}{c_*^2} \|\nabla v \cdot n_E\|_{L^2(\partial K_D)}^2 + \frac{c_*^2}{\delta h_K} \| [v] \|_{L^2(\partial K_D)}^2 \right\} \\
 & \leq \delta \sum_{K \in \mathcal{T}_h} \left\{ \|\nabla v\|_{L^2(K)}^2 + \frac{1}{2} \|\sqrt{\sigma_h} [v]\|_{L^2(\partial K_i)}^2 + 2 \|\sqrt{\sigma_h} [v]\|_{L^2(\partial K_D)}^2 \right\} \\
 & \leq \delta \|v\|_{1;h}^2.
 \end{aligned}$$

Dabei ist ∂K_i der Teil von ∂K , welcher im Inneren von Ω liegt und ∂K_D der Teil von ∂K , welcher am Rand von Ω liegt. Mit $c = 1 - \delta$ ergibt sich damit die Behauptung. \square

Wir geben noch eine Abschätzung für die Norm $\|\cdot\|_{1;h}^2$ des L^2 -Projektionsfehlers $u - \pi_h u$ an:

Lemma 4.14.

Sei $u \in H^s(\mathcal{T}_h)$ eine Lösung der Poissongleichung

$$-\Delta u = f \text{ auf } \Omega \tag{4.5.6}$$

mit $f \in L^2(\Omega)$ und Randbedingungen wie in 4.1.1 und $s \leq p + 1$, wobei p der maximale Polynomgrad in V_h ist. Dann ist a_{diff} die zugehörige Bilinearform der schwachen Formulierung und es gilt für den L^2 -Projektionsfehler die Abschätzung

$$\|u - \pi_h u\|_{1;h} \leq c h^{s-1} |u|_{H^s(\mathcal{T}_h)} \tag{4.5.7}$$

Beweis.

Dass a_{diff} die entsprechende Bilinearform ist, ist nach Konstruktion unmittelbar klar. Die Abschätzung folgt dann direkt aus (3.2.23) und den Spurgleichungen in Lemma 3.16. \square

Mit dieser Abschätzung des Projektionsfehlers können wir nun den Fehler $u - u_h$ abschätzen.

Satz 4.15.

Sei $u \in H^s(\mathcal{T}_h)$ eine Lösung der Poissongleichung (4.5.6) mit $2 < s \leq p + 1$ und p der maximale Polynomgrad in V_h . Sei weiterhin σ_h so gewählt wie in Lemma 4.13. Dann gilt für den Fehler $u - u_h$ die Abschätzung

$$\|u - u_h\|_{1;h} \leq c h^{s-1} |u|_{H^s(\mathcal{T}_h)}. \tag{4.5.8}$$

Beweis.

Wir schätzen zunächst mit der Dreiecksungleichung ab

$$\|u - u_h\|_{1;h} \leq \|\pi_h u - u_h\|_{1;h} + \|\pi_h u - u\|_{1;h}.$$

Für den zweiten Summanden haben wir die Abschätzung (4.5.7). Für den ersten Summanden

gilt

$$\begin{aligned}
 c \|\pi_h u - u_h\|_{1;h}^2 &\leq a_{\text{diff}}(\pi_h u - u_h, \pi_h u - u_h) \\
 &= a_{\text{diff}}(\pi_h u - u + u - u_h, \pi_h u - u_h) \\
 &= a_{\text{diff}}(\pi_h u - u, \pi_h u - u_h) + a_{\text{diff}}(u - u_h, \pi_h u - u_h) \\
 &= a_{\text{diff}}(\pi_h u - u, \pi_h u - u_h) \\
 &\leq \left| (\nabla(\pi_h u - u), \nabla(\pi_h u - u_h))_{L^2(\mathcal{T}_h)} \right| \\
 &\quad + \left| \int_{\Gamma_{\text{int}}} [\pi_h u - u] \langle \nabla(\pi_h u - u_h) \cdot n_E \rangle ds \right| \\
 &\quad + \left| \int_{\Gamma_{\text{int}}} \langle \nabla(\pi_h u - u) \cdot n_E \rangle [\pi_h u - u_h] ds \right| \\
 &\quad + \left| \sigma_h \int_{\Gamma_{\text{int}}} [\pi_h u - u] [\pi_h u - u_h] ds \right| \\
 &\quad + \left| \int_{\Gamma} \pi_h u - u (\nabla(\pi_h u - u_h) \cdot n) ds \right| \\
 &\quad + \left| \int_{\Gamma} (\nabla(\pi_h u - u) \cdot n) \pi_h u - u_h ds \right| \\
 &\quad + \left| 2\sigma_h \int_{\Gamma} \pi_h u - u \pi_h u - u_h ds \right|.
 \end{aligned}$$

Die erste Ungleichung gilt wegen der Koerzitivität von a_{diff} (Lemma 4.13). Im vorletzten Schritt haben wir die Galerkin–Orthogonalität ausgenutzt und im letzten Schritt den Betrag genommen und wieder die Dreiecksungleichung angewendet.

Die Hölder–Ungleichung, die Projektionsfehlerabschätzungen (3.2.23) und (3.2.29) sowie die inverse Ungleichung (3.2.24) liefern die folgenden Abschätzungen:

$$\begin{aligned}
 \left| (\nabla(\pi_h u - u), \nabla(\pi_h u - u_h))_{L^2(\mathcal{T}_h)} \right| &\leq \|\nabla(\pi_h u - u)\|_{L^2(\mathcal{T}_h)} \|\pi_h u - u_h\|_{1;h} \\
 &\leq c h^{s-1} \|u\|_{H^s(\mathcal{T}_h)} \|\pi_h u - u_h\|_{1;h} \\
 \left| \int_{\Gamma_{\text{int}}} [\pi_h u - u] \langle \nabla(\pi_h u - u_h) \cdot n_E \rangle ds \right| &\leq \|\sqrt{\sigma_h} [\pi_h u - u]\|_{L^2(\mathcal{E}_{\text{int}})} \\
 &\quad \cdot \left(\frac{1}{\sigma_h} \int_{\Gamma_{\text{int}}} \nabla(\pi_h u - u_h) \cdot n \nabla(\pi_h u - u_h) \cdot n ds \right)^{\frac{1}{2}} \\
 &\leq c h^{s-1} \|u\|_{H^s(\mathcal{T}_h)} \|\pi_h u - u_h\|_{1;h} \\
 \left| \int_{\Gamma_{\text{int}}} \langle \nabla(\pi_h u - u) \cdot n_E \rangle [\pi_h u - u_h] ds \right| &\leq \|\sigma_h^{-\frac{1}{2}} \langle \nabla(\pi_h u - u) \rangle\|_{L^2(\mathcal{E}_{\text{int}})} \\
 &\quad \cdot \left(\sigma_h \int_{\Gamma_{\text{int}}} [\pi_h u - u_h] [\pi_h u - u_h] ds \right)^{\frac{1}{2}} \\
 &\leq c h^{s-1} \|u\|_{H^s(\mathcal{T}_h)} \|\pi_h u - u_h\|_{1;h} \\
 \left| \sigma_h \int_{\Gamma_{\text{int}}} [\pi_h u - u] [\pi_h u - u_h] ds \right| &\leq \|\sqrt{\sigma_h} (\pi_h u - u)\|_{L^2(\mathcal{E}_{\text{int}})} \|\pi_h u - u_h\|_{1;h} \\
 &\leq c h^{s-1} \|u\|_{H^s(\mathcal{T}_h)} \|\pi_h u - u_h\|_{1;h}
 \end{aligned}$$

Wir haben hier zudem benutzt, dass $\sigma_h \sim h^{-1}$ gelte. Für die Integrale über den Rand Γ ergeben sich analoge Abschätzungen. Summiert man nun alle Abschätzungen auf und teilt durch $\|\pi_h u - u_h\|_{1;h}$, so erhält man die gewünschte Fehlerabschätzung. \square

Als Folge dieses Lemmas ergibt sich mit Bemerkung 3.9 (ii) folgendes

Korollar 4.16.

Unter denselben Voraussetzungen wie in Satz 4.15 ergibt sich für die L^2 -Norm des Fehlers die Abschätzung

$$\|u - u_h\|_{L^2(\mathcal{T}_h)} \leq c h^{s-1} |u|_{H^s(\mathcal{T}_h)}. \quad (4.5.9)$$

Für den Fall $s = 2$ lässt sich ebenfalls eine Abschätzung des Fehlers in der L^2 -Norm zeigen:

Satz 4.17.

Sei $u \in H^2(\mathcal{T}_h)$ eine Lösung des Variationsproblems

$$a_{\text{diff}}(u, v) = f(v) \quad \forall v \in H^1(\mathcal{T}_h)$$

und gelte für jede Lösung w der Poissongleichung (4.5.6) die Abschätzung $\|w\|_{H^2(\Omega)} \leq C \|f\|_{L^2(\Omega)}$ für beliebige $f \in L^2(\Omega)$. Dann gilt für den Fehler $u - u_h$ die Abschätzung

$$\|u - u_h\|_{L^2(\mathcal{T}_h)} \leq c h^2 |u|_{H^2(\mathcal{T}_h)}. \quad (4.5.10)$$

Beweis.

Der Beweis verläuft analog zum Beweis von Satz 3.22 (siehe Theorem 2.2.12 und 1.3.14 in (Kan07)). \square

Nachdem wir nun den Diffusionsterm behandelt haben, betrachten wir als nächstes den konvektiven Term. Da wir es im rein konvektiven Fall nur mit einer Differentialgleichung erster Ordnung zu tun haben, können wir nur auf dem Einströmbereich des Randes Γ Randbedingungen vorgeben.

Wir fassen alle Terme in a_{disc} , welche mit dem konvektiven Term zusammenhängen, in der Bilinearform a_{konv} zusammen. Damit diese wohldefiniert ist, müssen wir den Raum V_h etwas einschränken. Es sei daher

$$V = \{v \in L^2(\Omega) \mid \forall K \in \mathcal{T}_h : b \cdot \nabla v|_K \in L^2(K)\}.$$

Außerdem verlangen wir, dass $\nabla \cdot b \in L^\infty(\Omega)$ und dass fast überall $\nabla \cdot b(x) \leq 0$ gilt.

Lemma 4.18.

Die Bilinearform

$$\begin{aligned} a_{\text{konv}}(u, v) &:= \sum_{K \in \mathcal{T}_h} \left(\int_K (b \cdot \nabla u) v \, dx - \int_{\partial_- K \setminus \Gamma} (b \cdot n_K) [u] v^+ \, ds - \int_{\partial_- K \cap \Gamma} (b \cdot n_K) u^+ v^+ \, ds \right) \\ &= (b \cdot \nabla u, v)_{L^2(\mathcal{T}_h)} + \int_{\Gamma_{\text{int}}} |b \cdot n_K| [u] v^+ \, ds + \int_{\Gamma_-} |b \cdot n_K| u^+ v^+ \, ds \end{aligned} \quad (4.5.11)$$

ist auf $H^1(\mathcal{T}_h)$ positiv semidefinit.

Beweis.

Partielle Integration liefert

$$\begin{aligned} \int_K (b \cdot \nabla u) u \, dx &= - \int_K u (b \cdot \nabla u) \, dx - \int_K (\nabla \cdot bu) u \, dx \\ &\quad - \int_{\partial_- K \setminus \Gamma} |b \cdot n_K| u u \, ds + \int_{\partial_+ K \setminus \Gamma} |b \cdot n_K| u u \, ds \end{aligned}$$

beziehungsweise

$$2 \int_K (b \cdot \nabla u) u \, dx = - \int_K (\nabla \cdot bu) u \, dx - \int_{\partial_- K \setminus \Gamma} |b \cdot n_K| u u \, ds + \int_{\partial_+ K \setminus \Gamma} |b \cdot n_K| u u \, ds.$$

Summiert man dies über alle Zellen auf, so ergibt sich

$$2 (b \cdot \nabla u, u)_{L^2(\mathcal{T}_h)} = - (\nabla \cdot bu, u)_{L^2(\mathcal{T}_h)} - \int_{\Gamma_{\text{int}} \cup \Gamma_-} |b \cdot n_K| u^+ u^+ \, ds + \int_{\Gamma_{\text{int}} \cup \Gamma_+} |b \cdot n_K| u^- u^- \, ds.$$

Setzen wir dies ein, so erhalten wir

$$\begin{aligned} 2 a_{\text{konv}}(u, u) &= - (\nabla \cdot bu, u)_{L^2(\mathcal{T}_h)} - \int_{\Gamma_{\text{int}} \cup \Gamma_-} |b \cdot n_K| u^+ u^+ \, ds + \int_{\Gamma_{\text{int}} \cup \Gamma_+} |b \cdot n_K| u^- u^- \, ds \\ &\quad + 2 \int_{\Gamma_{\text{int}}} |b \cdot n_K| [u] u^+ \, ds + 2 \int_{\Gamma_-} |b \cdot n_K| u^+ u^+ \, ds \\ &= - (\nabla \cdot bu, u)_{L^2(\mathcal{T}_h)} + \int_{\Gamma_-} |b \cdot n_K| u^+ u^+ \, ds \\ &\quad + \int_{\Gamma_+} |b \cdot n_K| u^- u^- \, ds + \int_{\Gamma_{\text{int}}} |b \cdot n_K| [u] [u] \, ds \tag{4.5.12} \\ &\geq 0. \tag{□} \end{aligned}$$

Definition und Lemma 4.19.

Zu a_{konv} definieren wir auf V die Seminorm

$$\|u\|_{\text{konv}} := \left(a_{\text{konv}}(u, u) + \|h^{1/2} b \cdot \nabla u\|_{L^2(\mathcal{T}_h)}^2 \right)^{\frac{1}{2}}. \tag{4.5.13}$$

Beweis.

Als Folge des vorhergehenden Lemmas ist $\|u\|_{\text{konv}}$ eine Seminorm. Da der zusätzliche Term eine Norm ist, erhalten wir weiterhin eine Seminorm. □

Als nächstes weisen wir die Koerzitivität von a_{konv} nach.

Lemma 4.20.

Ist das Vektorfeld b konstant auf jeder Zelle K , so existieren Konstanten c und γ unabhängig von h , so dass für alle $u_h \in V_h$ gilt:

$$\|u_h\|_{\text{konv}} \leq c a_{\text{konv}}(u_h, u_h + \gamma h b \cdot \nabla u_h). \tag{4.5.14}$$

Beweis.

Wir ziehen die einzelnen Terme zunächst auseinander:

$$\begin{aligned} a_{\text{konv}}(u_h, u_h + \gamma hb \cdot \nabla u_h) &= a_{\text{konv}}(u_h, u_h) + \gamma \|h^{1/2} b \nabla u_h\|_{L^2(\mathcal{T}_h)} \\ &\quad + \gamma \int_{\Gamma_{\text{int}}} |b \cdot n_K| [u_h] hb \cdot \nabla u_h^+ ds \\ &\quad + \gamma \int_{\Gamma_-} |b \cdot n_K| u_h^+ hb \cdot \nabla u_h^+ ds \end{aligned}$$

Die Randintegrale schätzen wir betragsmäßig ab:

$$\begin{aligned} &\left| \gamma \int_{\Gamma_{\text{int}}} |b \cdot n_K| [u_h] hb \cdot \nabla u_h^+ ds \right| + \left| \gamma \int_{\Gamma_-} |b \cdot n_K| u_h^+ hb \cdot \nabla u_h^+ ds \right| \\ &\leq \gamma \| [u_h] \|_{L^2(\mathcal{E}_{\text{int}})} \| |b \cdot n_K| hb \cdot \nabla u_h^+ \|_{L^2(\mathcal{E}_{\text{int}})} + \gamma \| u_h^+ \|_{L^2(\Gamma_-)} \| |b \cdot n_K| hb \cdot \nabla u_h^+ \|_{L^2(\Gamma_-)} \\ &\leq \frac{\gamma}{4\delta} \| [|b \cdot n_K| u_h] \|_{L^2(\mathcal{E}_{\text{int}})}^2 + \delta \gamma \| hb \cdot \nabla u_h^+ \|_{L^2(\mathcal{E}_{\text{int}})}^2 \\ &\quad + \frac{\gamma}{4\delta} \| |b \cdot n_K| u_h^+ \|_{L^2(\Gamma_-)}^2 + \delta \gamma \| hb \cdot \nabla u_h^+ \|_{L^2(\Gamma_-)}^2 \\ &\leq \frac{\gamma}{4\delta} \left(\int_{\Gamma_{\text{int}}} |b \cdot n_K| [u_h] [u_h] ds + \int_{\Gamma_-} |b \cdot n_K| u_h^+ u_h^+ ds \right) + c\delta \gamma \| h^{1/2} b \cdot \nabla u_h \|_{L^2(\mathcal{T}_h)}^2. \end{aligned}$$

Wir haben hier der Reihe nach die Cauchy–Schwarz–Ungleichung, die Youngsche Ungleichung und die Spurgleichung (3.2.28) benutzt. Um $a_{\text{konv}}(u_h, u_h + \gamma hb \cdot \nabla u_h)$ nach unten abzuschätzen, setzen wir die zweite Abschätzung mit negativem Vorzeichen in die erste ein:

$$\begin{aligned} a_{\text{konv}}(u_h, u_h + \gamma hb \cdot \nabla u_h) &\geq a_{\text{konv}}(u_h, u_h) + \gamma \|h^{1/2} b \nabla u_h\|_{L^2(\mathcal{T}_h)} \\ &\quad - \frac{\gamma}{4\delta} \left(\int_{\Gamma_{\text{int}}} |b \cdot n_K| [u_h] [u_h] ds + \int_{\Gamma_-} |b \cdot n_K| u_h^+ u_h^+ ds \right) \\ &\quad - c\delta \gamma \|h^{1/2} b \cdot \nabla u_h\|_{L^2(\mathcal{T}_h)}^2. \end{aligned}$$

Wir wählen $\delta = \frac{\varepsilon}{2}$ und γ so, dass wir (4.5.12) nutzen können, um die gewünschte Abschätzung mit von h unabhängigem c zu erhalten. Man beachte hierbei, dass b auf jeder Zelle K als konstant angenommen wurde, so dass $\nabla \cdot b$ fast überall verschwindet. \square

Mit Hilfe der vorhergehenden Lemmata kann nun der Fehler $u - u_h$ für den rein konvektiven Fall abgeschätzt werden:

Satz 4.21.

Sei $u \in H^s(\Omega)$ eine Lösung der Konvektionsgleichung

$$\begin{aligned} b \cdot \nabla u &= f \text{ in } \Omega \\ u &= u^D \text{ auf } \Gamma_- \end{aligned} \tag{4.5.15}$$

mit $3/2 \leq s \leq p + 1$ und $u_h \in V_h$ eine Lösung des zugehörigen Discontinuous–Galerkin–Verfahrens. Hierbei sei $b|_K$ konstant auf jeder Zelle $K \in \mathcal{T}_h$ und p der maximale Polynomgrad der Polynome in V_h . Dann gilt für den Fehler $u - u_h$ die Abschätzung

$$\|u - u_h\|_{\text{konv}} \leq ch^{s-\frac{1}{2}} |u|_{H^s(\mathcal{T}_h)}. \tag{4.5.16}$$

Beweis.

Wir spalten zunächst wieder auf:

$$\|u - u_h\|_{\text{konv}} \leq \|\pi_h u - u_h\|_{\text{konv}} + \|u - \pi_h u\|_{\text{konv}}.$$

Für den Interpolationsfehler gilt folgende Abschätzung:

$$\begin{aligned} \|u - \pi_h u\|_{\text{konv}} &= \left(\frac{1}{2} \int_{\Gamma_-} |b \cdot n_K| (u - \pi_h u)^+ (u - \pi_h u)^+ ds \right. \\ &\quad + \frac{1}{2} \int_{\Gamma_+} |b \cdot n_K| (u - \pi_h u)^- (u - \pi_h u)^- ds \\ &\quad + \frac{1}{2} \int_{\Gamma_{\text{int}}} |b \cdot n_K| [(u - \pi_h u)] [(u - \pi_h u)] ds \\ &\quad \left. + \|h^{1/2} b \cdot \nabla (u - \pi_h u)\|_{L^2(\mathcal{T}_h)}^2 \right)^{\frac{1}{2}} \\ &\leq c \left(\int_{\Gamma_-} |b \cdot n_K| (u - \pi_h u) (u - \pi_h u) ds \right)^{\frac{1}{2}} \\ &\quad + c \left(\int_{\Gamma_+} |b \cdot n_K| (u - \pi_h u) (u - \pi_h u) ds \right)^{\frac{1}{2}} \\ &\quad + \|h^{1/2} b \cdot \nabla (u - \pi_h u)\|_{L^2(\mathcal{T}_h)} \\ &\leq c \|u - \pi_h u\|_{L^2(\mathcal{E}_{\text{int}} \cup \Gamma)} + \|h^{1/2} b \cdot \nabla (u - \pi_h u)\|_{L^2(\mathcal{T}_h)} \\ &\leq ch^{-\frac{1}{2}} \| (u - \pi_h u) \|_{L^2(\mathcal{T}_h)} + \tilde{c} h^{1/2} \| \nabla (u - \pi_h u) \|_{L^2(\mathcal{T}_h)} \\ &\leq ch^{-\frac{1}{2}} (\| (u - \pi_h u) \|_{L^2(\mathcal{T}_h)} + \| h \nabla (u - \pi_h u) \|_{L^2(\mathcal{T}_h)}) \\ &\leq ch^{-\frac{1}{2}} (\| (u - \pi_h u) \|_{L^2(\mathcal{T}_h)} + \| h \nabla (u - \pi_h u) \|_{L^2(\mathcal{T}_h)}) \\ &\leq ch^{-\frac{1}{2}} \| (u - \pi_h u) \|_{L^2(\mathcal{T}_h)} \\ &\leq ch^{s-\frac{1}{2}} |u|_{H^s(\mathcal{T}_h)}. \end{aligned} \tag{4.5.17}$$

Hier wurden die Ungleichung für die Wurzel einer Summe, die inverse Ungleichung (3.2.24), die Spurgleichung (3.2.27) und die Abschätzung (3.2.23) für den Projektionsfehler benutzt. Zudem haben wir verwendet, dass für $u - \pi_h u$ die Spuren mit der Funktion übereinstimmen und der Sprung verschwindet sowie das b auf jeder Zelle konstant ist.

Für den Fehler $\|\pi_h u - u_h\|_{\text{konv}}$ erhalten wir mit Hilfe von (4.5.14) und der Galerkin–Orthogonalität

$$\begin{aligned} \|\pi_h u - u_h\|_{\text{konv}} &\leq c a_{\text{konv}}(\pi_h u - u_h, \pi_h u - u_h + \gamma h b \cdot \nabla \pi_h u - u_h) \\ &= c a_{\text{konv}}(u - \pi_h u, \pi_h u - u_h + \gamma h b \cdot \nabla \pi_h u - u_h). \end{aligned}$$

Dies wiederum lässt sich folgendermaßen abschätzen:

$$\begin{aligned} &a_{\text{konv}}(u - \pi_h u, \pi_h u - u_h + \gamma h b \cdot \nabla \pi_h u - u_h) \\ &\leq (1 + \|b\|_{L^\infty(\mathcal{T}_h)}) \|h^{-\frac{1}{2}} (u - \pi_h u)\|_{L^2(\mathcal{T}_h)} \|h^{\frac{1}{2}} (\pi_h u - u_h)\|_{L^2(\mathcal{T}_h)} \\ &\quad + \|u - \pi_h u\|_{L^2(\mathcal{E}_{\text{int}})} \int_{\Gamma_{\text{int}}} |b \cdot n_K| [\pi_h u - u_h] [\pi_h u - u_h] ds \end{aligned}$$

$$\begin{aligned}
 & + \|u - \pi_h u\|_{L^2(\Gamma_+)} \int_{\Gamma_{\text{int}}} |b \cdot n_K| (\pi_h u - u_h)^- (\pi_h u - u_h)^- ds \\
 & + \gamma \|h^{1/2} b \cdot \nabla (u - \pi_h u)\|_{L^2(\mathcal{T}_h)} \|h^{1/2} b \cdot \nabla (\pi_h u - u_h)\|_{L^2(\mathcal{T}_h)} \\
 & + \gamma \|hb \cdot \nabla (\pi_h u^+ - u_h^+)\|_{L^2(\mathcal{E}_{\text{int}})} \int_{\Gamma_{\text{int}}} |b \cdot n_K| [u] [u] ds \\
 & + \gamma \|hb \cdot \nabla (\pi_h u^+ - u_h^+)\|_{L^2(\Gamma_-)} \int_{\Gamma_-} |b \cdot n_K| [u] [u] ds \\
 & \leq c \left(\|u - \pi_h u\|_{\text{konv}} + \|h^{-\frac{1}{2}} (u - \pi_h u)\|_{L^2(\mathcal{T}_h)} \right. \\
 & \quad \left. + \|u - \pi_h u\|_{L^2(\mathcal{E}_{\text{int}} \cup \Gamma_+)} \right) \|\pi_h u - u_h\|_{\text{konv}}.
 \end{aligned}$$

Den ersten Term in der Klammer haben wir soeben abgeschätzt. In ähnlicher Weise lassen sich die anderen beiden Terme abschätzen. Division durch $\|\pi_h u - u_h\|_{\text{konv}}$ liefert dann die behauptete Abschätzung. \square

Korollar 4.22.

Unter den Voraussetzungen von Satz 4.21 gilt für die L^2 -Norm des Fehlers

$$\|u - u_h\|_{L^2(\mathcal{T}_h)} \leq ch^{s-\frac{1}{2}} |u|_{H^s(\mathcal{T}_h)}. \quad (4.5.18)$$

Bemerkung 4.23.

Für kartesische Gitter lässt sich das letzte Ergebnis noch verbessern (LeS74):

Bestehe \mathcal{T}_h nur aus Rechtecken, sei b konstant auf Ω , $V_h = Q_k^{\text{disc}}$ und die stetige Lösung $u \in H^{k+2}(\Omega) \cap W^{k+1,\infty}(\Omega)$. Dann gilt die Abschätzung

$$\|u - u_h\|_{L^2(\mathcal{T}_h)} \leq Ch^{k+1} \|u\|_{H^{k+2}(\Omega)}. \quad (4.5.19)$$

Definition 4.24.

Als Energienorm zum SIPG–Verfahren für die Konvektions–Diffusions–Gleichung definieren wir

$$\|u\|_{\text{disc}}^2 := \epsilon \|u\|_{1;h}^2 + \|u\|_{\text{konv}} + c \|u\|_{L^2(\mathcal{T}_h)}. \quad (4.5.20)$$

c ist hierbei der entsprechende Koeffizient in a_{disc} .

Wegen des zusätzlichen Termes $\gamma hb \cdot \nabla u$ in den Abschätzungen zur Bilinearform a_{konv} können wir die Abschätzungen nicht einfach aufaddieren, sondern müssen erst sicherstellen, dass die übrigen Abschätzungen auch bei Hinzunahme dieses Terms gültig bleiben. Um Verwechslungen vorzubeugen, bezeichnen wir den Koeffizienten des reaktiven Terms in a_{disc} im Folgenden mit ρ statt c .

Lemma 4.25.

Es existieren Konstanten c und γ , so dass für alle $v \in V_h$ gilt:

$$\|v\|_{\text{disc}}^2 \leq c \left(\epsilon a_{\text{diff}}(v, v + \gamma hb \cdot \nabla v) + a_{\text{konv}}(v, v + \gamma hb \cdot \nabla v) + (\rho v, v + \gamma hb \cdot \nabla v)_{L^2(\mathcal{T}_h)} \right). \quad (4.5.21)$$

Beweis.

Analog zum Beweis von Lemma 4.20 lässt sich zeigen, dass es Konstanten c_k und γ_k gibt, so dass

$$\|v + \gamma_k hb \cdot \nabla v\|_{\text{konv}} \leq c_k a_{\text{konv}}(v, v + \gamma_k hb \cdot \nabla v) \quad (4.5.22)$$

gilt. Dabei existiert für alle $0 < \gamma \leq \gamma_k$ eine entsprechende Konstante c , so dass diese Abschätzung gilt.

Für den Reaktionsterm ergibt sich mittels einer inversen Abschätzung

$$(\rho v, \gamma_r h b \cdot \nabla v)_{L^2(\mathcal{T}_h)} \leq c \gamma_r (\rho v, v)_{L^2(\mathcal{T}_h)},$$

wobei wir γ_r so wählen, dass $c \gamma_r \leq 1/2$ gilt. Die Terme in a_{diff} lassen sich ebenso abschätzen mit einer Konstanten γ_d . Mit der Wahl $\gamma = \min(\gamma_k, \gamma_d, \gamma_r)$ ergibt sich die Behauptung. \square

Korollar 4.26.

Mit denselben inversen Abschätzungen wie im Beweis zu Lemma 4.25 ergibt sich für alle $v \in V_h$

$$\|v + \gamma h b \cdot \nabla v\|_{\text{disc}} \leq c \|v\|_{\text{disc}} \quad (4.5.23)$$

mit einer Konstanten c unabhängig von h .

Lemma 4.27.

Für alle Funktionen $v \in H^s(\mathcal{T}_h)$ gilt für den Fehler der L^2 -Projektion

$$\|v - \pi_h v\|_{\text{disc}} \leq \alpha |v|_{H^s(\mathcal{T}_h)} \quad (4.5.24)$$

mit $\alpha = \max\left(\sqrt{\epsilon} h^{s-1}, \sqrt{|b|} h^{s-\frac{1}{2}}, \sqrt{\rho} h^s\right)$.

Beweis.

Dies folgt einfach durch Summation von (3.2.23), (4.5.7) und (4.5.17). \square

Satz 4.28.

Sei $u \in H^s(\mathcal{T}_h)$ mit $3/2 \leq s \leq p + 1$ eine Lösung zu (4.1.1) und $u_h \in V_h$ eine Lösung des entsprechenden Variationsproblems (4.5.3), wobei p der maximale Polynomgrad der Ansatzfunktionen in V_h ist. Dann gilt für den Fehler $u - u_h$ die Abschätzung

$$\|u - u_h\|_{\text{disc}} \leq \alpha |u|_{H^s(\mathcal{T}_h)} \quad (4.5.25)$$

mit $\alpha = \max\left(\sqrt{\epsilon} h^{s-1}, \sqrt{|b|} h^{s-\frac{1}{2}}, \sqrt{\rho} h^s\right)$.

Beweis.

In Folge der vorhergehenden Lemmata und Korollare sind die Voraussetzungen von Satz 3.20 erfüllt. Damit ergibt sich die Behauptung. \square

Korollar 4.29.

Unter denselben Voraussetzungen wie in Satz 4.28 ergibt sich für die L^2 -Norm des Fehlers die Abschätzung

$$\|u - u_h\|_{L^2(\mathcal{T}_h)} \leq \alpha |u|_{H^s(\mathcal{T}_h)} \quad (4.5.26)$$

mit $\alpha = \max\left(\sqrt{\epsilon} h^{s-1}, \sqrt{|b|} h^{s-\frac{1}{2}}, \sqrt{\rho} h^s\right)$.

Beweis.

Dies lässt sich ebenso folgern wie Korollar 4.16 oder Korollar 4.22 wenn man zusätzlich beachtet, dass für $\rho \neq 0$ die Abschätzung $\sqrt{\rho} \|v\|_{L^2(\mathcal{T}_h)} \leq \|v\|_{\text{disc}}$ gilt. \square

4.5.2 Flussformulierung

Das im letzten Abschnitt vorgestellte Interior–Penalty–Verfahren beruht auf der Darstellung von 4.1.1 als eine Differentialgleichung zweiter Ordnung. Alternativ dazu kann man auch ein System von zwei Differentialgleichungen erster Ordnung betrachten. Dieses ist gegeben durch

$$\epsilon \nabla u = \zeta \text{ in } \Omega, \quad (4.5.27a)$$

$$-\nabla \cdot \zeta + b \cdot \nabla u + cu = f \text{ in } \Omega, \quad (4.5.27b)$$

$$u = u^D \text{ auf } \Gamma. \quad (4.5.27c)$$

Multiplikation mit Testfunktionen $\tau \in H^1(\mathcal{T}_h)^d$ beziehungsweise $v \in H^1(\mathcal{T}_h)$ und Integration über eine Zelle K einer Zerlegung \mathcal{T}_h ergibt

$$-\epsilon \int_K u \nabla \cdot \tau dx + \epsilon \int_{\partial K} u \tau \cdot n_K ds = \int_K \zeta \cdot \tau dx, \quad (4.5.28a)$$

$$\begin{aligned} \int_K \zeta \cdot \nabla v dx - \int_{\partial K} \zeta \cdot n_K v ds - \int_K u \nabla \cdot (bv) dx \\ + \int_{\partial K} uv (b \cdot n_K) ds + \int_K cuv dx = \int_K fv dx. \end{aligned} \quad (4.5.28b)$$

Formulieren wir nun damit ein Discontinuous–Galerkin–Verfahren, indem wir endlich–dimensionale Räume V_h und S_h sowie entsprechende Funktionen $u_h, v_h \in V_h, \zeta_h, \tau_h \in S_h$ wählen, so ist zu beachten, dass diese Funktionen keine wohldefinierten Werte auf den Kanten der Zerlegung \mathcal{T}_h haben. Für die Integrale über die Kanten wählt man daher so genannte numerische Flüsse. Es ergibt sich

$$-\epsilon \int_K u_h \nabla \cdot \tau_h dx + \epsilon \int_{\partial K} \hat{u}_h^d \tau \cdot n_K ds = \int_K \zeta \cdot \tau dx, \quad (4.5.29a)$$

$$\begin{aligned} \int_K \zeta_h \cdot \nabla v_h dx - \int_{\partial K} \hat{\zeta}_h \cdot n_K v ds - \int_K u_h \nabla \cdot (bv_h) dx \\ + \int_{\partial K} \hat{u}_h^c \hat{v}_h^c (b \cdot n_K) ds + \int_K c u_h v_h dx = \int_K f v_h dx, \end{aligned} \quad (4.5.29b)$$

mit den diffusiven Flüssen \hat{u}_h^d und $\hat{\zeta}_h$ sowie dem konvektiven Fluss \hat{u}_h^c . Allgemein kann man für die diffusiven Flüsse folgenden Ansatz wählen (Riv08):

$$\begin{aligned} \forall E \in \mathcal{E}_{\text{int}} : \quad \hat{u}_h^d|_E &= \langle u_h \rangle + \delta_3 \cdot n_E [u_h] - \delta_2 [\zeta_h] \cdot n_E, \\ \hat{\zeta}_h|_E &= \langle \zeta_h \rangle - \delta_1 [u_h] n_E - ([\zeta_h] \cdot n_E) \delta_3, \\ \forall E \in \Gamma : \quad \hat{u}_h^d|_E &= u^D, \\ \hat{\zeta}_h|_E &= \zeta_h - \delta_1 (u_h - u^D) n_E. \end{aligned}$$

Hierbei sind $\delta_1, \delta_2 \in \mathbb{R}$ und $\delta_3 \in \mathbb{R}^d$ und dürfen von x abhängen. Die Randbedingungen sind durch diese Wahl der numerischen Flüsse bereits schwach eingebunden.

Für den konvektiven Fluss \hat{u}_h^c wählen wir wieder den Standard–Upwind–Fluss (Bre04; Coc03).

Wir summieren die Beiträge (4.5.29) über alle Zellen auf und sortieren nach den verschie-

denen Funktionen um folgende Bilinear- und Linearformen zu erhalten:

$$a_{\text{flux}}(\zeta_h, \tau_h) = \int_{\Omega} \zeta_h \cdot \tau_h dx + \int_{\Gamma_{\text{int}}} \delta_2 [\zeta_h] \cdot n_E [\tau_h] \cdot n_E ds, \quad (4.5.31a)$$

$$b_{\text{flux}}(u_h, \tau_h) = \sum_{K \in \mathcal{T}_h} \int_K u_h \nabla \cdot \tau_h dx - \int_{\Gamma_{\text{int}}} (\langle u_h \rangle + \delta_3 \cdot n_E [u_h]) [\tau_h] ds, \quad (4.5.31b)$$

$$\begin{aligned} c_{\text{flux}}(u_h, v_h) &= \sum_{K \in \mathcal{T}_h} \int_K (b \cdot \nabla u + c u_h v_h) dx + \sum_{K \in \mathcal{T}_h} \int_{\partial_- K \setminus \Gamma} b \cdot n_K [u_h] v_h ds \\ &\quad + \sum_{K \in \mathcal{T}_h} \int_{\partial_- K \cap \Gamma} b \cdot n_K u_h^+ v_h ds + \int_{\Gamma_{\text{int}}} \delta_1 [u_h] [v_h] ds + \int_{\Gamma} \delta_1 u_h v_h ds, \end{aligned} \quad (4.5.31c)$$

$$g_{\text{flux}}(\tau_h) = \int_{\Gamma} u^D \tau_h \cdot n ds, \quad (4.5.31d)$$

$$f_{\text{flux}}(v_h) = \int_{\Omega} f v_h dx + \int_{\Gamma} \delta_1 u^D v_h ds - \int_{\Gamma_-} b \cdot n u^D v_h ds. \quad (4.5.31e)$$

Damit ergibt sich als Discontinuous–Galerkin–Verfahren

$$a_{\text{flux}}(\zeta_h, \tau_h) + b_{\text{flux}}(u_h, \tau_h) = g_{\text{flux}}(\tau_h), \quad (4.5.32a)$$

$$-b_{\text{flux}}(v_h, \zeta_h) + c_{\text{flux}}(u_h, v_h) = f_{\text{flux}}(v_h), \quad (4.5.32b)$$

wobei wir an die Räume V_h und S_h folgende Bedingungen stellen:

$$V_h = \{u \in L^2(\Omega) \mid u|_K \in \mathcal{V}(K), \forall K \in \mathcal{T}_h\}, \quad (4.5.33a)$$

$$S_h = \left\{ \zeta \in L^2(\Omega)^d \mid \zeta_i|_K \in \mathcal{S}(K), \forall K \in \mathcal{T}_h, 1 \leq i \leq d \right\} \quad (4.5.33b)$$

mit

$$\forall u \in \mathcal{V}(K) : \int_K \nabla u \cdot \tau dx = 0 \quad \forall \tau \in \mathcal{S}^d(K) \Rightarrow \nabla u = 0 \text{ auf } K. \quad (4.5.33c)$$

Meistens werden für $\mathcal{V}(K)$ und $\mathcal{S}(K)$ wieder Polynomräume gewählt.

Die Wahl von δ_1 , δ_2 und δ_3 ist durchaus nicht trivial und führt auf verschiedenen Methoden. Einen Überblick über mögliche Wahlen und einen Vergleich der entsprechenden Methoden für den rein diffusiven Fall, also für die Poissongleichung, findet man in (Arn02). Das von uns zuvor behandelte Interior–Penalty–Verfahren ergibt sich mit der Wahl

$$\begin{aligned} \hat{u}_h^d &= \langle u \rangle \text{ auf } \Gamma_{\text{int}}, \\ \hat{u}_h^d &= 0 \text{ auf } \Gamma, \\ \hat{\zeta}_h &= \langle \nabla u_h \rangle - \delta_1 [u_h]. \end{aligned}$$

Der hier noch auftretende Parameter δ_1 entspricht dem Stabilisierungsparameter σ in der Interior–Penalty–Formulierung.

Wir wollen auf die Flussformulierung nicht weiter eingehen, da sie in dieser Arbeit nicht benutzt wird. Weitere Ergebnisse finden sich beispielsweise in den bereits in diesem Abschnitt zitierten Arbeiten oder auch in (Coc98b; Coc01; Coc98a; Hes08; Kan07).

Kapitel 5

Numerische Ergebnisse

Wir werden in diesem Kapitel die mit Hilfe des Discontinuous–Galerkin–Verfahrens gewonnenen numerischen Ergebnisse diskutieren und mit Ergebnissen der Streamline–Diffusion–Methode vergleichen. Dabei wird sowohl der diffusions–dominante als auch der konvektions–dominante Fall der Konvektions–Diffusions–Gleichung betrachtet. Der reaktions–dominante Fall hingegen wird nicht betrachtet. Ziel ist dabei die Überprüfung der im vorherigen Kapitel präsentierten Aussagen zur Konvergenzordnung der Fehler und ein Vergleich der beiden Verfahren.

Zunächst sollen einige technische Aspekte erläutert werden. Diese betreffen die Implementierung des Discontinuous–Galerkin–Verfahrens, die Wahl des Parameters τ_k des Streamline–Diffusion–Verfahrens und die Wahl der Zerlegungen \mathcal{T}_h des Gebietes Ω für die von uns betrachteten Beispiele. Danach werden wir zunächst testen, ob die Anteile des diffusiven beziehungsweise des konvektiven Terms des Discontinuous–Galerkin–Verfahrens richtig implementiert sind. Im Anschluss daran betrachten wir verschiedene Testbeispiele zur Untersuchung des Verfahrens für die Konvektions–Diffusions–Gleichung. Diese lassen sich in drei Kategorien einteilen: Beispiele ohne Grenzschichten, Beispiele mit inneren Grenzschichten und Beispiele mit Randgrenzschichten.

5.1 Einige technische Details

5.1.1 Implementierungsaspekte

Um ein Discontinuous–Galerkin–Verfahren mit MooNMD nutzen zu können, ist die Implementierung der Kantenintegrale notwendig. Betrachten wir dazu die Formulierung (3.1.4) eines Galerkin–Verfahrens in Form eines linearen Gleichungssystems, so muss zunächst die Matrixstruktur erweitert werden, da nun auch Basisfunktionen aus benachbarten Gitterzellen, welche nicht lokale Basisfunktionen beider Gitterzellen sind, miteinander verkoppelt werden. Folglich treten weitere Nicht–Null–Einträge in der Matrix auf.

Bei der Berechnung der Einträge der Matrix und der rechten Seite können prinzipiell drei Typen von Kantenintegralen unterschieden werden:

- Integrale über innere Kanten, welche Test- und Ansatzfunktionen enthalten:

$$\begin{aligned} & - \epsilon \int_{\Gamma_{\text{int}}} ([u] \langle \nabla v \cdot n_E \rangle + \langle \nabla u \cdot n_E \rangle [v]) \, ds, \\ & - \sum_{K \in \mathcal{T}} \int_{\partial_{-K} \setminus \Gamma} (b \cdot n_K) [u] v^+ \, ds, \\ & + \sigma_h \int_{\Gamma_{\text{int}}} [u] [v] \, ds. \end{aligned}$$

Diese liefern einen Beitrag zur Matrix.

- Integrale über Randkanten, welche Test- und Ansatzfunktionen enthalten:

$$\begin{aligned}
 & - \epsilon \int_{\Gamma} (u (\nabla v \cdot n) + (\nabla u \cdot n) v) \, ds, \\
 & - \sum_{K \in \mathcal{T}} \int_{\partial_- K \cap \Gamma} (b \cdot n_K) u^+ v^+ \, ds, \\
 & + \sigma_h \int_{\Gamma} uv \, ds.
 \end{aligned}$$

Auch diese liefern einen Beitrag zur Matrix. Hier können wir einfach die Werte der Test- und Ansatzfunktionen am Rand einsetzen.

- Integrale über Randkanten, welche Ansatzfunktionen und Dirichlet-Randwerte enthalten:

$$\begin{aligned}
 & - \epsilon \int_{\Gamma} u^D (\nabla v \cdot n) \, ds, \\
 & - \sum_{K \in \mathcal{T}} \int_{\partial_- K \cap \Gamma} (b \cdot n_K) u^D v^+ \, ds, \\
 & + \sigma_h \int_{\Gamma} u^D v \, ds.
 \end{aligned}$$

Diese liefern einen Beitrag zur rechten Seite. Auch hier können wir einfach die Werte der Ansatzfunktionen am Rand und die Dirichlet-Randwerte einsetzen.

Es bleiben also noch die Integrale über die inneren Kanten. Um diese zu behandeln, führen wir folgenden Notationen ein:

- Ansatzfunktionen werden mit Φ_i bezeichnet. Es gilt $u = \sum_i u_i \Phi_i$.
- Als Testfunktionen benutzen wir die Basisfunktionen des entsprechenden Finite-Elemente-Raumes. Wir bezeichnen Testfunktionen mit Φ_j .
- Zur Berechnung und Assemblierung der Kantenintegrale wird eine Schleife über die Gitterzellen durchlaufen. Diese bezeichnen wir im Folgenden mit K . Funktionen, die mit der Gitterzelle K assoziiert sind, werden durch einen Index K gekennzeichnet.
- Neben den Funktionen auf einer Gitterzelle werden auch die Funktionen auf den Nachbarzellen der Gitterzelle gebraucht. Wir werden im Folgenden immer nur eine Nachbarzelle gleichzeitig betrachten. Die Nachbarzelle wird mit K' bezeichnet. Die mit ihr assoziierten Größen werden durch einen Index K' bezeichnet.
- Die Kante E ist der gemeinsame Teil der Ränder ∂K der Gitterzelle und $\partial K'$ der Nachbarzelle.

Bei der Berechnung der Integrale über die inneren Kanten treten Sprünge und Mittelwerte der Test- und Ansatzfunktionen auf. Um diese zu berechnen, berücksichtigen wir, dass bei unstetigen Methoden jede lokale Basisfunktion zugleich globale Basisfunktion ist. Jede globale Basisfunktion kann mit genau einer Zelle in Verbindung gebracht werden, innerhalb der sie

von 0 verschiedene Werte annimmt. Außerhalb dieser Zelle verschwindet die Basisfunktion identisch. Dies vereinfacht die Berechnung von Mittelwerten und Sprüngen. Wir betrachten dazu zwei Funktionen $\Phi_{i,K}$ und $\Phi_{j,K'}$. Für diese gilt:

$$\begin{aligned} [\Phi_{i,K}]_E &= \Phi_{i,K} - 0 = \Phi_{i,K}, & \langle \Phi_{i,K} \rangle &= \frac{1}{2} (\Phi_{i,K} + 0) = \frac{1}{2} \Phi_{i,K}, \\ [\Phi_{j,K'}]_E &= 0 - \Phi_{j,K'} = -\Phi_{j,K'}, & \langle \Phi_{j,K'} \rangle &= \frac{1}{2} (0 + \Phi_{j,K'}) = \frac{1}{2} \Phi_{j,K'}, \end{aligned}$$

und analog für Ableitungen der Funktionen. Damit können wir die Kantenintegrale zum Diffusionsterm und die Strafterme bereits vollständig behandeln:

Ansatzfunktion	Testfunktion	diffusives Kantenintegral
Zelle	Zelle	$-\frac{1}{2}\epsilon \int_E (\Phi_{i,K} (\nabla \Phi_{j,K} \cdot n_E) + (\nabla \Phi_{i,K} \cdot n_E) \Phi_{j,K}) ds$
Zelle	Nachbarn	$-\frac{1}{2}\epsilon \int_E (\Phi_{i,K} (\nabla \Phi_{j,K'} \cdot n_E) - (\nabla \Phi_{i,K} \cdot n_E) \Phi_{j,K'}) ds$
Nachbarn	Zelle	$-\frac{1}{2}\epsilon \int_E (-\Phi_{i,K'} (\nabla \Phi_{j,K} \cdot n_E) + (\nabla \Phi_{i,K'} \cdot n_E) \Phi_{j,K}) ds$
Nachbarn	Nachbarn	$\frac{1}{2}\epsilon \int_E (\Phi_{i,K'} (\nabla \Phi_{j,K'} \cdot n_E) + (\nabla \Phi_{i,K'} \cdot n_E) \Phi_{j,K'}) ds$

Konkrete Implementierung der Kantenintegrale zum diffusiven Term

(Tabelle 5.1)

Ansatzfunktion	Testfunktion	Strafterm
Zelle	Zelle	$\sigma \int_E \Phi_{i,K} \Phi_{j,K} ds$
Zelle	Nachbarn	$-\sigma \int_E \Phi_{i,K} \Phi_{j,K'} ds$
Nachbarn	Zelle	$-\sigma \int_E \Phi_{i,K'} \Phi_{j,K} ds$
Nachbarn	Nachbarn	$\sigma \int_E \Phi_{i,K'} \Phi_{j,K'} ds$

Konkrete Implementierung der Kantenintegrale zum Strafterm

(Tabelle 5.2)

Die bisher auftretenden Integrale lassen sich am einfachsten behandeln, indem man eine Schleife über die Gitterzellen durchläuft und innerhalb dieser eine Schleife über die Kanten der jeweiligen Gitterzelle. Innerhalb dieser Schleife ist eine Behandlung des Integrals, welches vom konvektiven Term herrührt, in seiner jetzigen Gestalt nicht möglich, da hier nur über

den Einströmbereich des Randes der Zelle integriert und die auftretende Summe daher über alle Zellen und nicht über alle Kanten geführt wird. Um dies zu berücksichtigen, wollen wir zunächst zur Vereinfachung annehmen, dass b auf jeder Kante konstant ist. Damit ist auch das Produkt $b \cdot n_k$ auf jeder Kante konstant, womit die Kante als Ganzes entweder zum Einströmbereich des Randes von K gehört oder nicht. Gehört sie zum Einströmbereich von K , ist also $(b \cdot n_k) < 0$, so tritt in der Summe daher das Integral

$$- \int_E (b \cdot n_K) [u]_K v_K^+ ds$$

auf. Gehört die Kante nicht zum Einströmbereich von K , so gehört sie zum Ausströmbereich von K und damit zum Einströmbereich der Nachbarzelle K' . In der Summe tritt daher das Integral

$$- \int_E (b \cdot n_{K'}) [u]_{K'} v_{K'}^+ ds = \int_E (b \cdot n_K) [u]_{K'} v_{K'}^+ ds$$

auf, wenn $(b \cdot n_K) > 0$ gilt. Wir haben hier die Beziehung $n_{K'} = -n_K$ benutzt. Den Fall $(b \cdot n_K) = 0$ brauchen wir nicht zu berücksichtigen, da sich in diesem Fall der Integrand zu Null ergibt. Lassen wir nun zu, dass sich das Vorzeichen des Produktes entlang einer Kante ändert, so können Integrale der beiden vorherigen Typen für eine Kante E auftreten. Der Beitrag der Kante lautet dann

$$- \int_{x \in E | (b \cdot n_K) < 0} (b \cdot n_K) [u]_K v_K^+ ds + \int_{x \in E | (b \cdot n_K) > 0} (b \cdot n_K) [u]_{K'} v_{K'}^+ ds.$$

Die tatsächlich auftretenden Terme können wir nun wieder in einer Tabelle auflisten:

Ansatzfunktion	Testfunktion	konvektives Kantenintegral
Zelle	Zelle	$-\int_{x \in E (b \cdot n_K) < 0} (b \cdot n_K) \Phi_{i,K} \Phi_{j,K} ds$
Zelle	Nachbarn	$-\int_{x \in E (b \cdot n_K) > 0} (b \cdot n_K) \Phi_{i,K} \Phi_{j,K'} ds$
Nachbarn	Zelle	$\int_{x \in E (b \cdot n_K) < 0} (b \cdot n_K) \Phi_{i,K'} \Phi_{j,K} ds$
Nachbarn	Nachbarn	$\int_{x \in E (b \cdot n_K) > 0} (b \cdot n_K) \Phi_{i,K'} \Phi_{j,K'} ds$

Konkrete Implementierung der Kantenintegrale zum konvektiven Term
(Tabelle 5.3)

Die Berechnung der Integrale erfolgt in MooNMD mittels Gaußscher Quadraturformeln. Bei der Implementierung der konvektiven Integrale wurde zur Zerlegung einer Kante E in Einström- und Ausströmbereich daher so vorgegangen, dass für jeden Quadraturpunkt getestet wird, ob dieser im Einström- oder Ausströmbereich liegt und sein Beitrag zum Integral über die Kante entsprechend gewertet.

Wie wir schon erwähnt haben, sorgen die Kantenintegrale dafür, dass sich in der Matrix weitere Nicht-Null-Einträge ergeben. Des Weiteren haben wir schon erwähnt, dass bei Verwendung von Räumen unstetiger Funktionen jede lokale Basisfunktion zugleich globale

Basisfunktion ist. Dies hat zur Folge, dass auch die Anzahl der Basisfunktionen gegenüber einer Methode mit Räumen stetiger Funktionen erhöht ist. Somit ist die Dimension der Matrix auch größer.

Zur Lösung des linearen Gleichungssystems setzen wir bei allen Beispielen den direkten Löser aus dem Paket UMFPACK ein (Dav04b).

5.1.2 Wahl des Parameters der Streamline–Diffusion–Methode

Wie wir in Abschnitt 4.4 gesehen haben, tritt bei der Streamline–Diffusion–Methode der Parameter τ_K auf. Eine mögliche Wahl dieses Parameters ist in (4.4.9) gegeben. Hierbei treten die Parameter τ_0 und τ_1 auf. Eine optimale Wahl dieser Parameter hängt vom jeweiligen Problem ab. Wir folgen einem Vorschlag von Stynes und Tobiska (Sty95) und setzen

$$\tau_K = \frac{\text{diam}(K)}{2\|b\|_{L^\infty(K)}} L(\text{Pe}_K/2) \quad \text{mit} \quad \text{Pe}_K = \frac{\|b\|_{L^\infty(K)} \text{diam}(K)}{2\varepsilon} \quad (5.1.1)$$

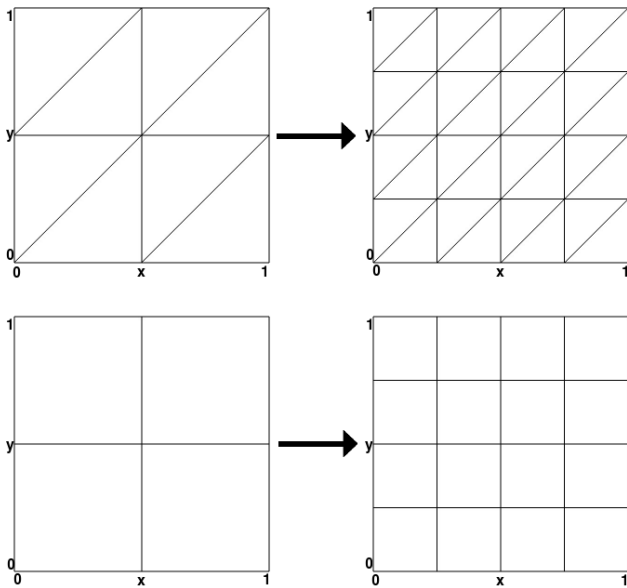
wobei $\text{diam}(K) = \sup(|x - y|; x, y \in K)$ und $L(\alpha)$ die Langevin–Funktion mit

$$L(\alpha) = \coth(\alpha) - \frac{1}{\alpha}$$

sind. Diese Wahl von τ_K ist in MooNMD bereits implementiert.

5.1.3 Wahl der Gebietszerlegung

Im Rahmen dieser Arbeit sollen Finite–Elemente auf Dreiecken und Rechtecken betrachtet werden. Als Gebiet Ω betrachten wir in allen Beispielen das Einheitsquadrat $[0, 1]^2$, welches wir in Dreiecke oder Quadrate zerlegen. Die Art der Zerlegung ist in Abbildung 5.1 dargestellt. Links ist dort die gröbste Zerlegung zu sehen, für welche wir die Konvektions–Diffusions–Gleichung lösen werden. Sie ist gegenüber der größtmöglichen Zerlegung schon um eine Stufe verfeinert und wird als Level–1–Gitter bezeichnet. Rechts ist die nächst feinere Zerlegung abgebildet, so dass die Art der Gitterverfeinerung klar wird. Die Anzahl der Gitterzellen vervierfacht sich dabei bei jeder Verfeinerung, so dass nach n Verfeinerungen ein Level– n –Gitter mit $2 \cdot 4^n$ Dreiecken beziehungsweise 4^n Quadraten vorliegt. Damit ergibt sich die Anzahl der Freiheitsgrade, welche zugleich der Dimension der Matrix entspricht, für die von uns betrachteten Finiten–Elemente entsprechend folgender Tabelle:



Zerlegung des Einheitsquadrates $[0, 1]^2$ in Dreiecke oder Quadrate. Die Abbildung zeigt die in dieser Arbeit benutzen Gitter auf den Leveln 1 und 2, wodurch auch die Art der Gitterverfeinerung beschrieben wird.

(Abbildung 5.1)

Element	Gittertyp	Freiheitsgrade SIPG	Freiheitsgrade SDFEM
P_1^{disc}	Dreiecke	$6 \cdot 4^n$	$(2^n + 1)^2$
P_1^{disc}	Quadrate	$3 \cdot 4^n$	$(2 \cdot 2^n + 1)^2$
P_2^{disc}	Dreiecke	$12 \cdot 4^n$	$(3 \cdot 2^n + 1)^2$
P_2^{disc}	Quadrate	$6 \cdot 4^n$	/
P_3^{disc}	Dreiecke	$20 \cdot 4^n$	/
P_3^{disc}	Quadrate	$10 \cdot 4^n$	/
Q_1^{disc}	Quadrate	$4 \cdot 4^n$	$(2^n + 1)^2$
Q_2^{disc}	Quadrate	$9 \cdot 4^n$	$(2 \cdot 2^n + 1)^2$
Q_3^{disc}	Quadrate	$16 \cdot 4^n$	$(3 \cdot 2^n + 1)^2$

Freiheitsgrade der verschiedenen Finite-Elemente-Räume

(Tabelle 5.4)

5.2 Rechnungen für verschiedene Beispiele

In diesem Abschnitt soll das SIPG-Verfahren näher untersucht werden. Zunächst werden wir an einfachen Beispielen die Implementierung der einzelnen Terme überprüfen. Anschließend werden wir an einem Beispiel mit glatter Lösung, das heißt ohne Grenzschichten, testen, ob auch die volle Konvektions-Diffusions-Gleichung erwartungsgemäß gelöst wird. Schließlich werden wir Beispiele mit Grenzschichten betrachten. Die Güte des Verfahrens soll an Hand der L^2 -Norm und der H^1 -Seminorm des Fehler $u - u_h$ sowie der zugehörigen Konvergenzordnungen bestimmt werden. Dabei wird der Parameter σ im Bereich von 1 bis $2 \cdot 10^6$ variiert. Zusätzlich betrachten wir Abbildungen der numerisch bestimmten Lösung bei ausgesuchten, möglichst optimalen Parametern. Zum Vergleich dient uns jeweils das Streamline-Diffusion-Verfahren. Um einen fairen Vergleich zu ziehen, wird hier auf SOLD-Terme zum shock-capturing verzichtet, da solche Terme für das SIPG-Verfahren noch nicht implementiert wurden.

Für alle im Folgenden aufgeführten Beispiele wurde das SIPG-Verfahren mit den Polynomräumen P_1^{disc} , P_2^{disc} und P_3^{disc} auf dem Dreiecksgitter und dem Quadratgitter sowie mit den Räumen Q_1^{disc} , Q_2^{disc} und Q_3^{disc} auf dem Quadratgitter angewendet. Für das Streamline-

Diffusion-Verfahren wurden die stetigen Räume P_1 , P_2 und P_3 auf dem Dreiecksgitter und die Räume Q_1 , Q_2 und Q_3 auf dem Quadratgitter benutzt. Den hochgestellten Index drücken wir im Folgenden. Es wurden jeweils Lösungen auf den Verfeinerungsleveln 1 bis 6 berechnet. Die Ergebnisse zum Level-1-Gitter werden im Folgenden nicht dargestellt. Sie sind vernachlässigbar, da die Konvergenzaussagen aus Kapitel 4 nur asymptotisch für eine gegen Null gehende Gitterweite h gelten.

Die angegebenen Konvergenzraten berechnen sich nach der Formel

$$\text{ord}_i = \log_2 \frac{\|u - u_{h,i}\|}{\|u - u_{h,i-1}\|}.$$

Hierbei ist ord_i die Konvergenzordnung auf Gitterlevel i , u die bekannte Lösung des jeweiligen Beispiels, $u_{h,i}$ die numerisch bestimmte Lösung auf Gitterlevel i und $\|\cdot\|$ eine der betrachteten Normen oder Halbnormen.

5.2.1 Test der Implementierung

Um sicherzustellen, dass die einzelnen Terme des SIPG-Verfahrens richtig implementiert sind, wurden zunächst die Poissongleichung (4.5.6) beziehungsweise die Konvektionsgleichung (4.5.15) betrachtet.

5.2.1.1 Diffusive Terme und Strafterme

Bei der Behandlung der Poissongleichung müssen auch die Strafterme mit berücksichtigt werden. Als Testbeispiele wählen wir hier zwei Beispiele aus (Riv08):

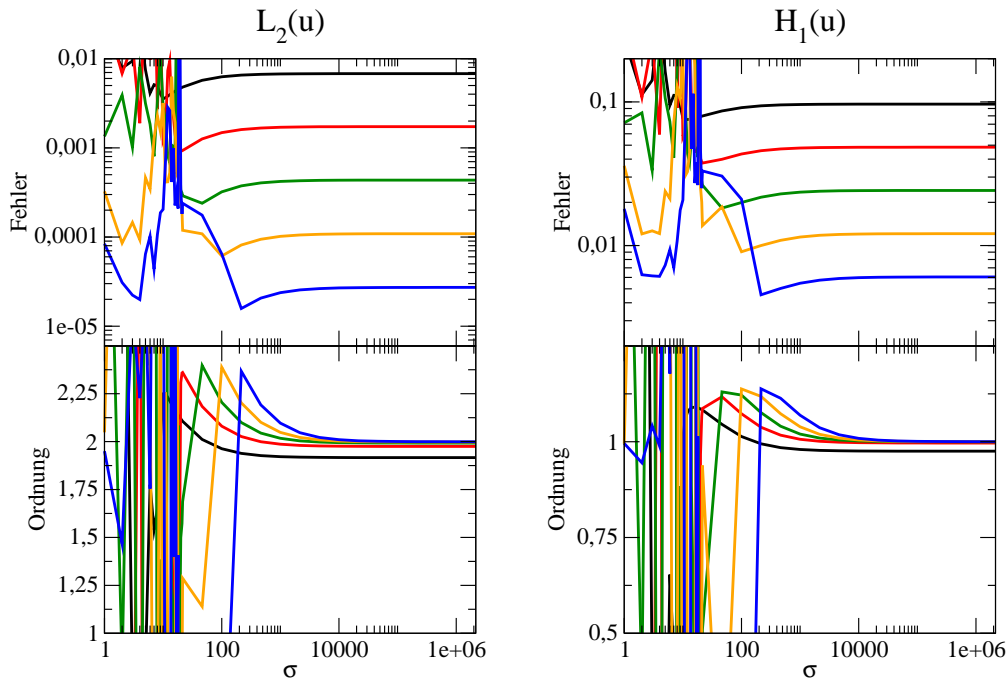
$$u = e^{-x-y^2} \quad \text{auf } \Omega = [0, 1]^2, \quad (5.2.1a)$$

$$u = x(x-1)y(y-1)e^{-x-y^2} \quad \text{auf } \Omega = [0, 1]^2. \quad (5.2.1b)$$

Die rechte Seite f von (4.5.6) und die Randbedingungen wurden dabei so gewählt, dass u die Gleichung löst. Wir haben beide Beispiele gewählt, da (5.2.1b) homogene Dirichlet-Randbedingungen erfüllt. Somit spielen die Integrale, welche die Randbedingungen schwach implementieren, hier keine Rolle. Das Beispiel (5.2.1a) hingegen besitzt inhomogene Randbedingungen, so dass sich mit diesem untersuchen lässt, ob die Randbedingungen richtig implementiert wurden.

Da die hier aufgeführten Beispiele nur dazu dienen, die Richtigkeit der Implementierung sicher zu stellen, beschränken wir uns auf die Darstellung der Fehler und Konvergenzraten für P_1 -Elemente auf Dreiecken und Q_1 -Elementen auf Quadraten. Ergebnisse für Räume mit höherem Polynomgrad werden wir im Rahmen der vollen Konvektions-Diffusions-Gleichung behandeln. Es sei aber erwähnt, dass diese für die beiden diskutierten Beispiele qualitativ ähnliche Ergebnisse zeigen. Ebenso verzichten wir an dieser Stelle auf einen Vergleich mit dem Streamline-Diffusion-Verfahren.

Abbildung 5.2 zeigt für Beispiel (5.2.1b) die L^2 -Norm und die H^1 -Seminorm des Fehlers $u - u_h$ bei Verwendung von P_1 -Elementen auf dem Dreiecksgitter und die dazugehörigen Konvergenzordnungen. Die verschiedenfarbigen Linien beziehen sich auf die Gitterlevel 2 (schwarz), 3 (rot), 4 (grün), 5 (orange) und 6 (blau). Wie wir sehen, ist das Verhalten des SIPG-Verfahrens hier für kleine Werte von σ , das heißt bis ungefähr 20, sehr uneinheitlich. Ab diesem Wert werden die Fehler dann kleiner, erreichen ein Minimum, wachsen wieder ein wenig an und nehmen für weiter wachsendes σ einen festen Wert an. Die Position des



Fehler und Konvergenzordnungen Finiterelemente erster Ordnung auf dem Dreiecksgitter (P_1) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.1b). In der oberen Reihe werden der L^2 - und der H^1 -Fehler gezeigt, in der unteren die zugehörigen Ordnungen. Beides wird dargestellt für die Gitterlevel 2 (schwarz), 3 (rot), 4 (grün), 5 (orange) und 6 (blau).

(Abbildung 5.2)

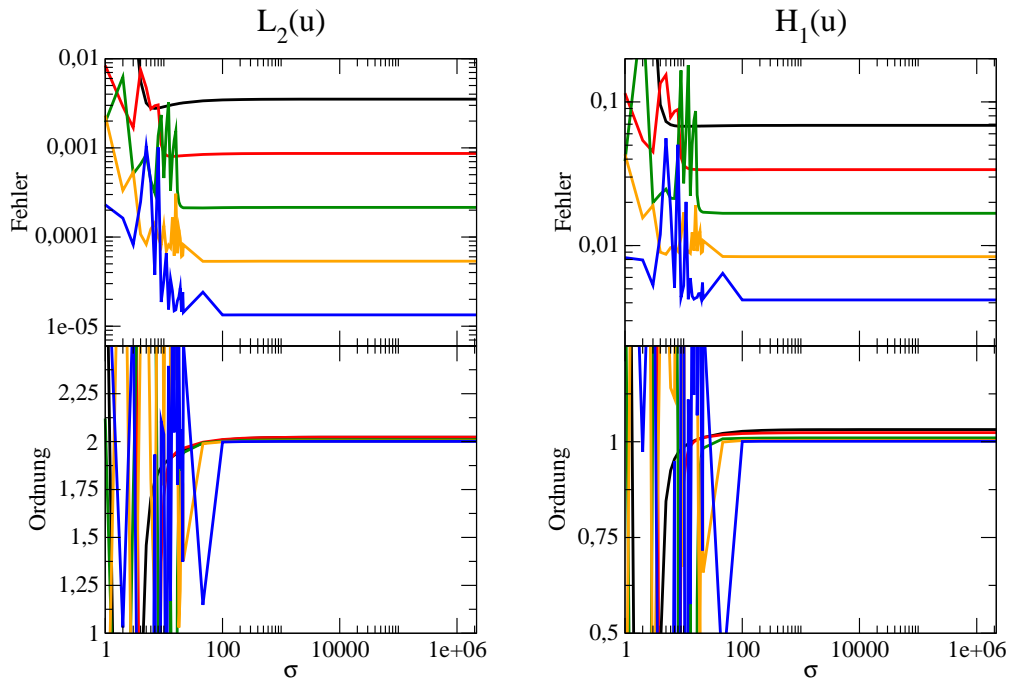
Minimum hängt dabei vom Gitterlevel ab. Auf Grund des sprunghaften Verhaltens der Fehlerwerte sind auch die Konvergenzordnungen für kleine σ sehr unterschiedlich. Da sich die Position des Minimums der Fehler verschiebt, kommt es jeweils zu einer Spitze in der Konvergenzordnung, bevor diese dann mit weiter wachsendem σ konstant wird und die aus der Fehlerabschätzung zu erwartenden Werte annimmt.

Die L^2 -Norm und H^1 -Seminorm sowie die zugehörigen Konvergenzordnungen für Beispiel (5.2.1b) bei Verwendung des Q_1 -Elementes auf dem Quadratgitter sind in Abbildung 5.3 dargestellt. Wie zu erkennen, zeigen diese ein Verhalten, welches sehr ähnlich zum P_1 -Element auf Dreiecken ist. Im Unterschied zu diesem entspricht der konstante Wert, welcher für große σ angenommen wird, jedoch dem Minimum der Fehler. Dieses wird schon für leicht kleinere Werte von σ als beim P_1 -Element angenommen. Da der Fehler hier ab Erreichen des Minimums nicht mehr wächst, kommt es auch ab diesem zu keinen Spitzen mehr in den Konvergenzordnungen. Diese erreichen wieder die von der Theorie vorhergesagten Werte.

Nachdem wir nun ein Beispiel mit homogenen Dirichlet-Randwerten behandelt haben, wollen wir an Hand eines Beispiels mit inhomogenen Dirichlet-Randwerten die schwache Implementierung der Randbedingungen für das Dreiecksgitter untersuchen.

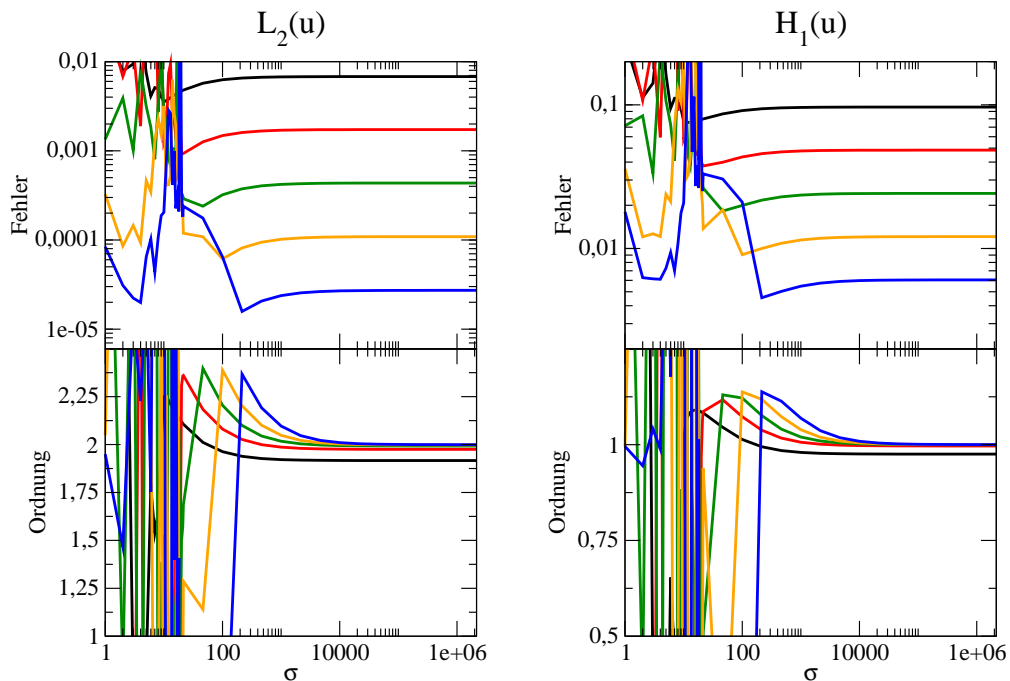
In Abbildung 5.4 sind die schon bekannten Fehler und zugehörigen Konvergenzordnungen für das Beispiel (5.2.1a) aufgetragen. Es zeigt sich ein praktisch zu Abbildung 5.2 identisches Bild. Dies ist nicht verwunderlich, da das Verhalten wesentlich von der Exponentialfunktion, welche beiden Beispielen gemein ist, bestimmt wird. Dadurch zeigt sich, dass auch die Randintegrale richtig implementiert sind. Abbildung 5.5 bestätigt uns dies nochmal für das Quadratgitter. Auch hier ist kein Unterschied zu Abbildung 5.3 zu erkennen.

Neben den hier gezeigten Beispielen wurden auch Lösungen betrachtet, welche Polynome



Fehler und Fehlerordnungen Finiter-Elemente erster Ordnung auf dem Quadratgitter (Q_1) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.1b). Für weitere Erklärungen siehe Abbildung 5.2.

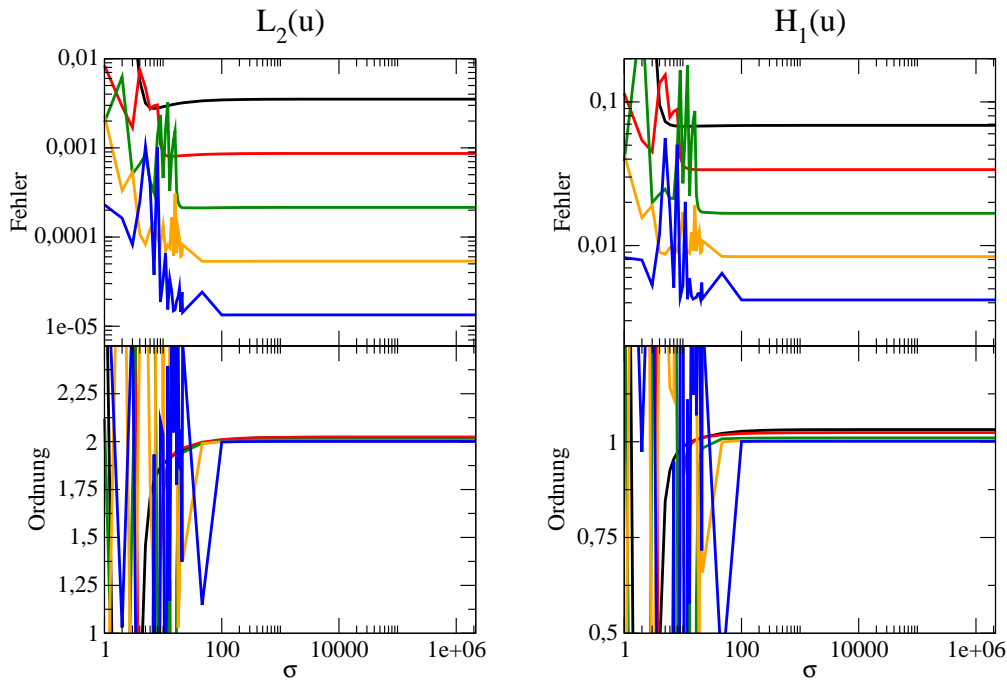
(Abbildung 5.3)



Fehler und Fehlerordnungen Finiter-Elemente erster Ordnung auf dem Dreiecksgitter (P_1) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.1a). Für weitere Erklärungen siehe Abbildung 5.2.

(Abbildung 5.4)

darstellen und im Ansatzraum liegen. Diese konnten numerisch exakt, das heißt mit Fehlern in der Größenordnung von 10^{-15} reproduziert werden.



Fehler und Fehlerordnungen Finiter-Elemente erster Ordnung auf dem Quadratgitter (Q_1) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.1a). Für weitere Erklärungen siehe Abbildung 5.2.

(Abbildung 5.5)

5.2.1.2 Konvektive Term und Strafterme

Als nächstes wollen wir die Implementierung der konvektiven Terme untersuchen. Dazu betrachten wir die reine Konvektionsgleichung (4.5.15) mit $b = (1, 2)^T$ und homogenen Dirichlet-Randwerten auf dem Einströmrand $\Gamma_- = \{(x, y) \in [0, 1]^2 \mid x = 0 \vee y = 0\}$. Wir haben hier den Vektor b derart gewählt, dass er zu keiner vorkommenden Kante parallel ist. Die rechte Seite bestimmen wir so, dass

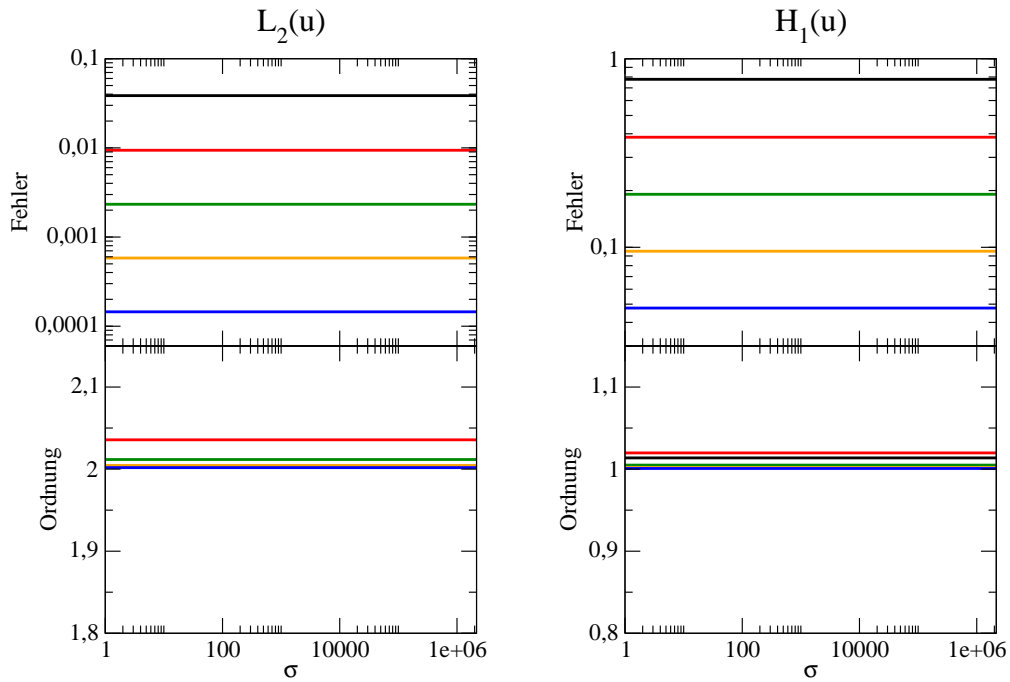
$$u = \sin(\pi x) \sin(\pi y) \quad (5.2.2)$$

die Gleichung löst. Zusätzlich zur Bilinearform a_{konv} berücksichtigen wir hier auch noch die Strafterme, um ihren Einfluss im konvektiven Fall zu testen.

Abbildung 5.6 zeigt für Beispiel (5.2.2) die L^2 -Norm und die H^1 -Seminorm des Fehlers $u - u_h$ bei Verwendung von P_1 -Elementen auf dem Dreiecksgitter und die dazugehörigen Konvergenzordnungen. Die Konvergenzordnungen entsprechen den Erwartungen. Insbesondere sind Fehler und Ordnungen unabhängig von σ . Dies ist im Einklang mit den Fehlerabschätzungen zur Bilinearform a_{konv} in Abschnitt 4.5, für welche die Strafterme nicht gebraucht werden.

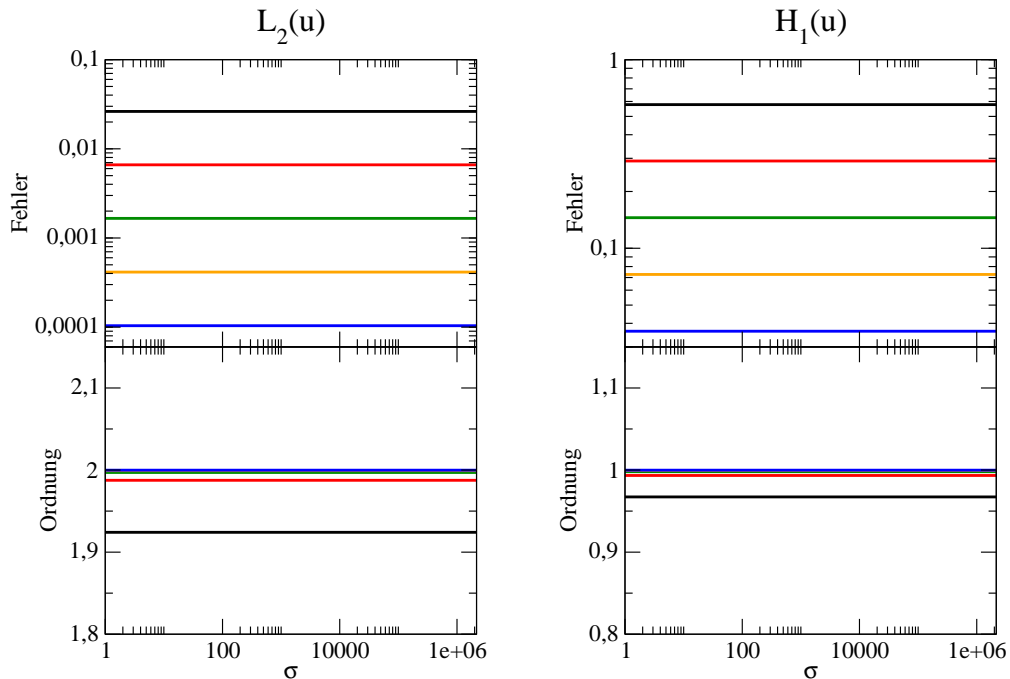
Abbildung 5.7 zeigt die Fehler und Konvergenzordnungen für das Beispiel (5.2.2) bei Verwendung von Q_1 -Elementen auf dem Quadratgitter. Die Fehler sind hier ein wenig größer als für die P_1 -Elemente auf Dreiecken. Der Vergleich zwischen den beiden Elementen fällt also hier umgekehrt gegenüber dem Vergleich bei den rein diffusiven Beispielen aus. Die Konvergenzordnungen der Q_1 -Elemente entsprechen wieder den Erwartungen. Auch hier zeigt sich keine Abhängigkeit der Ergebnisse von σ .

Für den rein konvektiven Fall wurden ebenfalls zusätzlich Lösungen betrachtet, welche



Fehler und Fehlerordnungen Finiter-Elemente erster Ordnung auf dem Dreiecksgitter (P_1) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.2). Für weitere Erklärungen siehe Abbildung 5.2.

(Abbildung 5.6)



Fehler und Fehlerordnungen Finiter-Elemente erster Ordnung auf dem Quadratgitter (Q_1) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.2). Für weitere Erklärungen siehe Abbildung 5.2.

(Abbildung 5.7)

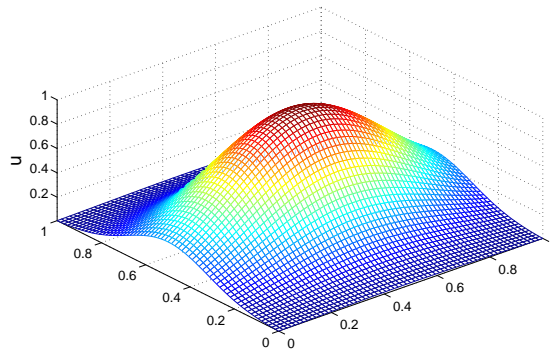
Polynome darstellen und im Ansatzraum liegen. Auch in diesem Fall konnten die Polynome numerisch exakt reproduziert werden.

5.2.2 Beispiel ohne Grenzschichten

Nachdem wir die einzelnen Terme auf ihre richtige Implementierung untersucht haben, wollen wir nun klären, ob wir auch die volle Konvektions–Diffusions–Gleichung erwartungsgemäß lösen können. Dazu betrachten wir hier folgende exponentielle Funktion, welche auch in (Bur04) als Testbeispiel verwendet wurde:

$$u(x, y) = e^{-5(x-0.5)^2 - 15(y-0.5)^2} \quad (5.2.3)$$

auf dem Einheitsquadrat $\Omega = [0, 1]^2$. Die Vorgaben auf Ω lauten $\epsilon = 10^{-6}$, $b = (1, 0)^T$ und $c = 1$. Wir befinden uns also im konvektions–dominanten Fall. Zu diesen Vorgaben wird die rechte Seite f der Konvektions–Diffusions–Gleichung so angepasst, dass u die Gleichung löst.



Darstellung der exakten Lösung für das Beispiel (5.2.3).

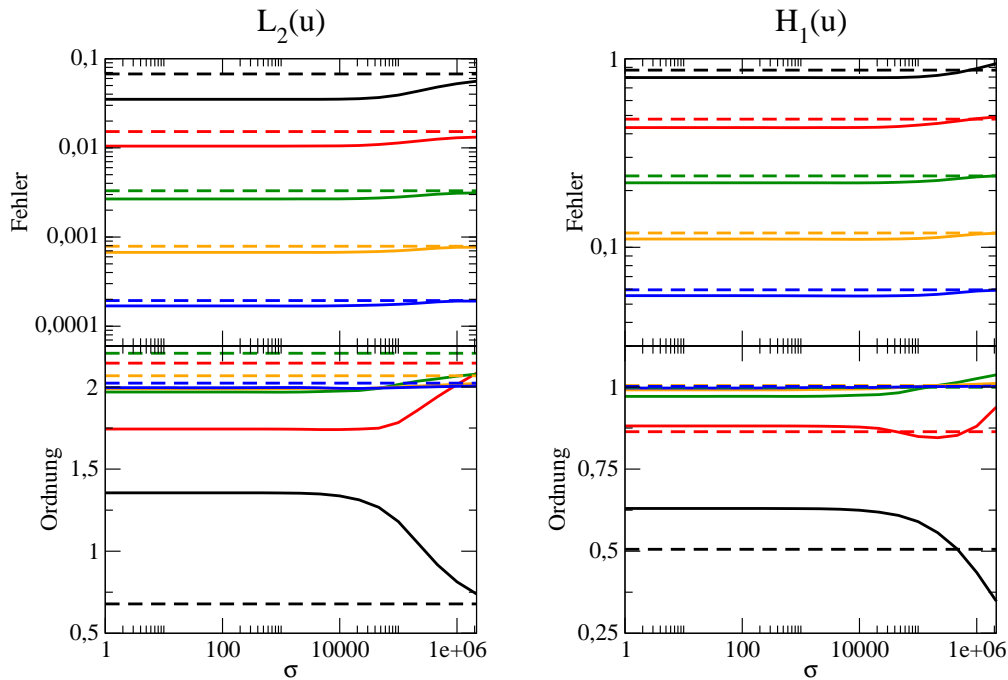
(Abbildung 5.8)

In Abbildung 5.8 ist u über dem Einheitsquadrat dargestellt. Wie wir sehen, besitzt die Funktion keine Grenzschichten, sondern ändert sich sehr gleichmäßig ohne steile Gradienten.

Abbildung 5.9 zeigt die L^2 -Norm und die H^1 -Seminorm des Fehlers $u - u_h$ bei Verwendung von P_1 -Elementen auf dem Dreiecksgitter und die zugehörigen Konvergenzordnungen. Der Farbcode entspricht dem von Abbildung 5.2. Zum Vergleich sind hier die Fehler und Konvergenzordnungen, welche sich für das Streamline–Diffusion–Verfahren ergeben, mit gestrichelten Linien aufgenommen. Beide Verfahren zeigen auf allen Levels vergleichbare Fehler und Konvergenzordnungen, wobei die Fehler des SIPG–Verfahrens für sehr große σ etwas anwachsen. Im Gegensatz zum rein diffusiven Fall in der Poissongleichung lassen sich hier auch bei kleinen Werten von σ gute Ergebnisse erzielen.

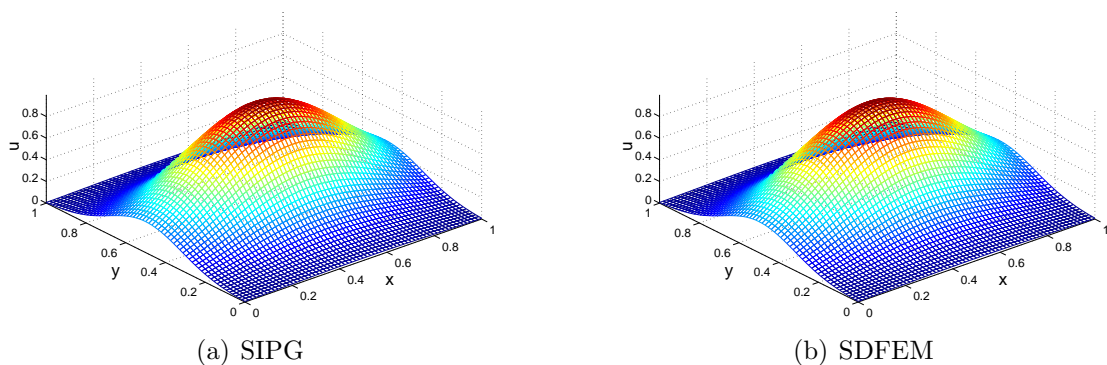
Einen direkten Vergleich der numerisch bestimmten Lösungen zeigt Abbildung 5.10. Links ist die mit Hilfe des SIPG–Verfahrens berechnete Lösung abgebildet, rechts diejenige des Streamline–Diffusion–Verfahrens. Beide Lösungen beziehen sich auf Gitterlevel 6. Ein Vergleich mit Abbildung 5.8 zeigt, dass beide Verfahren die exakte Lösung gut wiedergeben. Auch zwischen den Verfahren lässt sich mit bloßem Auge kein Unterschied erkennen. Ein Vergleich der numerisch bestimmten Lösungen für Elemente höherer Ordnung und für Rechnungen mit dem Quadratgitter fällt ebenso aus. Da sich mit bloßem Auge keine Unterschiede erkennen lassen, verzichten wir an dieser Stelle und im Folgenden darauf neben der Lösung für P_1 -Elemente auf dem Dreiecksgitter weitere numerisch bestimmte Lösungen zu den Beispielen zu zeigen.

Abbildung 5.11 zeigt die Ergebnisse, welche sich für P_1 -Elemente auf dem Quadratgitter ergeben. Ein Vergleich mit der Streamline–Diffusion–Methode ist hier nicht möglich, da stetige P_1 -Elemente auf Quadraten nicht existieren. Die Fehler auf dem Quadratgitter sind



Fehler und Fehlerordnungen Finites-Elemente erster Ordnung auf dem Dreiecksgitter (P_1) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.3) bei einer Diffusionskonstanten von $\epsilon = 10^{-6}$. In der oberen Reihe werden der L^2 - und der H^1 -Fehler gezeigt, in der unteren die zugehörigen Ordnungen. Beides wird dargestellt für die Gitterlevel 2 (schwarz), 3 (rot), 4 (grün), 5 (orange) und 6 (blau). Die durchgezogenen Linien gehören zum SIPG-Verfahren. Die gestrichelten Linien geben die entsprechenden Größen des Streamline-Diffusion-Verfahrens an.

(Abbildung 5.9)

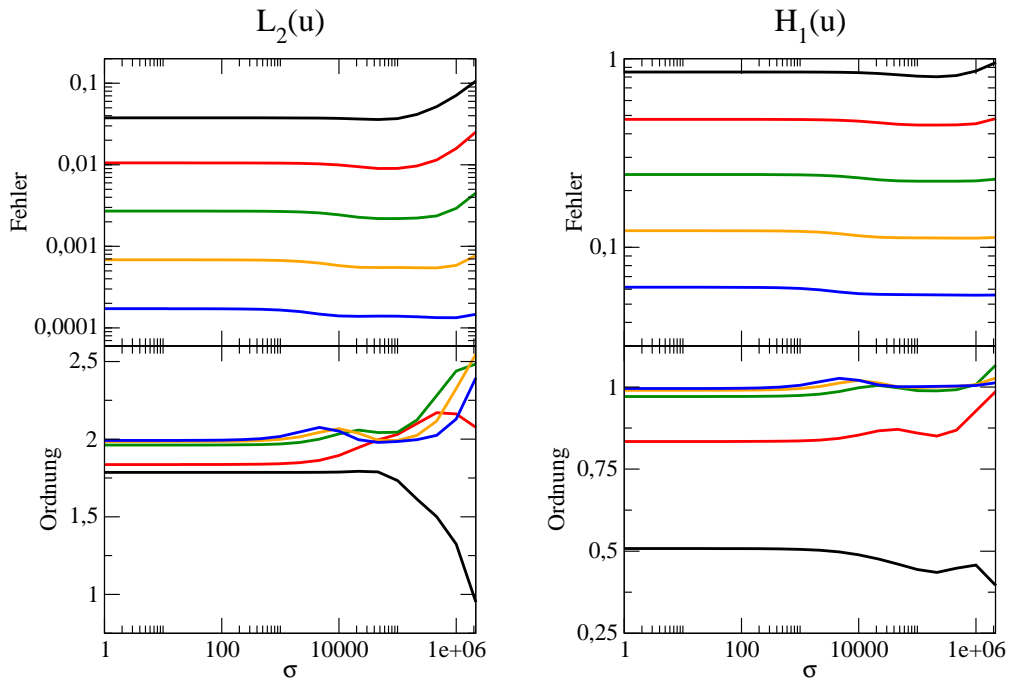


Numerische Lösung für das Beispiel (5.2.3) bei einer Diffusionskonstanten von $\epsilon = 10^{-6}$ auf Gitterlevel 6 bei $\sigma = 1000$ für das SIPG-Verfahren (links) und für das Streamline-Diffusion-Verfahren (rechts) bei Verwendung von P_1 -Elementen auf dem Dreiecksgitter. Die Wahl des Parameters erfolgte so, dass möglichst kleine Fehler und gute Konvergenz zu erwarten sind.

(Abbildung 5.10)

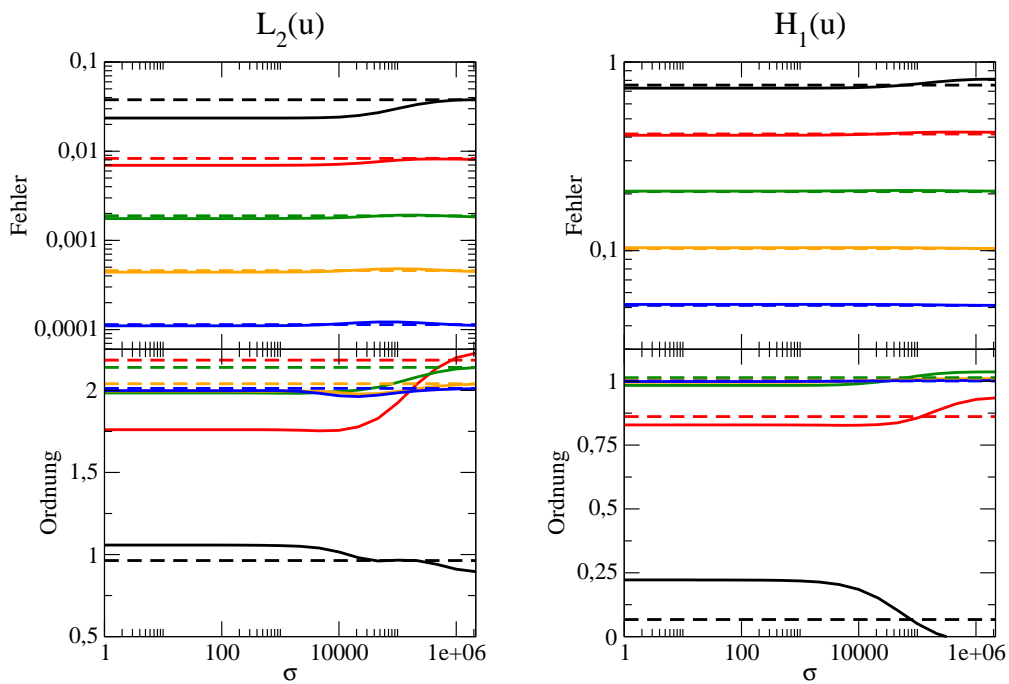
vergleichbar mit den Fehlern auf dem Dreiecksgitter. Für große σ werden die Fehler hier bei großen Gitterweiten auch größer, für entsprechend feine Gitter jedoch kleiner.

Die Fehler und Konvergenzordnungen für Q_1 -Elemente auf dem Quadratgitter sind in



Fehler und Fehlerordnungen Finiter-Elemente erster Ordnung auf dem Quadratgitter (P_1) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.3) bei einer Diffusionskonstanten von $\epsilon = 10^{-6}$. Für weitere Erklärungen siehe Abbildung 5.9.

(Abbildung 5.11)



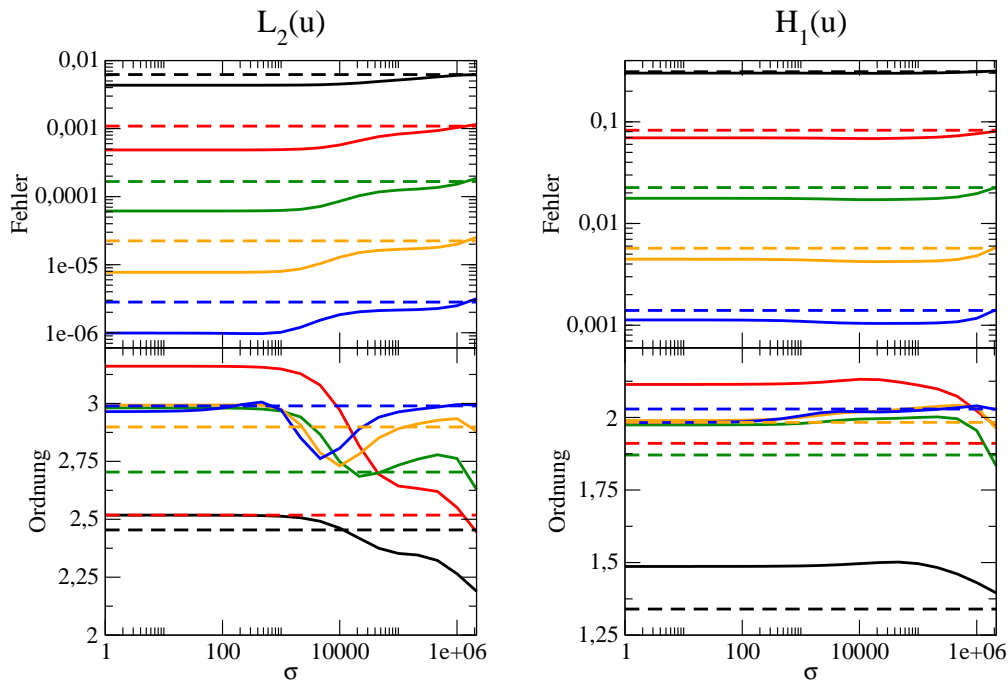
Fehler und Fehlerordnungen Finiter-Elemente erster Ordnung auf dem Quadratgitter (Q_1) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.3) bei einer Diffusionskonstanten von $\epsilon = 10^{-6}$. Für weitere Erklärungen siehe Abbildung 5.9.

(Abbildung 5.12)

Abbildung 5.12 dargestellt. Die Fehler sind hier bei genügend feinen Gittern praktisch unabhängig vom Stabilisierungsparameter σ und auf allen Gitterleveln mit den Fehlern der Streamline–Diffusion–Methode nahezu identisch. Im Vergleich mit den P_1 -Elementen liefern die Q_1 -Elemente auf dem Quadratgitter geringfügig kleinere Fehler. Die Konvergenzordnung entspricht wieder den Erwartungen.

Vergleichen wir die bisherigen Ergebnisse diese Abschnittes mit den Ergebnissen der beiden vorhergehenden Abschnitte, so lässt sich sagen, dass sich die Dominanz des konvektiven Terms hier auch darin zeigt, dass sich für kleine Werte von σ systematisch kleine Fehler ergeben, statt der unsystematischen Ergebnisse im rein diffusiven Fall.

Als nächstes wollen wir nun auch Elemente höherer Ordnung betrachten.

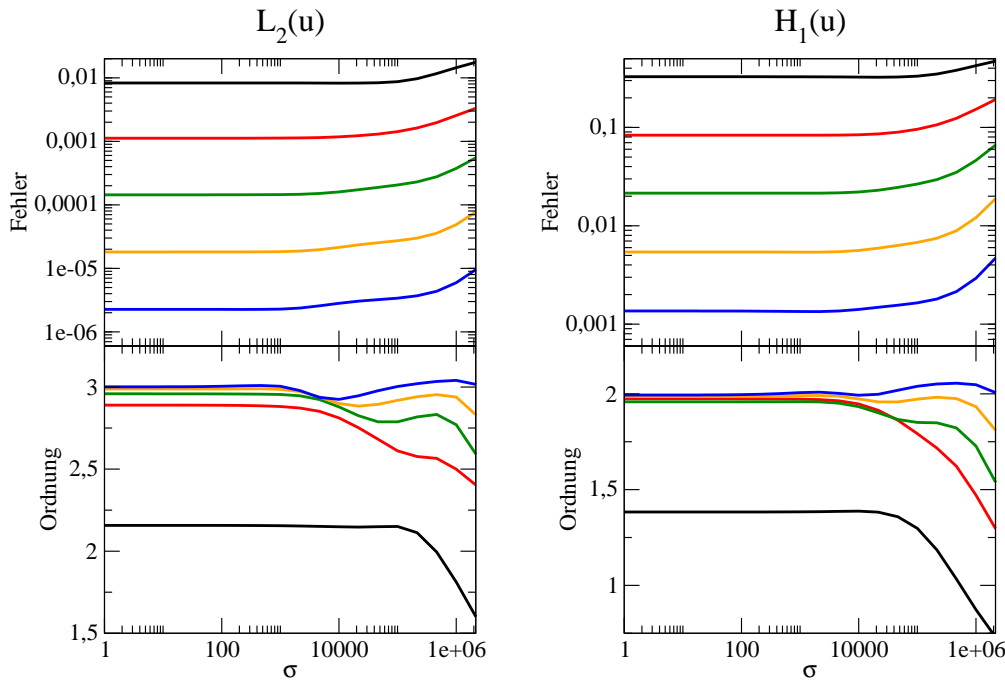


Fehler und Fehlerordnungen Finiter–Elemente zweiter Ordnung auf dem Dreiecksgitter (P_2) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.3) bei einer Diffusionskonstanten von $\epsilon = 10^{-6}$. Für weitere Erklärungen siehe Abbildung 5.9.

(Abbildung 5.13)

In Abbildung 5.13 sind die Fehler und Konvergenzordnungen bei Verwendung von P_2 -Elementen auf dem Dreiecksgitter zu sehen. Im Vergleich zu den P_1 -Elementen zeigt sich hier eine stärkere Abhängigkeit des Fehlers von σ . Ab einem genügend großen Wert von σ wachsen die L^2 -Fehler mit wachsendem σ . Die H^1 -Fehler wachsen ebenfalls, allerdings erst bei größeren Werten von σ . Was dabei “genügend groß” bedeutet, hängt vom Gitterlevel ab. Mit zunehmendem Gitterlevel wandert die Schwelle, ab welcher die Fehler größer werden, zu kleineren σ . Dies spiegelt sich auch in der Konvergenzordnung wieder, da diese bei den entsprechenden Werten ein Minimum annimmt. Abgesehen davon entsprechen die Konvergenzordnungen wieder den Erwartungen. Die Fehler des SIPG–Verfahrens sind hier nicht größer als die Fehler des Streamline–Diffusion–Verfahrens.

Abbildung 5.14 zeigt die Ergebnisse zu P_2 -Elementen auf dem Quadratgitter. Auch hier führt die Verwendung großer Stabilisierungsparameter zu größeren Fehlern. Insgesamt fallen die Fehler hier ein wenig größer aus als bei den P_2 -Elementen auf dem Dreiecksgitter. Dies ist nicht verwunderlich. Da die Anzahl der Basisfunktionen pro Gitterzelle für P_p -Elemente



Fehler und Fehlerordnungen Finite-Elemente zweiter Ordnung auf dem Quadratgitter (P_2) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.3) bei einer Diffusionskonstanten von $\epsilon = 10^{-6}$. Für weitere Erklärungen siehe Abbildung 5.9.

(Abbildung 5.14)

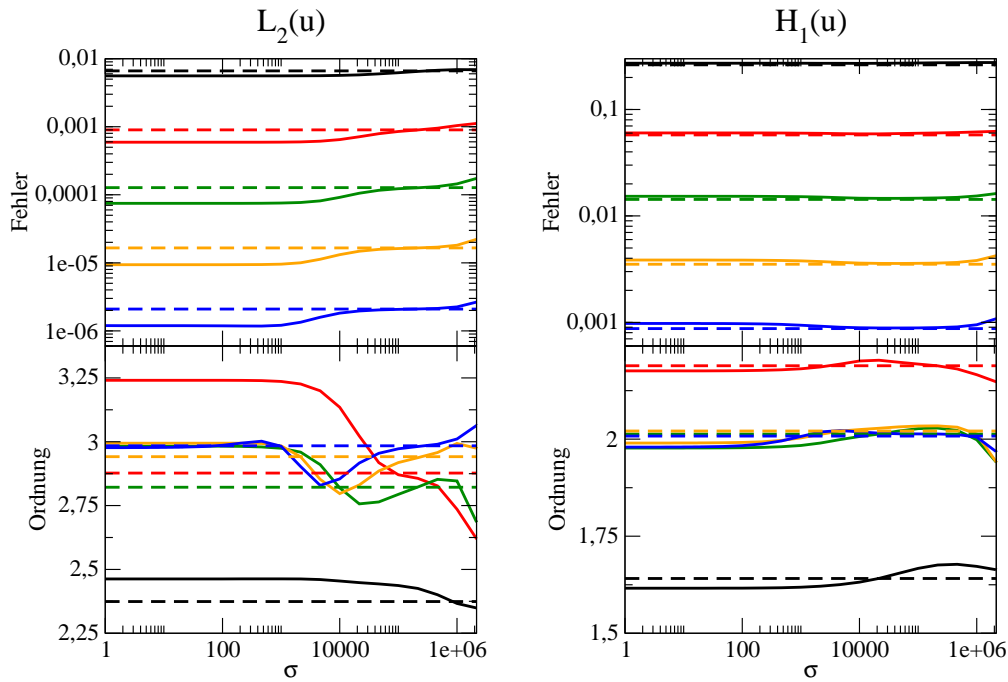
auf Dreiecken und Quadraten gleich ist, das Dreiecksgitter aber doppelt so viele Gitterzellen wie das Quadratgitter gleichen Levels enthält, stehen auf dem Dreiecksgitter insgesamt mehr Basisfunktionen zur Verfügung, was eine bessere Näherung der exakten Lösung erlaubt. Die Konvergenzordnung erfüllt auch für P_2 -Elemente auf dem Quadratgitter die Erwartungen.

Die Fehler und Konvergenzordnungen, welche sich für Beispiel (5.2.3) bei Verwendung von Q_2 -Elementen auf dem Quadratgitter ergeben, sind in Abbildung 5.15 aufgetragen. Sie zeigen eine größere Abhängigkeit der Fehler vom Stabilisierungsparameter als die Q_1 -Elemente. Der Verlauf der Fehler mit σ ist ähnlich zu den P_2 -Elementen auf dem Dreiecksgitter. Die Konvergenzordnungen entsprechen den Erwartungen. Im Vergleich mit dem Streamline-Diffusion-Verfahren sind die L^2 -Fehler des SIPG-Verfahrens hier leicht kleiner. Der Unterschied in den H^1 -Fehlern ist vernachlässigbar.

Schließlich betrachten wir noch Polynomräume der Ordnung 3. Abbildung 5.16 zeigt die bekannten Größen für P_3 -Elemente auf Dreiecken. Ihr Verhalten ist dem für P_1 -Elemente ähnlicher als dem für P_2 -Elemente. Die Fehler sind nahezu konstant und nahezu identisch mit den Fehlern des Streamline-Diffusion-Verfahrens. Auch die Konvergenzordnungen fallen erwartungsgemäß aus.

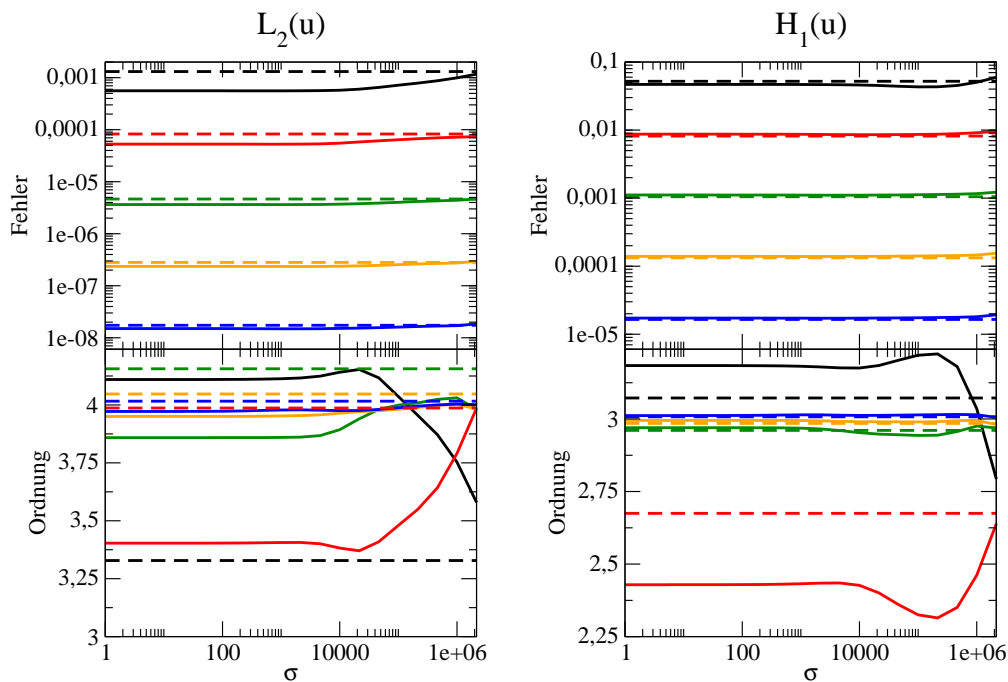
Auch auf Quadraten ähnelt das Verhalten der Fehler und Konvergenzordnungen für P_3 -Elemente eher dem für P_1 -Elemente als dem für P_2 -Elemente, was aus Abbildung 5.17 ersichtlich wird. Jedoch wachsen die Fehler hier auf allen abgebildeten Gitterlevels bei sehr großen Werten von σ leicht an.

Fehler und Konvergenzordnungen für Q_3 -Elemente sind in Abbildung 5.18 abgebildet. Die Fehler sind hier praktisch unabhängig vom Stabilisierungsparameter und annähernd identisch mit den Fehlern des Streamline-Diffusion-Verfahrens. Die Konvergenzordnungen entsprechen wieder den Erwartungen.



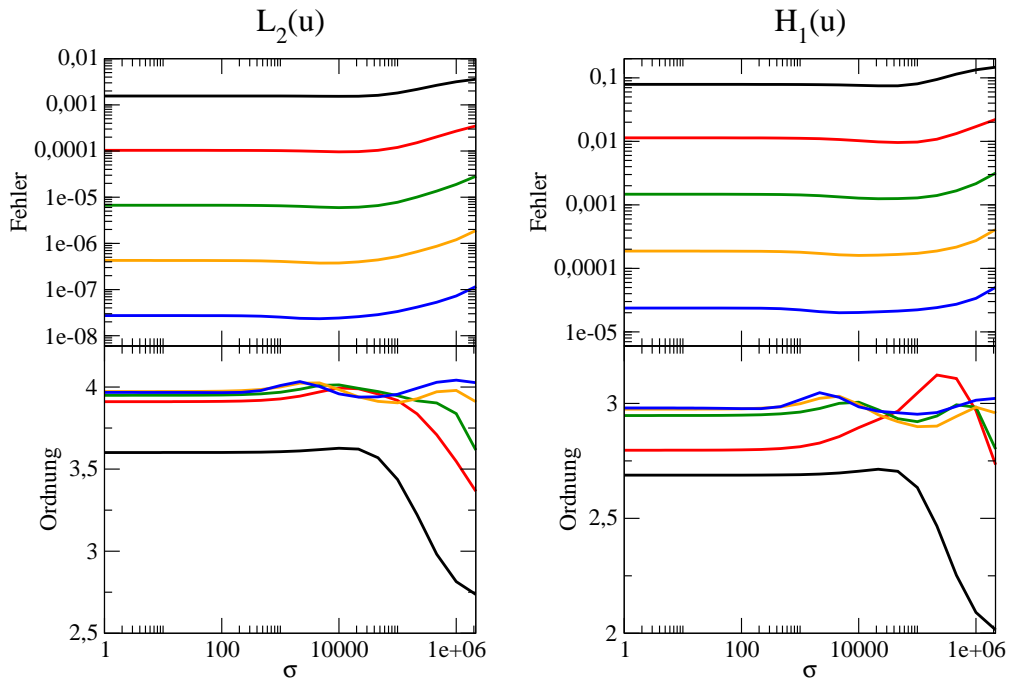
Fehler und Fehlerordnungen Finiter-Elemente zweiter Ordnung auf dem Quadratgitter (Q_2) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.3) bei einer Diffusionskonstanten von $\epsilon = 10^{-6}$. Für weitere Erklärungen siehe Abbildung 5.9.

(Abbildung 5.15)



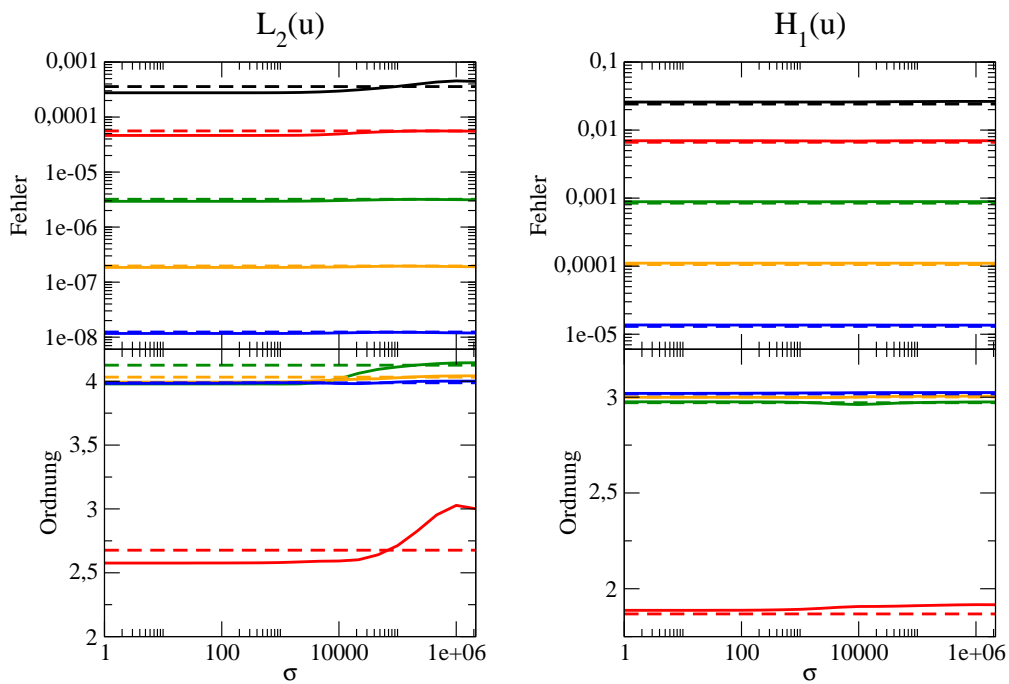
Fehler und Fehlerordnungen Finiter-Elemente dritter Ordnung auf dem Dreiecksgitter (P_3) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.3) bei einer Diffusionskonstanten von $\epsilon = 10^{-6}$. Für weitere Erklärungen siehe Abbildung 5.9.

(Abbildung 5.16)



Fehler und Fehlerordnungen Finiter-Elemente dritter Ordnung auf dem Quadratgitter (P_3) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.3) bei einer Diffusionskonstanten von $\epsilon = 10^{-6}$. Für weitere Erklärungen siehe Abbildung 5.9.

(Abbildung 5.17)



Fehler und Fehlerordnungen Finiter-Elemente dritter Ordnung auf dem Quadratgitter (Q_3) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.3) bei einer Diffusionskonstanten von $\epsilon = 10^{-6}$. Für weitere Erklärungen siehe Abbildung 5.9.

(Abbildung 5.18)

Element	Gitter	σ	L^2 -Fehler	Ordnung	H^1 -Fehler	Ordnung
P_1	Dreiecke	1000	$1.69 \cdot 10^{-4}$	2.00	$5.54 \cdot 10^{-2}$	1.00
P_2	Dreiecke	464	$9.69 \cdot 10^{-7}$	3.01	$1.11 \cdot 10^{-3}$	2.00
P_3	Dreiecke	1000	$1.49 \cdot 10^{-8}$	3.98	$1.72 \cdot 10^{-5}$	3.01
P_1	Quadrate	2154	$1.39 \cdot 10^{-4}$	2.00	$5.63 \cdot 10^{-2}$	1.00
P_2	Quadrate	464	$2.26 \cdot 10^{-6}$	3.00	$1.35 \cdot 10^{-3}$	2.00
P_3	Quadrate	2154	$2.38 \cdot 10^{-8}$	4.03	$2.08 \cdot 10^{-5}$	3.05
Q_1	Quadrate	100	$1.10 \cdot 10^{-4}$	2.00	$5.19 \cdot 10^{-2}$	1.00
Q_2	Quadrate	464	$1.18 \cdot 10^{-6}$	3.00	$9.57 \cdot 10^{-4}$	2.00
Q_3	Quadrate	1000	$1.16 \cdot 10^{-8}$	3.99	$1.36 \cdot 10^{-5}$	3.02

Optimaler Stabilisierungsparameter σ sowie zugehörige Fehler und Konvergenzordnungen der verschiedenen Finiten-Elemente für das Beispiel (5.2.3).

(Tabelle 5.5)

Element	SIPG	SDFEM	Verhältnis
P_1	1.07	0.30	3.51
P_2	5.83	0.96	6.10
P_3	21.31	4.01	5.31
Q_1	0.89	0.14	6.34
Q_2	7.31	1.90	3.85
Q_3	36.07	9.78	3.69

Rechenzeiten in Sekunden für die verschiedenen Finiten-Elemente des SIPG- und des Streamline-Diffusion-Verfahren für das Beispiel (5.2.3).

(Tabelle 5.6)

Tabelle 5.2.2 zeigt die L^2 -Norm und die H^1 -Seminorm des Fehlers sowie die zugehörigen Konvergenzordnungen des SIPG-Verfahrens für alle hier behandelten Finiten-Elemente bei möglichst optimaler Wahl des Stabilisierungsparameters σ . Möglichst optimal bedeutet dabei, dass sowohl auf möglichst kleine Fehlerwerte als auch auf möglichst große Konvergenzordnungen bezogen auf Gitterlevel 6 geachtet wurde. Dabei sei erwähnt, dass sich in allen Fällen auch mit den meisten anderen getesteten Werten von σ ähnlich gute Ergebnisse erzielen lassen, wie aus den Abbildungen 5.9 bis 5.18 hervorgeht.

Wie schon mehrfach erwähnt, führt die Verwendung von Räumen unstetiger Polynome dazu, dass sich die Gesamtzahl der Basisfunktionen erhöht. Dadurch werden die Matrizen größer. Zudem enthalten sie durch die Verkopplung von Basisfunktionen, welche zu verschiedenen Zellen gehören mehr von Null verschiedenen Einträge. Dies alles schlägt sich in der Rechenzeit nieder. Tabelle 5.6 gibt für das Beispiel (5.2.3) die Rechenzeiten zur Lösung des Gleichungssystems auf Gitterlevel 6 für die verschiedenen Finiten-Elemente des SIPG- und des Streamline-Diffusion-Verfahren sowie das Verhältnis der Laufzeiten zueinander an. Hierbei wurde für σ jeweils der Wert aus Tabelle 5.2.2 gewählt. Wie erwartet ist die Rechenzeit beim SIPG-Verfahren länger. Das Verhältnis der Laufzeiten liegt zwischen 3.5 und 6.34.

Als Fazit dieses Abschnittes lässt sich sagen, dass die Ergebnisse des SIPG-Verfahren bei Abwesenheit von Grenzschichten die gleiche Qualität erreichen wie die Ergebnisse des Streamline-Diffusion-Verfahrens. Jedoch sind die Rechenzeiten des SIPG-Verfahren deutlich länger.

5.2.3 Beispiel mit inneren Grenzschichten

Im letzten Abschnitt haben wir gesehen, dass das SIPG-Verfahren gute Ergebnisse liefert, falls die Lösung keine Grenzschichten aufweist. Wir wollen nun sehen, wie gut dieses Verfahren bei Vorhandensein innerer Grenzschichten ist. Dazu betrachten wir ein Beispiel aus (Joh97):

$$u = \frac{16}{\pi} \left(\arctan \left(\frac{2}{\sqrt{\epsilon}} (0.0625 - (x - 0.5)^2 - (y - 0.5)^2) \right) + \frac{\pi}{2} \right) x(x-1)y(y-1) \quad (5.2.4)$$

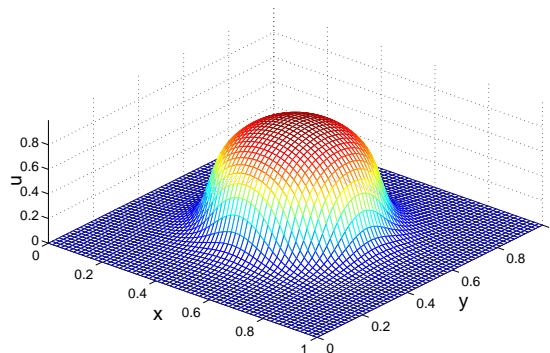
auf $\Omega = [0, 1]^2$. Hierbei sind $b = (2, 3)^T$, $c = 2$ und die rechte Seite f passend gewählt, so dass (5.2.4) die Konvektions-Diffusions-Gleichung löst.

Die Lösung (5.2.4) besitzt eine kreisförmige innere Grenzschicht. Wir werden dieses Beispiel für zwei Werte von ϵ betrachten. Bei $\epsilon = 10^{-3}$ ist die Grenzschicht noch moderat, während wir es bei $\epsilon = 10^{-6}$ mit einer steilen Grenzschicht zu tun haben.

Das vorherige Beispiel hat uns gezeigt, dass Q_p -Elemente auf dem Quadratgitter bessere Ergebnisse zeigen als P_p -Elemente. Mit dem Hinweis, dass dies auch für Beispiele mit Grenzschichten so ist, verzichten wir daher im Folgenden auf die Darstellung der Ergebnisse zu P_p -Elementen auf Quadraten.

5.2.3.1 Moderate innere Grenzschichten

Das Aussehen von (5.2.4) für $\epsilon = 10^{-3}$ auf dem Einheitsquadrat ist in Abbildung 5.19 dargestellt.

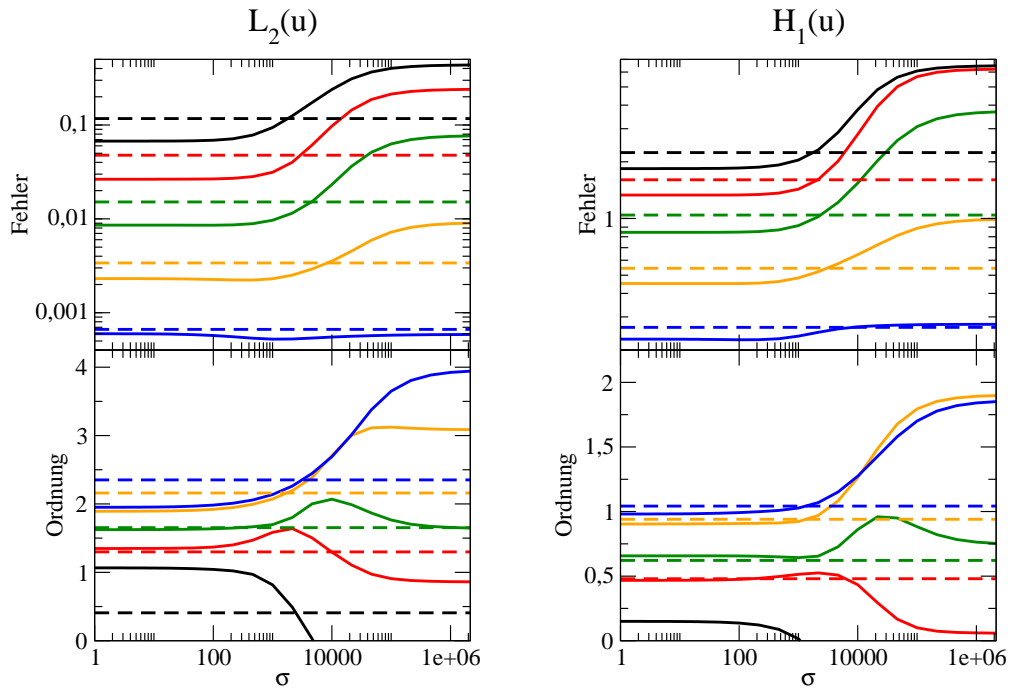


Darstellung der exakten Lösung für das Beispiel (5.2.4) bei $\epsilon = 10^{-3}$.

(Abbildung 5.19)

Abbildung 5.20 zeigt die L^2 -Norm und die H^1 -Seminorm des Fehlers zum Beispiel (5.2.4) mit $\epsilon = 10^{-3}$ bei Verwendung von P_1 -Elementen auf dem Dreiecksgitter. Es zeigt sich, dass die Fehler des SIPG-Verfahrens bis Gitterlevel 5 bei kleinen Werten von σ kleiner sind, als die Fehler des Streamline-Diffusion-Verfahrens, während sie bei großen σ größer sind. Der Schwellenwert, bei dem der Übergang von kleineren zu größeren Fehlern liegt, wandert dabei mit zunehmendem Level zu größeren σ . Auf Level 6 sind die Fehler beider Verfahren schließlich nahezu gleich groß. Die Konvergenzordnungen entsprechen den Erwartungen. Die sehr großen Konvergenzraten bei großen σ lassen sich durch die dort größeren Fehler auf den größeren Gittern erklären. Daher ist ihnen keine größere Bedeutung zuzumessen.

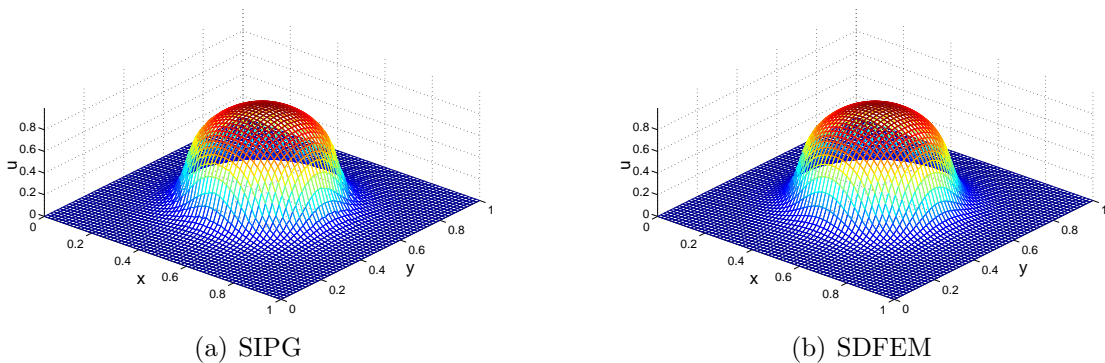
Die numerisch bestimmten Lösungen beider Verfahren für P_1 -Elemente auf dem Dreiecksgitter sind in Abbildung 5.21 zu sehen. Beide sind einander sehr ähnlich und geben die exakte



Fehler und Fehlerordnungen Finiterelemente erster Ordnung auf dem Dreiecksgitter (P_1) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.4) bei einer Diffusionskonstanten von $\epsilon = 10^{-3}$. Für weitere Erklärungen siehe Abbildung 5.9.

(Abbildung 5.20)

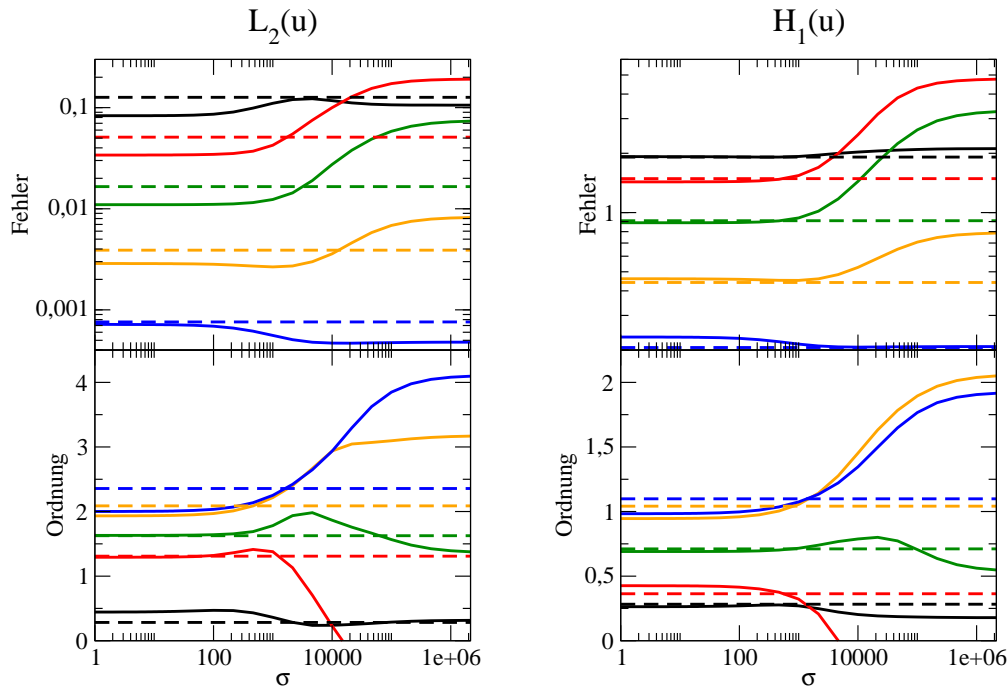
Lösung gut wieder. Die Lösung des SIPG-Verfahrens zeigt im unteren Teil der Grenzschicht einen leicht steileren Gradienten.



Numerische Lösung für das Beispiel (5.2.4) bei einer Diffusionskonstanten von $\epsilon = 10^{-3}$ bei $\sigma = 464$ für das DG-Verfahren (links) und für das Streamline-Diffusion-Verfahren (rechts) bei Verwendung von P_1 -Elementen auf dem Dreiecksgitter. Für weitere Erklärungen siehe Abbildung 5.10.

(Abbildung 5.21)

Die Ergebnisse zu Q_1 -Elementen auf dem Quadratgitter sind in Abbildung 5.22 gezeigt. Auch hier schneidet das SIPG-Verfahren bis Level 5 bei kleineren σ besser ab, auf Level 6 jedoch ergeben sich bei größeren σ kleinere Fehler. Daran ist zu erkennen, dass die Wahl eines optimalen Parameters auch von der Feinheit des Gitters abhängen kann. Insgesamt nehmen auch hier die Fehler des SIPG-Verfahrens und des Streamline-Diffusion-Verfahrens ähnliche



Fehler und Fehlerordnungen Finite-Elemente erster Ordnung auf dem Quadratgitter (Q_1) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.4) bei einer Diffusionskonstanten von $\epsilon = 10^{-3}$. Für weitere Erklärungen siehe Abbildung 5.9.

(Abbildung 5.22)

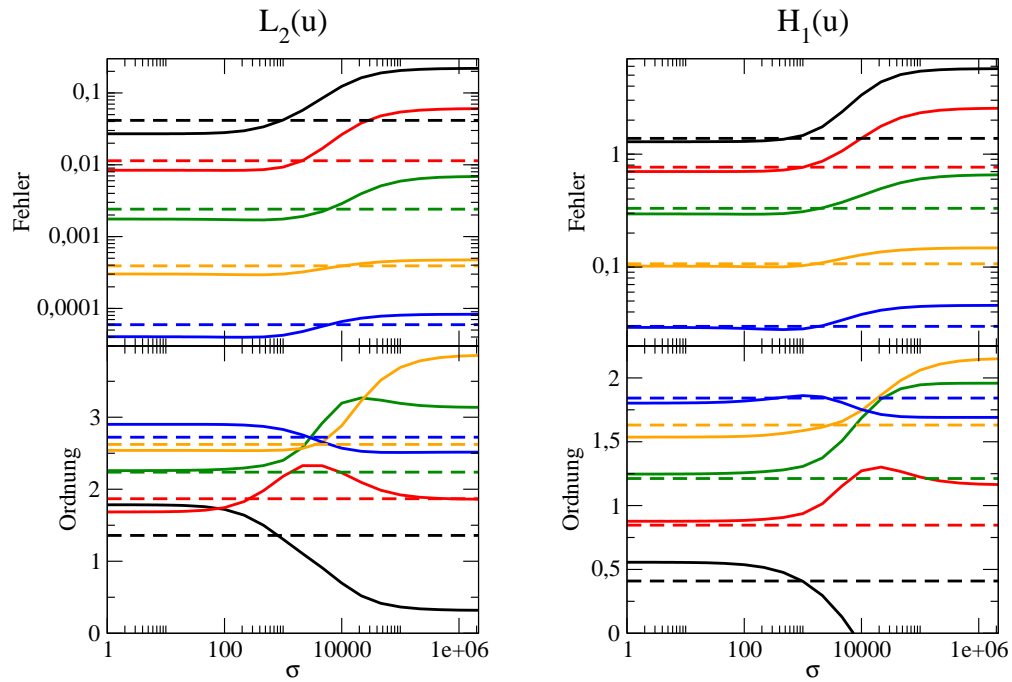
Werte an. Für die Konvergenzordnungen gilt hier dasselbe wie schon für die P_1 -Elemente.

In Abbildung 5.23 sind die Ergebnisse zu P_2 -Elementen auf dem Dreiecksgitter dargestellt. Die Fehler verhalten sich hier ähnlich wie bei den P_1 -Elementen, jedoch sind auch bei Gitterlevel 6 die Fehler für kleinere Werte von σ kleiner als für größere Werte. Die Konvergenzordnungen liegen leicht unter den Erwartungen. Allerdings sind sowohl die Fehler als auch die Konvergenzordnungen von SIPG- und Streamline-Diffusion-Verfahren wieder fast identisch.

Auf dem Quadratgitter ergibt sich für Q_2 -Elemente ein ähnliches Bild wie für die P_2 -Elemente auf dem Dreiecksgitter, was sich aus Abbildung 5.24 erkennen lässt. Auch hier führen kleinere Werte von σ zu kleineren Fehlern. Die Konvergenzordnungen des L^2 -Fehlers kommen hier den Erwartungen näher. Die H^1 -Fehler des SIPG-Verfahrens sind für Q_2 -Elemente etwas größer als die H^1 -Fehler des Streamline-Diffusion-Verfahrens.

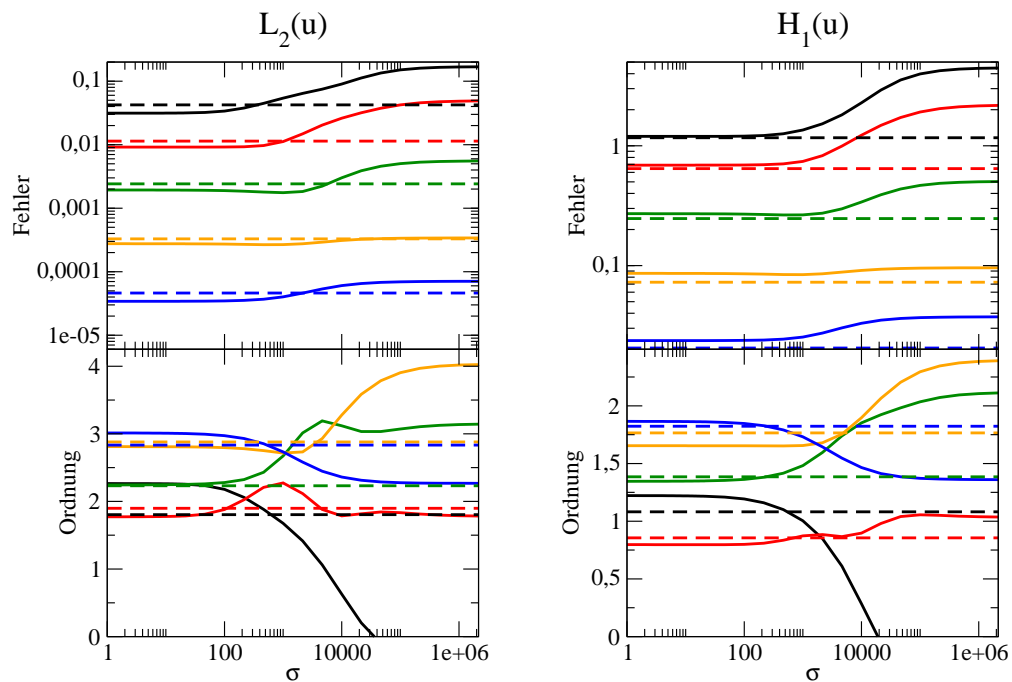
Die Ergebnisse zu P_3 -Elementen auf dem Dreiecksgitter sind in Abbildung 5.25 dargestellt. Hier fällt sofort der sprunghafte Anstieg der Fehler auf Gitterlevel 6 für kleine Werte von σ auf. Dieser ist eine Folge der schlechten Konditionierung des zugehörigen linearen Gleichungssystems, für welches der verwendete Löser keine Lösung mehr findet. Es ist zu vermuten, dass sich, wenn das Gleichungssystem gelöst würde, auch hier für kleinere Werte von σ leicht bessere Fehler ergeben würden und die Konvergenzordnungen den Erwartungen entsprächen.

Abbildung 5.26 zeigt die Fehler und Konvergenzordnungen, welche sich bei Verwendung von Q_3 -Elementen auf dem Quadratgitter ergeben. In diesem Fall ergeben sich keine Probleme durch die Kondition des Gleichungssystems. Die L^2 -Fehler zeigen hier auf Gitterlevel 6 praktisch keine Abhängigkeit mehr vom Stabilisierungsparameter und sind nahezu identisch



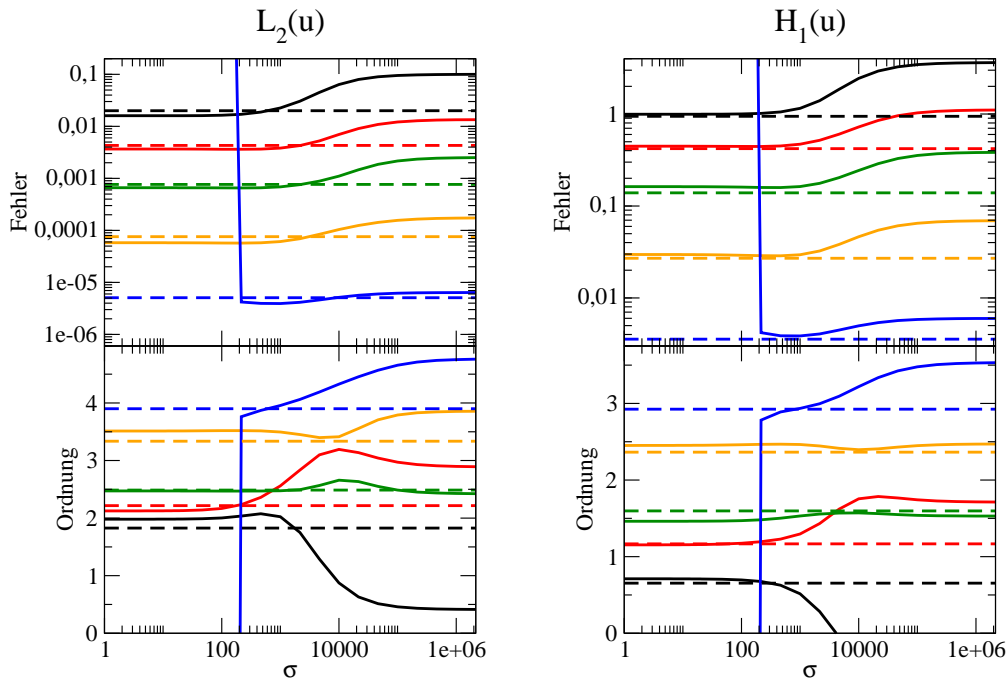
Fehler und Fehlerordnungen Finiter-Elemente zweiter Ordnung auf dem Dreiecksgitter (P_2) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.4) bei einer Diffusionskonstanten von $\epsilon = 10^{-3}$. Für weitere Erklärungen siehe Abbildung 5.9.

(Abbildung 5.23)



Fehler und Fehlerordnungen Finiter-Elemente zweiter Ordnung auf dem Quadratgitter (Q_2) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.4) bei einer Diffusionskonstanten von $\epsilon = 10^{-3}$. Für weitere Erklärungen siehe Abbildung 5.9.

(Abbildung 5.24)



Fehler und Fehlerordnungen Finiten-Elemente dritter Ordnung auf dem Dreiecksgitter (P_3) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.4) bei einer Diffusionskonstanten von $\epsilon = 10^{-3}$. Für weitere Erklärungen siehe Abbildung 5.9.

(Abbildung 5.25)

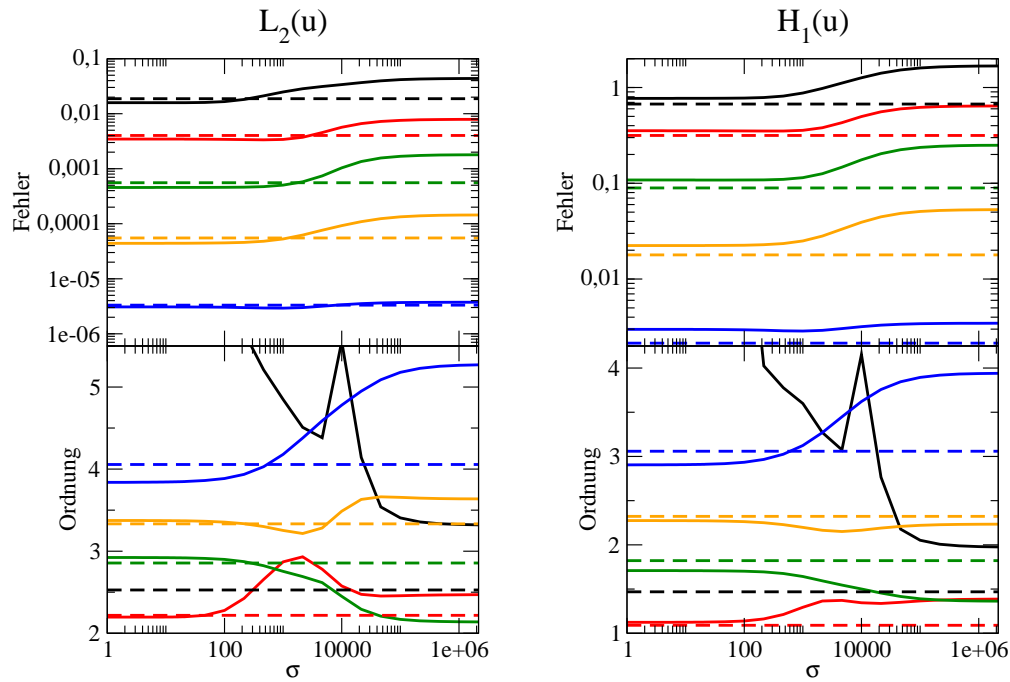
für SIPG- und Streamline-Diffusion-Verfahren. Bei den H^1 -Fehlern schneidet das SIPG-Verfahren etwas schlechter ab. Die Konvergenzordnungen liegen auch hier in etwa bei den erwarteten Werten.

Element	Gitter	σ	L^2 -Fehler	Ordnung	H^1 -Fehler	Ordnung
P_1	Dreiecke	464	$5.35 \cdot 10^{-4}$	2.06	$2.30 \cdot 10^{-1}$	1.01
P_2	Dreiecke	100	$3.98 \cdot 10^{-5}$	2.90	$2.86 \cdot 10^{-2}$	1.82
P_3	Dreiecke	1000	$3.93 \cdot 10^{-6}$	3.96	$3.85 \cdot 10^{-3}$	2.94
Q_1	Quadrate	10000	$4.68 \cdot 10^{-4}$	2.93	$2.10 \cdot 10^{-1}$	1.35
Q_2	Quadrate	20	$3.43 \cdot 10^{-5}$	3.00	$2.37 \cdot 10^{-2}$	1.86
Q_3	Quadrate	1000	$2.92 \cdot 10^{-6}$	4.18	$2.86 \cdot 10^{-3}$	3.12

Optimaler Stabilisierungsparameter σ sowie zugehörige Fehler und Konvergenzordnungen der verschiedenen Finiten-Elemente für das Beispiel (5.2.4) bei einer Diffusionskonstanten $\epsilon = 10^{-3}$.

(Tabelle 5.7)

Tabelle 5.7 listet die möglichst optimalen Werte für σ zur Behandlung des Beispiels (5.2.4) für den Fall moderater Grenzschichten und die zugehörigen Fehler sowie Konvergenzordnungen auf Gitterlevel 6 für alle hier behandelten Finiten-Elemente auf. Das Verfahren kommt den erwarteten Konvergenzordnungen dabei sehr nahe beziehungsweise übertrifft diese sogar im Fall der Q_3 -Elemente. Hier wäre das Verhalten bei weiterer Verfeinerung des Gitters zu betrachten, was wir auf Grund der zur Verfügung stehenden Rechenkapazitäten nicht untersuchen konnten.



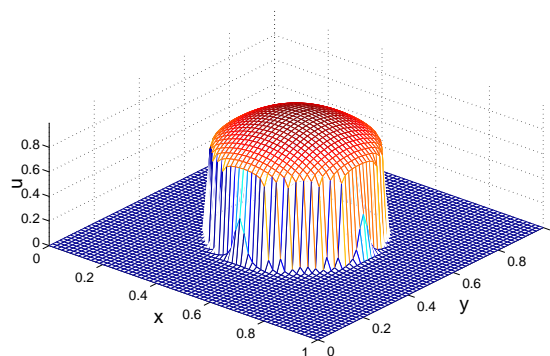
Fehler und Fehlerordnungen Finiter-Elemente dritter Ordnung auf dem Quadratgitter (Q_3) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.4) bei einer Diffusionskonstanten von $\epsilon = 10^{-3}$. Für weitere Erklärungen siehe Abbildung 5.9.

(Abbildung 5.26)

Wir haben in diesem Abschnitt gesehen, dass das SIPG-Verfahren im Fall moderater innerer Grenzschichten gute Ergebnisse liefert und dass diese mit den Ergebnissen des Streamline-Diffusion-Verfahren konkurrieren können.

5.2.3.2 Steile innere Grenzschichten

Die Funktion aus Beispiel (5.2.4) für $\epsilon = 10^{-6}$ ist in Abbildung 5.27 gezeigt. Wie man sieht, sind die Grenzschichten hier steiler als in Abbildung 5.19. Daher ist es prinzipiell schwieriger, diese Lösung mit Hilfe numerischer Verfahren zu berechnen.

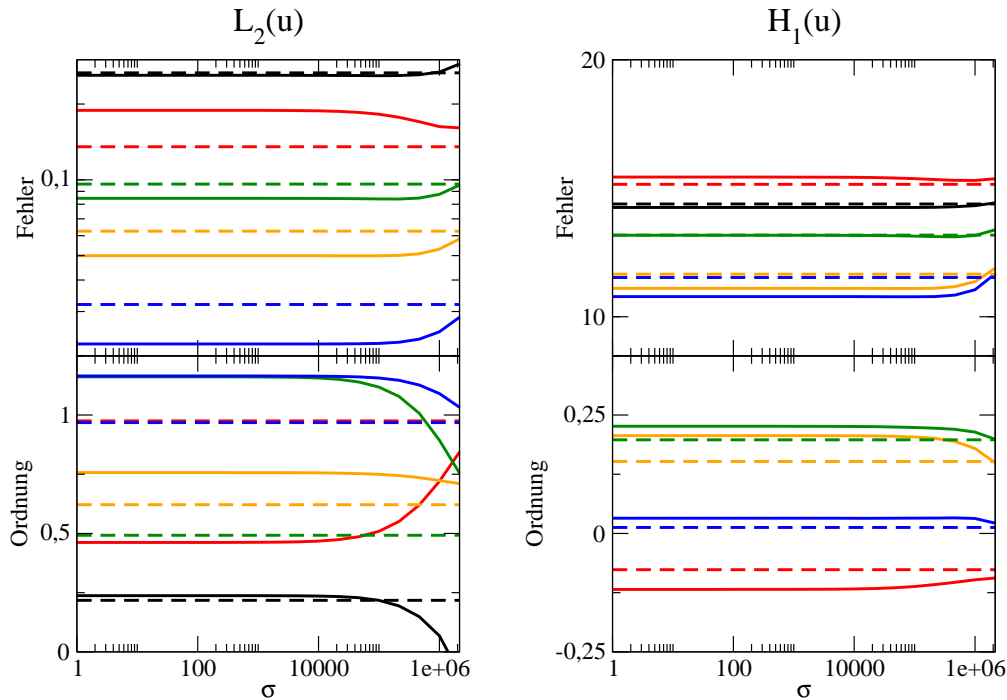


Darstellung der exakten Lösung für das Beispiel (5.2.4) bei $\epsilon = 10^{-6}$.

(Abbildung 5.27)

Abbildung 5.28 zeigt die bekannten Fehler und zugehörigen Konvergenzordnungen für den Fall steiler innerer Grenzschichten bei Verwendung von P_1 -Elementen auf dem Dreiecksgit-

ter. Die Fehler sind hier deutlich größer als in Abbildung 5.20, insbesondere die H^1 -Fehler. Die Konvergenz in der L^2 -Norm ist um 1 kleiner als bei den bisherigen Beispielen. In der H^1 -Seminorm stagniert der Fehler sogar fast. Fehler und Konvergenzordnungen sind in einem weiten Bereich unabhängig von σ und werden nur für sehr große Werte etwas schlechter. Jedoch sind die Ergebnisse des SIPG-Verfahrens bei allen untersuchten Werten des Stabilisierungsparameters besser als die des Streamline-Diffusion-Verfahrens.



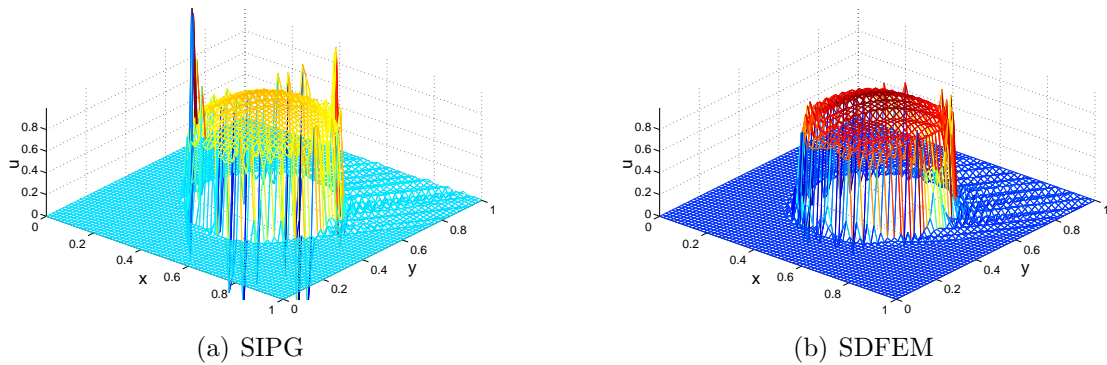
Fehler und Fehlerordnungen Finiter-Elemente erster Ordnung auf dem Dreiecksgitter (P_1) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.4) bei einer Diffusionskonstanten von $\epsilon = 10^{-6}$. Für weitere Erklärungen siehe Abbildung 5.9.

(Abbildung 5.28)

Die numerisch bestimmten Lösungen sind in Abbildung 5.29 zu sehen. Während das SIPG-Verfahren bei den Fehlern leicht besser abschneidet als das Streamline-Diffusion-Verfahren, zeigt sich hier, dass es größere Oszillationen innerhalb der Grenzschicht zeigt. Vor allem zeigen sich auch Spitzen in den negativen Bereich hinein. Das Streamline-Diffusion-Verfahren zeigt auch Oszillationen, jedoch sind diese innerhalb der Grenzschicht kleiner. Bei beiden Verfahren werden die Oszillationen in Richtung des Advektionsvektors b weiter in das Gebiet hineingetragen.

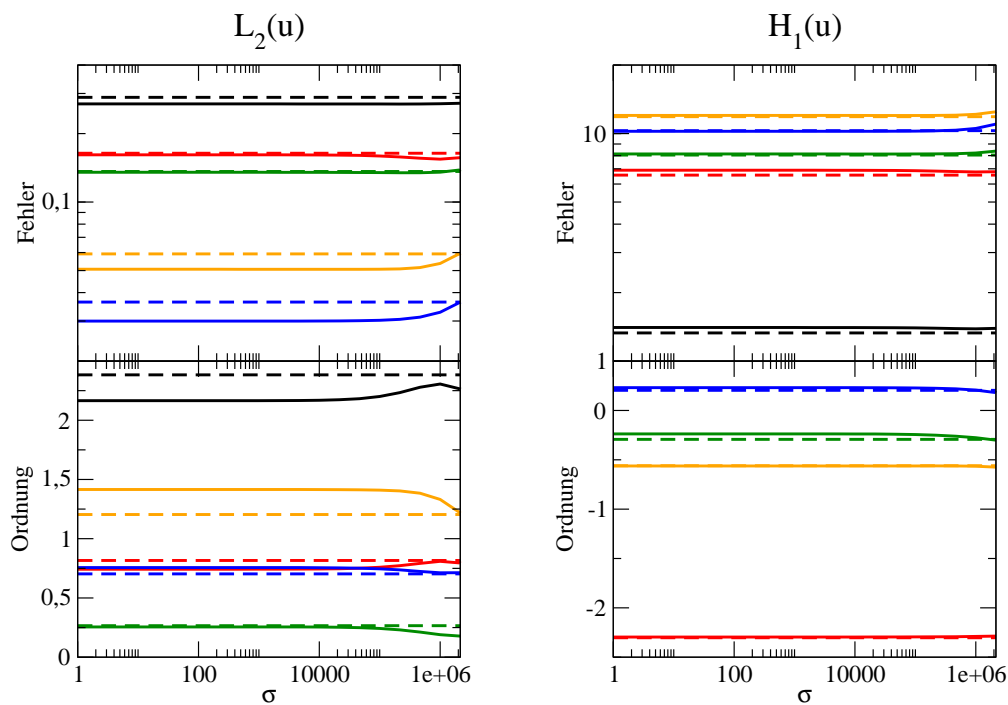
Die Ergebnisse bei Verwendung von Q_1 -Elementen auf dem Quadratgitter sind in Abbildung 5.30 zu sehen. Hier zeigt sich ein ähnliches Bild wie zuvor bei den P_1 -Elementen. Die L^2 -Norm des Fehlers ist hier etwas größer und die zugehörige Konvergenzordnung etwas kleiner. Die H^1 -Fehler stagnieren wiederum nahezu. Die Ergebnisse sind hier für das SIPG-Verfahren kaum besser als für das Streamline-Diffusion-Verfahren. Die numerisch bestimmten Lösungen sind ähnlich zu denen der P_1 -Elemente, einschließlich der Spitzen im positiven und negativen Bereich, welche das SIPG-Verfahren zeigt, und daher hier nicht abgebildet.

In Abbildung 5.31 sind die Fehler und Konvergenzordnungen zu sehen, welche sich bei Verwendung von P_2 -Elementen auf dem Dreiecksgitter ergeben. Die L^2 -Fehler zeigen im



Numerische Lösung für das Beispiel (5.2.4) bei einer Diffusionskonstanten von $\epsilon = 10^{-6}$ bei $\sigma = 1000$ für das DG-Verfahren (links) und für das Streamline-Diffusion-Verfahren (rechts) bei Verwendung von P_1 -Elementen auf dem Dreiecksgitter. Für weitere Erklärungen siehe Abbildung 5.10.

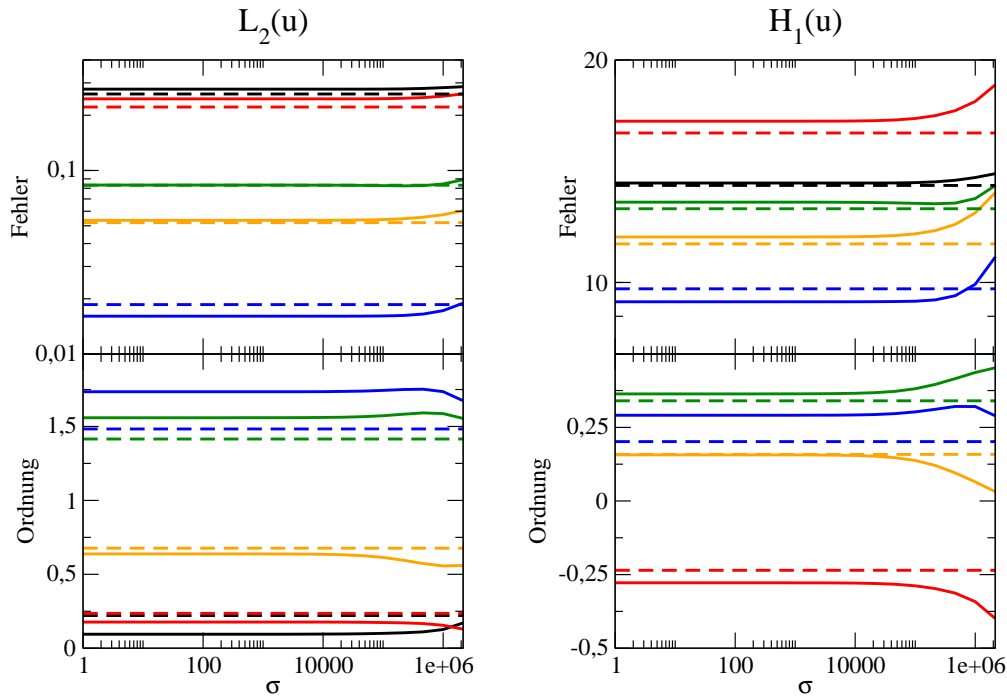
(Abbildung 5.29)



Fehler und Fehlerordnungen Finite-Elemente erster Ordnung auf dem Quadratgitter (Q_1) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.4) bei einer Diffusionskonstanten von $\epsilon = 10^{-6}$. Für weitere Erklärungen siehe Abbildung 5.9.

(Abbildung 5.30)

Vergleich zu den P_1 -Elementen eine leichte Verbesserung, während die H^1 -Fehler nur unwesentlich kleiner sind. Die Konvergenzordnungen sind hier etwas besser, bleiben aber auch hinter den Erwartungen zurück. So ist die Konvergenzordnung in der L^2 -Norm nur etwa halb so groß wie bei den vorherigen Beispielen. Die Fehler für das SIPG-Verfahren sind leicht kleiner als diejenigen des Streamline-Diffusion-Verfahrens. Für die numerisch bestimmten Lösungen ergibt sich aber wieder ein ähnliches Bild wie schon für die P_1 -Elemente. Insgesamt sind die Ergebnisse wieder nahezu unabhängig von σ .



Fehler und Fehlerordnungen Finiten-Elemente zweiter Ordnung auf dem Dreiecksgitter (P_2) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.4) bei einer Diffusionskonstanten von $\epsilon = 10^{-6}$. Für weitere Erklärungen siehe Abbildung 5.9.

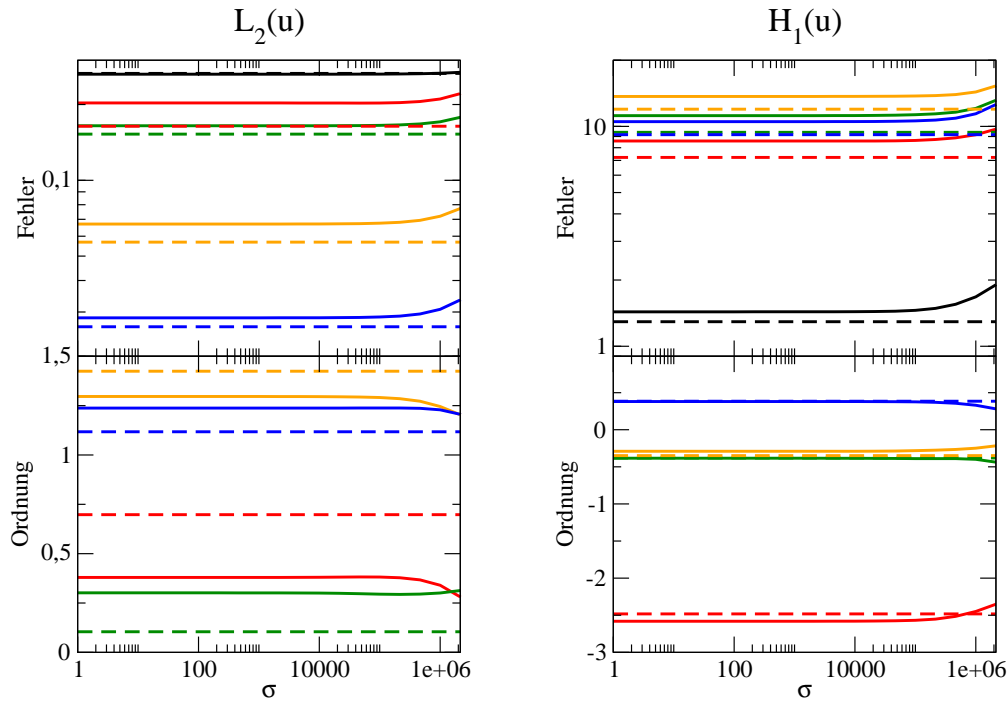
(Abbildung 5.31)

Auch bei Verwendung von Q_2 -Elementen auf dem Quadratgitter ergibt sich keine Verbesserung, wie Abbildung 5.32 zeigt. Die Fehler sind hier etwas größer als für die P_2 -Elemente und fallen für das SIPG-Verfahren größer aus als für das Streamline-Diffusion-Verfahren. Auch hier zeigt sich nur für große Werte von σ eine Abhängigkeit der Fehler vom Stabilisierungsparameter.

Abbildung 5.33 zeigt die Fehler und Konvergenzordnungen bei Verwendung von P_3 -Elementen auf dem Dreiecksgitter. Gegenüber den P_2 -Elementen zeigt sich hier keine Verbesserung. Die Konvergenzordnung in der L^2 -Norm fällt wieder ungefähr halb so groß aus, wie in den vorherigen Beispielen; die in der H^1 -Seminorm ungefähr ein Viertel so groß. Die Fehler beider Verfahren sind praktisch identisch. Auch bei Verwendung von P_3 -Elementen zeigt sich keine Abhängigkeit der Ergebnisse vom Stabilisierungsparameter außer im Falle sehr großer σ .

In Abbildung 5.34 sind die Ergebnisse bei Verwendung von Q_3 -Elementen auf dem Quadratgitter dargestellt. Es zeigt sich ein ähnliches Bild wie bei den schon besprochenen Finiten-Elementen. Die Ergebnisse sind weitestgehend unabhängig von σ und im Vergleich zu den Q_2 -Elementen etwas schlechter. Auch bei den Q_3 -Elementen zeigt das Streamline-Diffusion-Verfahren leicht kleinere Fehler als das SIPG-Verfahren.

Tabelle 5.8 gibt die optimalen Werte der Fehler und Konvergenzordnungen des SIPG-Verfahrens auf Gitterlevel 6 bei der Lösung des Beispiels (5.2.4) mit einer Diffusionskonstanten $\epsilon = 10^{-6}$ an. Da die Fehler weitestgehend unabhängig von σ sind, haben wir hier für alle Finiten-Elemente den gleichen Stabilisierungsparameter gewählt. Eine andere Wahl von σ verbessert die Ergebnisse nicht signifikant. Es sei an dieser Stelle darauf hingewiesen, dass die relativ großen Werte der H^1 -Fehler eine Folge großer Quadraturfehler auf Grund



Fehler und Fehlerordnungen Finiten-Elemente zweiter Ordnung auf dem Quadratgitter (Q_2) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.4) bei einer Diffusionskonstanten von $\epsilon = 10^{-6}$. Für weitere Erklärungen siehe Abbildung 5.9.

(Abbildung 5.32)

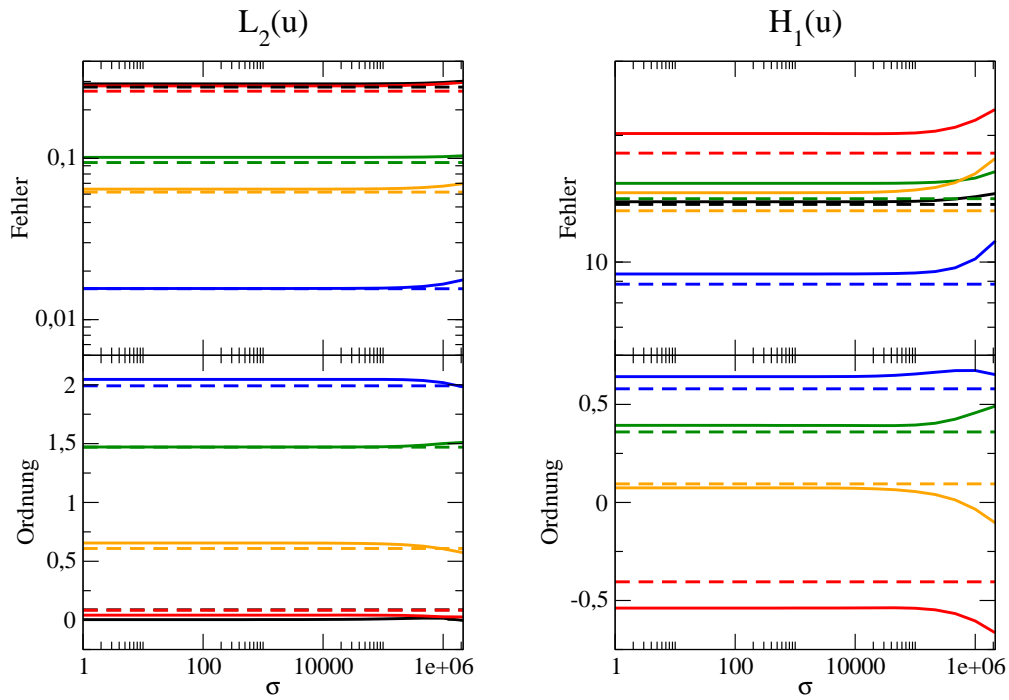
Element	Gitter	σ	L^2 -Fehler	Ordnung	H^1 -Fehler	Ordnung
P_1	Dreiecke	1000	$2.23 \cdot 10^{-2}$	1.16	10.56	0.03
P_2	Dreiecke	1000	$1.61 \cdot 10^{-2}$	1.74	9.42	0.29
P_3	Dreiecke	1000	$1.56 \cdot 10^{-2}$	2.05	9.37	0.64
Q_1	Quadrate	1000	$3.00 \cdot 10^{-2}$	0.76	10.23	0.23
Q_2	Quadrate	1000	$2.84 \cdot 10^{-2}$	1.24	10.51	0.38
Q_3	Quadrate	1000	$4.99 \cdot 10^{-2}$	1.25	18.56	0.42

Optimaler Stabilisierungsparameter σ sowie zugehörige Fehler und Konvergenzordnungen der verschiedenen Finiten-Elemente für das Beispiel (5.2.4) bei einer Diffusionskonstanten $\epsilon = 10^{-6}$.

(Tabelle 5.8)

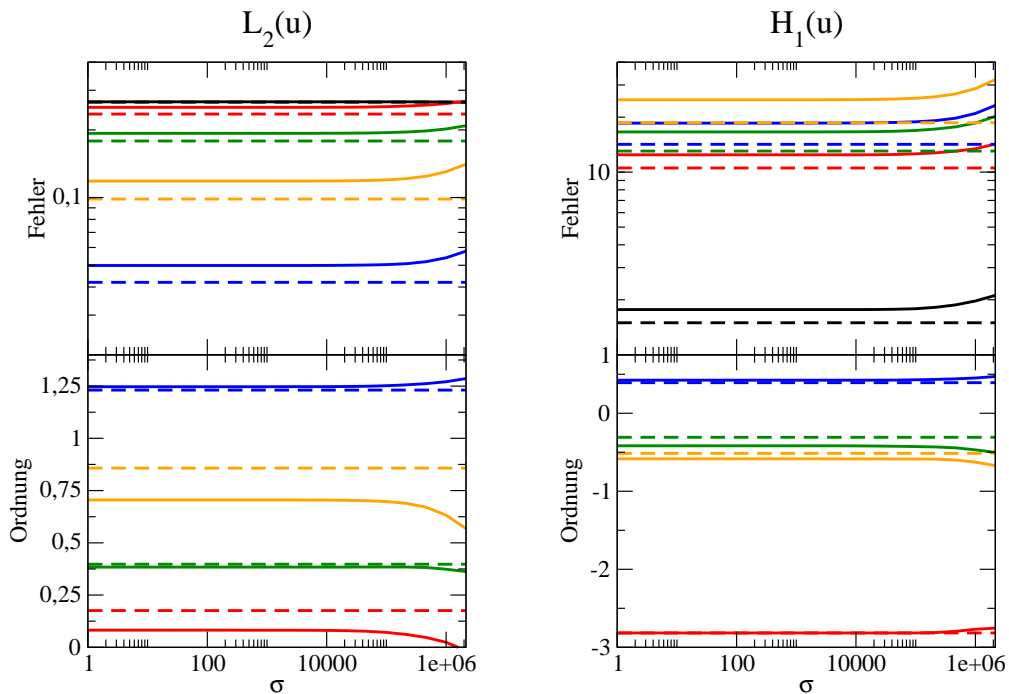
der Oszillationen sein können.

Insgesamt lässt sich sagen, dass das SIPG-Verfahren in der hier behandelten Form nicht in der Lage ist, Probleme mit steilen inneren Grenzschichten mit befriedigender Genauigkeit zu behandeln. Insbesondere transportiert es wie das Streamline-Diffusion-Verfahren Oszillationen aus den Grenzschichten weiter in das Gebiet hinein.



Fehler und Fehlerordnungen Finiter-Elemente dritter Ordnung auf dem Dreiecksgitter (P_3) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.4) bei einer Diffusionskonstanten von $\epsilon = 10^{-6}$. Für weitere Erklärungen siehe Abbildung 5.9.

(Abbildung 5.33)



Fehler und Fehlerordnungen Finiter-Elemente dritter Ordnung auf dem Quadratgitter (Q_3) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.4) bei einer Diffusionskonstanten von $\epsilon = 10^{-6}$. Für weitere Erklärungen siehe Abbildung 5.9.

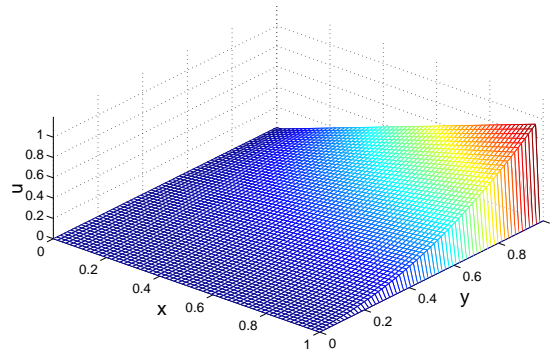
(Abbildung 5.34)

5.2.4 Beispiel mit Randgrenzschichten

Auch für Randgrenzschichten ist in (Joh97) ein Beispiel angegeben. Dieses lautet

$$u = xy^2 - y^2 e^{-\frac{2}{\epsilon}(1-x)} - x e^{-\frac{3}{\epsilon}(1-y)} + e^{-\frac{1}{\epsilon}(5-2x-3y)} \quad (5.2.5)$$

auf $\Omega = [0, 1]^2$ mit $b = (2, 3)^T$, $c = 1$ und passend gewählter rechter Seite f . Das Beispiel besitzt exponentielle Grenzschichten bei $x = 1$ und $y = 1$. Wir betrachten es für $\epsilon = 10^{-3}$. Bei diesem Wert zeigen sich bereits deutliche Randgrenzschichten, wie in Abbildung 5.35 zu sehen ist. Ein Vergleich beider Verfahren für ein Beispiel mit moderaten Randgrenzschichten fällt ähnlich aus wie die Vergleiche für moderate innere Grenzschichten oder das Beispiel ohne Grenzschichten, weshalb wir auf die Darstellung eines solchen Beispiels verzichten.



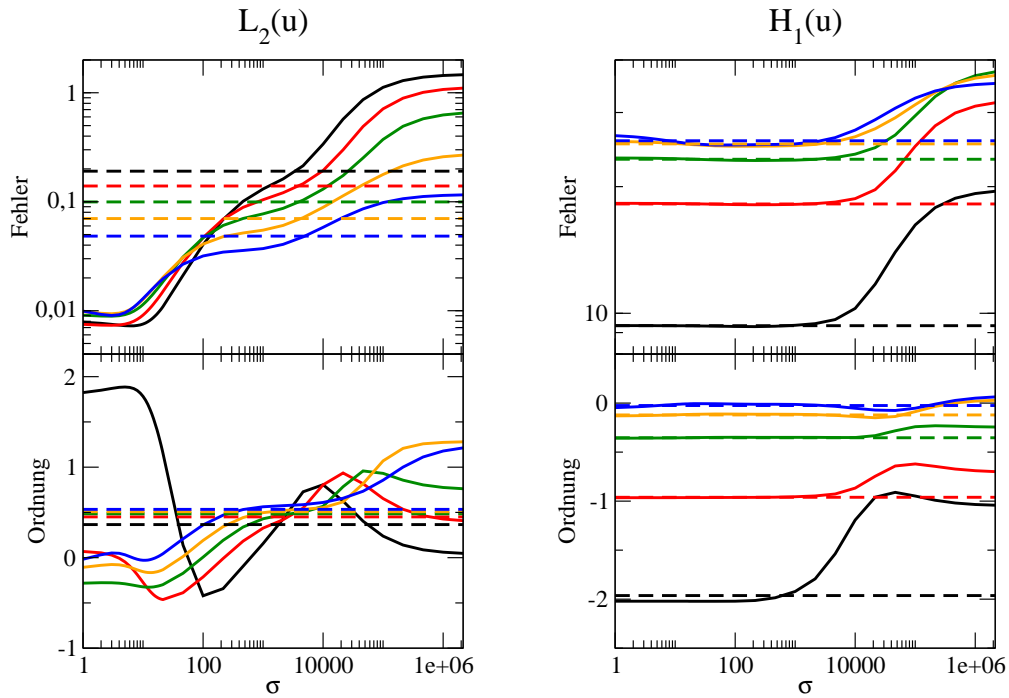
Darstellung der exakten Lösung für das Beispiel (5.2.5) bei $\epsilon = 10^{-3}$.
(Abbildung 5.35)

Abbildung 5.36 zeigt Fehler und Konvergenzordnungen bei Verwendung von P_1 -Elementen auf dem Dreiecksgitter. Bei diesem Beispiel liefern die kleineren σ bessere Ergebnisse. Bei sehr kleinen σ wachsen die L^2 -Fehler jedoch mit feiner werdendem Gitter. Die H^1 -Fehler sind in einem großen Parameterbereich konstant. Sie wachsen allerdings für die meisten getesteten σ mit zunehmendem Gitterlevel. Dies ist auch beim Streamline-Diffusion-Verfahren der Fall. Bei passend gewähltem Stabilisierungsparameter lassen sich kleinere Fehler als mit dem Streamline-Diffusion-Verfahren erreichen. Die Konvergenzordnungen der L^2 -Fehler schwanken im Bereich kleiner Fehler um 0 und wachsen mit größerem σ und damit mit wachsendem Fehler bis circa 1.25. Damit bleiben sie im gesamten Bereich hinter den Erwartungen zurück. Die H^1 -Fehler stagnieren bestenfalls.

Die numerischen Lösungen sind in Abbildung 5.37 zu sehen. Das SIPG-Verfahren nimmt nahe $x = y = 1$ leicht zu große Werte an und zeigt zudem einen Ausreißer in den negativen Bereich. Das Streamline-Diffusion-Verfahren zeigt zwar keine Ausreißer, kann aber die Lösung in der Grenzschicht entlang des Randes $x = 1$ auch nicht richtig wiedergeben.

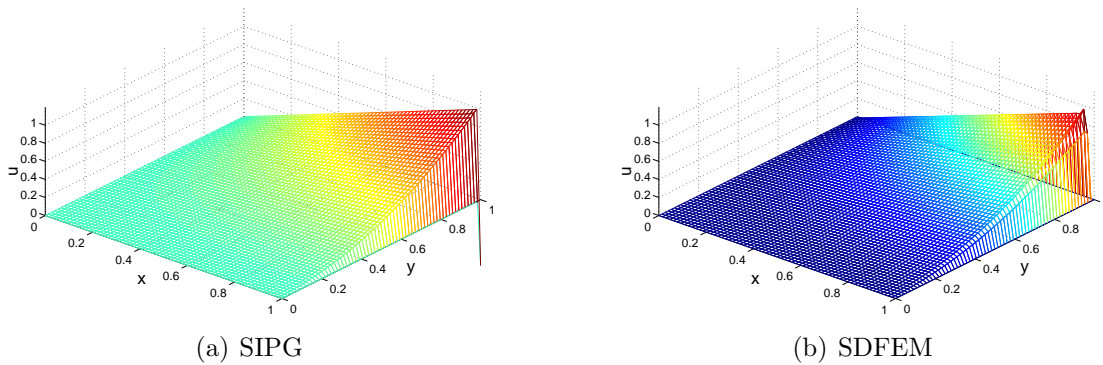
In Abbildung 5.38 sind die Ergebnisse bei Verwendung von Q_1 -Elementen auf dem Quadratgitter gezeigt. Die L^2 -Fehler zeigen ein ähnliches Verhalten wie bei Verwendung von P_1 -Elementen auf dem Dreiecksgitter, während die H^1 -Fehler eine stärkere Parameterabhängigkeit aufweisen. Jedoch wachsen letztere auch hier mit wachsendem Stabilisierungsparameter sowie feiner werdendem Gitter. Die Konvergenzordnungen der L^2 -Fehler liegen auf dem feinsten Gitter zwischen -1.1 für kleine σ und 1 für große σ . Die Konvergenzordnungen der H^1 -Fehler sind durchweg negativ mit Werten zwischen -2 und 0.

Die Ergebnisse bei Verwendung von P_2 -Elementen auf dem Dreiecksgitter sind in Abbildung 5.39 zu sehen. Die Verwendung Finiter-Elemente zweiter Ordnung führt hier zu



Fehler und Fehlerordnungen Finiter-Elemente erster Ordnung auf dem Dreiecksgitter (P_1) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.5) bei einer Diffusionskonstanten von $\epsilon = 10^{-3}$. Für weitere Erklärungen siehe Abbildung 5.9.

(Abbildung 5.36)

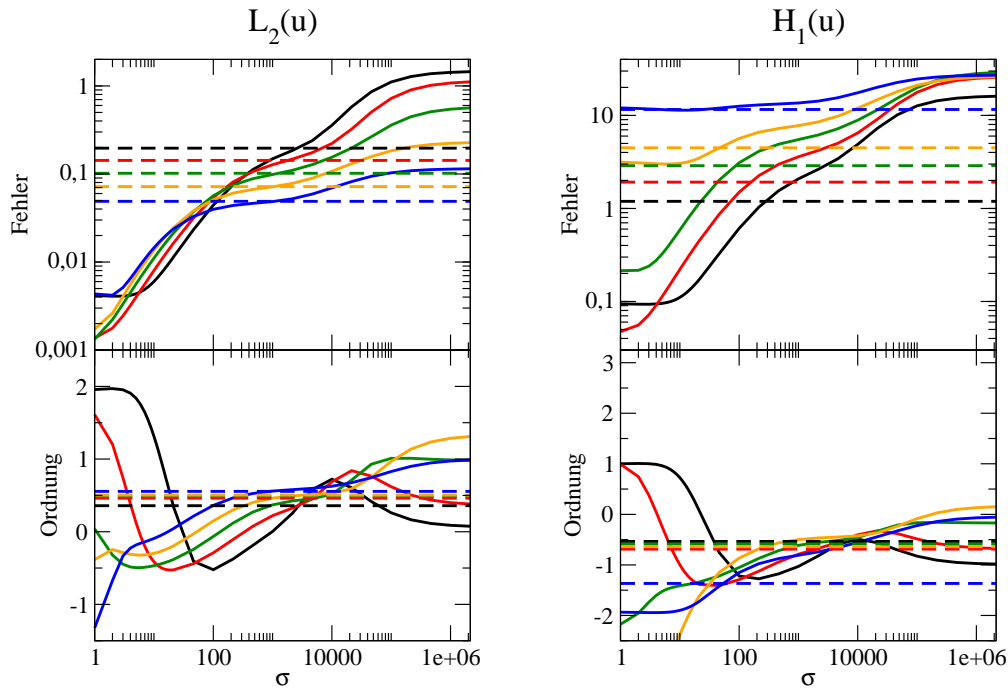


Numerische Lösung für das Beispiel (5.2.5) bei einer Diffusionskonstanten von $\epsilon = 10^{-3}$ bei $\sigma = 3$ für das DG-Verfahren (links) und für das Streamline-Diffusion-Verfahren (rechts) bei Verwendung von P_1 -Elementen auf dem Dreiecksgitter. Für weitere Erklärungen siehe Abbildung 5.10.

(Abbildung 5.37)

keinen Verbesserungen, wie ein Vergleich mit Abbildung 5.36 zeigt. Insgesamt sieht man ein ähnliches Verhalten der Fehler, welche wieder sehr groß ausfallen.

Auch auf dem Quadratgitter führt eine Erhöhung des Polynomgrades nicht zu Verbesserungen. In Abbildung 5.40 sind die Fehler und Konvergenzordnungen für Q_2 -Elemente auf dem Quadratgitter dargestellt. Im Bereich kleiner σ sind die Fehler sogar größer als bei Verwendung von Q_1 -Elementen. Die Konvergenzordnungen der L^2 -Fehler wachsen mit zunehmendem Stabilisierungsparameter von -1 bis 1, die der H^1 -Fehler von -2 auf circa 0.06.



Fehler und Fehlerordnungen Finite-Elemente erster Ordnung auf dem Quadratgitter (Q_1) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.5) bei einer Diffusionskonstanten von $\epsilon = 10^{-3}$. Für weitere Erklärungen siehe Abbildung 5.9.

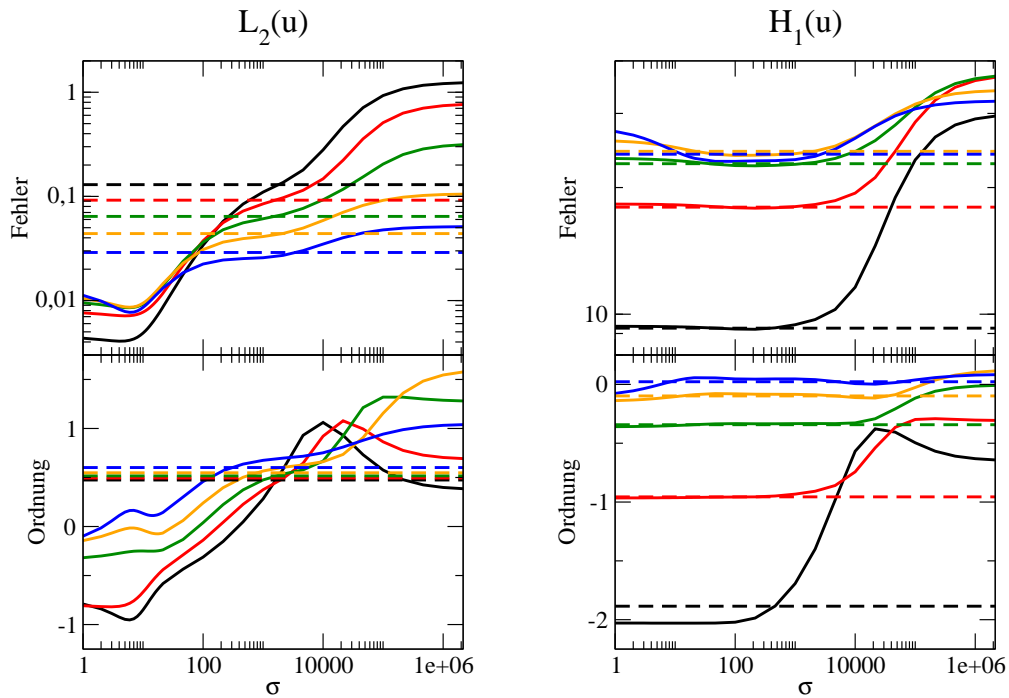
(Abbildung 5.38)

Die Ergebnisse, welche sich bei Verwendung von P_3 -Elementen auf dem Dreiecksgitter ergeben, sind in Abbildung 5.41 dargestellt. Hier zeigt sich wieder das Problem eines schlecht konditionierten linearen Gleichungssystems, weshalb die Fehler für Werte von σ unterhalb von etwa 200 divergieren. Abgesehen davon ergibt sich wieder ein ähnliches Bild wie zuvor und keine Verbesserung der Fehler oder der Konvergenzordnungen.

Auch auf dem Quadratgitter führt die Verwendung von Q_3 -Elementen nicht zu kleineren Fehlern oder besseren Konvergenzordnungen, wie Abbildung 5.42 zeigt. Zwar rücken die L^2 -Fehler im Bereich großer σ näher an die Fehler des Streamline-Diffusion-Verfahrens heran, sind aber immer noch vergleichsweise groß. Die H^1 -Fehler können die Ergebnisse des Streamline-Diffusion-Verfahrens auf Gitterlevel 6 nicht erreichen. Für die Konvergenzordnungen gelten im wesentlichen die gleichen Aussagen, wie bereits für die Q_2 -Elemente.

Tabelle 5.9 gibt die bestmöglichen Fehlerwerte und Konvergenzordnungen für Beispiel (5.2.5) bei einer Diffusionskonstante von $\epsilon = 10^{-3}$ und möglichst optimaler Wahl von σ wieder. Die Ergebnisse zeigen, dass die Größenordnung der Fehler unabhängig vom Grad der verwendeten Polynome ist.

Als Fazit dieses Abschnittes halten wir fest, dass das SIPG-Verfahren in der für diese Arbeit benutzten Form bei der Berechnung von Lösungen zu Problemen mit steilen Randgrenzschichten innerhalb akzeptabler Fehlerschranken wiederzugeben. Insbesondere führt die bestenfalls geringe Konvergenzordnung dazu, dass eine Verfeinerung des Gitters keine wesentliche Verbesserung der Fehler mit sich bringt.



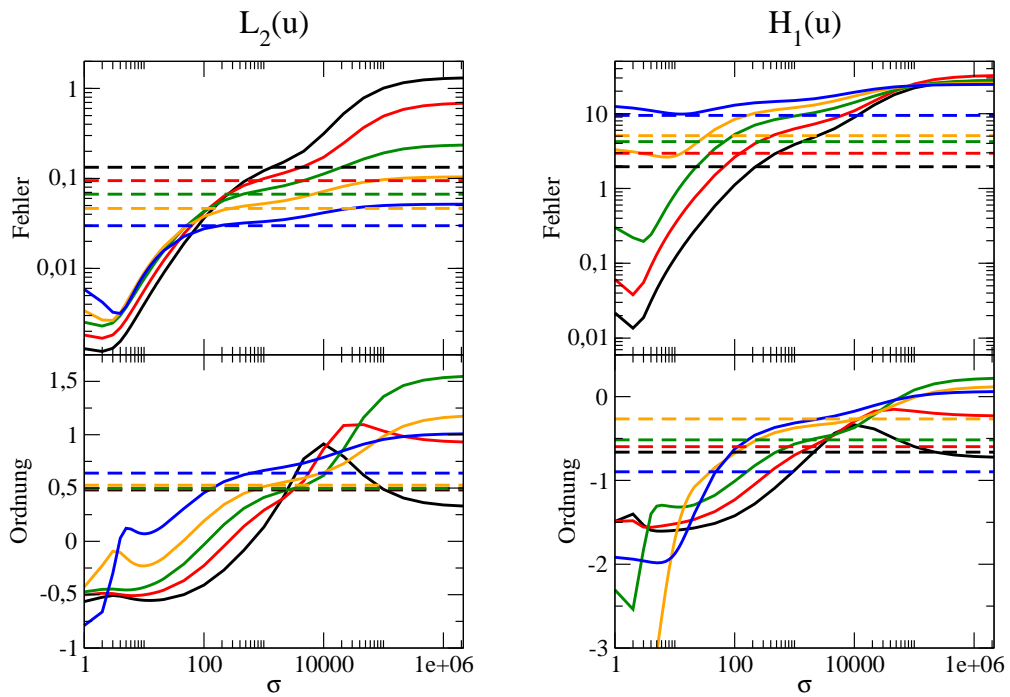
Fehler und Fehlerordnungen Finiten-Elemente zweiter Ordnung auf dem Dreiecksgitter (P_2) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.5) bei einer Diffusionskonstanten von $\epsilon = 10^{-3}$. Für weitere Erklärungen siehe Abbildung 5.9.

(Abbildung 5.39)

Element	Gitter	σ	L^2 -Fehler	Ordnung	H^1 -Fehler	Ordnung
P_1	Dreiecke	3	$9.04 \cdot 10^{-3}$	0.05	26.14	-0.03
P_2	Dreiecke	6	$7.70 \cdot 10^{-3}$	0.16	25.18	0.00
P_3	Dreiecke	215	$1.82 \cdot 10^{-2}$	0.67	21.01	0.09
Q_1	Quadrate	2	$4.19 \cdot 10^{-3}$	-0.67	11.86	-1.94
Q_2	Quadrate	4	$3.16 \cdot 10^{-3}$	0.03	11.00	-1.98
Q_3	Quadrate	6	$2.58 \cdot 10^{-3}$	0.55	9.74	-2.05

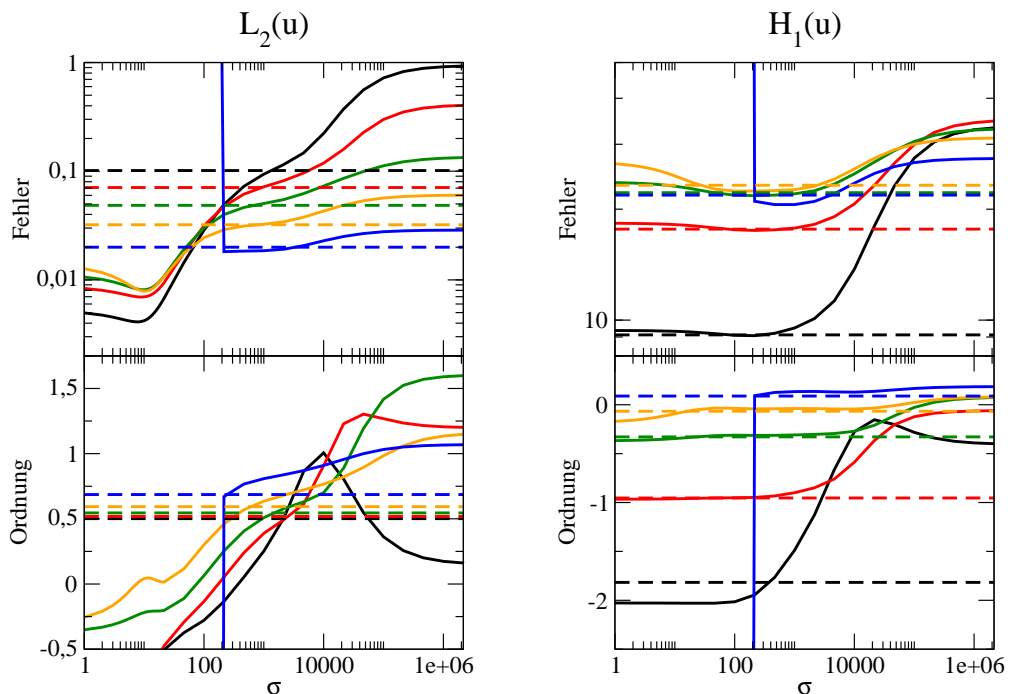
Optimaler Stabilisierungsparameter σ sowie zugehörige Fehler und Konvergenzordnungen der verschiedenen Finiten-Elemente für das Beispiel (5.2.5) bei einer Diffusionskonstanten $\epsilon = 10^{-3}$.

(Tabelle 5.9)



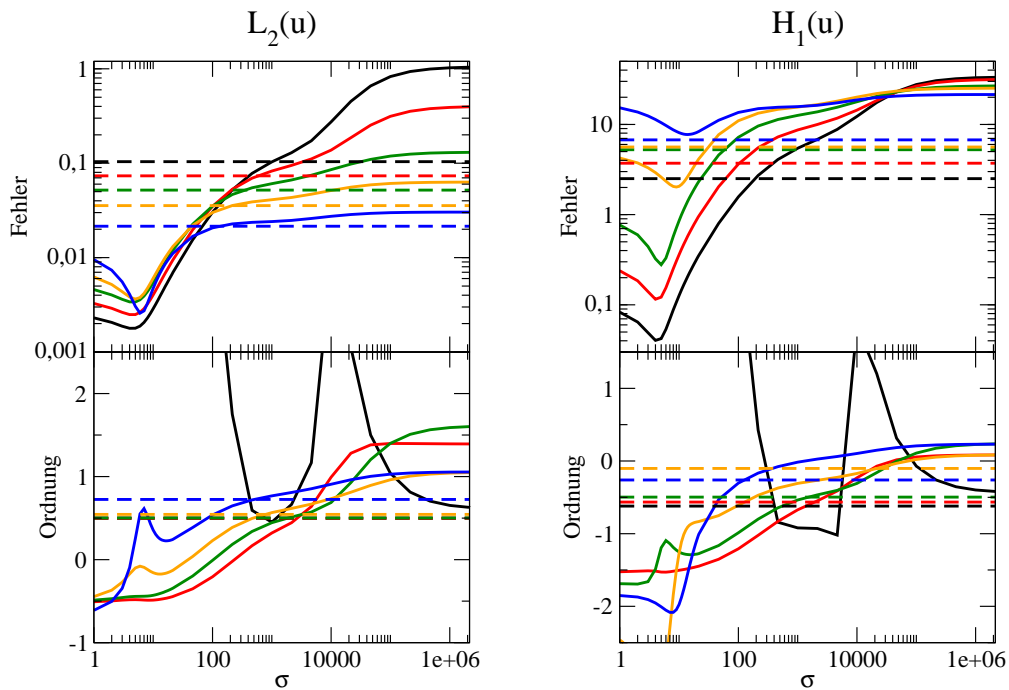
Fehler und Fehlerordnungen Finiter-Elemente zweiter Ordnung auf dem Quadratgitter (Q_2) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.5) bei einer Diffusionskonstanten von $\epsilon = 10^{-3}$. Für weitere Erklärungen siehe Abbildung 5.9.

(Abbildung 5.40)



Fehler und Fehlerordnungen Finiter-Elemente dritter Ordnung auf dem Dreiecksgitter (P_3) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.5) bei einer Diffusionskonstanten von $\epsilon = 10^{-3}$. Für weitere Erklärungen siehe Abbildung 5.9.

(Abbildung 5.41)

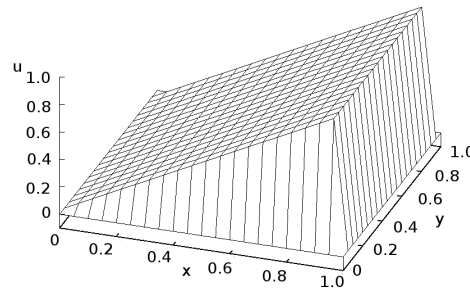


Fehler und Fehlerordnungen Finite-Elemente dritter Ordnung auf dem Quadratgitter (Q_3) aufgetragen gegen den Stabilisierungsparameter σ für das Beispiel (5.2.5) bei einer Diffusionskonstanten von $\epsilon = 10^{-3}$. Für weitere Erklärungen siehe Abbildung 5.9.

(Abbildung 5.42)

5.2.5 Qualität der Randbedingungen

Wie wir schon öfter erwähnt haben, sind die Randbedingungen beim SIPG-Verfahren nur schwach implementiert. Daher stellt sich die Frage, wie gut diese von der numerisch bestimmten Lösung wiedergegeben werden können, insbesondere im Falle von Randgrenzschichten. Um dies zu untersuchen, betrachten wir ein Beispiel aus (Joh07). Dabei wählen wir $\epsilon = 10^{-6}$, $b = (0, 1)^T$, $c = 0$, $f = 1$ und homogene Dirichlet-Randbedingungen auf dem Gebiet $\Omega = [0, 1]^2$. Abbildung 5.43 zeigt die zugehörige Lösung. Diese besitzt eine exponentielle Randgrenzschicht bei $x = 1$ und parabolische Randgrenzschichten bei $y = 0$ und $y = 1$. Außerhalb der Randgrenzschichten ist die Lösung sehr ähnlich zu $u(x, y) = x$.



Darstellung der exakten Lösung für das Beispiel mit den Daten $\epsilon = 10^{-6}$, $b = (0, 1)^T$, $c = 0$, $f = 1$ und homogene Dirichlet-Randbedingungen auf $\Omega = [0, 1]^2$ (Joh07).

(Abbildung 5.43)

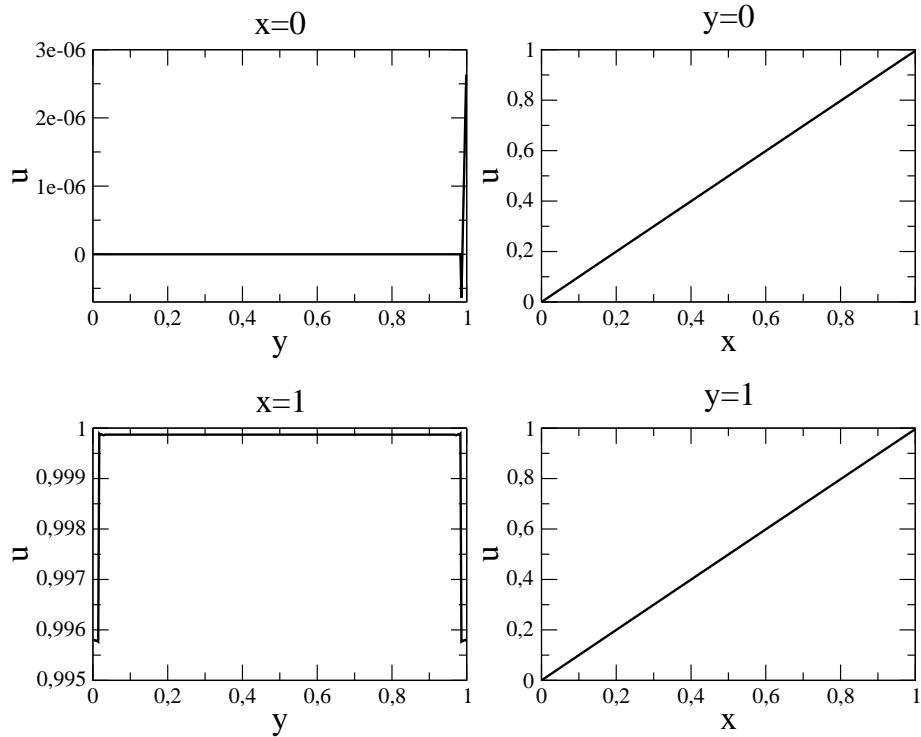
Abbildung 5.44(a) zeigt die Randwerte der mit Hilfe des SIPG-Verfahrens bestimmten Lösung u_h bei Verwendung von P_1 -Elementen auf dem Dreiecksgitter. Für $x = 0$ werden die Randwerte gut wiedergegeben. Hier liegt jedoch auch keine Grenzschicht vor. Für $x = 1$ hingegen nimmt u_h fast konstant den Wert 1 an, obwohl die Randbedingungen hier den Wert 0 vorgeben. Die Randbedingung ist also nicht erfüllt. Gleiches gilt für die Randwerte bei $y = 0$ und $y = 1$. Hier scheint eher $u_h(x, y) = x$ zu gelten. Insgesamt orientiert sich die numerische Lösung also mehr an der Lösung im Inneren als an den Randwerten.

Die Randwerte der numerisch bestimmten Lösung u_h bei Verwendung von Q_1 -Elementen auf dem Quadratgitter sind in Abbildung 5.44(b) dargestellt. Es ergibt sich ein ähnliches Bild wie für die P_1 -Elemente auf dem Dreiecksgitter. Auch auf dem Quadratgitter werden die Randwerte bei Verwendung linearer Finiten-Elemente also nicht richtig wiedergegeben.

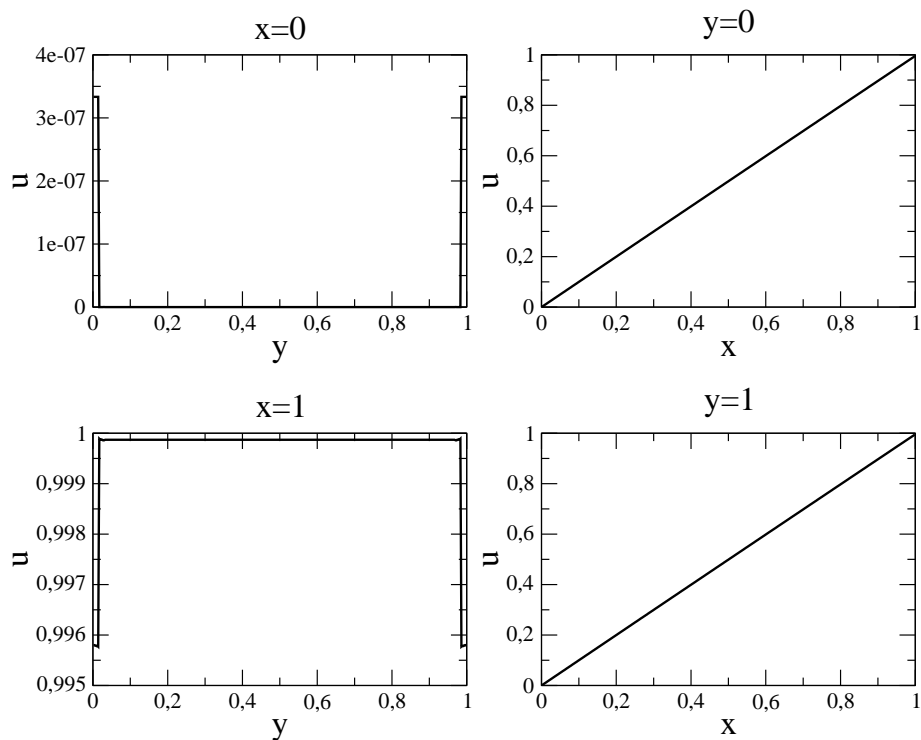
Die Verwendung von P_2 -Elementen auf dem Dreiecksgitter führt zu den in Abbildung 5.45 gezeigten Randwerten. Wie zu erkennen, führt eine Erhöhung des Polynomgrades nicht zu einer besseren Wiedergabe der vorgegebenen Randbedingungen. Am Rand $x = 1$ treten hier zusätzlich leichte Oszillationen nahe $y = 0$ und $y = 1$ auf. Die Steigung der Geraden, welche sich bei $y = 0$ und $y = 1$ zeigen, ist etwas kleiner als 1.

Für die anderen im Rahmen dieser Arbeit verwendeten Finiten-Elemente für das SIPG-Verfahren ergeben sich ähnliche Ergebnisse, weshalb wir auf die weitere Darstellung hier verzichten.

Insgesamt können wir festhalten, dass Randbedingungen an Rändern, an welchen steile Grenzschichten vorliegen, nicht erfüllt werden. In Folge dessen werden auch die Randgrenzschichten nicht wiedergegeben.



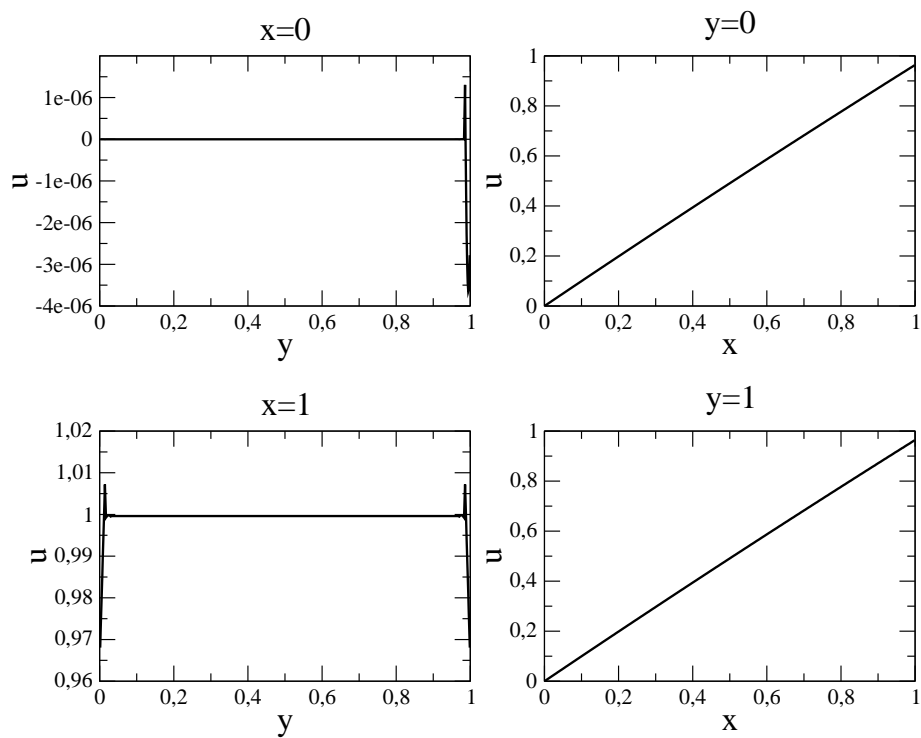
(a) P_1 -Elemente auf dem Dreiecksgitter.



(b) Q_1 -Elemente auf dem Quadratgitter.

Randwerte der mittels des SIPG-Verfahrens bestimmten Lösung zum Beispiel aus Abbildung 5.43 bei Verwendung von Q_1 -Elementen auf dem Quadratgitter. Dargestellt sind die Werte von u_h für $x = 0$ und $x = 1$ gegen y (links oben beziehungsweise unten) sowie für $y = 0$ und $y = 1$ gegen x (rechts oben beziehungsweise unten).

(Abbildung 5.44)



Randwerte der mittels des SIPG-Verfahrens bestimmten Lösung zum Beispiel aus Abbildung 5.43 bei Verwendung von P_2 -Elementen auf dem Dreiecksgitter. Dargestellt sind die Werte von u_h für $x = 0$ und $x = 1$ gegen y (links oben beziehungsweise unten) sowie für $y = 0$ und $y = 1$ gegen x (rechts oben beziehungsweise unten).

(Abbildung 5.45)

Kapitel 6

Zusammenfassung und Ausblick

Ziel dieser Arbeit war die Implementierung eines Discontinuous–Galerkin–Verfahrens für die Konvektions–Diffusions–Gleichung im Programmpaket MooNMD und die Untersuchung dieses Verfahrens an Hand geeigneter Testbeispiele. Da die Konvektions–Diffusions–Gleichung im Programmpaket MooNMD als eine Differentialgleichung zweiter Ordnung implementiert ist, lag die Verwendung eines Interior–Penalty–Verfahrens näher als die Verwendung eines Verfahrens in der so genannten Flussformulierung, da hier mit einem System zweier Differentialgleichungen erster Ordnung gearbeitet wird. Aus den verschiedenen Interior–Penalty–Verfahren wurde das symmetrische Verfahren SIPG gewählt. Theoretische Ergebnisse zu diesem Verfahren wie Fehlerabschätzungen und Angaben zur Konvergenzordnung der Fehler werden ausführlich in (Kan07) behandelt. Dieses Buch diente sowohl als theoretische Referenz als auch als Ausgangsbasis für die Implementierung.

Neu implementiert werden mussten die Terme, welche Integrale über die Kanten enthalten und so Basisfunktionen aus verschiedenen Zellen verknüpfen. Des Weiteren musste die Matrixstruktur erweitert werden, um diese Terme berücksichtigen zu können. Bei den hinzukommenden Termen kann man zwischen solchen, welche vom diffusiven Anteil der Gleichung herrühren und solchen, welche mit dem konvektiven Anteil zusammenhängen, unterscheiden. Für beide Terme wurde zunächst getrennt getestet, ob diese richtig implementiert wurden. Dabei zeigt sich im rein diffusiven Fall ein uneinheitliches Verhalten der Fehler im Bereich kleiner Werte von σ . Für hinreichend große Werte werden aber die Fehler und Konvergenzordnungen konstant. Die theoretisch zu erwartenden Werte der Konvergenzordnungen werden erreicht. Im rein konvektiven Fall erhielten wir die erwarteten Ergebnisse bezüglich der Konvergenzordnungen und die Unabhängigkeit vom Stabilisierungsparameter σ .

Bei der Untersuchung des SIPG–Verfahrens für die Konvektions–Diffusions–Gleichung wurden verschiedene Fälle betrachtet:

- (i) ein exponentielles Beispiel ohne Grenzschichten aus (Bur04),
- (ii) ein Beispiel mit moderaten inneren Grenzschichten aus (Joh97),
- (iii) dasselbe Beispiel mit steilen inneren Grenzschichten, welche sich bei einer kleineren Diffusionskonstante ergeben,
- (iv) ein Beispiel mit Randgrenzschichten aus (Joh97).

Als Gebiet diente in allen Fällen das Einheitsquadrat $[0, 1]^2$. Für alle Beispiele wurden jeweils Lösungen unter Verwendung von $P_1^{\text{disc-}}$, $P_2^{\text{disc-}}$ und $P_3^{\text{disc-}}$ -Elementen auf einem Dreiecksgitter und einem Quadratgitter sowie unter Verwendung von $Q_1^{\text{disc-}}$, $Q_2^{\text{disc-}}$ und $Q_3^{\text{disc-}}$ -Elementen auf einem Quadratgitter berechnet. Die Feinheit der Gitter wurde dabei bis auf Gitterlevel 6 erhöht. Zum Vergleich dienten jeweils Ergebnisse, welche mittels eines Streamline–Diffusion–Verfahrens berechnet wurden. Zur Bewertung der Verfahren haben wir die L^2 -Norm und die H^1 -Seminorm der Fehler $u - u_h$ sowie die zugehörigen Konvergenzordnungen betrachtet.

Für das Beispiel ohne Grenzschichten zeigt das SIPG–Verfahren gute Ergebnisse. Die Konvergenzordnungen entsprechen hier den Erwartungen und fallen ebenso wie die Fehler bei beiden Verfahren ungefähr gleich aus. Es stellt sich heraus, dass die Ergebnisse für lineare und kubische Elemente fast unabhängig von σ sind, wohingegen große Werte von σ bei quadratischen Elementen zu etwas schlechteren Ergebnissen führen. Diese Aussagen treffen sowohl auf die Elemente auf dem Dreiecks- als auch auf diejenigen auf dem Quadratgitter zu. Auf dem Quadratgitter zeigt sich, dass die Q -Elemente bei jeder Ordnung bessere Ergebnisse liefern, als die P -Elemente. Zwischen P -Elementen auf dem Dreiecksgitter und Q -Elementen auf dem Quadratgitter zeigt sich kein wesentlicher Unterschied.

Neben der Qualität der Verfahren wurde auch ihre Laufzeit an Hand des Beispiels ohne Grenzschichten verglichen. Da das SIPG–Verfahren mehr Basisfunktionen als das Streamline–Diffusion–Verfahren besitzt, ist die zugehörige Matrix größer. Zudem enthält sie prozentual mehr von Null verschiedene Einträge, da auch Basisfunktionen aus benachbarten Zellen über die Kantenintegrale miteinander verkoppelt sind. Daher sind die Laufzeiten des SIPG–Verfahrens zwischen 3.5 und 6.34-mal so lang wie die Laufzeiten des Streamline–Diffusion–Verfahrens.

Für das Beispiel mit moderaten inneren Grenzschichten lassen sich mit dem SIPG–Verfahren ebenfalls gute Ergebnisse erzielen. Die Abhängigkeit von σ ist hier vor allem auf den größeren Gittern stärker ausgeprägt. Kleine Werte von σ liefern dabei meist bessere Ergebnisse. Eine Ausnahme ist hier das Q_1 -Element auf Gitterlevel 6. Wie schon beim Beispiel ohne Grenzschichten sind die Ergebnisse von SIPG- und Streamline–Diffusion–Verfahren ähnlich. Die Konvergenzordnungen entsprechen wieder den theoretischen Vorhersagen.

Beim Beispiel mit steilen inneren Grenzschichten zeigen SIPG- und Streamline–Diffusion–Verfahren dieselbe Schwäche: Beide zeigen Oszillationen innerhalb der Grenzschicht, welche beim SIPG–Verfahren sogar stärker ausfallen. Diese Oszillationen werden bei beiden Verfahren in Richtung des Advektionsvektors b in das Gebiet hinein getragen. In Folge dessen zeigen beide Verfahren ähnlich große Fehler, welche beim SIPG–Verfahren für die meisten Parameter konstant sind und nur für sehr große σ leicht anwachsen. Zudem sind die Konvergenzordnungen für die L^2 -Fehler sehr klein und für die H^1 -Fehler nahe Null, so dass eine Verfeinerung des Gitters die Ergebnisse nicht merklich verbessert.

Im Fall des Beispiel mit Randgrenzschichten stellt sich ebenfalls heraus, dass das SIPG–Verfahren diese nicht adäquat wiedergeben kann. Im Gegensatz zu den vorherigen Beispielen zeigt sich hier eine ausgeprägte Abhängigkeit der L^2 -Fehler vom Parameter σ , jedoch sind diese auch bei einer optimalen Wahl von σ noch sehr groß. Die Konvergenzordnungen bleiben hinter den Erwartungen zurück. Im Bereich kleiner L^2 -Fehler zeigen sich sogar teilweise negative Konvergenzordnungen. Die H^1 -Fehler sind im gesamten Parameterbereich sehr groß und wachsen mit zunehmend feinerem Gitter. Das Streamline–Diffusion–Verfahren liefert bei diesem Beispiel jedoch auch keine besseren Ergebnisse.

Da die Randbedingungen im SIPG–Verfahren nur schwach implementiert sind, wurde außerdem getestet, inwiefern die berechnete Lösung die Randbedingungen erfüllt. Hierzu diente ein Beispiel mit homogenen Dirichlet–Randbedingungen, welches sowohl parabolische als auch exponentielle Randgrenzschichten zeigt (Joh07). Es stellte sich heraus, dass die Randbedingungen in diesem Fall nicht erfüllt sind. Vielmehr folgt die berechnete Lösung dem Verhalten der exakten Lösung im Inneren des Gebietes. Dies bedeutet leider auch, dass die Randgrenzschichten nicht wiedergegeben werden.

Ausblick

Ausgehend von dieser Arbeit könnten folgende Fragestellungen untersucht werden:

- Wir haben hier nur das SIPG–Verfahren betrachtet. Ein Vergleich mit dem NIPG–Verfahren oder das IIPG–Verfahren steht noch aus; ebenso wie ein Vergleich mit der Flussformulierung aus Abschnitt 4.5.2.
- Der in dieser Arbeit verwendete direkte Löser aus dem Programmpaket UMFPACK (Dav04a) konnte für kleine Werte von σ bei zwei Beispielen keine Lösung bestimmen. Es stellt sich daher die Frage, ob die Verwendung eines anderen Löser hier Abhilfe schaffen kann und ob sich dadurch die Laufzeiten des SIPG–Verfahrens verbessern lassen.
- Eine Möglichkeit um Grenzschichten besser aufzulösen, wäre die adaptive Anpassung der Gitterweite (h -Methoden). Diese ist auch mit einer Anpassung des Polynomgrades (hp -Methoden) kombinierbar. Bei stetigen Basisfunktionen ergibt sich dabei das Problem, dass man an den Kanten der Gitterzellen entsprechende Anschlussbedingungen erfüllen muss, um die Stetigkeit zu garantieren. Dies ist bei Discontinuous–Galerkin–Verfahren nicht der Fall, wodurch sich die Implementierung einer hp -Methode deutlich einfacher gestaltet.
- Neben einem Term, welcher den Sprung in den Funktionswerten beim Übergang über eine Kante bestraft, wird beispielsweise in (Riv08) vorgeschlagen, auch den Sprung in den Gradienten zu bestrafen. Ein solcher Term wird in der Literatur auch zur Stabilisierung für stetige Finite–Elemente vorgeschlagen (Bur04).

Literaturverzeichnis

- [Ada75] ADAMS, R.A.: *Sobolev spaces*, Academic Press, New York (1975)
- [Alt85] ALT, H.W.: *Lineare Funktionalanalysis. Eine anwendungsorientierte Einführung*, Springer Berlin (1985)
- [Arn82] ARNOLD, D. N.: An interior penalty finite element method with discontinuous elements. *SIAM J. Numer. Anal.* (1982), Bd. 19: S. 724 – 760
- [Arn02] ARNOLD, D. N.; BREZZI, F.; COCKBURN, B. und MARINI, L. D.: Unified Analysis of Discontinuous Galerkin Methods for Elliptic Problems. *SIAM J. Numer. Anal.* (2002), Bd. 39: S. 1749 – 1779
- [Azz80] AZZAM, A.: On differentiability properties of solutions of elliptic differential equations. *J. Math. Anal. Appl.* (1980), Bd. 75: S. 431–440
- [Bab02] BABUSKA, I. M.; BANERJEE, U. und OSBORN, J.E.: On principles for the selection of shape functions for the Generalized Finite Element Method. *Computer methods in applied mechanics and engineering* (2002), Bd. 191(49-50): S. 5595–5629
- [Bra07] BRAESS, D.: *Finite Elemente: Theorie, schnelle Löser und Anwendungen in der Elastizitätstheorie*, Springer Verlag, Heidelberg (2007)
- [Bre94] BRENNER, S.C. und SCOTT, L.R.: *The Mathematical Theory of Finite Element Methods*, Bd. 15 von *Texts in Applied Mathematics*, Springer-Verlag New York (1994)
- [Bre04] BREZZI, F.; MARINI, L. D. und SÜLI, E.: Discontinuous Galerkin Methods for first-order hyperbolic problems. *Math. Models Methods Appl. Sci.* (2004), Bd. 14: S. 1893 – 1903
- [Bur04] BURMAN, E. und HANSBO, P.: Edge stabilization for Galerkin approximations of convection–diffusion–reaction problems. *Comput. Methods Appl. Mech. Engrg.* (2004), Bd. 193: S. 1437 – 1453
- [Chr76] CHRISTIE, I.; GRIFFITHS, D.F.; MITCHELL, A.R. und ZIENKIEWICZ, O.C.: Finite element methods for second order differential equations with significant first derivatives. *Int. J. Numer. Meth. Engrg.* (1976), Bd. 10: S. 1389 – 1396
- [Cia72] CIARLET, P.G. und RAVIART, P.-A.: General Lagrange and Hermite Interpolation in R^n with Applications to Finite Element Methods. *Arch. Rat. Mech. Anal.* (1972), Bd. 46: S. 177–199
- [Cia78] CIARLET, P.G.: *The finite element method for elliptic problems*, North-Holland Publishing Company, Amsterdam – New York – Oxford (1978)

- [Cio99] CIORANESCU, D.: *An Introduction to Homogenization*, Nr. 17 in Oxford Lecture Series in Mathematics and Its Applications, Oxford University Press (1999)
- [Coc98a] COCKBURN, B.: An introduction to the discontinuous Galerkin method for convection-dominated problems, in: A. Quarteroni (Herausgeber) *Advanced Numerical Approximation of Nonlinear Hyperbolic Equations*, Bd. 1697 von *Lecture Notes in Mathematics*, Springer, Berlin / Heidelberg (1998), S. 151 – 268
- [Coc98b] COCKBURN, B. und SHU, C.-W.: The local discontinuous Galerkin method für time-dependent convection-diffusion systems. *SIAM J. Numer. Anal.* (1998), Bd. 35: S. 2440 – 2463
- [Coc00] COCKBURN, B.; KARNIADAKIS, G. und SHU, C. W.: The Development of Discontinuous Galerkin Methods, in: B. Cockburn; G. Karniadakis und C. W. Shu (Herausgeber) *Discontinuous Galerkin Methods: Theory, Computation and Applications*, Lecture Notes in Computational Science and Engineering, Springer (2000), S. 3 – 50
- [Coc01] COCKBURN, B. und SHU, C.-W.: Runge-Kutta discontinuous Galerkin methods for convection-dominated problems. *J. of Sci. Comp.* (2001), Bd. 16: S. 173 – 261
- [Coc03] COCKBURN, B.; KANSCHAT, G. und SCHÖTZAU, D.: The local discontinuous Galerkin method for the Oseen equations. *Math. of Comp.* (2003), Bd. 73: S. 569 – 593
- [Dav04a] DAVIDSON, P.A.: *Turbulence, An Introduction for Scientists and Engineers*, Oxford University Press (2004)
- [Dav04b] DAVIS, T.A.: Algorithm 832: UMFPACK V4.3—an unsymmetric-pattern multifrontal method. *ACM Trans. Math. Software* (2004), Bd. 30: S. 196 – 199
- [Goe83] GOERING, H.; FELGENHAUER, A.; LUBE, G.; ROOS, H. G. und TOBISKA, L.: *Singularly perturbed differential equations*, Akademie Verlag, Berlin (1983)
- [Gri85] GRISVARD, P.: *Elliptic problems in nonsmooth domains.*, Pitman, Boston (1985)
- [Gro05] GROSSMANN, C. und ROOS, H.-G.: *Numerische Behandlung partieller Differentialgleichungen*, Teubner-Verlag (2005)
- [Heg08] HEGARTY, A.F.; KOPTEVA, N.; O’RIORDAN, E. und STYNES, M.: *BAIL 2008: Boundary and Interior Layers*, Bd. 69 von *Lecture Notes in Computational Science and Engineering*, Springer Verlag, Berlin (2008)
- [Hei77] HEINRICH, J.C.; HUYAKORN, P.S.; ZIENKIEWICZ, O.C und MITCHELL, A.R.: An ‘upwind’ finite element scheme for two-dimensional convective transport equation. *Int. J. Numer. Methods Engrg.* (1977), Bd. 11: S. 131 – 143
- [Hes08] HESTHAVEN, J.S. und WARBURTON, T.: *Nodal Discontinuous Galerkin Methods: Algorithms, Analysis, and Applications*, Nr. 54 in Texts in Applied Mathematics, Springer, Berlin (2008)
- [Hou01] HOUSTON, P. und SÜLI, E.: Stabilised hp-finite element approximations of partial differential equations with nonnegative characteristic form. *Computing* (2001), Bd. 66: S. 99–119

- [Hug79] HUGHES, T.J.R. und BROOKS, A.N.: A Multidimensional Upwind Scheme with no Crosswind Diffusion, in: T.J.R. Hughes (Herausgeber) *Finite Element Methods for Convection Dominated Flows, AMD vol.34*, ASME, New York (1979), S. 19 – 35
- [Joh87] JOHNSON, C. SCHATZ, A.H. und WAHLBIN, L.B. Crosswind smear and pointwise errors in streamline diffusion finite element methods. *Math. Comput.* (1987), Bd. 49: S. 25 – 38
- [Joh97] JOHN, V.; MAUBACH, J.M. und TOBISKA, L.: Nonconforming Streamline-Diffusion-Finite-Element-Methods for Convection-Diffusion Problems. *Numer. Math.* (1997), Bd. 78: S. 165 – 188
- [Joh07] JOHN, V. und KNOBLOCH, P.: A comparison of spurious oscillations at layers diminishing (SOLD) methods for convection–diffusion equations: PART I – a review. *Comput. Methods Appl. Mech. Engrg.* (2007), Bd. 196: S. 2197 – 2215
- [Joh08] JOHN, V. und KNOBLOCH, P.: A comparison of spurious oscillations at layers diminishing (SOLD) methods for convection–diffusion equations: PART II – Analysis for P_1 and Q_1 finite elements. *Comput. Methods Appl. Mech. Engrg.* (2008), Bd. 197: S. 1997 – 2014
- [Kan07] KANSCHAT, G.: *Discontinuous Galerkin Methods for Viscous Incompressible Flow*, Advances in Numerical Mathematics, Teubner Research, 1 Aufl. (2007)
- [LeS74] LESAIN, P. und RAVIART, P.-A.: On a finite element method for solving the neutron transport equations, in: C. de Boor (Herausgeber) *Mathematical aspects of finite elements in partial differential equations*, Academic Press, New York (1974), S. 89–123
- [Mic77] MICHAEL, J. H.: A general theory for linear elliptic partial differential equations. *J. diff. equations* (1977), Bd. 23: S. 1–29
- [Näv92] NÄVERT, U.: *A finite element method for convection-diffusion problems*, Dissertation, Chalmers university of Technology, Göteborg (1992)
- [Nit72] NITSCHKE, J.: *Über ein Variationsproblem zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind.*, Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg, Springer Berlin / Heidelberg (1972)
- [Ree73] REED, W.H. und HILL, T.R.: Triangular mesh methods for the neutron transport equation, Techn. Ber. LA-UR-73-479, Los Alamos Scientific Laboratory (1973)
- [Riv08] RIVIERE, B.: *Discontinuous Galerkin Methods For Solving Elliptic And Parabolic Equations: Theory and Implementation*, Nr. 35 in *Frontiers in Applied Mathematics*, SIAM (2008)
- [Roo08] ROOS, H.-G.; STYNES, M. und TOBISKA, L.: *Robust Numerical Methods for Singularly Perturbed Differential Equations*, Bd. 24 von *Springer Series in Computational Mathematics*, Springer, 2 Aufl. (2008)

- [Sch98] SCHWAB, C.: *P- and hp- Finite Element Methods: Theory and Applications in Solid and Fluid Dynamics*, Calderon Press, Oxford (1998)
- [Sev07] SEVILLA, R.; FERNANDÉZ, S. und HUERTA, A.: NURBS-enhanced finite element method (NEFEM). *International Journal for Numerical Methods in Engineering* (2007), Bd. 76(1): S. 56–83
- [Sty95] STYNES, M. und TOBISKA, L.: Necessary L^2 -uniform convergence conditions for difference schemes for two-dimensional convection-diffusion problems. *Comput. Math. Appl.* (1995), Bd. 29: S. 45 – 53
- [Sty05] STYNES, M.: Steady-state convection-diffusion problems. *Acta Numerica* (2005), Bd. 15
- [Tab77] TABATA, M.: A finite element approximation corresponding to the upwind differencing. *Memoirs of Numerical Mathematics* (1977), Bd. 1: S. 47 – 63
- [Tho97] THOMÉE, V.: *Galerkin finite element methods for parabolic problems*, Bd. 25 von *Springer Series in Computational Mathematics*, Springer-Verlag, Berlin (1997)
- [Tre75] TREVES, F.: *Basic Linear Partial Differential Equations*, Academic Press, New York (1975)
- [Wer95] WERNER, D.: *Functional Analysis*, Springer-Verlag (1995)
- [Wig70] WIGLEY, N.M.: Mixed boundary value problems in domains with corners. *Math. Z.* (1970), Bd. 115: S. 33–52
- [Wlo82] WLOKA, J.: *Partielle Differentialgleichungen*, B.G. Teubner Stuttgart (1982)
- [Yos65] YOSHIDA, K.: *Functional Analysis*, Springer-Verlag (1965)
- [Zho96] ZHOU, G. und RANNACHER, R.: Pointwise superconvergence of the streamline diffusion finite element method. *Numer. Methods Partial Differ. Equations* (1996), Bd. 12(1): S. 123–145
- [Zho97] ZHOU, G.: How accurate is the streamline diffusion finite element method? *Math. Comp.* (1997), Bd. 66: S. 31 – 44