DIPLOMARBEIT

Stabilisierte Finite–Element Verfahren für die Konvektions–Diffusionsgleichung und die Oseen–Gleichung

zur Erlangung des akademischen Grades

DIPLOM-MATHEMATIKERS

Durchgeführt am

Institut für Angewandte Mathematik,
der Universität des Saarlandes

vorgelegt von

Rudolf Helmut Umla

Hiermit versichere ich an Eides statt, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel verwendet habe.

Saarbrücken, den

Ort, Datum

Rudolf Helmut Umla

Danksagung

Mein Dank gilt

- Herrn Prof. Dr. V. John für die Vergabe des interessanten Themas und die Betreuung meiner Arbeit.
- meinen Eltern, Silvia und Helmut Umla, sowie meiner Großmutter, Johanna Aubertin, für die finanzielle Unterstützung während des Studiums.
- meinem Vater und Matthias Augustin für das Korrekturlesen der Arbeit.
- Marc Neef für die freundschaftliche Zusammenarbeit während des Studiums.
- meiner Freundin Anne für die Geduld.

Inhaltsverzeichnis

Symbolverzeichnis

Symbol	Bezeichnung	Seite
a_{CIP}	Bilinearform der CIP–Methode	61,101
a_{RB}	Bilinearform der residualen Stabilisierung	95
a_{SD}	Bilinearform der Streamline–Diffusion–Methode	55
a_{SG}	Bilinearform des Standard–Galerkin–Verfahrens zur Oseen–Gleichung	93
a_W	Bilinearform für schwache Randbedingungen	53,94
b	Advektionsvektor	45
b_h	Stückweise lineare Approximation an den Advektionsvektor b	62
b_W	Bilinearform für schwache Randbedingungen der Oseen–Gleichung	94
c_0	Konstante und untere Schranke bei der Konvektions–Diffusionsgleichung	51
C_F	Friedrichs–Konstante	13
C^k	Raum der k-mal stetig differenzierbaren Funktionen	5
$C^{k,s}$	Raum der k-mal stetig differenzierbaren Hölder–stetigen Funktionen mit Exponent s	5
$\mathcal{C}_h^p$	Clément-Interpolationsoperator XX	42
h	Maximaler Zellendurchmesser	51
$\widetilde{h}$	Funktion in Abhängigkeit des Zellendurchmessers	94
h_E	Länge der Kante E	44
h_K	Durchmesser der Gitterzelle K	40
H^k	Sobolev–Raum $W^{k,2}$	9
$H^2(\mathcal{T}_h)$	Raum der stückweisen H^2 Funktionen bezüglich $\mathcal{T}_h$	62
i_h^p	Scott-Zhang–Interpolationsoperator	43
I_h^p	Lagrange–Interpolationsoperator	41
j_h	Stabilisierungsterm der CIP–Methode	61
j_p, j_u	Stabilisierungsterm der CIP–Methode für die Oseen–Gleichung	101
$\ker(T)$	Kern des Operators T	13
L^p	Raum der Lebesgue–Funktionen	7
$L^2_*(\Omega)$	Raum der L^2-Funktionen mit Mittelwert Null auf Ω	90
L^1_{para}	Fehlergröße für parabolische Grenzschicht	86
L^1_{expo}	Fehlergröße für exponentielle Grenzschicht	87
$\mathcal{L}(V,Q)$	Raum der linear stetigen Abbildungen von V nach Q	13
n	Normalenvektor	46
P^p	Finite–Element–Räume auf Dreiecken	33
$\mathbb{P}^p$	Raum der Polynome mit Grad kleiner gleich p	33
Pe_K	Lokale Pécletzahl	59
Q^p	Finite–Element–Räume auf Rechtecken	36

$\mathbb{Q}^p$	Raum der Polynome mit Grad kleiner gleich p in jeder Variablen	36
r	Elementordnung	92
Re_K	Lokale Reynoldszahl	101
$R(T)$	Bild des Operators T	13
R_V	Rieszscher Darstellungsoperator auf dem Raum V	17
T'	Dualoperator des Operators T	13
$\mathcal{T}_h$	Gebietszerlegung	32
V'	Dualraum des Raums V	7
$W^{k,p}$	Sobolev–Räume	9
$W_0^{k,p}(\Omega)$	Sobolev–Räume mit kompaktem Träger auf Ω	10
α	Multiindex	11
α	Freier Parameter in der Norm der residualen Stabilisierung	96
δ_0, δ_1	Stabilisierungsparameter der Streamline–Diffusion–Methode	59
δ_{ij}	Diracsche Deltadistribution	31
δ_K	Stabilisierungsparameter der Streamline–Diffusion–Methode	55
δ_K	Stabilisierungsparameter der residualen Stabilisierung	95
ϵ	Diffusionskonstante	45
γ	Spuroperator	11
γ	Freier Parameter bei schwachen Randbedingungen	53
γ_K	Stabilisierungsparameter der residualen Stabilisierung	95
Γ	Rand des Gebiets Ω	45
Γ_-	Einströmrand von Ω	53
μ_{inv}	Konstante aus der inversen Ungleichung	44
ν	Viskosität	89
π	Projektionsoperator	15
π_h	L^2-Projektionsoperator	41
π^*	Interpolationsoperator der CIP–Methode	62,105
τ	Stabilisierungsparameter der CIP–Methode	61,101
τ_{grad}, τ_{div}, τ_p	Stabilisierungsparameter der CIP–Methode	112
ζ	Minimumsfunktion der CIP–Methode	101
$\partial\Omega$	Rand des Gebiets Ω	6
∇u	Gradient von u	45
$\nabla \cdot u$	Divergenz von u	89
Δu	Laplace von u	45
$(\cdot,\cdot)$	L^2-Skalarprodukt über dem Gebiet Ω	8
$(\cdot,\cdot)_X$	L^2-Produkt über der Menge X	48
$\langle\cdot,\cdot\rangle$	Duales Produkt	13
$\lvert\cdot\rvert_{W^{k,p}}$	Sobolev–Halbnormen	9
$\lVert a\rVert$	Norm der Bilinearform a	14
$\lVert T\rVert$	Operatornorm des Operators T	14
$\lVert\cdot\rVert_a$	Durch die Bilinearform a induzierte Norm	26
$\lVert\cdot\rVert_{W^{k,p}}$	Sobolev–Normen	9
$\lvert\!\lvert\!\lvert\cdot\rvert\!\rvert\!\rvert_{\mathrm{CIP}}$	Norm der CIP–Methode	62,102
$\lvert\!\lvert\!\lvert\cdot\rvert\!\rvert\!\rvert_{\mathrm{RB}}$	Norm der residualen Stabilisierung	96
$\lvert\!\lvert\!\lvert\cdot\rvert\!\rvert\!\rvert_{\mathrm{SD}}$	Norm der Streamline–Diffusion–Methode	56
$X \hookrightarrow Y$	Einbettung von Raum X in Raum Y	9

X°	Annihilator der Menge X	13
$X^\perp$	Orthogonales Komplement der Menge X	15
$[u]_E$	Sprung der Funktion u über die Kante E	61

Kapitel 1

Einleitung

Ziel dieser Arbeit ist die numerische Untersuchung stabilisierter Finite–Elemente–Verfahren für die Konvektions–Diffusions- und die Oseen–Gleichung. Bei beiden Gleichungen handelt es sich um lineare partielle Differentialgleichungen. Die Konvektions–Diffusionsgleichung lautet

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

Die skalare Funktion u ist auf dem Gebiet $\Omega \subset \mathbb{R}^d$, $d \in \mathbb{N}$, gesucht, während c, b und f vorgegebene, ausreichend reguläre Funktionen sind. Gehorcht die Funktion u geeigneten Randbedingungen auf dem Gebietsrand von Ω, so besitzt die Konvektions–Diffusionsgleichung eine eindeutige Lösung. Die Lösung zu bestimmen ist für eine Vielzahl von Anwendungen von Interesse. Dies ist im Allgemeinen nicht allein mit analytischen Rechnungen möglich. Daher werden numerische Verfahren verwandt wie beispielsweise die Methode der Finiten–Elemente, der Finiten–Differenzen sowie der Finiten–Volumen. In dieser Arbeit werden wir die Methode der Finiten–Elemente untersuchen. Die Stärken dieser Methode sind unter anderem eine relativ große Genauigkeit und ihre Flexibilität gegenüber dem vorliegenden Lösungsgebiet. Aufgrund ihrer Stärken zählt die Finite–Element–Methode zu den beliebtesten numerischen Verfahren zur Lösung der Konvektions–Diffusionsgleichung.

In der vorliegenden Arbeit wird die Konvektions–Diffusionsgleichung vor allem für den konvektions–dominanten Fall mit $\epsilon \ll \|b\|_{L^\infty(\Omega)}$ betrachtet. In diesem Fall liefert die Standard–Variante der Finiten-Element–Methode oft keine zufriedenstellenden Ergebnisse mehr. Zu beobachten sind unphysikalische Oszillationen in der berechneten Lösung. Um diese Oszillationen zu reduzieren, wird die Finite–Element–Methode stabilisiert. Hierzu untersuchen wir zwei Stabilisierungsverfahren: die Streamline–Diffusion–Methode und die Kanten–Stabilisierung.

Die Streamline–Diffusion–Methode kann als das klassische Stabilisierungsverfahren angesehen werden. Nachdem durch [HB79] die Idee der Streamline–Diffusion–Methode bekannt wurde, ist die Methode über Jahrzehnte in einer Vielzahl von Arbeiten numerisch wie analytisch untersucht worden. Die Kanten–Stabilisierung ist hingegen vergleichsweise wenig erforscht. Obwohl die Idee dieser Methode in [DD75] ebenfalls früh publiziert wurde, hat die Kanten–Stabilisierung erst neuerdings durch die Arbeiten von Burman und Hansbo an Bedeutung gewonnen [BH04], [BH06], [BFH06].

Das erste Ziel der vorliegenden Arbeit ist ein numerischer Vergleich beider Stabilisierungsverfahren. Zur Simulation benutzen wir das Programmpaket “Mathematics and object oriented Numerics in Magdeburg”, kurz MooNMD. Die Streamline–Diffusion–Methode ist dort bereits vorimplementiert. Im Gegensatz dazu ist die Kanten–Stabilisierung im Rah-

men dieser Arbeit in den Code eingearbeitet worden.

Neben der Konvektions-Diffusionsgleichung betrachten wir die Oseen–Gleichung:

$$\begin{aligned} -\nu\Delta u + (b \cdot \nabla u) + cu + \nabla p &= f \qquad \text{in } \Omega, \\ \nabla \cdot u &= 0 \qquad \text{in } \Omega. \end{aligned}$$

Gesucht sind hier die vektorwertige Größe u und die skalare Größe p auf dem Gebiet Ω, während b, c und f vorgegebene, hinreichend reguläre Funktionen sind. Auch zur Lösung der Oseen–Gleichung werden oft Finite–Element–Methoden eingesetzt. Das Standard–Verfahren ist hier in der Regel instabil, was eine Stabilisierung erforderlich macht. In Analogie zur Konvektions–Diffusionsgleichung untersuchen wir zwei Stabilisierungsverfahren: die klassische residuale Stabilisierung und eine erweiterte Form der Kanten–Stabilisierung. Zu der klassischen residualen Stabilisierung findet man in der Literatur zahlreiche numerische Ergebnisse, siehe zum Beispiel [TL91], [FF92] und [MLR09]. Die Kanten–Stabilisierung ist numerisch hingegen wenig erforscht (die Ergebnisse beschränken sich im Wesentlichen auf die Arbeit [BFH06]).

Das zweite Ziel der vorliegenden Arbeit besteht darin, die Kanten–Stabilisierung für die Oseen–Gleichung detaillierter zu untersuchen und die Ergebnisse mit der klassischen residualen Stabilisierung zu vergleichen. Dazu ist der Code der Kanten–Stabilisierung für die Konvektions–Diffusionsgleichung entsprechend in MooNMD erweitert worden. Die klassische residuale Stabilisierung ist hingegen bereits in MooNMD implementiert und konnte ohne weiteren Aufwand genutzt werden.

Die Arbeit gliedert sich wie folgt:

In **Kapitel 2** werden die für diese Arbeit relevanten funktionalanalytische Grundlagen wiederholt. Dazu gehört die Theorie der Lebesgue- und Sobolevräume, der Satz vom abgeschlossenen Bild sowie der Projektionssatz.

Aufbauend auf dieser Theorie untersuchen wir in **Kapitel 3** die eindeutige Lösbarkeit von Variationsproblemen. Variationsprobleme treten bei der schwachen Formulierung der Konvektions–Diffusions- beziehungsweise der Oseen–Gleichung auf und stellen den theoretischen Ausgangspunkt für die Finite–Element–Methoden dar.

Variationsprobleme können mit Hilfe von Galerkin–Verfahren approximiert werden.
Galerkin–Verfahren führen wir in **Kapitel 4** ein, untersuchen deren Lösbarkeit und geben Abschätzungen für die Güte der Approximation an.

Als Spezialfall der Galerkin–Verfahren behandeln wir in **Kapitel 5** die Finiten–Element–Methode. Betrachtet werden im Wesentlichen simpliziale Finite–Elemente sowie Finite–Elemente auf Rechtecken beziehungsweise auf Hexaedern. Nachdem die grundlegenden Eigenschaften dieser Finiten–Elemente diskutiert worden sind, werden Hilfsmittel wie Interpolationsfehler–Abschätzungen, inverse Ungleichungen und Spurungleichungen zur Verfügung gestellt.

In **Kapitel 6** betrachten wir die theoretischen Grundlagen zur Konvektions-Diffusionsgleichung. Zunächst wird die Lösbarkeit der Gleichung im klassischen Sinne untersucht, dann wird die Gleichung in ihre schwache Formulierung überführt. Diese entspricht einem Variationsproblem. Anschließend analysieren wir die Lösbarkeit dieses Problems mit Hilfe

der Resultate aus Kapitel 2. Zur Approximation einer Lösung stellen wir im Anschluss das Standard–Galerkin–Verfahren, die Streamline–Diffusion-Methode und die Kanten–Stabilisierung vor und leiten a-priori Abschätzungen für den Verfahrensfehler ab.

In **Kapitel 7** führen wir numerische Rechnungen zur Konvektions-Diffusionsgleichung mit dem Standard–Galerkin–Verfahren, der Streamline–Diffusion-Methode und der Kanten–Stabilisierung durch. Unter anderem überprüfen wir, ob die von der Analysis vorausgesagten Konvergenzraten aus Kapitel 6 durch die Verfahren erreicht werden und welche der beiden Stabilisierungsverfahren gemessen an den Fehlerwerten und der Laufzeit besser abschneidet.

Kapitel 8 verläuft weitgehend analog zu Kapitel 5 mit dem Unterschied, dass die Oseen–Gleichung statt der Konvektions-Diffusionsgleichung betrachtet wird. Als Stabilisierungsverfahren führen wir die klassische residuale Stabilisierung sowie eine erweiterte Version der Kanten–Stabilisierung ein.

Numerische Ergebnisse zur Oseen–Gleichung werden in **Kapitel 9** diskutiert. Dabei werden die residuale und die Kanten–Stabilisierung anhand der Fehlerwerte und Laufzeit miteinander verglichen und die Fehlerordnungen beider Verfahren den Voraussagen aus Kapitel 8 gegenübergestellt. Zudem wird untersucht, wie die Stabilisierungsparameter der Kanten–Stabilisierung optimalerweise gewählt werden sollten. Schließlich werden Ergebnisse für die Kanten–Stabilisierung bei der Verwendung von Taylor–Hood Elementen vorgestellt.

Die Ergebnisse der Arbeit werden in **Kapitel 10** zusammengefasst.

Kapitel 2

Mathematische Grundlagen

In diesem Kapitel werden die mathematischen Grundlagen dieser Arbeit gelegt. Zunächst erinnern wir an die Definition einiger klassischer Funktionenräume. Dann führen wir die Lebesgue- und Sobolevräume ein und geben nützliche Eigenschaften dieser Räume an, wie zum Beispiel Einbettungseingenschaften, Reflexivität und die Behandlung von Randbedingungen. Abschließend nennen wir Ergebnisse der linearen Funktionalanalysis, welche zur Behandlung von Variationsproblemen eingesetzt werden sollen. Dazu gehören die Eigenschaften dualer Operatoren und der Satz vom abgeschlossenen Bild.

2.1 Klassische Funktionenräume

Um Differentialgleichungen im klassischen Sinn behandeln zu können, benötigt man Funktionen die ausreichend oft differenzierbar sind. Die geeigneten Funktionenräume findet man in folgendem Lemma:

Lemma 2.1.
Sei Ω eine beschränkte offene Menge aus $\mathbb{R}^d$, $d \in \mathbb{N}$, $\alpha = (\alpha_1, ..., \alpha_d) \in \mathbb{N}_0^d$ ein Multiindex und $k \in \mathbb{N}_0$. Der Raum der k-mal stetig differenzierbaren Funktionen auf dem Abschluss von Ω sei definiert durch

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

Der Raum $C^k(\bar{\Omega})$ wird mit der Norm

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

zum Banachraum. $D^\alpha v$ steht dabei für die partielle Ableitung von v der Form

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

Beweis. Die Rechnung für den eindimensionalen Fall findet man zum Beispiel in [Wer95] auf Seite 6. Für höhere Dimensionen rechnet man analog. □

Manchmal benötigen wir zusätzlich das Konzept der Hölder–Stetigkeit.

Lemma 2.2.
Seien die Voraussetzungen von Lemma 2.1 gegeben. Der Raum der Hölder-stetigen Funktionen $C^{k,s}(\bar{\Omega})$, $0 < s \leq 1$, sei der Teilraum von $C^k(\bar{\Omega})$ bestehend aus den Funktionen v,

für welche $D^\alpha v$, $0 \leq |\alpha| \leq k$, Hölder-stetig mit Exponent s ist, für welche es also eine Konstante $C > 0$ mit gibt:

$$|D^\alpha v(x) - D^\alpha v(y)| \leq C\,|x-y|^s \qquad \forall x, y \in \Omega.$$

Dann wird $C^{k,s}(\bar{\Omega})$ mit der Norm

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

zum Banachraum. Dabei gilt für $r \geq s$ die Inklusion $C^{k,r}(\Omega) \subset C^{k,s}(\Omega)$.

Beweis. Den Beweis, dass $C^{k,s}(\bar{\Omega})$ ein Banachraum ist, rechnet man wieder ähnlich wie im Beweis zu Lemma 2.1 nach. Zum Beweis der Inklusion siehe [Ada75] Theorem 1.31.

Im Fall $s = 1$ spricht man von Lipschitz–Stetigkeit. Wir werden des Öfteren voraussetzen, dass das Gebiet Ω einen Lipschitz–Rand besitzt.

Definition 2.3. (**Lipschitz–Rand**)
Sei $\Omega \subset \mathbb{R}^d$, $d \in \mathbb{N}$, ein beschränktes Gebiet. Dann sagen wir, Ω besitzt einen Lipschitz–Rand $\partial\Omega$ oder Ω ist ein Lipschitzgebiet, wenn es endlich viele Mengen $w_1, ..., w_n \subset \mathbb{R}^d$ gibt mit:

(1.) die Ω überdeckden,

(2.) $\partial\Omega \cap \omega_i$ ist für alle $i = 1, ..., n$, Graph einer Lipschitz-stetigen Funktion,

(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.

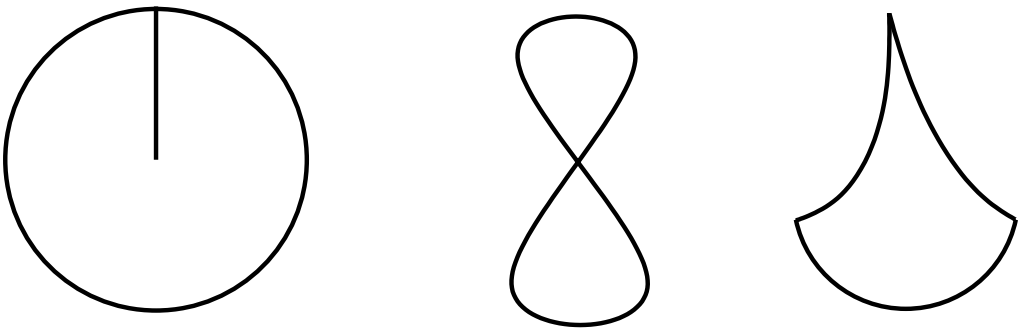

Abbildung 2.1: Beispiele für Gebiete mit *nicht* Lipschitz–stetigem Rand. Der Grund für die fehlende Regularität ist links der Schlitz im Gebietsrand, in der Mitte die Überschneidung und rechts der Scheitel.

2.2 Lebesgue–Räume

Die im letzten Abschnitt vorgestellten klassischen Funktionenräume sind für das Folgende oft zu speziell. Wir führen daher allgemeinere Räume ein, die so genannten Lebesgue–Räume. Die Beweise zu den Aussagen in diesem Kapitel können beispielsweise in [Ada75] Kapitel 2 nachgeschlagen werden. Ausgangspunkt für die Definition von Lebesgue–Räume ist die folgende Definition:

Definition 2.4.
Seien μ das Lebesgue–Maß auf dem $\mathbb{R}^n$, $\Omega \subset \mathbb{R}^n$ eine beschränkte Menge und $f : \Omega \to \mathbb{R}$ eine messbare Abbildung. Zu diesen Vorgaben definieren wir die Halbnormen $\|\cdot\|_{L^p(\Omega)}$ durch

$$\|f\|_{L^p(\Omega)} = \left(\int_\Omega |f|^p \, d\mu\right)^{1/p}, \quad 1 \le p < \infty,$$
$$\|f\|_{L^\infty(\Omega)} = \inf_{\mu(\mathcal{N})=0} \sup_{x \in \Omega \setminus \mathcal{N}} |f(x)|$$

und mit den Halbnormen für $1 \le p \le \infty$ die Räume

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

Es handelt sich dabei um Halbnormen, da auch Funktionen, welche nicht identisch Null sind, durch $\|\cdot\|_{L^p(\Omega)}$ auf Null abgebildet werden, wie zum Beispiel die Funktion $f : [0,1] \to \mathbb{R}$ mit

$$f(x) = \begin{cases} 1, & \text{falls } x = 0.5, \\ 0, & \text{sonst.} \end{cases}$$

Um aus den $\mathcal{L}^p$-Räumen die Lebesgue–Räume $L^p(\Omega)$ zu erhalten, unterteilt man die $\mathcal{L}^p$-Räume in Äquivalenzklassen, indem alle Funktionen zusammengefasst werden, die bezüglich des Lebesgue–Maßes bis auf eine Nullmenge gleich sind:

Definition 2.5. Lebesgue–Räume
Sei

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

Dann sind die Lebesgue–Räume definiert durch den Faktorraum

$$L^p(\Omega) = \mathcal{L}^p(\Omega)/\mathcal{N}^p. \tag{2.1}$$

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, denn 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 Funktion nie der gleichen Äquivalenzklasse angehören.

Das folgende Lemma fasst einige Eigenschaften der L^p-Räume zusammen:

Lemma 2.6. Sei $1 \le p, q \le \infty$.
(i) Für $f, g \in L^p(\Omega)$ gelten $f + g \in L^p(\Omega)$ und die *Minkowski–Ungleichung*:

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

Die Minkowski–Ungleichung ist gerade die Dreiecksungleichung für die L^p-Normen.

(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:

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

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

(iv) Die L^p-Räume sind vollständig, das heißt, sie sind Banachräume.

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

$$\begin{aligned} &(L^p(\Omega))' = L^q \quad \text{für } 1 < p, q < \infty,\ \frac{1}{p} + \frac{1}{q} = 1, \\ &(L^1(\Omega))' = L^\infty, \\ &(L^\infty(\Omega))' \neq L^1. \end{aligned} \tag{2.2}$$

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

Eine besondere Rolle spielt in dieser Arbeit der Raum $L^2(\Omega)$. Durch

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

ist dort ein Skalarprodukt definiert. Damit ist $L^2(\Omega)$ ein Hilbertraum. Das L^2-Skalarprodukt kürzen wir im Folgenden mit $(\cdot, \cdot)$ ab. Insbesondere für den nächsten Abschnitt benötigen wir einen noch etwas allgemeineren Funktionenraum:

Definition 2.7.
Der Raum der lokal integrierbaren Funktionen ist definiert durch

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

2.3 Sobolev–Räume

Für unsere Zwecke sind die L^p-Räume oftmals noch zu allgemein. Wir wollen in Räumen arbeiten, in denen der Ableitungsbegriff zumindest in einem schwachen Sinne definiert werden kann. Dazu führen wir die Sobolev–Räume ein. Falls nicht anders erwähnt, findet man die Beweise zu den Aussagen dieses Abschnitts in [Ada75] Kapitel 3 und 5. Die folgende Definition erklärt, was eine Ableitung im schwachen Sinne ist.

Definition 2.8. (Schwache Ableitung)
Sei Ω ein beschränktes Gebiet in $\mathbb{R}^d$, $d \in \mathbb{N}$ und $f \in L^1_{\text{loc}}(\Omega)$ und α ein Multiindex. Falls eine Funktion $g \in L^1_{\text{loc}}$ existiert, so dass

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

für alle v aus C_0^∞ gilt, so heißt g die schwache Ableitung von f der Ordnung α.

Dabei bezeichnet D die klassische Ableitung und $C_0^\infty(\Omega)$ den Raum aller beliebig oft differenzierbaren Funktionen mit kompaktem Träger in Ω. Die Definition ist so gewählt, dass eine im klassischen Sinne differenzierbare Funktion auch schwach ableitbar ist und dass ihre schwache Ableitung fast überall mit ihrer klassischen übereinstimmt. Analog zur klassischen Ableitung schreiben wir $D^\alpha f$ für die schwache Ableitung von f der Ordnung α. Mit dem gerade eingeführten Konzept der schwachen Ableitung können wir die Sobolev–Räume definieren.

Definition 2.9. (Sobolev–Räume)
Seien $k \in \mathbb{N}_0$ und $1 \leq p \leq \infty$ und Ω ein beschränktes Gebiet in $\mathbb{R}^d$, $d \in \mathbb{N}$. Die Sobolev–Räume sind definiert durch

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

Mit den Normen

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

werden die Sobolev–Räume zu Banachräumen. Zudem sind die Sobolev–Räume mit $p = 2$ sogar Hilberträume. Das zugehörige Skalarprodukt $(\cdot,\cdot)_{W^{k,2}(\Omega)} : W^{k,2}(\Omega) \times W^{k,2}(\Omega) \to \mathbb{R}$ ist definiert durch

$$(u, v)_{W^{k,2}(\Omega)} = \sum_{|\alpha| \leq k} (D^\alpha u, D^\alpha v). \tag{2.3}$$

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

Neben obiger Definition der Sobolev–Räume findet man in der Literatur manchmal eine andere. Dort wird $W^{k,p}(\Omega)$ durch den Abschluss von $C^\infty(\Omega)$ bezüglich der $\|\cdot\|_{W^{k,p}(\Omega)}$-Norm definiert. Beide Definitionen sind unabhängig vom Gebiet Ω äquivalent zueinander.

Zusätzlich zu den $\|\cdot\|_{W^{k,p}(\Omega)}$-Normen werden wir des Öfteren von folgenden Halbnormen Gebrauch machen:

$$|f|_{W^{k,p}(\Omega)} = \left(\sum_{|\alpha| = k} \left(\|D^\alpha f\|_{L^p(\Omega)} \right)^p \right)^{1/p} \qquad \text{für } 1 \leq p < \infty,$$
$$|f|_{W^{k,\infty}(\Omega)} = \sum_{|\alpha| = k} \|D^\alpha f\|_{L^\infty(\Omega)}. \tag{2.4}$$

Als nächstes untersuchen wir, wie die Sobolev–Räume untereinander und mit den klassischen Funktionenräume zusammenhängen. Der Zusammenhang besteht dabei im Sinne von Einbettungen.

Definition 2.10. (**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 $I : X \to Y,\ Iv = v$ für alle $v \in X$ ist stetig.
Eine Einbettung deuten wir durch $X \hookrightarrow Y$ an.

Für Einbettungen von Sobolev–Räumen untereinander kann diese Definition direkt verwandt werden. Insbesondere ist hierbei $W^{k,p}(\Omega) \hookrightarrow W^{l,q}(\Omega)$ äquivalent zu $W^{k,p}(\Omega) \subset W^{l,q}(\Omega)$. Was wir hingegen unter einer Einbettung in die klassischen Funktionenräume $C^l(\Omega)$ verstehen, muss noch geklärt werden, da es sich bei den $W^{k,p}$–Räumen streng genommen nicht um Funktionen, sondern um Äquivalenzklassen von Funktionen handelt. 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.

Satz 2.11. (**Sobolevsche Einbettungstheoreme**)
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:

Fall 1) Für $mp < d$:

$$W^{j+m,p}(\Omega) \hookrightarrow W^{j,q}(\Omega), \qquad \text{falls zusätzlich } q \leq \frac{dp}{d-mp}.$$

Fall 2) 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 $q = \infty$.

Fall 3) Für $mp > d$:

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

Dieses Resultat wird für unsere Zwecke genügen. Verallgemeinerungen bezüglich des Gebiets findet man wieder in [Ada75]. Wichtige Spezialfälle der Sobolevschen Einbettungstheoreme sind beispielsweise die Aussagen $H^2(\Omega) \subset C(\Omega)$ für $d = 2, 3$ und $W^{1,1}(\Omega) \subset L^2(\Omega)$ für $d = 2$.
Es sei außerdem erwähnt, dass manche Aussagen der Sobolevschen Einbettungstheoreme nicht verstärkt werden können: Ist die Ungleichung $q \leq dp/(d-mp)$ bei Fall 1 verletzt, so kann sogar für beliebige Gebiete $\Omega \subset \mathbb{R}^d$ gezeigt werden, dass keine Einbettung mehr existiert (siehe Beispiel 5.25 [Ada75]). Ebenfalls keine Einbettung existiert im Fall 2, wenn $p > 1$ und $q = \infty$ gilt (siehe Beispiel 5.26 [Ada75]).
Ein weiterer Zusammenhang zwischen den Sobolev– und den klassischen Funktionenräumen besteht in folgender Dichtheitsaussage:

Satz 2.12.
Sei Ω ein Gebiet mit Lipschitz–Rand und $C_0^k(\Omega)$ der Raum der k-mal stetig differenzierbaren Funktionen mit kompaktem Träger in Ω. Falls $k \geq l$ ist, liegt $C_0^k(\Omega)$ dicht in $H^l(\Omega)$.

Beweis. Siehe [Wlo82] Satz 3.6. □

In dieser Arbeit werden wir Randwertprobleme in den Sobolev–Räumen behandeln. Allerdings ist der Rand $\partial\Omega$ eines Gebiets $\Omega \subset \mathbb{R}^d$ bezüglich des d–dimensionalen Lebesgue–Maßes eine Nullmenge, so dass derselben Äquivalenzklasse in $W^{k,p}(\Omega)$ Funktionen mit beliebigen Randwerten angehören. Daher macht es keinen Sinn, Randbedingungen im klassischen Sinne zu fordern. Stattdessen verlangen wir, dass die Randbedingungen zumindest in einem schwachen Sinne erfüllt sind. Wir diskutieren zunächst den Fall homogener Dirichlet–Randbedingungen

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

Im schwachen Sinne soll dann gelten, dass u in $W_0^{k,p}(\Omega)$ liegt:

Definition 2.13.
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 $H_0^k(\Omega)$.

Vieles von der Struktur der Sobolev–Räume überträgt sich auf die so definierten Räume. Insbesondere ist $W_0^{k,p}(\Omega)$ ein Banachraum und $H_0^k(\Omega)$ ein Hilbertraum.
Im Fall inhomogener Dirichlet–Randbedingungen

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

wählt man einen anderen Zugang, welcher durch den folgenden Satz ermöglicht wird:

Satz 2.14. (**Spuroperator**)
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 $\gamma : W^{k,p}(\Omega) \to L^2(\partial\Omega)$, der die Randwerte von Funktionen u aus $W^{1,p} \cap C(\bar{\Omega})$ unverändert lässt:

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

γ wird Spuroperator genannt.

Beweis. Siehe [Wlo82] Satz 8.7.

Mit Hilfe des Spuroperators definiert man nun die Randwerte einer Sobolev–Funktion über das folgende Korollar.

Korollar 2.15. (**Spur einer Sobolev–Funktion**)
Es gelten die Voraussetzungen von Satz 2.14 und es sei u eine Funktion aus $W^{k,p}(\Omega)$, $k \geq 1$. Dann wähle eine Funktionenfolge $\{u_n\}_{n=1}^\infty$ aus $C(\bar{\Omega})$ mit Grenzwert u und definiere mit Hilfe des Spuroperators

$$\lim_{n\to\infty} (\gamma u_n) = \gamma u.$$

Über diesen Grenzwert wird der Funktion u eindeutig die Spur γu auf $\partial\Omega$ zugeordnet.

Beweis. Gemäß Satz (2.12) gibt es mindestens eine Folge aus $C(\bar{\Omega})$ mit Grenzwert u. Damit sichert die Existenz des Spuroperators γ zugleich die Existenz der Spur. Die Eindeutigkeit der Spur ergibt sich aus der Stetigkeit und Eindeutigkeit des Spuroperators. Dazu betrachte das Kriterium der Folgenstetigkeit. □

Insbesondere ist obige Konstruktion kompatibel mit Definition 2.13 für homogene Dirichlet–Randwerte, denn es gilt:

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

Für $k \geq 2$ kann man zusätzlich eine Aussage über die Randwerte der Ableitungen von $W_0^{k,p}(\Omega)$-Funktionen machen:

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

Bemerkung 2.16.
Für diese Bemerkung gelten die gleichen Voraussetzungen wie in Satz 2.14. In Satz 2.14 haben wir das Bild des Spuroperators als $L^2(\partial\Omega)$ angegeben. Nach dem Beweis in [Wlo82] besitzt die Randfunktion sogar größere Regularität. Eine Funktion aus $W^{k,p}(\Omega)$ wird nämlich vom Spuroperator auf den Funktionenraum $H^{k-1/2}(\partial\Omega)$ abgebildet. Der Raum $H^{k-1/2}(\partial\Omega)$ mit nicht ganzahligem Exponenten $k - 1/2$ ist ein Unterraum von $H^{k-1}(\partial\Omega)$

(siehe [Wlo82] Satz 6.1). Die Sobolev–Räume mit nicht ganzahligem Exponenten werden als Sobolev–Slobodecki–Räume bezeichnet. Die Definition der Sobolev–Slobodecki–Räume lautet

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

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

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

Gemäß [Wlo82] Satz 3.1 ist $H^{k+s}(\Omega)$ ein Hilbertraum. Für weitere Resultate zu den Sobolev–Slobodecki–Räume siehe [Wlo82] oder beispielsweise [Tre75].

Durch die Räume $W_0^{k,p}(\Omega)$ lassen sich Sobolev–Räume eines neuen Typs motivieren, nämlich Sobolev–Räume mit negativem Exponenten k.

Definition 2.17.
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

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

Den Zusammenhang zwischen $W^{-k,q}(\Omega)$ und $W_0^{k,p}(\Omega)$ gibt folgendes Lemma an.

Lemma 2.18.
Seien Ω ein beschränktes Gebiet und $1 < p, q < \infty$ mit $1/p + 1/q = 1$. Dann ist $W^{-k,q}(\Omega)$ der Dualraum von $W_0^{k,p}(\Omega)$.

Auch vektor- und tensorwertige Sobolev–Funktionen werden in dieser Arbeit vorkommen. Diese Funktionen ergeben sich auf natürliche Weise durch Produktbildung von skalaren Sobolev–Funktionen. Die Struktur der Sobolev–Räume bleibt dabei erhalten. So sind die Räume $[W^{k,p}(\Omega)]^n$ sowie $[W^{k,p}(\Omega)]^{n\times n}$, $n \in \mathbb{N}$, weiterhin Banachräume und die Räume $[H^k(\Omega)]^n$ sowie $[H^k(\Omega)]^{n\times n}$ weiterhin Hilberträume. Die zugehörigen Normen und Skalarprodukte sind komponentenweise definiert durch:

$$\|u\|_{[W^{k,p}(\Omega)]^n} := \sum_{i=1}^n \|u_i\|_{W^{k,p}(\Omega)}, \qquad \|u\|_{[W^{k,p}(\Omega)]^{n\times n}} := \sum_{i,j=1}^n \|u_{i,j}\|_{W^{k,p}(\Omega)},$$

$$(u,v)_{[H^k(\Omega)]^n} := \sum_{i=1}^n (u_i, v_i)_{H^k(\Omega)}, \qquad (u,v)_{[H^k(\Omega)]^{n\times n}} := \sum_{i,j=1}^n (u_{i,j}, v_{i,j})_{H^k(\Omega)}.$$

Im Folgenden schreiben wir abkürzend $\|u\|_{W^{k,p}(\Omega)}$ statt $\|u\|_{[W^{k,p}(\Omega)]^n}$ oder $\|u\|_{[W^{k,p}(\Omega)]^{n\times n}}$ und $(u,v)_{H^k(\Omega)}$ statt $(u,v)_{[H^k(\Omega)]^n}$ oder $(u,v)_{[H^k(\Omega)]^{n\times n}}$. Obige Aussagen gelten insbesondere auch für vektor- oder tensorwertige Lebesgue–Räume.

Für spätere Zwecke benötigen wir noch die folgende Ungleichung:

Lemma 2.19. (**Friedrichs–Ungleichung**)
Seien $\Omega \subset \mathbb{R}^d$, $d \in \mathbb{N}$, ein beschränktes Gebiet und $1 \leq p \leq \infty$. Dann gilt für alle $u \in W_0^{k,p}(\Omega)$

$$\|u\|_{W^{k,p}(\Omega)} \leq C_F \, |u|_{W^{k,p}(\Omega)} \tag{2.5}$$

mit einer Konstanten $C_F > 0$ die nur von p, d und dem Durchmesser von Ω abhängt. C_F wird auch als Friedrichs–Konstante bezeichnet.

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

2.4 Hilfsmittel der Funktionalanalysis

Insbesondere für das nächste Kapitel, in welchem die Lösbarkeit von Variationsproblemen untersucht wird, benötigen wir einige Hilfsmittel der linearen Funktionalanalysis. Sei V ein normierter Vektorraum. Den Dualraum von V bezeichnen wir mit V' und das duale Produkt zwischen $v \in V$ und $v' \in V'$ mit

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

Seien Q ein weiterer normierter Vektorraum. Zur Unterscheidung versehen wir im Folgenden die Normen und die dualen Produkte mit dem Index V oder Q. Zudem schreiben wir $\mathcal{L}(V, Q)$ für den Raum aller linearen, stetigen Operatoren, die von V nach Q abbilden. Hauptresultat dieses Abschnitts ist der Satz vom abgeschlossenen Bild. Für einen Operator $T \in \mathcal{L}(V, Q)$ macht dieser Satz unter anderem eine Aussage über den dualen Operator T':

Lemma 2.20. (**Dualer Operator**)
Seien V, Q normierte Räume. Zu einem Operator $T \in \mathcal{L}(V, Q)$ gibt es genau einen Operator $T' : \; Q' \to V'$ mit

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

Den Operator $T' : \; Q' \to V'$ nennen wir den zu T dualen Operator. Der duale Operator ist stetig und linear. Zuweilen bezeichnet man T' auch als den zu T adjungierten Operator oder kurz als die Adjungierte.

Beweis. Siehe [Yos65] Seite 193 bis 195 oder [Alt85] Seite 262. □

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

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

Mit $\ker(T)$ bezeichnen wir den Kern des Operators T und mit $R(T)$ seinen Bildbereich. Dann lautet der Satz vom abgeschlossenen Bild:

Satz 2.21. (Satz vom abgeschlossenen Bild)
Seien V, 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:
(1.) $R(T)$ ist abgeschlossen in Q.
(2.) $R(T')$ ist abgeschlossen in V'.
(3.) $R(T) = \ker(T')^\circ = \{q \in Q : \; \langle q', q \rangle = 0 \text{ für alle } q' \in \ker(T')\}$.
(4.) $R(T') = \ker(T)^\circ = \{v \in V' : \; \langle v', v \rangle = 0 \text{ für alle } v \in \ker(T)\}$.

Beweis. Siehe [Yos65] von Seite 205 bis 207. □

Im nächsten Kapitel wird der Satz vom abgeschlossenen Bild benutzt, um die Lösbarkeit von Variationsproblemen nachzuweisen. Die im Satz aufgeführten Mengen $\ker(T')^\circ$ und $\ker(T)^\circ$ bezeichnet man auch als Annihilatoren von $\ker(T')$ und $\ker(T)$. Neben dem Satz vom abgeschlossenen Bild benötigen wir noch folgende Eigenschaft des dualen Operators:

Lemma 2.22.
Seien V, Q Banachräume und $T \in \mathcal{L}(V, Q)$. Dann existiert die Inverse $T^{-1} \in \mathcal{L}(Q, V)$ 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. □

Der Operator T ist in diesem Abschnitt als Element von $\mathcal{L}(V, Q)$ vorausgesetzt worden. Auf dem Raum $\mathcal{L}(V, Q)$ gibt es eine Norm, die so genannte Operatornorm.

Lemma 2.23. (**Operatornorm**)
Seien V, Q normierte Vektorräume. 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 wird Operatornorm genannt. Mit der Operatornorm wird $\mathcal{L}(V, Q)$ zum normierten Raum. Ist Q zusätzlich ein Banachraum, so auch $\mathcal{L}(V, Q)$.

Beweis. Siehe [Wer95] Seite 47. □

Einige weitere nützliche Eigenschaften der Operatornorm sind:

Lemma 2.24. (**Operatornorm**)
Seien V, Q normierte Vektorräume sowie $T \in \mathcal{L}(V, Q)$ und $S \in \mathcal{L}(Q, V)$.

(i) Die Operatornorm ist submultiplikativ, das heißt, es gilt

$$\|TS\| \leq \|T\|\|S\|.$$

(ii) Der Operator T und sein dualer Operator T' besitzen die gleiche Operatornorm, sprich $\|T\| = \|T'\|$.

Beweis. Für Teil (i) siehe [Wer95] Seite 48 und für Teil (ii) [Yos65] Seite 195. □

Auch einem Skalarprodukt können wir eine Art Operatornorm zuordnen:

Lemma 2.25.
Seien V, Q normierte Räume und $a(\cdot, \cdot) : V \times Q \to \mathbb{R}$ 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. Diese Norm bezeichnen wir als die Norm von a.

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

Die Beschränktheit von a ist äquivalent zur Aussage $\|a\| = C < \infty$. Für die Bilinearform a werden wir auch folgendes Resultat verwenden:

Lemma 2.26.
Seien V, Q normierte Räume und $a(\cdot,\cdot) : V \times Q \to \mathbb{R}$ eine stetige Bilinearform. Dann gibt es genau einen stetigen linearen Operator $A : V \to Q'$, so dass für alle $v \in V$ und alle $q \in Q$ gilt:

$$a(v, q) = \langle Av, q\rangle_Q. \tag{2.6}$$

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

Beweis. Man definiert $A : V \to Q'$, indem man das Bild Av vermöge der Abbildungsvorschrift $(Av)(q) = a(v, q)$ spezifiziert. Damit ist A wohldefiniert sowie wegen der Voraussetzungen an a stetig und linear. Eine andere Wahl von A widerspricht der Forderung (2.6), das heißt, A ist eindeutig. □

Ein weiteres Hilfsmittel wird der Projektionssatz sein.

Satz 2.27. (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

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

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

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

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

Beweis. Siehe [Yos65] Seite 82 und 83. □

Abschließend sei noch an eine Aussage der klassischen Analysis erinnert:

Lemma 2.28. (**Young–Ungleichung**)
Seien $a, b \in \mathbb{R}$. Dann gilt für alle $\delta > 0$ die Young–Ungleichung:

$$ab \le \frac{1}{2\delta}a^2 + \frac{\delta}{2}b^2.$$

Beweis. Elementares Nachrechnen. □

Kapitel 3

Variationsprobleme in Hilberträumen

In dieser Arbeit treten Variationsprobleme in Hilberträumen bei der schwachen Formulierung von partiellen Differentialgleichungen auf. Zunächst werden wir den Variationsproblem–Typ vorstellen, welcher bei der Behandlung von Konvektions–Diffusionsgleichungen auftritt. Dann diskutieren wir gemischte Variationsprobleme, wie sie bei der schwachen Formulierung der Oseen–Gleichung auftreten. Zudem werden Sätze zur Verfügung gestellt, mit welchen die eindeutige Lösbarkeit der Variationsprobleme gesichert werden kann.

3.1 Variationsprobleme

Seien V ein reeller Hilbertraum mit dem Skalarprodukt $(\cdot,\cdot)_V$ sowie der induzierten Norm $\|\cdot\|_V = (\cdot,\cdot)_V^{1/2}$ und $f \in V'$. Sei ferner $a : V \times V \to \mathbb{R}$ eine stetige und bezüglich der Norm $\|\cdot\|_V$ strikt koerzitive Bilinearform, das heißt, es gibt zwei Konstanten $C, \widetilde{C} > 0$ mit:

$$\begin{aligned} a(v,v) &\geq C\|v\|_V^2 && \forall v \in V \quad \text{(strikte Koerzitivität)},\\ a(u,v) &\leq \widetilde{C}\|u\|_V\|v\|_V && \forall u,v \in V \quad \text{(Stetigkeit)}. \end{aligned}$$

Das folgende Problem wird als Variationsproblem bezeichnet: Finde ein $u \in V$, so dass für alle $v \in V$ gilt:

$$a(u,v) = f(v). \tag{3.1}$$

Falls a zusätzlich symmetrisch ist, folgt die eindeutige Lösbarkeit dieses Variationsproblems aus dem Rieszschen Darstellungssatz:

Satz 3.1. (Rieszscher Darstellungssatz)
Sei V ein Hilbertraum und a ein Skalarprodukt auf V. Zu jedem beschränkten linearen Funktional $w' \in V'$ gibt es ein eindeutig bestimmtes $u \in V$ mit

$$a(u,v) = \langle w', v\rangle \qquad \forall v \in V.$$

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

Beweis. Siehe [Yos65] Seite 90. □

Bemerkung 3.2.
(1.) Aus funktionalanalytischer Sicht besagt der Rieszschen Darstellungssatz, dass die Elemente eines Hilbertraums eindeutig mit Elementen aus dem Dualraum identifiziert werden können.

(2.) Oft tritt obiges Variationsproblem in der folgenden äquivalenten Formulierung auf: Finde u aus V, so dass gilt:

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

bezüglich des Funktionals

$$J(v) := \frac{1}{2}\|v\|_V^2 - f(v).$$

Den Beweis zur Äquivalenz beider Probleme findet man beispielsweise in [Cia78], Theorem 1.1.2.

Der Rieszsche Darstellungssatz ist für unsere Zwecke oft zu speziell. Wir benötigen eine Verallgemeinerung für unsymmetrische Bilinearformen. Beschränktheit und strikte Koerzitivität bezüglich $\|\cdot\|_V$ behalten wir hingegen bei. Das zugehörige Variationsproblem ist dann immer noch eindeutig lösbar, wie der Satz von Lax–Milgram zeigt:

Satz 3.3. (Satz von Lax–Milgram)
Sei $a(\cdot,\cdot) : V \times V \to \mathbb{R}$ eine beschränkte Bilinearform und sei a strikt koerzitiv bezüglich $\|\cdot\|_V$. Zu jedem beschränkten linearen Funktional $v' \in V'$ gibt es ein eindeutig bestimmtes $u \in V$ mit

$$a(u,v) = \langle v', v\rangle \qquad \forall v \in V.$$

Beweis. Der Beweis erfolgt mit Hilfe des Rieszschen Darstellungssatzes, siehe [Yos65] Seite 92. □

Mit Hilfe des Satzes von Lax–Milgram werden wir die eindeutige Lösbarkeit der Konvektions–Diffusionsgleichung untersuchen.

Bemerkung 3.4.
Eine äquivalente Aussage des Satzes ist, dass es einen eindeutig bestimmten linearen, stetigen Isomorphismus A von V nach V' gibt. Dies ist stärker als die Aussage von Lemma 2.26, wo auf die Koerzitivität von a verzichtet worden ist. Aus Lemma 2.26 folgt nur die eindeutige Existenz eines stetigen, linearen Operators A von $V \to Q'$, nicht jedoch dessen Injektivität oder Surjektivität.

3.2 Gemischte Variationsprobleme

Im letzten Abschnitt ist die Bilinearform a immer über dem Produktraum $V \times V$ definiert worden. Bei gemischten Variationsproblemen ist a hingegen über dem Produktraum $V \times Q$ von V mit einem weiteren Hilbertraum Q gegeben. Das verallgemeinerte Variationsproblem lautet hierzu:

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

$$a(u,q) = f(q). \tag{3.2}$$

Dabei ist wieder $f \in V'$. Sei A der gemäß Lemma 2.26 zur Bilinearform a zugeordnete Operator. Dann ist die eindeutige Lösbarkeit des obigen Variationsproblems äquivalent zu der Aussage, dass A ein Isomorphismus ist. Mit Hilfe des nächsten Satzes werden wir ein hinreichendes Kriterium für die Isomorphie von A angeben:

Satz 3.5.
Seien V, Q reelle Hilberträume, $a(\cdot,\cdot) : V \times Q \to \mathbb{R}$ eine stetige Bilinearform, $A \in \mathcal{L}(V, Q')$ der Darstellungsoperator von a und I der Identitätsoperator auf Q'. Für alle $C > 0$ sind folgende drei Aussagen äquivalent:

(i) $\sup_{0\neq v\in V} \frac{a(v,q)}{\|v\|_V} \geq C\|q\|_Q \qquad \forall q \in Q.$

(ii) $\|A'q\| \geq C\|q\|_Q \qquad \forall q \in Q.$

(iii) Es gibt $S \in \mathcal{L}(Q', V)$, so dass $AS = I$ auf Q' und $\|S\| \leq C^{-1}$, wobei I für die Identität steht.

Beweis. Der Beweis orientiert sich an [Bre74] Theorem 0.1.
Die Äquivalenz von (i) und (ii) folgt aus der Gleichheit der linken Seiten:

$$C\|q\|_Q \leq \sup_{0\neq v\in V} \frac{a(v,q)}{\|v\|_V} = \sup_{0\neq v\in V} \frac{\langle Av, q\rangle_Q}{\|v\|_V} = \sup_{0\neq v\in V} \frac{\langle A'q, v\rangle_V}{\|v\|_V} = \|A'q\|. \tag{3.3}$$

Die erste Gleichung ergibt sich aus der Definition des Darstellungsoperators von a (siehe Lemma 2.26), die zweite aus der Definition des dualen Operators (siehe Lemma 2.20) und die dritte aus der Definition der Operatornorm (siehe Lemma 2.23).
Für die Implikation (iii)⇒ (ii) schätzen wir mit Hilfe des Rieszsche Darstellungsoperator R_Q auf Q ab:

$$\begin{aligned}\sup_{0\neq v\in V} \frac{a(v,q)}{\|v\|_V} &\geq \frac{a(SR_Q^{-1}q, q)}{\|SR_Q^{-1}q\|_V} = \frac{\langle ASR_{Q^{-1}}q, q\rangle_Q}{\|SR_Q^{-1}q\|_V} \\ &= \frac{\langle R_Q^{-1}q, q\rangle_Q}{\|SR_Q^{-1}q\|_V} = \frac{(q,q)_Q}{\|SR_Q^{-1}q\|_V} \\ &= \frac{\|q\|_Q^2}{\|SR_Q^{-1}q\|_V}.\end{aligned} \tag{3.4}$$

Der Reihe nach folgen die Relationen aus der der Tatsache, dass $SR_Q^{-1}q$ für alle $q \in Q$ Element von V ist, der Definition des Darstellungsoperators von a, der Voraussetzung $AS = I$, dem Rieszschen Darstellungssatz (Satz 3.1) sowie dem Zusammenhang zwischen Skalarprodukt und induzierter Norm. Als nächstes schätzen wir den Nenner von (3.4) nach oben ab:

$$\|SR_Q^{-1}q\|_V \leq \|S\|\|R_Q^{-1}q\|_{Q'} \leq \|S\|\|q\|_Q \leq k^{-1}\|q\|_Q. \tag{3.5}$$

Die erste Ungleichung ergibt sich, da für einen linearen stetigen Operator $T : V \to Q$ die Ungleichung $\|Tv\|_Q \leq \|T\|\|v\|_V$ gilt (betrachte die Definition der Operatornorm). Die zweite Ungleichung folgt wegen der Isometrie des Rieszschen Darstellungsoperators (siehe Satz 3.1) und die letzte wegen der Voraussetzung $\|S\| \leq C^{-1}$. Kombiniert man nun die Abschätzungen (3.4) und(3.5), so folgt die gewünschte Aussage:

$$\|A'q\| \geq C\|q\|_Q \qquad \forall q \in Q.$$

Als nächstes beweisen wir (ii) ⇒ (iii). Wir definieren A_E als die Einschränkung von A auf das orthogonale Komplement von $\ker(A)$, welches gegeben ist durch

$$(\ker A)^\perp := \{v \in V : (v, \zeta) = 0 \text{ falls } A\zeta = 0\}. \tag{3.6}$$

Nach Konstruktion ist $A_E(v) = 0 \Leftrightarrow v = 0$, also ist A_E injektiv. Nun zeigen wir mit Hilfe des Satzes vom abgeschlossenen Bild, dass A_E surjektiv ist. Daraus folgt als Zwischenresultat, dass A_E ein Isomorphismus von $(\ker A)^\perp$ auf Q' ist (wegen der Linearität und Injektivität von A_E). Der Satz vom abgeschlossenen Bild soll in Richtung (2.) nach (3.) verwandt werden. Wir wollen also zeigen, dass $R(A'_E)$ eine in V' abgeschlossene Menge ist. Betrachte hierzu eine Cauchy–Folge $\{v'_n\}_{n=1}^\infty$ aus $R(A'_E)$ mit Grenzwert v'_0. Mit q_n kürzen wir das Urbild des n-ten Folgengliedes ab, also $A'_E(q_n) = v'_n$. Aus (ii) folgt nun

$$\|q_n - q_m\|_Q \leq \frac{1}{C}\|v'_n - v'_m\| \qquad \forall n, m \in \mathbb{N}.$$

Damit ist auch $\{q_n\}_{n=1}^\infty$ eine Cauchy–Folge. Ihren Grenzwert bezeichnen wir mit q_0. Aufgrund der Stetigkeit von A'_E haben wir $v'_n = A'_E(q_n) \to A'_E(q_0)$ beim Übergang $n \to \infty$. Dies bedeutet $v'_0 = A'_E(q_0) \in R(A'_E)$. Mit anderen Worten: Jede Cauchy–Folge aus $R(A'_E)$ besitzt einen Grenzwert in dieser Menge. Damit ist $R(A'_E)$ eine in V' abgeschlossene Menge. Nun verwenden wir den Satz vom abgeschlossenen Bild in der Richtung (2.) nach (3.) und erhalten $R(A_E) = \ker(A'_E)^\circ$. Gemäß (ii) haben wir $A'_E(q) = 0 \Leftrightarrow q = 0$, also $\ker(A'_E) = \{0\}$. Daraus folgt wegen $R(A_E) = \ker(A'_E)^\circ = Q'$ wie gewünscht die Surjektivität von A_E. Wir wissen jetzt, dass A_E ein Isomorphismus ist. Seine Inverse sei A_E^{-1}. Gemäß Konstruktion gilt für die Inverse $AA_E^{-1} = I$ auf Q'. Mit der Wahl $S := A_E^{-1} = I$ folgt die erste Aussage von (iii).
Bleibt noch, die Ungleichung $\|A_E^{-1}\| \leq 1/C$ zu zeigen. Mit A_E besitzt wegen Lemma 2.22 auch der duale Operator A'_E eine Inverse $(A'_E)^{-1}$. Zu einem beliebigen $v' \in R(A'_E) \setminus \{0\}$ sei q das Urbild, also $q = (A'_E)^{-1}v'$. Dies in (ii) eingesetzt, ergibt

$$\|A'_E(A'_E)^{-1}v'\| \geq C\|(A'_E)^{-1}v'\| \Longleftrightarrow \frac{1}{C} \geq \frac{\|(A'_E)^{-1}v'\|_Q}{\|v'\|}.$$

Da v' beliebig aus $R(A_{E'}) \setminus \{0\}$ gewählt war, bleibt die Ungleichung beim Übergang zum Supremum bestehen:

$$\sup_{0 \neq v' \in R(A_{E'})} \frac{\|(A'_E)^{-1}v'\|_Q}{\|v'\|} \leq \frac{1}{C}.$$

Nach Definition der Operatornorm steht auf der linken Seite $\|(A'_E)^{-1}\|$. Es folgt:

$$\|(A'_E)^{-1}\| = \|(A_E^{-1})'\| = \|A_E^{-1}\| \leq \frac{1}{C}.$$

Die erste Gleichheit ergibt sich aus der Tatsache, dass für einen Isomorphismus wegen Lemma 2.22 $(A'_E)^{-1} = (A_E^{-1})'$ gilt, die zweite aus Lemma 2.24. Damit ist auch die zweite Aussage von (iii) gezeigt. □

Für endlichdimensionale Hilberträume V, Q gleicher Dimension folgt aus Satz 3.5 bereits, dass A ein Isomorphismus ist. Der allgemeinere Fall wird durch folgendes Korollar abgedeckt:

Korollar 3.6.
Seien die Voraussetzungen von Satz 3.5 erfüllt, dass heißt, seien V, Q reelle Hilberträume, $a(\cdot,\cdot) : V \times Q \to \mathbb{R}$ eine stetige Bilinearform und $A \in \mathcal{L}(V, Q')$ der Darstellungsoperator von a . Für alle $C > 0$ und $\widetilde{C} > 0$ sind dann folgende drei Aussagen äquivalent:

(i) $\sup_{0\neq v\in V} \frac{a(v,q)}{\|v\|_V} \geq C\|q\|_Q \quad \forall q \in Q$ und $\sup_{0\neq q\in Q} \frac{a(v,q)}{\|q\|_Q} \geq \widetilde{C}\|v\|_V \quad \forall v \in V.$

(ii) $\|A'q\| \geq C\|q\|_Q \quad \forall q \in Q$ und $\quad \|Av\| \geq \widetilde{C}\|v\|_V \quad \forall v \in V.$

(iii) A ist ein Isomorphismus von V auf Q' mit $\|A^{-1}\| \leq C^{-1}$ und $\|(A')^{-1}\| \leq \widetilde{C}^{-1}$.

Beweis. Die Aussage ergibt sich, indem man den Satz 3.5 einmal auf die Bilinearform a anwendet und einmal auf die Bilinearform $\widetilde{a} : Q \times V$ mit $\widetilde{a}(q,v) := a(v,q)$. □

Die Aussage in Richtung von (i) nach (iii) lässt sich nochmals verstärken:

Satz 3.7.
Seien die Voraussetzungen von Satz 3.5 erfüllt und a genüge der Bedingung

$$\sup_{0\neq q\in Q} \frac{a(v,q)}{\|v\|_V\|q\|_Q} \leq C\|v\|_V \qquad \forall v \in V. \tag{3.7}$$

Ferner gelte:

$$\text{Für alle } q \in Q \setminus \{0\} \quad \exists v \in V \quad \text{mit } a(v,q) \neq 0. \tag{3.8}$$

Dann ist der Darstellungsoperator A von a ein Isomorphismus von V nach Q'.

Beweis. Siehe [BA72] Theorem 5.2.1. □

Bemerkung 3.8.
(1.) Die Bedingung aus Satz 3.7 an a wird auch als Babuška–Brezzi–Bedingung bezeichnet. Sie ist offenbar äquivalent zur Aussage

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

Aufgrund dieser Äquivalenz nennt man die Babuška–Brezzi–Bedingung manchmal auch inf-sup Bedingung.

(2.) Auch im Spezialfall $V = Q$ ist der obige Satz allgemeiner als der Satz von Lax–Milgram: Falls a strikt koerzitiv ist, erfüllt a die Babuška–Brezzi–Bedingung sowie Bedingung (3.8) mit der Wahl $v = q$.

Bei der Behandlung der Oseen–Gleichung wird noch ein anderer Typ eines gemischten Variationsproblems auftreten. Seien V,Q wieder reelle Hilberträume, $a : V \times V \to \mathbb{R}$ sowie $b : V \times Q \to \mathbb{R}$ stetige Bilinearformen und $f \in V'$, $g \in Q'$. Das folgende Problem wird ebenfalls als gemischtes Variationsproblem bezeichnet: Finde $(u,p) \in V \times Q$, so dass für alle $(v,q) \in V \times Q$ gilt:

$$\begin{aligned} a(u,v) + b(v,p) &= f(v) \\ b(u,q) &= g(q). \end{aligned} \tag{3.9}$$

Dieses gemischte Variationsproblem kann im Produktraum $\Sigma := V \times Q$ in ein Variationsproblem vom Typ (3.1) überführt werden: Da (3.9) für alle $(v,q) \in V \times Q$ zu gelten hat, können dort beide Gleichungen zusammengefasst werden zu:

$$a(u,v) + b(v,p) + b(u,q) = f(v) + g(q). \tag{3.10}$$

Zu dieser Gleichung definieren wir $F \in V' \times Q'$

$$F(u,p) := f(v) + g(q) \tag{3.11}$$

und durch

$$d((u,p),(v,q)) := a(u,v) + b(v,p) + b(u,q) \tag{3.12}$$

eine Bilinearform auf $V \times Q$. Sei außerdem $U := (u,p)$ und $V := (v,q)$. Die zu (3.9) äquivalente Formulierung vom Typ (3.1) lautet damit: Finde $U \in \Sigma$, so dass für alle V aus Σ gilt:

$$d(U,V) = F(V). \tag{3.13}$$

Prinzipiell kann also das gemischte Variationsproblem (3.9) wie das bereits diskutierte Variationsproblem (3.1) oder wie das gemischte Variationsproblem (3.2) behandelt werden. Das Problem hierbei ist allerdings, dass der Bilinearform d manchmal die nötigen Eigenschaften fehlen, um die obigen Resultate verwenden zu können. Daher werden wir noch einen anderen Zugang beschreiben. Es soll ein hinreichendes Kriterium für die eindeutige Lösbarkeit von (3.9) formuliert werden, welches noch schwächere Voraussetzungen an a und b stellt. Seien A, B die Darstellungsoperatoren von a und b und sei der Operator $\Lambda : V \times Q \to V' \times Q'$ definiert durch

$$\Lambda(v,q) := (Av + B'q, Bq). \tag{3.14}$$

Dann ist die eindeutige Lösbarkeit von (3.9) äquivalent zur Aussage, dass Λ ein Isomorphismus ist.

Satz 3.9. Sei Z der Kern des Operators B, Z' der zugehörige Dualraum und π die orthogonale Projektion von V' auf Z'. Der dem gemischten Variationsproblem (3.9) zugeordnete Operator Λ ist genau dann ein Isomorphismus von $V \times Q$ auf $V' \times Q'$, wenn die folgenden beiden Bedingungen erfüllt sind:

(i) πA ist ein Isomorphismus von Z auf Z'.

(ii) Es gibt $C > 0$, so dass gilt

$$\|B'q\| \geq C\|q\|_Q \qquad \text{für alle } q \in Q.$$

Beweis. Der Beweis folgt mit Hilfe von Satz 3.5 und dem Satz vom abgeschlossenen Bild gemäß Theorem 1.1 von [Bre74].

Der obige Satz ist recht abstrakt. Für unsere Zwecke genügt eine abgeschwächte Variante:

Satz 3.10. Seien a, b wie oben und es gebe zusätzlich $\delta, \gamma > 0$, so dass gilt:

$$\forall v \in Z: \quad a(v,v) \geq \delta\|v\|_V \qquad \text{(Z-Elliptizität)},$$

$$\inf_{0\neq q\in Q} \sup_{0\neq v\in V} \frac{b(v,q)}{\|v\|_V \|q\|_Q} \geq \gamma \qquad \text{(Babuška–Brezzi–Bedingung)},$$

wobei $Z := \{v \in V : \ b(v,q) = 0 \quad \forall q \in Q\}$. Dann besitzt das gemischte Variationsproblem (3.9) eine eindeutige Lösung.

Beweis. Die Babuška–Brezzi–Bedingung ist wegen der Aussage von Satz 3.5 in Richtung (i) nach (ii) äquivalent zu Teil (ii) von Satz 3.9. Um Satz 3.9 anwenden zu können, muss also nur noch gezeigt werden, dass πA ein Isomorphismus von Z nach Z' ist. Zunächst vergewissern wir uns, dass der Projektionsoperator π existiert und zu den linear stetigen Abbildungen von V' nach Z' gehört. Da B ein stetiger Operator ist, ist Z als Kern von B ein abgeschlossener Unterraum von V. Damit ist auch Z' ein abgeschlossener Unterraum

von V' und wir können den Projektionssatz 2.27 anwenden. Dieser liefert die Existenz des Projektionsoperators $\pi : V' \to Z'$ und belegt insbesondere dessen Stetigkeit und Linearität. Damit ist πA als Verkettung zweier linear stetiger Operatoren ebenfalls linear und stetig. Als nächstes zeigen wir, dass πA surjektiv ist. Sei dazu $f \in Z'$ beliebig vorgegeben. Die Bilinearform a ist als Z-elliptisch vorausgesetzt, was der strikten Koerzitivität von a auf Z entspricht. Der Satz von Lax–Milgram ergibt die eindeutige Existenz von $u \in Z$ mit

$$a(u, v) = \langle f, v \rangle \qquad \forall v \in Z,$$

was äquivalent ist zu

$$(\pi A)(u) = f.$$

Da $f \in Z'$ beliebig vorgeben war, impliziert diese Aussage die Surjektivität von πA auf Z'. Bleibt noch die Injektivität von πA zu zeigen. Da πA linear und stetig ist, genügt es für alle $u \in Z$, die Aussage $(\pi A)(u) = 0 \Rightarrow u = 0$ zu beweisen. Sei also u aus Z mit $(\pi A)(u) = 0$. Nach Definition von π gilt

$$0 = \langle (\pi A)(u), v \rangle = \langle Au, v \rangle \qquad \forall v \in Z,$$

also insbesondere

$$0 = \langle Au, u \rangle = a(u, u).$$

Aus der Z-Elliptizität folgt daraus wie gewünscht $u = 0$. Insgesamt ist πA also ein Isomorphismus und Satz 3.9 ergibt die eindeutige Lösbarkeit von (3.9) . $\square$

Den letzten Satz werden wir anwenden, wenn die Lösbarkeit der Oseen–Gleichungen untersucht wird.

Kapitel 4

Galerkin–Verfahren

Analytisch können die Variationsprobleme aus dem letzten Abschnitt oft nicht oder nur mit großem Aufwand gelöst werden. Daher nutzt man numerische Näherungsverfahren. In dieser Arbeit verwenden wir dazu Finite–Elemente–Methoden, welche auf Galerkin–Verfahren basieren. Als nächstes werden wir klären, wie Galerkin–Verfahren auf lineare Gleichungssysteme führen und wann es eine eindeutige Lösung gibt. Außerdem wird untersucht, wie gut die aus einem Galerkin–Verfahren bestimmte Näherungslösung die Lösung des Variationsproblems approximiert.

4.1 Galerkin–Verfahren für Variationsprobleme

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

$$a(u, v) = f(v). \tag{4.1}$$

Ist $a : V \times V \to \mathbb{R}$ eine stetige und bezüglich der Norm $\|\cdot\|_V$ strikt koerzitive Bilinearform, so besitzt obiges Variationsproblem eine eindeutige Lösung. Dies ist die Aussage des Satz von Lax–Milgram (Satz 3.3).
In dieser Arbeit werden wir die Lösung von Variationsproblemen des obigen Typs in einem endlichdimensionalen Teilraum $V_h \subset V$ mit Hilfe des Galerkin–Verfahrens nähern. Eine Näherungslösung erhält man dabei aus der analogen Formulierung von (4.1) auf dem Raum V_h, nämlich: Finde $u_h \in V_h$, so dass:

$$a(u_h, v_h) = f(v_h) \quad \forall v_h \in V_h. \tag{4.2}$$

Auch das Galerkin–Verfahren besitzt nach dem Satz von Lax–Milgram eine eindeutige Lösung. Die Voraussetzungen sind erfüllt, da die Koerzitivität von a auch im Unterraum V_h gilt und jeder endliche Unterraum eines Hilbertraums selbst ein Hilbertraum ist.
Die Lösung des Galerkin–Verfahrens führt auf ein endlichdimensionales lineares Gleichungssystem. Um ein lineares Gleichungssystem aufzustellen, wählen wir eine beliebige Basis $\{\phi_i\}_{i=1}^N$ von V_h aus.
Bezüglich dieser Basis kann (4.2) wegen der Linearität von a und f umformuliert werden zu: Finde $u_h \in V_h$, so dass:

$$a(u_h, \phi_i) = f(\phi_i) \quad \text{für } i = 1, ..., N. \tag{4.3}$$

Demnach genügt es, die vermeintliche Lösung gegen die Basisfunktionen ϕ_i zu testen statt gegen alle Funktionen aus V_h.

Wir stellen nun die gesuchte Lösung u_h als Linearkombination der Basisfunktionen mit den noch unbekannten Koeffizienten u_j dar

$$u_h = \sum_{j=1}^{N} u_j \phi_j. \tag{4.4}$$

Einsetzen in (4.3) ergibt

$$a\left(\sum_{j=1}^{N} u_j \phi_j, \phi_i\right) = f(\phi_i) \quad \text{für } i = 1, ..., N \tag{4.5}$$

oder bei Ausnutzung der Linearität von a im ersten Argument

$$\sum_{j=1}^{N} u_j a(\phi_j, \phi_i) = f(\phi_i) \quad \text{für } i = 1, ..., N. \tag{4.6}$$

Identifizieren wir

$$\mathbf{A} := \{a(\phi_j, \phi_i)\}_{i,j=1}^{N}, \quad \mathbf{u} := (u_i)_{i=1}^{N} \quad \text{und} \quad \mathbf{f} := (f(\phi_i))_{i=1}^{N}, \tag{4.7}$$

so ergibt sich als äquivalente Formulierung zu (4.6)

$$\mathbf{A}\mathbf{u} = \mathbf{f}. \tag{4.8}$$

Aus diesem linearen Gleichungssystem können die unbekannten Koeffizienten u_j bestimmt werden, woraus die Lösung u_h folgt. Bei der Lösung kann die positive Definitheit von $\mathbf{A}$ nützlich sein, welche aus der strikten Koerzitivität von a folgt.
Nachdem nun klar ist, wie eine Näherungslösung u_h bestimmt werden kann, fragt sich, wie gut u_h die Lösung u des ursprünglichen Variationsproblems (4.1) approximiert. Genauer kann diese Frage erst dann geklärt werden, wenn der Hilbertraum V und der endliche Teilraum V_h spezifiziert werden, was in Kapitel 5 geschehen wird. Aber auch schon ohne eine konkrete Wahl ist die folgende Aussage möglich (zunächst unter stärkeren Voraussetzungen an a):

Lemma 4.1.
Die Bilinearform a sei stetig, symmetrisch und strikt koerzitiv bezüglich der von a induzierten Energienorm $\|\cdot\|_a = a(\cdot,\cdot)^{1/2}$. Sei ferner u die Lösung von (4.1) und u_h die zugehörige Näherungslösung berechnet aus (4.2). Dann gilt:

$$\|u - u_h\|_a = \min_{v_h \in V_h} \|u - v_h\|_a. \tag{4.9}$$

Beweis. Wähle in (4.1) und (4.2) mit v_h eine beliebige Testfunktion aus V_h. Subtrahiert man (4.2) von (4.1), so ergibt sich:

$$a(u - u_h, v_h) = 0 \quad \forall v_h \in V_h. \tag{4.10}$$

Es folgt für alle $v_h \in V_h$

$$\|u - u_h\|_a^2 = a(u - u_h, u - u_h) = a(u - u_h, u - v_h) \leq \|u - u_h\|_a \|u - v_h\|_a, \tag{4.11}$$

wobei sich die zweite Identität aus (4.10) ergibt und die Abschätzung nach oben aus der Cauchy–Schwarz–Ungleichung. Division durch $\|u - u_h\|_a$ und Übergang zum Infimum vervollständigt den Beweis. □

Bemerkung 4.2.
Der im Beweis auftretende Zusammenhang

$$a(u - u_h, v_h) = 0 \quad \forall v_h \in V_h \tag{4.12}$$

wird als Galerkin–Orthogonalität bezeichnet. Er bleibt auch dann gültig, wenn auf die Symmetrie von a verzichtet wird. Auch für weitere Approximations–Sätze stellt die Galerkin–Orthogonalität eine nützliche Eigenschaft des Galerkin–Verfahrens dar.

Obiges Lemma besagt, dass u_h die bestmögliche Approximation an u in V_h bzgl. der von a induzierten Norm darstellt. Allerdings benötigt man dazu die Symmetrie von a, was oft eine zu starke Annahme ist. Lassen wir die Forderung nach Symmetrie fallen, so gilt folgendes Resultat:

Lemma 4.3. (Lemma von Cea)
Die Bilinearform a sei stetig und strikt koerzitiv bezüglich $\|\cdot\|_V$. Sei zudem u die Lösung von (4.1) und u_h diejenige von (4.2). Dann gilt

$$\|u - u_h\|_V \leq C \min_{v_h \in V_h} \|u - v_h\|_V \tag{4.13}$$

mit $C > 0$.

Beweis. Ähnlich zum Beweis von Lemma 4.1 können wir für alle v_h aus V_h schreiben:

$$\delta\|u - u_h\|_V^2 \leq a(u - u_h, u - u_h) = a(u - u_h, u - v_h) \leq c\|u - u_h\|_V\|u - v_h\|_V. \tag{4.14}$$

Der Reihe nach wurde die strikte Koerzitivität von a mit der Koerzitivitäts–Konstanten δ genutzt, die Galerkin–Orthogonalität und die Stetigkeit von a. Die Behauptung folgt nach Division durch $\|u - u_h\|_V$, Übergang zum Infimum und der Definition $C := c/\delta$. $\square$

Gemäß dem Lemma von Cea wird die Lösung u in jeder zu $\|\cdot\|_V$ äquivalenten Norm durch die Näherungslösung u_h quasi–optimal approximiert, das heißt, bestmöglich bis auf Multiplikation mit einer Konstanten.

4.2 Galerkin–Verfahren für gemischte Variationsprobleme

Seien V,Q Hilberträume, $a : V \times V \to \mathbb{R}$ sowie $b : V \times Q \to \mathbb{R}$ stetige Bilinearformen und $f \in V'$, $g \in Q'$. Hierzu sei das folgende gemischte Variationsproblem gegeben: Finde $(u, p) \in V \times Q$, so dass für alle $(v, q) \in V \times Q$ gilt:

$$\begin{aligned} a(u, v) + b(v, p) &= f(v) \\ b(u, q) &= g(q). \end{aligned} \tag{4.15}$$

Falls b der Babuška–Brezzi–Bedingung genügt und a Z-elliptisch ist mit

$$Z := \{v \in V : \ b(v, q) = 0 \quad \forall q \in Q\}, \tag{4.16}$$

erhalten wir die eindeutige Lösbarkeit des obigen Variationsproblems aus Satz 3.10.
Auch gemischte Variationsprobleme werden in dieser Arbeit mit Galerkin–Verfahren gelöst, welche in diesem Zusammenhang als gemischte Galerkin–Verfahren bezeichnet werden. Dabei wird die Lösung von (4.15) in einem endlichdimensionalen Teilraum $V_h \times Q_h \subset$

$V \times Q$ approximiert. Analog zu (4.2) lautet das gemischte Galerkin–Verfahren: Finde ein $(u_h, p_h) \in V_h \times Q_h$, so dass für alle (v_h, q_h) aus $V_h \times Q_h$ gilt:

$$\begin{aligned} a(u_h, v_h) + b(v_h, p_h) &= f(v_h) \\ b(u_h, q_h) &= g(q_h). \end{aligned} \tag{4.17}$$

Für die eindeutige Lösbarkeit des gemischten Galerkin–Verfahrens gilt die folgende hinreichende Bedingung:

Lemma 4.4.
Seien a,b wie oben und $\delta, \gamma > 0$, so dass gelten:

$$\forall v_h \in Z_h : \quad a(v_h, v_h) \geq \delta \|v_h\|_{V_h} \qquad (Z_h\text{-Elliptizität}), \tag{4.18}$$

$$\inf_{0 \neq q_h \in Q_h} \sup_{0 \neq v_h \in V_h} \frac{b(v_h, q_h)}{\|v_h\|_{V_h} \|q_h\|_{Q_h}} \geq \gamma \qquad (\text{diskrete Babuška–Brezzi–Bedingung}), \tag{4.19}$$

wobei $Z_h := \{v_h \in V_h : \ b(v_h, q_h) = 0 \quad \forall q_h \in Q_h\}$. Dann besitzt das gemischte Galerkin Verfahren (4.17) eine eindeutige Lösung.

Beweis. Siehe [Bre74] Korollar 2.1.

Die Voraussetzungen des Lemmas sind so gewählt, dass der Beweis von Satz 3.10 übernommen werden kann. In der Tat ist es notwendig die Voraussetzungen zu verlangen, ungeachtet dessen, ob die Annahmen von Satz 3.10 gelten. So folgt die Z_h–Elliptizität von a nicht aus der Z–Elliptizität von a, da Z_h im Allgemeinen Elemente enthält, welche nicht in Z liegen. Außerdem kann die diskrete Babuška–Brezzi–Bedingung verletzt sein, obwohl das ursprüngliche Problem (4.15) der Babuška–Brezzi–Bedingung genügt. Die diskrete Babuška–Brezzi–Bedingung muss eigens durch eine geeignete Wahl von V_h und Q_h gesichert werden.
Die Voraussetzungen des letzten Lemmas führen nicht nur auf die eindeutige Existenz einer Lösung des gemischten Galerkin–Verfahrens, sondern auch auf die Quasi–Optimalität dieser Lösung.

Lemma 4.5.
Unter denselben Annahmen wie in Lemma 4.4 gilt:

$$\|u - u_h\| + \|p - p_h\| \leq C \left(\inf_{v_h \in V_h} \|u - v_h\|_V + \inf_{q_h \in Q_h} \|p - q_h\|_Q \right). \tag{4.20}$$

Beweis. Siehe [Bre74] Korollar 2.1.

Die Lösung des gemischten Galerkin–Verfahrens werden wir aus einem endlichdimensionalen linearen Gleichungssystem bestimmen. Das Gleichungssystem kann ähnlich wie bei dem zuvor behandelten Galerkin–Verfahren aufgestellt werden. Seien $\{\phi_i\}_{i=1}^N$ und $\{\psi_i\}_{i=1}^M$

Basen von V_h und Q_h. Wegen der Linearität von a und b ist das gemischte Galerkin–Verfahren äquivalent zu:
Finde $u_h = \sum_{j=1}^{N} u_j \phi_j \in V_h$ und $p_h = \sum_{j=1}^{M} p_j \psi_j \in Q_h$, so dass

$$\sum_{j=1}^{N} u_j a(\phi_j, \phi_i) + \sum_{j=1}^{M} p_j b(\psi_j, \phi_i) = f(\phi_i) \quad \text{für } i = 1, ..., N, \tag{4.21}$$

$$\sum_{j=1}^{N} u_j b(\phi_j, \psi_i) = g(\psi_i) \quad \text{für } i = 1, ..., M. \tag{4.22}$$

Das Gleichungssystem (4.22) lässt sich mit den Definitionen

$$\mathbf{A} := \{a(\phi_j, \phi_i)\}_{i,j=1}^{N}, \quad \mathbf{B} := \{b(\phi_j, \psi_i)\}_{i=1,...,M}^{j=1,...,N}, \tag{4.23}$$

$$\mathbf{u} := (u_i)_{i=1}^{N}, \quad \mathbf{p} := (p_i)_{i=1}^{M}, \quad \mathbf{f} := (f(\phi_i))_{i=1}^{N} \quad \text{und} \quad \mathbf{g} := (g(\psi_i))_{i=1}^{M} \tag{4.24}$$

in die folgende algebraische Form bringen:

$$\begin{pmatrix} \mathbf{A} & \mathbf{B}^T \\ \mathbf{B} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{p} \end{pmatrix} = \begin{pmatrix} \mathbf{f} \\ \mathbf{g} \end{pmatrix}. \tag{4.25}$$

Kapitel 5

Finite–Elemente

Wie im letzten Kapitel angedeutet, werden wir die Lösung von Variationsproblemen mit Galerkin–Verfahren annähern. Bei Konvektions–Diffusionsgleichungen wird die Näherungslösung in einem endlichdimensionalen Hilbertraum V_h berechnet und bei der Oseen–Gleichung in einem endlichdimensionalen Produktraum $V_h \times Q_h$. Als nächstes führen wir die für diese Arbeit relevanten Räume V_h und $V_h \times Q_h$ ein. Bei der Konstruktion dieser Räume folgen wir dem Finite–Element–Konzept, das zunächst vorgestellt und motiviert wird. Anschließend führen wir spezielle Finite–Elemente auf Dreiecken und Rechtecken ein und diskutieren, wie gut die exakte Lösung in diesen Räumen genähert werden kann.

5.1 Motivation und Grundidee

Wie wir im letzten Kapitel gesehen haben, wird bei Galerkin–Verfahren ein lineares Gleichungssystem mittels Basisfunktionen des Raums V_h aufgestellt. Sowohl der Aufwand zum Aufstellen des Gleichungssystems als auch seine Lösung werden durch den Raum V_h sowie durch die dort gewählte Basis bestimmt. Um den Aufwand gering zu halten, unterteilt man bei Finite–Elementen das Gebiet Ω, auf dem das Galerkin–Verfahren angewendet werden soll, in endlich viele Teilgebiete $K_1, ..., K_n$. Die Basisfunktionen werden dann so konstruiert, dass sie auf fast allen Teilgebieten verschwinden. Dies hat den folgenden Vorteil: Zum Aufstellen des lineares Gleichungssystem müssen Skalarprodukte zwischen den Basisfunktionen berechnet werden, die sich in der Regel über das gesamte Gebiet erstrecken. Verschwinden die Basisfunktionen aber auf fast allen K_i, so verschwinden die Skalarprodukte ebenfalls fast überall und müssen folglich nur noch in wenigen K_i ausgewertet werden.
Die Wahl der Basisfunktionen erfolgt mit Hilfe von Funktionalen $\Phi_i : V_h \to \mathbb{R}$. Die Funktionale werden als linear unabhängig voneinander, linear und stetig vorausgesetzt. Weiterhin verlangen wir , dass die Funktionale Φ_i unisolvent bezüglich V_h sind.

Definition 5.1. (Unisolvenz)
Sei N die Dimension von V_h. Die Funktionale $\Phi_1, ..., \Phi_N$ werden unisolvent bezüglich des Raums V_h genannt, falls es zu jedem $\mathbf{a} = (a_1, ..., a_1)^T \in \mathbb{R}^N$ genau ein $u \in V_h$ gibt mit

$$\Phi_i(u) = a_i \quad \text{für } i = 1, ..., N. \tag{5.1}$$

Wählt man für die Vektoren $\mathbf{a}$ die Standard-Einheitsvektoren des $\mathbb{R}^N$, so folgt aus der Unisolvenz der Funktionale, dass es einen eindeutig bestimmten Satz $\{\phi_j\}_{j=1}^N$ von Funktionen aus V_h gibt mit

$$\Phi_i(\phi_j) = \delta_{ij} \qquad \text{für alle } 1 \leq i, j \leq N. \tag{5.2}$$

δ_{ij} ist dabei die Diracsche Deltadistribution, definiert durch

$$\delta_{ij} = \begin{cases} 1, & \text{falls } i = j, \\ 0, & \text{sonst.} \end{cases}$$

Die ϕ_j spannen nach Konstruktion eine Basis von V_h auf. Um welche Basis es sich handelt, wird durch die Funktionale Φ_i entschieden.
Inzwischen haben wir alle nötigen Begriffe zur formalen Definition eines Finite–Elements eingeführt.

Definition 5.2. (Finites–Element)
Das Tripel (Ω, V_h, Σ) wird als Finites–Element bezeichnet. Dabei ist

- Ω ein nichtleeres, abgeschlossenes Gebiet des $\mathbb{R}^d$ $(d \geq 1)$.
- V_h ein N-dimensionaler Raum von Funktionen mit Definitionsgebiet Ω und Bildbereich in $\mathbb{R}$. V_h wird in diesem Zusammenhang Finite–Elemente–Raum genannt.
- $\Sigma = \{\Phi_i\}_{i=1}^N$ ein Satz linear unabhängiger, linearer, stetiger und bezüglich V_h unisolventer Funktionale.

Obige Definition für ein Finites–Element ist recht allgemein und soll nun spezifiziert werden. Die wohl beliebteste Wahl bei Finite–Elementen ist, den Raum V_h aus Funktionen aufzubauen, deren Einschränkung auf die Teilgebiete K_i Polynome sind. Polynomräume sind besonders günstig, da sich Polynome leicht integrieren und ableiten lassen, was die Berechnung der oben erwähnten Skalarprodukte vereinfacht. Außerdem kann eine geeignete Basis mit recht wenig Aufwand konstruiert werden und die Matrix des zu lösenden Gleichungssystems wird nur schwach besetzt sein. Konkret werden wir Beispiele für Polynomräume in den beiden folgenden Abschnitten einführen, zunächst auf Dreieckselementen, dann auf Rechtecken. Neben Polynomräumen werden manchmal auch Räume von stückweise rationalen [SFH07] oder stückweise trigonometrische Funktionen [BBO02] verwendet. Allerdings beschränken wir uns in dieser Arbeit auf Polynomräume.
Gebräuchliche Funktionale sind zum Beispiel Punktauswertungen $\Phi(u) = u(\mathbf{x})$, Auswertungen der ersten Ableitung $\Phi(u) = \partial_i u(\mathbf{x})$ oder Integralmittelwerte über einzelne Zellen der Gebietszerlegung $\Phi(u) = \int_K u(\mathbf{x})$. In dieser Arbeit werden die Funktionale immer Punktauswertungen sein. Die zugehörigen Finite–Elemente werden in diesem Fall als Lagrange–Elemente bezeichnet.

5.2 Finite–Elemente auf Dreiecken

Sei Ω ein polygonal berandetes Gebiet in $\mathbb{R}^2$, welches mit einer zulässige Zerlegung $\mathcal{T}$ aus Dreiecken trianguliert ist.

Definition 5.3. (Zulässig Zerlegung)
Eine Zerlegung $\mathcal{T} = \{K_1, ..., K_M\}$ von $\Omega \subset \mathbb{R}^2$ mit abgeschlossenen Gitterzellen K_i heißt zulässig genau dann, wenn gilt:

(i) Die Zerlegung überdeckt das ganze Gebiet, das heißt $\Omega = \cup_{i=1}^M K_i$.

(i) 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.

Insbesondere sind Zerlegungen unzulässig, die, wie in Abbildung 5.1 gezeigt, einen hängenden Knoten beinhalten (roter Punkt). Obige Defintion lässt sich unmittelbar auf $\Omega \subset \mathbb{R}^3$ erweitern, wenn man neben Kanten bei Punkt (iii) auch Seitenflächen zulässt.

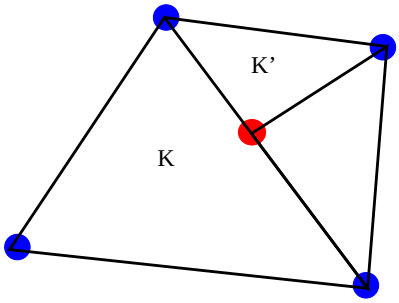

Abbildung 5.1: Unzulässig Triangulierung mit einem hängendem Knoten (roter Punkt). Die Schnittmenge der Zellen K und K' besteht aus mehr als einem Punkt. Aufgrund des hängenden Knotens teilen sich beide Zellen dennoch keine ganze Kante (das heißt Punkt (iii) aus Definition 5.3 ist verletzt).

Die zulässige Zerlegung $\mathcal{T}$ stellt den Ausgangspunkt für nachfolgende Definition der Finite–Element–Räume auf Dreiecken dar:

$$P^p := \left\{ v \in \mathcal{C}(\bar{\Omega}) :\ v|_K \in \mathbb{P}_p(K)\ \forall K \in \mathcal{T} \right\}, \quad p \geq 1, \tag{5.3}$$

wobei $\mathbb{P}^p(K)$ den Raum der Polynome mit Grad kleiner gleich p bezeichnet:

$$\mathbb{P}^p(K) := \left\{ p : K \to \mathbb{R} :\ p(\mathbf{x}) = \sum_{|\alpha| \leq p} \mu_\alpha \mathbf{x}^\alpha \right\}, \qquad \mu_\alpha \in \mathbb{R},\ \alpha = \text{Multiindex}.$$

Die Finite–Element-Räume P^p bestehen also aus auf $\bar{\Omega}$ stetigen Funktionen, welche eingeschränkt auf eine einzelne Dreieckszelle Polynome sind.
Bisher haben wir das Gebiet Ω klassifiziert und die Finite–Element–Räume P^p vorgestellt. Um ein vollständiges Finites–Element gemäß Definition 5.2 zu erhalten, fehlen noch die zugehörigen Sätze unisolventer Funktionale $\{\Phi_i\}_{i=1}^N$. Als Funktionale wählen wir Punktauswertungen. Ausgewertet wird jeweils in einem der Punkte, die für jede Dreieckszelle gemäß dem in Abbildung 5.2 gezeigten Schema angeordnet sind. Also zum Beispiel für die stückweise linearen Finite–Elemente P^1 in einem Eckpunkt oder für die stückweise quadratischen Finite–Elemente P^2 in einem Eck- oder Mittelpunkt der Kanten. Die Punkte, in denen ausgewertet wird, bezeichnen wir als Knoten.

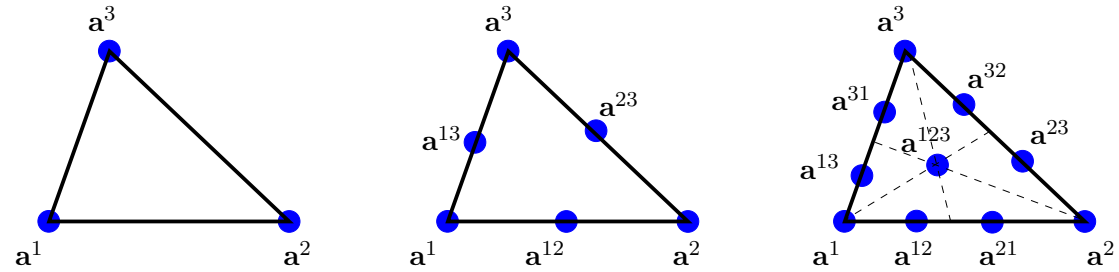

Abbildung 5.2: Verteilung der Knoten bei den P^p-Elementen. Links für $p = 1$, in der Mitte für $p = 2$ und rechts für $p = 3$.

Wie im letzten Abschnitt erwähnt, kann über die Funktionale eine Basis des zugrundeliegenden Finite–Element–Raums abgeleitet werden. Dies werden wir nun für die Räume

P^1, P^2 und P^3 tun. Der gesuchte Satz von Basisfunktionen $\{\phi_j\}_{j=1}^N$ hat bezüglich den Funktionalen Φ_i der Bedingung

$$\Phi_i(\phi_j) = \delta_{ij} \qquad \text{für alle } 1 \leq i, j \leq N \tag{5.4}$$

zu genügen. Wenn wir diese Basis angeben und zudem ihre Eindeutigkeit zeigen, beweisen wir auch die Unisolvenz der Funktionale. Die zugehörigen Basisfunktionen werden auch Standard–Knoten–Basisfunktionen genannt.
Besonders elegant lassen sich die Basisfunktionen in den so genannten baryzentrischen Koordinaten darstellen. Die baryzentrischen Koordinaten sind lokal auf jedem Dreieck K definiert.

Definition 5.4. (baryzentrische Koordinaten)
Seien $\mathbf{a}^1, \mathbf{a}^2, \mathbf{a}^3$ die Eckpunkte des Dreiecks K. Als baryzentrische Koordinaten des Punktes $\mathbf{x} = (x_1, x_2) \in K$ bezeichnen wir die Koeffizienten $\lambda_1, \lambda_2, \lambda_3 \geq 0$ in der Darstellung

$$x_j = \sum_{i=1}^{3} \lambda_i a_j^i, \quad 1 \leq j \leq 2, \quad \text{mit } \sum_{i=1}^{3} \lambda_i = 1. \tag{5.5}$$

Jeder Punkt aus dem Dreieck K besitzt eine Darstellung in den baryzentrischen Koordinaten, da K der konvexen Hülle seiner Eckpunkte $\mathbf{a}^j$ entspricht. Zudem ist die Darstellung eindeutig, was sofort aus der paarweisen linearen Unabhängigkeit der Vektoren $\mathbf{a}^1 - \mathbf{a}^2$, $\mathbf{a}^1 - \mathbf{a}^3$, $\mathbf{a}^2 - \mathbf{a}^3$ gefolgert werden kann.
Statt eine Basisfunktion ϕ direkt auf ganz Ω anzugeben, stellen wir ihre Einschränkung auf jedem Dreieck der Zerlegung in den baryzentrischen Koordinaten dar. Dazu unterscheiden wir zunächst für jedes Dreieck K zwei Fälle. Gemäß (5.4) ist eine Basisfunktion nur in genau einem Knoten ungleich 0. Den zugehörigen Knoten nennen wir 1-Knoten. Im ersten Fall liege der 1-Knoten nicht in K. Durch Lösen des zugehörigen Gleichungssystem folgt dann, dass die Einschränkung $\phi|_K$ der Basisfunktion der Nullfunktion auf K entspricht.[1]
In dem anderen Fall liege der 1-Knoten in K. In diesem Fall ist $\phi|_K$ verschieden von der Nullfunktion und lautet in den baryzentrischen Koordinaten wie folgt:

- Für die stetigen, stückweise linearen Finite–Elemente P^1:

$$\phi|_K(\lambda) = \zeta_i(\lambda) = \lambda_i, \tag{5.6}$$

  wenn der 1-Knoten die Ecke $\mathbf{a}^i$ ist.

- Für die stetigen, stückweise quadratischen Finite–Elemente P^2:

$$\phi|_K(\lambda) = \zeta_i(\lambda) = \lambda_i(2\lambda_i - 1), \tag{5.7}$$

  wenn der 1-Knoten die Ecke $\mathbf{a}^i$ ist und

$$\phi|_K(\lambda) = \zeta_{ij}(\lambda) = 4\lambda_i\lambda_j, \tag{5.8}$$

  wenn der 1-Knoten der Mittelpunkt $\mathbf{a}^{ij}$ der Kante mit den Eckpunkten $\mathbf{a}^i$ und $\mathbf{a}^j$ ist.

- Für die stetigen, stückweise kubischen Finite–Elemente P^3:

$$\phi|_K(\lambda) = \zeta_i(\lambda) = \frac{1}{2}\lambda_i(3\lambda_i - 1)(3\lambda_i - 2), \tag{5.9}$$

[1] Dies impliziert die gewünschte Eigenschaft der Basisfunktionen, auf fast allen Zellen der Zerlegung zu verschwinden.

wenn der 1-Knoten die Ecke $\mathbf{a}^i$ ist,

$$\phi|_K(\lambda) = \zeta_{ij}(\lambda) = \frac{9}{2}\lambda_i\lambda_j(3\lambda_i - 1), \tag{5.10}$$

wenn der 1-Knoten im Punkt $\mathbf{a}^{ij}$ liegt (siehe Abbildung 5.2) und

$$\phi|_K(\lambda) = \zeta_{ijk}(\lambda) = 27\lambda_i\lambda_j\lambda_k, \tag{5.11}$$

wenn der 1-Knoten im Schwerpunkt $\mathbf{a}^{123}$ des Dreiecks liegt.

Durch direktes Nachrechnen kann man sich vergewissern, dass es keine andere Darstellung für $\phi|_K$ gibt. Damit haben wir auf jeder Zelle eine eindeutige Darstellung für die Einschränkung einer Basisfunktion gefunden und somit auch für die Basisfunktionen selbst.

Bemerkung 5.5.
(i) Die Basisfunktionen sind wie gewünscht stetig, insbesondere an den Kanten der Dreiecke. Dies wird durch die Vorgabe der Funktionswerte in den Kantenknoten gewährleistet.

(ii) Die Funktionen ζ spannen auch lokal auf der jeweiligen Zelle eine Basis auf und werden daher lokale Basisfunktionen genannt. Zur Unterscheidung bezeichnet man die Basisfunktionen ϕ gelegentlich als globale Basisfunktionen.

Bei der numerischen Implementierung von Finite–Elementen hat sich der Gebrauch von parametrischen Elementen als effizient erwiesen. Auch das in dieser Arbeit verwandte Programmpaket MooNMD benutzt parametrische Finite–Elemente. Die Idee ist dabei wie folgt: Zunächst werden die lokalen Basisfunktionen auf einer Referenzzelle $\hat{K}$ berechnet. Bei Dreieckselementen wählt man per Konvention das Dreieck mit den Kanten $(0,0)$, $(1,0)$ und $(0,1)$. Nun wählt man eine so genannte Referenztransformation $F_K : \hat{K} \to K$, welche die Referenzzelle $\hat{K}$ auf die Zelle K abbildet. Mit dieser Referenztransformation konstruiert man die lokalen Basisfunktionen $\{\phi\}$ auf den Zellen K der Gebietszerlegung aus den lokalen Basisfunktionen $\{\hat{\phi}\}$ auf der Referenzzelle durch

$$\phi(\mathbf{x}) = \hat{\phi} \circ F_K^{-1}(\mathbf{x}). \tag{5.12}$$

Wie oben gesehen, setzen sich aus diesen lokalen Basisfunktion die globalen zusammen. Der Vorteil von parametrischen Finite–Elementen ist ihre Speichereffizienz: Statt die lokalen Basisfunktionen auf allen Zellen abzuspeichern, speichert man nur die lokalen Basisfunktionen auf der Referenzzelle und die Transformationen von der Referenzzelle auf die Zellen der Gebietszerlegung. Zudem wird ein parametrischer Zugang oft in der Analysis mit folgender Beweisidee benutzt: Zunächst wird von den Gitterzellen auf die Referenzzelle transformiert. Dann wird die Behauptung für die Referenzzelle bewiesen. Anschließend transformiert man wieder auf die Zellen der Zerlegung zurück.
Eine mögliche Referenztransformation für das Dreieck K mit den Eckpunkten $\mathbf{a}^1, \mathbf{a}^2$, $\mathbf{a}^3$ ist die affine Transformation:

$$F_K(\hat{\mathbf{x}}) = \mathbf{B}\hat{\mathbf{x}} + \mathbf{b}, \quad \text{wobei} \tag{5.13}$$

$$\mathbf{B} = \begin{pmatrix} a_1^2 - a_1^1 & a_2^2 - a_2^1 \\ a_1^3 - a_1^1 & a_2^3 - a_2^1 \end{pmatrix} \quad \text{und} \quad \mathbf{b} = \begin{pmatrix} a_1^1 \\ a_2^1 \end{pmatrix}. \tag{5.14}$$

Die Transformation F_K ist so gewählt, dass $(0,0)$ auf $\mathbf{a}^1$ abgebildet wird, $(1,0)$ auf $\mathbf{a}^2$ und $(0,1)$ auf $\mathbf{a}^3$. Da ein Dreieck der konvexen Hülle seiner Eckpunkte entspricht, folgt daraus bereits $F_K(\hat{K}) = K$.

Die Variationsprobleme, welche aus der Konvektions–Diffusionsgleichung hervorgehen, werden üblicherweise im Hilbertraum $H_0^1(\Omega)$ gestellt sein. Diese Probleme können in den soeben eingeführten P^p-Räumen approximiert werden. Die Bedingung, dass die Lösung auf dem Rand $\partial\Omega$ verschwindet, kann dabei einfach in die P^p-Räume eingearbeitet werden. Dazu nimmt man den Unterraum aus P^p heraus, der durch die Basisfunktionen mit 1-Knoten auf dem Gebietsrand aufgespannt wird, oder mit anderen Worten: Man bildet den Durchschnitt $P^p \cap H_0^1(\Omega)$.

Die gemischten Variationsproblemen, welche aus der Oseen–Gleichung hervorgehen, werden in Produkträumen gestellt sein, zum Beispiel in $[H_0^1(\Omega)]^2 \times L^2(\Omega)$. Wir nähern ein solches Variationsprobleme in den P^p-Räumen wie folgt: Die beiden Komponenten von $[H_0^1(\Omega)]^2$ werden jeweils in einem Finite–Element–Raum P^p vom gleichen Typ approximiert. Der Approximationsraum zu $L^2(\Omega)$ kann hingegen unterschiedlich sein. Für ihn verwenden wir zwar die gleiche Gebietszerlegung, erlauben jedoch P^q-Räume mit einem anderen Polynomgrad q.

Dreieckselemente können offenbar nur genutzt werden, wenn das Problem auf $\Omega \subset \mathbb{R}^2$ gestellt ist. Für höherdimensionale Gebiete $\Omega \subset \mathbb{R}^d$, $d > 2$, lassen sich P^p–Element analog zu den Dreieckselementen konstruieren. Die Gitterzellen sind Simplizes. Demnach spricht man von simplizialen Finite–Elementen. Für eine allgemeinere Diskussion von simplizialen Finite–Elementen sei auf [Cia78] verwiesen.

Abschließend sei hervorgehoben, dass die besprochenen P^p-Elementen eine recht spezielle Klasse von Finite–Elementen sind. In [Cia78] werden weitere Beispiele von Finite–Elementen auf Simplizes diskutiert. Dazu gehören beispielsweise die Hermite– und Crouzeix–Raviart–Elemente.

5.3 Finite–Elemente auf Rechtecken

Sei Ω ein Gebiet aus $\mathbb{R}^2$, zu dem die zulässige Zerlegung $\mathcal{T}$ in Rechtecke K existiert. Zu $\mathcal{T}$ definiert man die Finite–Elemente–Räume auf Rechtecken durch:

$$Q^p := \left\{ v \in \mathcal{C}(\bar{\Omega}) : \ v|_K \in \mathbb{Q}_p(K) \ \forall K \in \mathcal{T} \right\}, \quad p \geq 1, \tag{5.15}$$

wobei $\mathbb{Q}_p(K)$ gegeben ist durch die Polynome mit Grad kleiner gleich p in jeder Variablen:

$$\mathbb{Q}_p(K) := \left\{ p : K \to \mathbb{R} : \ p(\mathbf{x}) = \sum_{\{\alpha:\ \alpha_1,\alpha_2 \leq p\}} \mu_\alpha \mathbf{x}^\alpha \right\}, \qquad \mu_\alpha \in \mathbb{R},\ \alpha = \text{Multiindex}. \tag{5.16}$$

Die Finite–Element–Räume Q^p bestehen also aus auf $\bar{\Omega}$ stetigen Funktionen, welche eingeschränkt auf eine einzelne Rechteckzelle Polynomen entsprechen. Als Funktionale verwenden wir erneut ausschließlich Punktauswertungen. Ausgewertet wird jeweils in einem der Punkte, die für jede Rechteckzelle gemäß dem in Abbildung 5.3 gezeigten Schema angeordnet sind. Zum Beispiel liegen die Knoten der Q^1-Elemente in den Eckpunkten und die Knoten der Q^2-Elemente in den Eck- oder Mittelpunkten der Kanten sowie im Mittelpunkt des Rechtecks.

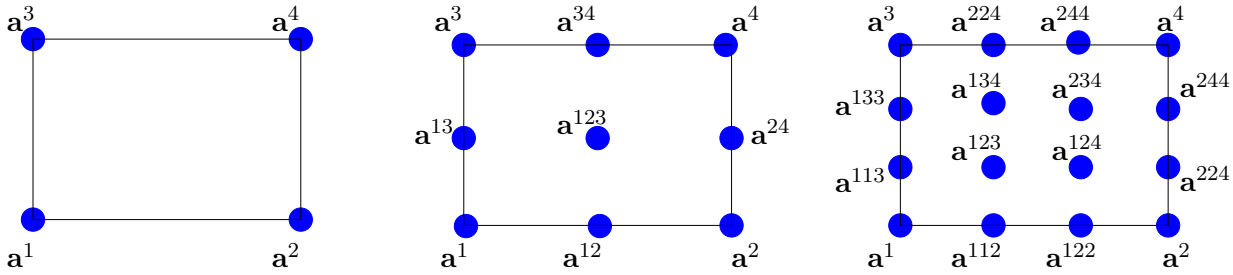

Abbildung 5.3: Verteilung der Knoten bei den Q^p-Elementen. Links für $p = 1$, in der Mitte für $p = 2$ und rechts für $p = 3$.

Als nächstes wollen wir die Standard–Knoten–Basis der Elemente Q^1,Q^2,Q^3 angeben. Im Gegensatz zu Dreiecken stehen hierzu keine baryzentrischen Koordinaten zur Verfügung. Stattdessen wählen wir einen parametrischen Zugang. Dabei bestimmen wir die lokalen Basisfunktionen $\{\hat{\phi}\}$ auf einem Referenzrechteck $\hat{K}$ und einen Satz von Referenztransformationen $\{F_K\}_{K\in\mathcal{T}}$ mit $F(\hat{K}) = K$. Daraus lässt sich dann analog zu den Dreieckselementen die gesuchte Basis darstellen.
Als Referenzrechteck wird in der Literatur entweder das Einheitsquadrat $[0, 1]^2$ oder das große Einheitsquadrat $[-1, 1]^2$ gewählt. In MooNMD wird das große Einheitsquadrat benutzt. Für dieses Referenzrechteck lautet eine mögliche Referenztransformation auf das Rechteck mit den Ecken $\mathbf{a}^1, \mathbf{a}^2, \mathbf{a}^3, \mathbf{a}^4$:

$$F_K(\hat{\mathbf{x}}) = \mathbf{B}\hat{\mathbf{x}} + \mathbf{b}, \quad \text{wobei} \tag{5.17}$$

$$\mathbf{B} = \frac{1}{2}\begin{pmatrix} a_1^4 - a_1^1 & 0 \\ 0 & a_1^4 - a_2^1 \end{pmatrix} \quad \text{und} \quad \mathbf{b} = \frac{1}{2}\begin{pmatrix} a_1^1 + a_1^4 \\ a_2^1 + a_2^4 \end{pmatrix}. \tag{5.18}$$

Eine Vereinfachung bei Finite–Elementen auf Rechtecken ergibt sich aus der Tatsache, dass sie als Tensorprodukt eindimensionaler Finite–Elemente geschrieben werden können. Demzufolge können auch die gesuchten Basisfunktionen als Produkt eindimensionaler Funktionen geschrieben werden.
Bei eindimensionalen Finite–Elementen ist Ω ein Intervall, welches in Teilintervalle zerlegt wird. Die eindimensionalen Finite–Element–Räume, welche wir zur Konstruktion der Basisfunktionen nutzen werden, bestehen aus stetige Funktionen, die eingeschränkt auf die Intervalle Polynomen entsprechen. Die zugehörigen Funktionale sind Punktauswertungen, wobei die Knoten nach Abbildung 5.4 verteilt sind. Aus den Funktionalen folgt eine Basis der eindimensionalen Finite–Element–Räume. Als Referenzintervall für eindimensionale Finite–Elemente wählt man üblicherweise das Einheitsintervall $[0, 1]$ oder das große Einheitsintervall $[-1, 1]$. Die Idee ist nun, die lokalen Basisfunktionen auf dem Referenzintervall $[-1, 1]$ anzugeben und aus diesen durch Multiplikation untereinander die lokalen Basisfunktionen auf dem großen Referenzrechteck zu konstruieren.

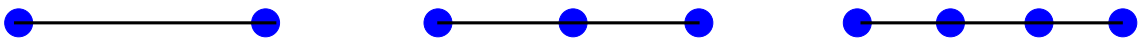

Abbildung 5.4: Übliche Verteilung der Knoten bei stetigen eindimensionalen Finite–Elementen. Links lineare Elemente, in der Mitte quadratische, rechts kubische.

- Die Basisfunktionen für stetige, stückweise lineare eindimensionale Finite–Elemente lauten auf dem großen Referenzintervall:

$$\zeta_1(\hat{x}) = \frac{1}{2}(1-\hat{x}), \qquad \zeta_2(\hat{x}) = \frac{1}{2}(1+\hat{x}). \tag{5.19}$$

Die Basisfunktionen $\hat{\phi}_i = \hat{\phi}_i(\hat{x},\hat{y})$, $i = 1,...,4$, von Q^1 auf dem großen Referenzrechteck sind dann:

$$\hat{\phi}_1 = \zeta_1(\hat{x})\zeta_1(\hat{y}), \quad \hat{\phi}_2 = \zeta_2(\hat{x})\zeta_1(\hat{y}), \quad \hat{\phi}_3 = \zeta_1(\hat{x})\zeta_2(\hat{y}), \quad \hat{\phi}_4 = \zeta_2(\hat{x})\zeta_2(\hat{y}). \tag{5.20}$$

Die Basisfunktion $\hat{\phi}_i$ hat dabei ihren 1-Knoten im Eckpunkt $\mathbf{a}_i$

- Die Basisfunktionen für stetige, stückweise quadratische eindimensionale Finite–Elemente lauten auf dem großen Referenzintervall:

$$\zeta_1(\hat{x}) = -\frac{1}{2}\hat{x}(1-\hat{x}), \qquad \zeta_2(\hat{x}) = (1+\hat{x})(1-\hat{x}), \qquad \zeta_3(\hat{x}) = \frac{1}{2}\hat{x}(1+\hat{x}). \tag{5.21}$$

Die 9 Basisfunktionen $\hat{\phi} = \hat{\phi}(\hat{x},\hat{y})$ von Q^2 auf dem großen Referenzrechteck sind dann:

$$\hat{\phi}_1 = \zeta_1(\hat{x})\zeta_1(\hat{y}), \quad \hat{\phi}_2 = \zeta_3(\hat{x})\zeta_1(\hat{y}), \quad \hat{\phi}_3 = \zeta_1(\hat{x})\zeta_3(\hat{y}), \quad \hat{\phi}_4 = \zeta_3(\hat{x})\zeta_3(\hat{y}),$$
$$\hat{\phi}_{12} = \zeta_2(\hat{x})\zeta_1(\hat{y}), \quad \hat{\phi}_{13} = \zeta_1(\hat{x})\zeta_2(\hat{y}), \quad \hat{\phi}_{24} = \zeta_3(\hat{x})\zeta_2(\hat{y}), \quad \hat{\phi}_{34} = \zeta_2(\hat{x})\zeta_3(\hat{y}),$$
$$\hat{\phi}_M = \zeta_2(\hat{x})\zeta_2(\hat{y}).$$

Die Basisfunktion $\hat{\phi}_i$ hat ihren 1-Knoten 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 lauten auf dem großen Referenzintervall:

$$\zeta_1(\hat{x}) = -\frac{1}{16}(3\hat{x}+1)(3\hat{x}-1)(\hat{x}-1), \tag{5.22}$$
$$\zeta_2(\hat{x}) = \frac{9}{16}(3\hat{x}-1)(\hat{x}+1)(\hat{x}-1), \tag{5.23}$$
$$\zeta_3(\hat{x}) = -\frac{9}{16}(3\hat{x}+1)(\hat{x}+1)(\hat{x}-1), \tag{5.24}$$
$$\zeta_4(\hat{x}) = \frac{1}{16}(3\hat{x}+1)(3\hat{x}-1)(\hat{x}+1). \tag{5.25}$$

Die 16 Basisfunktionen $\hat{\phi} = \hat{\phi}(\hat{x},\hat{y})$ von Q^3 auf dem großen Referenzrechteck sind dann:

$$\hat{\phi}_1 = \zeta_1(\hat{x})\zeta_1(\hat{y}), \quad \hat{\phi}_2 = \zeta_4(\hat{x})\zeta_1(\hat{y}), \quad \hat{\phi}_3 = \zeta_1(\hat{x})\zeta_4(\hat{y}), \quad \hat{\phi}_4 = \zeta_4(\hat{x})\zeta_4(\hat{y}),$$
$$\hat{\phi}_{112} = \zeta_3(\hat{x})\zeta_1(\hat{y}), \quad \hat{\phi}_{122} = \zeta_2(\hat{x})\zeta_1(\hat{y}), \quad \hat{\phi}_{113} = \zeta_1(\hat{x})\zeta_3(\hat{y}), \quad \hat{\phi}_{133} = \zeta_1(\hat{x})\zeta_2(\hat{y}),$$
$$\hat{\phi}_{334} = \zeta_3(\hat{x})\zeta_4(\hat{y}), \quad \hat{\phi}_{344} = \zeta_2(\hat{x})\zeta_4(\hat{y}), \quad \hat{\phi}_{224} = \zeta_4(\hat{x})\zeta_3(\hat{y}), \quad \hat{\phi}_{244} = \zeta_4(\hat{x})\zeta_2(\hat{y}),$$
$$\hat{\phi}_{123} = \zeta_3(\hat{x})\zeta_3(\hat{y}), \quad \hat{\phi}_{124} = \zeta_2(\hat{x})\zeta_3(\hat{y}), \quad \hat{\phi}_{134} = \zeta_3(\hat{x})\zeta_2(\hat{y}), \quad \hat{\phi}_{234} = \zeta_2(\hat{x})\zeta_2(\hat{y}).$$

Die Basisfunktionen $\hat{\phi}$ haben ihren 1-Knoten in dem gleich-indizierten Punkt $\mathbf{a}$ von Abbildung 5.3.

Bemerkung 5.6.
(1.) In der Praxis untersucht man häufig Gebiete, welche polygonal berandet sind und dennoch nicht in Rechtecke zerlegt werden können. Solche Gebiete kann man beispielsweise mit Dreiecken oder mit allgemeinen Vierecken überdecken. Die parametrischen Finite–Elemente–Räume zu den allgemeinen Vierecken bestehen allerdings nicht mehr unbedingt aus stückweise polynomialen Funktionen. Da hier die Referenztransformation bilinear ist, setzen sich die Finite–Elemente–Räume aus stückweise rationalen Funktionen zusammen.

(2.) Analog zu den Finite–Elementen auf Dreiecken lassen sich aus den Q^p-Räumen passende Approximationsräume zur Konvektions–Diffusions- und Oseen–Gleichung konstruieren (siehe hierzu Ende Abschnitt 5.2).

(3.) Die dreidimensionale Verallgemeinerung der Rechteckelemente sind Finite–Elemente auf Hexaedern. Die Finite–Element–Räume bestehen weiterhin aus stückweise polynomialen Funktionen und die zugehörigen Funktionale sind Punktauswertungen. Beispielsweise für stückweise lineare Elemente wertet man die Funktionen in den Eckpunkten der Hexaeder aus und für stückweise quadratische Elemente in den Ecken, den Mittelpunkten der Kanten und Seitenflächen sowie im Schwerpunkt der Hexaeder.

(4.) Neben den Q^p-Elementen gibt es weitere gebräuchliche Finite–Elemente auf Rechtecken beziehungsweise Hexaedern wie die nichtkonformen Q^p_{nc}- oder die diskontinuierlichen Q^p_{disc}-Elemente (für Literatur hierzu siehe beispielsweise [MS07] oder [Riv08]).

5.4 Abschätzung des Interpolationsfehlers

Nachdem wir die Finite–Elemente–Räume eingeführt haben, stellt sich die Frage, wie gut in ihnen die exakte Lösung auf dem Lösungsgebiet Ω approximiert werden kann. Dazu werden wir in diesem Abschnitt Interpolationsfehler–Abschätzungen der folgenden Form angeben:

$$\|u - Iu\|_{X(\mathcal{A})} \leq Ch^s \|u\|_{Y(\mathcal{B})}. \tag{5.26}$$

Dabei ist u die exakte Lösung, Iu ihre Interpolation in einem Finite–Element–Raum, I der zugehörige Interpolationsoperator, $\mathcal{A}, \mathcal{B} \subset \Omega$, $X(\mathcal{A})$ sowie $Y(\mathcal{B})$ geeignete Räume mit Normen $\|\cdot\|_{X(\mathcal{A})}$,$\|\cdot\|_{Y(\mathcal{B})}$, $s \in \mathbb{R}^+$, $C > 0$ eine gitterunabhängige Konstante und h eine gitterabhängige Größe wie zum Beispiel der Durchmesser einer Gitterzelle oder die Länge einer Kante. Die Aussagekraft solcher Abschätzungen liegt für uns vor allem darin, dass der Interpolationsfehler $\|u - Iu\|$ beim Übergang $h \to 0$ mit der Konvergenzordnung s gegen Null strebt. In dieser Hinsicht sind die folgenden Interpolationsfehler–Abschätzungen optimal, das heißt, man erhält durch eine andere Wahl des Interpolationsoperators I im Allgemeinen keine bessere Konvergenzordnung $\widetilde{s} > s$. Die Konstante C in der Abschätzung ist hingegen oft suboptimal und kann durch eine geschickte Wahl des Interpolationsoperators verkleinert werden.

Um brauchbare Approximations–Resultate zu erhalten, muss die Zerlegung des Lösungsgebiets ausreichend regulär sein. Wir fordern im Folgenden, dass die Gebietszerlegung quasi–uniform ist. Diese Eigenschaft ist bei den in dieser Arbeit verwandten Finite–Elementen immer gegeben. Zudem genügt sie als Voraussetzung für die Interpolationsfehler–Abschätzungen in dieser Arbeit.

Definition 5.7. (Quasi–uniforme Triangulierung)

Eine zulässige Triangulierung $\mathcal{T}_h$ von $\Omega \subset \mathbb{R}^2$ bezeichnet man als quasi–uniform, wenn

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

$$\max_{K \in \mathcal{T}_h} \frac{\sigma_K}{\rho_K} \leq C \qquad (\textit{Formregularität}), \tag{5.27}$$

(2.) wenn alle Zellen von der gleichen Größenordnung sind, das heißt, wenn es eine Konstante $\xi > 0$ gibt, so dass

$$\min_{K \in \mathcal{T}_h} h_K \leq \xi \max_{K \in \mathcal{T}_h} h_K \quad \text{mit } h_K := \operatorname{diam}(K) \quad (\textit{Größenregularität}). \tag{5.28}$$

Für $\Omega \subset \mathbb{R}^3$ wird lediglich Punkt 2 der obigen Definition modifiziert. Statt des Umkreisradius wählt man für ρ_K den Radius der kleinsten Kugel, welche die Gitterzelle K enhält, und statt des Innkreisradius σ_K den Radius der größten Kugel, welche in der Gitterzelle liegt. Abbildung 5.5 zeigt Gebietszerlegungen, welche nicht quasiuniform sind.

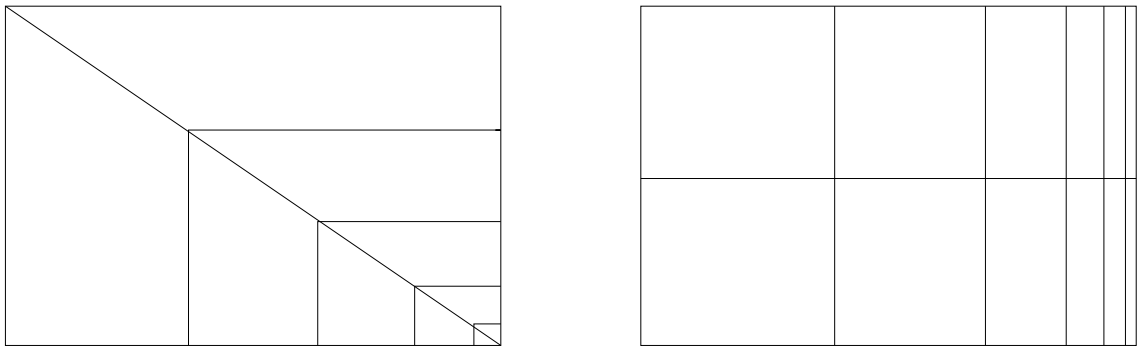

Abbildung 5.5: Denkt man sich die Gebietszerlegung wie in der Abbildung angedeutet fortgesetzt, erhält man links eine Gebietszerlegung, welche nicht größenregulär ist, und rechts eine Gebietszerlegung, welche nicht formregulär ist.

Die folgenden Interpolationsfehler–Abschätzungen gelten gleichermaßen für Finite–Elemente auf Dreiecken und Rechtecken beziehungsweise in drei Dimensionen für Tetraeder und Hexaeder. Wir kürzen daher die Finite–Element–Räume P_h^p und Q_h^p mit demselben Symbol V_h^p ab. Falls später benötigt, beziehen wir auch die Elementordnung p in die Interpolationsfehler–Abschätzungen mit ein.

Die erste Interpolationsfehler–Abschätzung benutzt den Lagrange–Interpolationsoperator I_h^p. Dieser ist auf K wie folgt definiert: Sei $\{\phi_j\}_{j=1}^N$ die lokale Standard–Knoten–Basis bezüglich der Knoten p_j des Finite–Elemente–Raumes $V_h^p(\Omega)$ auf der Zelle K. Nun definieren wir

$$I_h^p(u)(x) = \sum_{j=1}^{N} a_j \phi_j(x), \tag{5.29}$$

wobei die Koeffizieten a_j durch $a_j := u(p_j)$ bestimmt werden. Nach Konstruktion erhält der Lagrange–Interpolationsoperator insbesondere homogene Randbedingungen.

Lemma 5.8. (**Lagrange–Interpolation**)
Seien $\mathcal{T}_h$ eine quasi–uniforme Zerlegung von $\Omega \subset \mathbb{R}^d$, $d \geq 2$, und $I_h^p : H^k(K) \to V_h^p(K)$, $k > d/2$, der Lagrange–Interpolationsoperator auf $K \in \mathcal{T}_h$. Dann gilt für $u \in H^k(K)$ die lokale Interpolationsfehler–Abschätzung:

$$\|u - I_h^p u\|_{H^m(K)} \leq C \frac{h_K^{l-m}}{p^{k-m}} \|u\|_{H^k(K)}, \tag{5.30}$$

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

Beweis. Siehe [HS01] Abschnitt 4. Die Abhängigkeit von der Elementordnung p folgt in Kombination mit [Sch98] Korollar 4.68.

Folgendes Lemma ergibt eine Interpolationsfehler–Abschätzung bezüglich der Seminormen der Sobolevräume:

Lemma 5.9.
Sei $\mathcal{T}_h$ eine quasi–uniforme Zerlegung von $\Omega \subset \mathbb{R}^d$, $d \geq 2$, und sei $I_h^{k-1} : H^k(K) \to V_h^{k-1}(K)$, $k > d/2$, der Lagrange–Interpolationsoperator auf $K \in \mathcal{T}_h$. Dann gilt für $u \in H^k(K)$ die lokalen Interpolationsfehler–Abschätzung:

$$\left|u - I_h^p u\right|_{H^m(K)} \leq C h_K^{k-m} \left|u\right|_{H^k(K)}, \tag{5.31}$$

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

Beweis. Siehe [CR72] Theorem 2.

Die Voraussetzungen sind hier etwas spezieller als im vorausgehenden Lemma, da der Polynomgrad p auf $k-1$ fixiert ist, wenn gegen die Seminorm $|\cdot|_{H^{k-1}(K)}$ abgeschätzt wird. [2] Als nächstes geben wir Interpolationsfehler–Abschätzungen für die globale L^2–Projektion $\pi_h : L^2(\Omega) \to V_h^p(\Omega)$ an. Insbesondere erhalten wir dabei Abschätzungen auf dem Gebietsrand $\partial\Omega$. Die L^2–Projektion auf $V_h^p(\Omega)$ ist definiert durch $\pi_h(u) = u_h$ mit u_h aus

$$(u - u_h, v_h) = 0 \qquad \forall v_h \in V_h^p(\Omega). \tag{5.32}$$

Lemma 5.10. (L^2–**Projektion**)
Sei $\mathcal{T}_h$ eine quasi–uniforme Zerlegung von $\Omega \subset \mathbb{R}^d$, $d \geq 2$, und sei $\pi_h : L^2(\Omega) \to V_h^p(\Omega)$ die globale L^2–Projektion. Dann gelten für $u \in H^k(\Omega)$, $k \geq 2$, die globale Interpolationsfehler–Abschätzungen

$$\|u - \pi_h u\|_{H^m(\Omega)}^2 \leq C \sum_{K \in \mathcal{T}_h} h_K^{2(l-m)} \left|u\right|_{H^k(K)}, \tag{5.33}$$

$$\|u - \pi_h u\|_{L^2(\partial\Omega)}^2 \leq C \sum_{K \in \mathcal{T}_h} h_K^{2l} \left|u\right|_{H^k(\partial\Omega)} \tag{5.34}$$

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

Beweis. Siehe [RST08] Lemma 3.83 und Bemerkung 3.84. Der Beweis benutzt die Interpolationsfehler–Abschätzung aus Lemma 5.9.

[2] Ersetzt man die Seminormen in Lemma 5.9 durch Normen, so erhält man wieder die Aussage von Lemma 5.30. Man vergewissert sich dabei schnell, dass insbesondere der Polynomgrad nicht mehr auf $k-1$ fixiert sein muss.

Bemerkung 5.11.
Die L^2–Projektion erhält gegenüber der Lagrange–Interpolation im Allgemeinen keine homogene Randbedingungen.

Bisher haben wir ausschließlich Interpolationsfehler–Abschätzungen für Funktionen u aus $H^2(\Omega)$ angegeben. Der Clément–Interpolationsoperator macht auch eine Aussage für $u \in H^1(\Omega)$. Die Konstruktion des Clément–Interpolationsoperators $\mathcal{C}_h^p$ verläuft auf $\Omega \subset \mathbb{R}^2$ wie folgt (für höhere Dimensionen geht es analog): Sei $\{\phi_j\}_{j=1}^N$ die Standard–Knoten-Basis bezüglich der Knoten p_j im Finite–Elemente Raum $V_h^p(\Omega)$. Dann definieren wir

$$\mathcal{C}_h^p(u)(x) = \sum_{j=1}^{N} a_j \phi_j(x). \tag{5.35}$$

Die Koeffizienten a_j sind gegeben durch $a_j := (\pi_{w_j} u)(p_j)$, wobei $\pi_{w_j} : L^2(w_j) \to V_h^p(w_j)$ die L^2–Projektion auf die Vereinigung aller Gitterzellen w_j ist, zu welchen der Knoten p_j gehört (siehe Abbildung 5.6). Für spätere Zwecke sei bemerkt, dass gemäß der obigen Konstruktion der Clément–Interpolationsoperator auch auf Funktionen aus $L^2(\Omega)$ angewandt werden kann. In diesem Fall erhält man allerdings keine brauchbare Interpolations–Fehlerabschätzung mehr.

Lemma 5.12. (**Clément–Interpolation**)
Sei $\mathcal{T}_h$ eine quasi–uniforme Zerlegung von $\Omega \subset \mathbb{R}^d$, $d \geq 2$, und sei $\mathcal{C}_h^p : H^k(\Omega) \to V_h^p(\Omega)$, $k \geq 1$ der Clément–Interpolationsoperator. Dann gilt auf $K \in \mathcal{T}_h$ für $u \in H^k(\Omega)$ die lokale Interpolationsfehler–Abschätzung

$$\|u - \mathcal{I}_h^p u\|_{H^m(K)} \leq C h_K^{l-m} \|u\|_{H^k(w(K))}, \tag{5.36}$$

$$\text{mit } 0 \leq m \leq l := \min(p+1, k).$$

Dabei ist $C > 0$ eine Konstante und $w(K)$ die Vereinigung aller Zellen, welche mit Zelle K mindestens einen Punkt gemeinsam haben.

Beweis. Siehe [Clé75] oder [BG09]. □

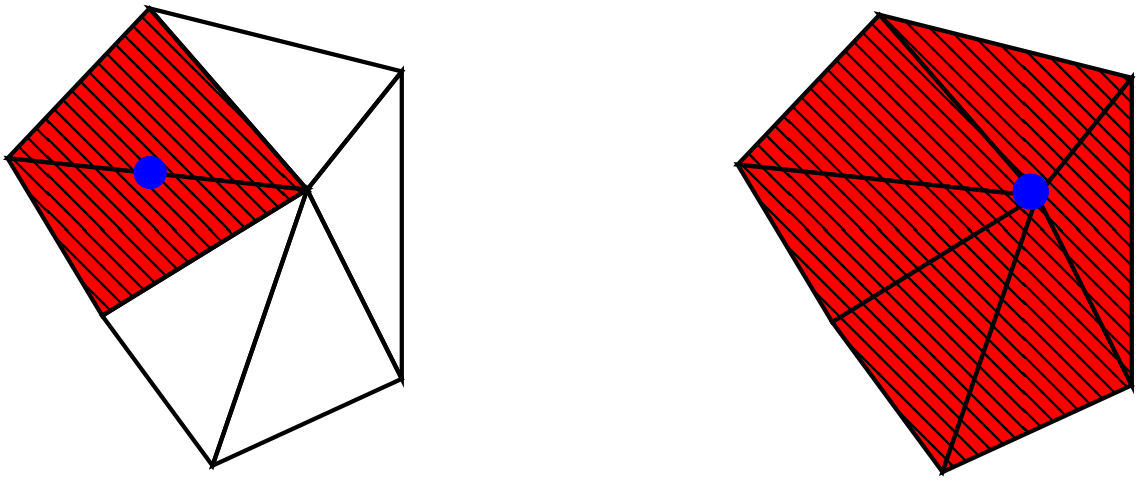

Abbildung 5.6: Schematische Skizze zur Konstruktion des Clément–Interpolationsoperators. Der blaue Punkt kennzeichne den Knoten p_j. Der schattierte Bereich gibt dann die zugehörige Menge w_j von Gitterzellen an.

Ein weiterer Interpolationsoperator, welcher ebenfalls für Funktionen mit H^1–Regularität benutzt werden kann, ist der Scott–Zhang–Interpolationsoperator. Dieser ist zwar nur für Funktionen aus $H_0^1(\Omega)$ definiert, besitzt jedoch gegenüber dem Clément–Interpolationsoperator den Vorteil, dass er homogene Randbedingungen erhält.
Der Scott–Zhang–Interpolationsoperator i_h^p wird auf $\Omega \subset \mathbb{R}^2$ wie folgt konstruiert (für höhere Dimensionen verläuft die Konstruktion analog): Sei $\{\phi_j\}_{j=1}^N$ die Standard–Knoten-Basis bezüglich den Knoten p_j im Finite–Elemente Raum $V_h^p(\Omega)$. Dann definieren wir

$$i_h^p(u)(x) = \sum_{j=1}^{N} a_j \phi_j(x). \tag{5.37}$$

Für innere Knoten p_j einer Zelle K setzen wir $a_j := (\pi_K u)(p_j)$, wobei $\pi_K : L^2(K) \to V_h^p(K)$ die L^2–Projektion in K ist. Für einen Knoten p_j, der auf einer Kante mindestens einer Zelle liegt, wählen wir eine dieser Kanten E_j aus und setzen $a_j := (\pi_{E_j} u)(p_j)$, wobei $\pi_{E_j} : L^2(E_j) \to V_h^p(E_j)$ die L^2–Projektion auf die Kante E_j ist. Liegt der Knoten p_j auf dem Gebietsrand $\partial\Omega$, so darf nur eine Randkante $E_j \subset \partial\Omega$ gewählt werden (siehe Abbildung 5.7).

Lemma 5.13. (**Scott-Zhang–Interpolation**)
Sei $\mathcal{T}_h$ eine quasi–uniforme Zerlegung von $\Omega \subset \mathbb{R}^d$, $d \geq 2$, und sei $i_h^p : H_0^k(\Omega) \to V_h^p(\Omega) \cap H_0^1(\Omega)$ der Scott–Zhang–Interpolationsoperator. Dann gilt auf $K \in \mathcal{T}_h$ für alle $u \in H_0^k(\Omega)$ die lokale Interpolationsfehler–Abschätzung

$$\|u - \mathcal{I}_h^p u\|_{H^m(K)} \leq C \frac{h_K^{l-m}}{p^{k-m}} \|u\|_{H^k(w(K))}, \tag{5.38}$$

$$\text{mit } 0 \leq m \leq l := \min(p+1, k).$$

Dabei ist $C > 0$ eine Konstante und $w(K)$ die Vereinigung aller Zellen, welche mit Zelle K mindestens einen Punkt gemeinsam haben.

Beweis. Siehe [SZ90] Theorem 4.1. Die Abhängigkeit von der Elementordnung p folgt erneut in Kombination mit [Sch98] Korollar 4.68.

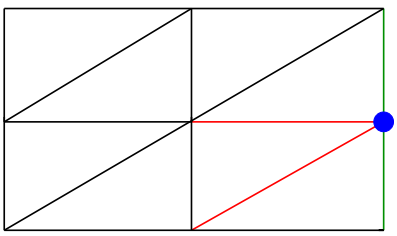

Abbildung 5.7: Schematische Skizze zur Konstruktion des Scott–Zhang–Interpolationsoperators. Entspricht der blaue Punkt dem Knoten p_j, so projiziert man auf eine der grünen Kanten nicht aber auf eine der roten (weitere Details siehe Text).

Für eine ausführliche Einführung in die Interpolationstheorie von Funktionen aus $H^1(\Omega)$ sei auf [EG04] verwiesen.
Die Interpolationsfehler–Abschätzungen in diesem Abschnitt gelten auch, wenn vektorwertige Funktionen aus $[H^k(\Omega)]^d$, $d \in \mathbb{N}$, durch Funktionen aus $[V_h^p]^d$ interpoliert werden. Dazu wendet man obige Interpolationen komponentenweise an und schätzt jede Komponente einzeln durch die Sätze ab.

5.5 Spur- und inverse Ungleichungen

Neben Interpolationsfehler–Abschätzungen benötigen wir inverse Ungleichungen. Bei inversen Ungleichungen werden Ableitungen von Finite–Element–Funktionen nach oben gegen niedrigere Ableitungen sowie gegen die Funktionswerte selbst abgeschätzt.

Lemma 5.14. (**Inverse Ungleichungen**)
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–Konstante aus Definition 5.7 abhängt, so dass für alle $u_h \in V_h^p(\Omega)$ und alle Gitterzellen $K \in \mathcal{T}_h$ gilt:

$$\|\nabla u_h\|_{L^2(K)} \leq \mu_{\text{inv}} h_K^{-1} p^2 \|u_h\|_{L^2(K)} \tag{5.39}$$

Zudem gilt für $u_h \in [V_h^p]^d$:

$$\|\nabla \cdot u_h\|_{L^2(K)} \leq \|\nabla u_h\|_{L^2(K)} \leq \mu_{\text{inv}} h_K^{-1} p^2 \|u_h\|_{L^2(K)}. \tag{5.40}$$

Beweis. Siehe [Bra07] Kapitel 2, Satz 6.8. Die Abhängigkeit von der Elementordnung p folgt in Kombination mit [Sch98] Theorem 4.76.

Ein weiteres Hilfsmittel in dieser Arbeit sind Spurungleichungen. Mit Spurungleichungen können die Werte von Sobolev–Funktionen auf dem Rand einer Gitterzelle nach oben gegen die Funktionswerte im Inneren der Zelle abgeschätzt werden.

Lemma 5.15. (**Spurungleichungen**)
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 $v \in H^1(\Omega)$ die Spurungleichungen

$$\|v\|_{L^2(\partial K)}^2 \leq C \left(h_K^{-1} \|v\|_{L^2(K)}^2 + h_K \|v\|_{H^1(K)}^2 \right), \tag{5.41}$$

$$h_E^{1/2} \|v\|_{L^2(\partial K)}^2 \leq C \left(\|v\|_{L^2(K)}^2 + h_K \|v\|_{H^1(K)}^2 \right) \tag{5.42}$$

mit einer Konstanten $C > 0$ und der Länge h_E der Kante E (beziehungsweise dem Flächeninhalt h_E der Seitenfläche E).

Beweis. Den Beweis zur Ungleichung (5.41) findet man in [Tho97] Seite 26. Die Ungleichung (5.42) wird in [RST08] angegeben - allerdings ohne Beweis. □

Kapitel 6

Die Konvektions–Diffusionsgleichung

Die Konvektions–Diffusionsgleichung beschreibt einerseits eine Vielzahl physikalischer Vorgänge, andererseits tritt sie des Öfteren als Hilfsgleichung bei komplizierteren Modellen auf wie zum Beispiel bei bestimmten Turbulenzmodellen [LR06]. Mit diesem Kapitel wird eine Einführung in die Theorie der Konvektions–Diffusionsgleichungen gegeben. Dazu betrachten wir zunächst die Lösbarkeit der Gleichung im klassischen Sinne und beschreiben typische Merkmale einer Lösung. Dann formulieren wir die Konvektions–Diffusionsgleichung in ein Variationsproblem um und untersuchen dessen Lösbarkeit. Wie sich zeigt, besitzt die variationelle Formulierung unter recht allgemeinen Voraussetzungen eine eindeutige Lösung. Zur Approximation dieser Lösung führen wir das Standard–Galerkin–Verfahren ein. Dieses Verfahren liefert jedoch oft schlechte Resultate. Die Ergebnisse lassen sich mit Stabilisierungverfahren verbessern. Zwei solcher Stabilisierungsverfahren werden als nächstes vorgestellt: die Streamline–Diffusion–Methode und die Kanten–Stabilisierung. Ähnlich zu [RST08] diskutieren wir, unter welchen Voraussetzungen die Stabilisierungsverfahren eine eindeutige Lösung besitzen, und leiten eine a-priori Abschätzung für den Verfahrensfehler ab.

6.1 Grundlegende Eigenschaften und analytischer Zugang

Sei Ω ein beschränktes offenes Gebiet in $\mathbb{R}^d$ mit $d = 2$ oder $d = 3$. Eine skalare Konvektions–Diffusionsgleichung für die Funktion $u : \bar{\Omega} \to \mathbb{R}$ ist von der Form:

$$-\epsilon \Delta u + b(x) \cdot \nabla u + c(x)u = f(x) \quad \text{für } x \in \Omega \tag{6.1}$$

mit $\epsilon > 0$, passenden Randbedingungen auf dem Gebietsrand $\partial\Omega$ und ausreichend glatten Funktionen $b : \Omega \to \mathbb{R}^2$ sowie $c, f : \Omega \to \mathbb{R}$. Skalare Konvektions–Diffusionsgleichungen beschreiben die Verteilung einer skalaren Größe u wie zum Beispiel der Konzentration oder Temperatur in Ω. Für diese Verteilung berücksichtigt die Konvektions–Diffusionsgleichung (6.1) verschiedene Prozesse: Diffusion (erster Term in (6.1)), Konvektion mit einem Geschwindigkeitsfeld b (zweiter Term), Erzeugung oder Vernichtung der Größe u beispielsweise durch chemische Reaktionen (dritter Term) und Einfluss von Volumenkräften wie zum Beispiel der Schwerkraft (vierter Term).
In der Praxis lassen sich die Randbedingungen zumeist durch folgenden Ansatz beschreiben:

$$\begin{aligned} u &= g_1 \quad \text{auf } \Gamma_1 && \text{(Dirichlet–Randbedingung)}, \\ n \cdot \nabla u &= g_2 \quad \text{auf } \Gamma_2 && \text{(Neumann–Randbedingung)}, \\ \beta u + n \cdot \nabla u &= g_3 \quad \text{auf } \Gamma_3 && \text{(Robin–Randbedingung)}. \end{aligned}$$

Dabei sind Γ_i, $i = 1, 2, 3$, disjunkte Teilstücke des Gebietsrandes $\partial\Omega$ mit $\partial\Omega = \Gamma_1 \cup \Gamma_2 \cup \Gamma_3$, $n = n(x)$ der äußere Normalenvektor von $\partial\Omega$ an der Stelle x, β eine stetige Funktion auf Γ_3 und g_i stetige Funktionen auf Γ_i. Aus mathematischer Sicht kann die Neumann–Randbedingung auch als Spezialfall der Robin–Randbedingung behandelt werden. Der Einfachheit halber beschränken wir uns im Folgenden zumeist auf homogene Dirichlet–Randbedingungen, das heißt, auf die Vorgabe

$$u = 0 \quad \text{auf } \partial\Omega. \tag{6.2}$$

Weiterhin wollen wir in dieser Arbeit zwei Fälle bei Konvektions–Diffusionsgleichungen unterscheiden: den diffusions–dominanten Fall mit

$$\epsilon > \|b\|_{L^\infty(\Omega)} \text{ und } \epsilon \gtrsim \|c\|_{L^\infty(\Omega)} \tag{6.3}$$

sowie den konvektions–dominaten Fall mit

$$\epsilon \ll \|b\|_{L^\infty(\Omega)} \text{ und } \|b\|_{L^\infty(\Omega)} \gtrsim \|c\|_{L^\infty(\Omega)}. \tag{6.4}$$

Auf dem konvektions-dominanten Fall wird unser Hauptaugenmerk liegen. In diesem Fall zeichnet sich die Verteilung u üblicherweise durch das Vorhandensein von Grenzschichten aus. Im diffusions–dominanten Fall sind Grenzschichten hingegen untypisch. Grenzschichten sind kleine Teilgebiete von Ω, in denen u einen relativ großen Gradienten hat. Man unterscheidet drei Typen von Grenzschichten:

(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. 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 6.1. 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 $\partial\Omega$. Gemäß diesen Vorgaben erhalten wir parabolische Grenzschichten bei $y = 0$ und $y = 1$.

(ii) Exponentielle Grenzschichten
Exponentielle Grenzschichten treten an Dirichlet–Rändern auf, welche im Ausflussbereich der Strömung liegen, das heißt, wo $n(x) \cdot b(x) > 0$ gilt. Ihre Breite ist von der Ordnung $\mathcal{O}(\epsilon\ln(1/\epsilon))$. Damit sind exponentiellen Grenzschichten noch schmaler als parabolische, was ihre numerische Behandlung weiter erschwert. Ein Beispiel für eine exponentielle Grenzschicht wird ebenfalls durch den Graphen auf der linken Seite von Abbildung 6.1 gegeben. Gemäß den obigen Vorgaben befindet sich die exponentielle Grenzschicht bei $x = 1$. Auf der rechten Seite der Abbildung findet man ein weiteres Beispiele für eine exponentielle Grenzschicht. Betrachtet wird dort die Konvektions–Diffusionsgleichung zu den Vorgaben $\epsilon = 10^{-8}$, $b = (\cos(-\pi/3), \sin(-\pi/3))^T$, $c = 0$, $f = 0$ auf $\Omega = [0, 1]^2$ und den Randwerten

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

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

(iii) Innere Grenzschichten
Wie der Name vermuten lässt, treten innere Grenzschichten im Inneren von Ω auf. Verursacht werden sie durch Unstetigkeiten in den Dirichlet–Randwerten. Ihre Struktur ähnelt

derjenigen von parabolischen Grenzschichten, weswegen sie manchmal auch als innere parabolische Grenzschichten bezeichnet werden. Der Graph auf der rechten Seite von Abbildung 6.1 zeigt beispielhaft 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 Problemdaten plausibel gemacht werden, siehe hierzu [Sty05].

An Dirichlet–Rändern, welche im Einströmbereich liegen, das heißt wo $n(x) \cdot b(x) < 0$ gilt, gibt es hingegen üblicherweise keine Grenzschichten. Für weitere Details zu Grenzschichten sei auf [HKOS08] verwiesen.

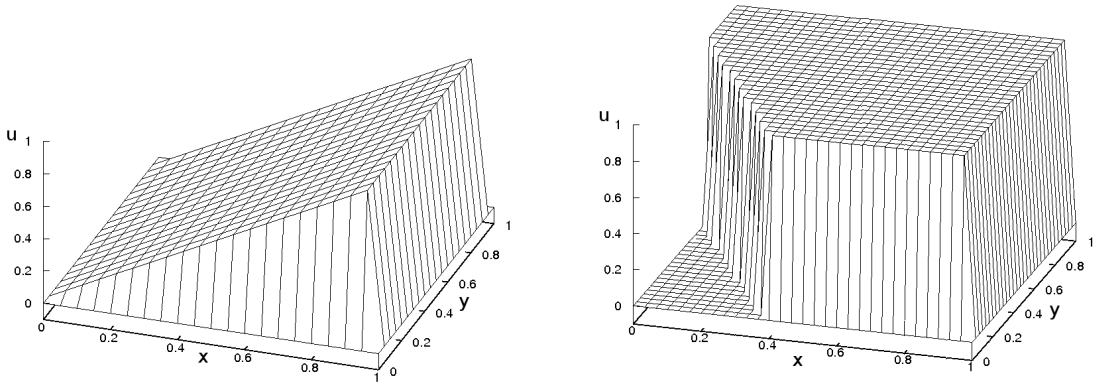

Abbildung 6.1: Lösung u der Konvektions–Diffusionsgleichung mit Grenzschichten (übernommen von [JK07a]). Für weitere Details siehe Text.

Als nächstes diskutieren wir die klassische Lösbarkeit der Konvektions–Diffusionsgleichung. Für die Existenz einer klassischen Lösung gibt der folgende Satz eine hinreichende Bedingung:

Satz 6.1.
Sei Ω ein beschränktes Gebiet mit Lipschitz–stetigem Rand $\partial\Omega$ und seien b,c und f Hölder–stetig auf $\bar{\Omega}$ mit $c(x) \geq 0$ für alle $x \in \bar{\Omega}$. Dann besitzt die skalare Konvektions–Diffusionsgleichung (6.1) versehen mit der homogenen Randbedingnung (6.2) eine eindeutige Lösung u aus $C(\bar{\Omega}) \cap C^2(\Omega)$.

Beweis. Siehe [Mic77]. □

Bemerkung 6.2.

(i) Der Satz kann auch die eindeutige Lösbarkeit sicherstellen, wenn durch $u = g$ auf $\partial\Omega$ inhomogene Dirichlet–Randbedingungen bezüglich einer ausreichend glatten Funktion g zu erfüllen sind. Die Idee dabei ist die folgende: Man setzt die Funktion g vom Gebietsrand auf eine Funktion $\widetilde{g} \in C(\bar{\Omega}) \cap C^2(\Omega)$ fort. Mit Hilfe dieser Fortsetzung definiert man die Hilfsfunktion $\widetilde{u} = u - \widetilde{g}$ und setzt $u = \widetilde{u} + g$ in (6.1) ein. Die zusätzlichen Terme bringt man anschließend auf die rechte Seite der Gleichung. Dadurch erhält man zur Bestimmung von $\widetilde{u}$ wieder eine Konvektions–Diffusionsgleichung mit homogenen Dirichlet–Randbedingungen. Für diese kann dann der Satz angewandtwerden, woraus die eindeutige Lösbarkeit folgt. Diese Vorgehensweise bezeichnet man auch als Homogenisierung. Ob die Homogenisierung gelingt, hängt davon ab, ob eine Fortsetzung $\widetilde{g}$ existiert. Die Existenz der Fortsetzung ist dabei offenbar nur von dem

Gebietsrand und der Randbedingung g abhängig. Ausführlich wird die Homogenisierung von Differentialgleichungen zum Beispiel in [Cio99] behandelt.

(ii) Beispielsweise in [Wig70] wird der Fall von Neumann- oder Robin–Randbedingungen untersucht sowie Situationen, in denen der Typ der Randbedingung auf $\partial\Omega$ variiert.

(iii) Möchte man eine höhere Regularität der Lösung u sicherstellen (beispielsweise $u \in C^{2,\alpha}(\bar{\Omega})$), werden zusätzlich Kompatibilitäts-Bedingungen an den Ecken des Gebietes nötig (siehe hierzu [Azz80] oder [Gri85]).

Als Ausgangspunkt für das Studium weitere analytische Eigenschaften der Konvektions–Diffusionsgleichung sei [RST08] oder [GFL$^+$83] empfohlen.

6.2 Die Schwache Formulierung

Mit Satz 6.1 aus dem letzten Abschnitt kann die eindeutige Lösbarkeit der Konvektions–-Diffusionsgleichung nachgewiesen werden. In der Praxis sind allerdings oft die Voraussetzungen des Satzes verletzt und es fällt schwer, auf analytischem Weg Aussagen über die Lösbarkeit zu machen. Nicht selten existiert dann keine klassische Lösung mehr, sprich eine Lösung aus $C(\bar{\Omega}) \cap C^2(\Omega)$. Hinzu kommt: Selbst wenn eine klassische Lösung existiert, so ist es zumeist schwierig, diese mit rein analytischen Hilfsmitteln zu bestimmen.
In diesem Abschnitt werden wir die Konvektions–Diffusionsgleichung in ihre schwache Formulierung überführen. Die Lösungen des schwachen Problems sind im Allgemeinen nicht mehr Funktionen aus $C(\bar{\Omega}) \cap C^2(\Omega)$, sondern Sobolev– oder Lebesgue–Funktionen. Für solche Funktionenräume gibt es eine deutlich allgemeinere Lösbarkeitstheorie. Zudem entspricht die schwache Formulierung einem Variationsproblem. Wie in den letzten Kapiteln gesehen, können Variationsprobleme mit Galerkin–Verfahren approximiert werden. Diese bilden wiederum die Grundlage für die Finite–Elemente–Methoden. Somit dient die schwache Formulierung zugleich als Ausgangspunkt für Finite–Elemente–Methoden.

Wir gehen von der klassischen Formulierung der Konvektions–Diffusionsgleichung versehen mit relativ allgemeinen Randbedingungen aus:

$$-\epsilon\Delta u + b \cdot \nabla u + cu = f(x) \quad \text{in } \Omega, \tag{6.5a}$$
$$u = g_1 \quad \text{auf } \Gamma_1, \tag{6.5b}$$
$$\beta u + n \cdot \nabla u = g_2 \quad \text{auf } \Gamma_2. \tag{6.5c}$$

Die Bezeichnungen sind dabei wie im letzten Abschnitt. Die Neumann–Randbedingung behandeln wir diesmal nur als Spezialfall der Robin–Randbedingung. Die Daten seien für die schwache Formulierung ausreichend regulär vorausgesetzt, das heißt: $b, c \in L^\infty(\Omega)$, $f \in L^2(\Omega)$, $\beta \in L^\infty(\Omega)$, $g_i \in L^2(\Gamma_i)$.
Es soll eine geeignete Bestimmungsgleichung für eine schwache Lösung $u \in H^1(\Omega)$ aufgestellt werden. Multiplikation von (6.5a) mit einer beliebigen Testfunktion $v \in H^1(\Omega)$ mit $v|_{\Gamma_1} = 0$, Integration über Ω und Anwendung der partiellen Integration auf das Integral $(-\epsilon\Delta u, v)$ ergibt:

$$\epsilon(\nabla u, \nabla v) - \epsilon(n \cdot \nabla u, v)_{\Gamma_1} - \epsilon(n \cdot \nabla u, v)_{\Gamma_2} + (b \cdot \nabla u, v) + (cu, v) = (f, v). \tag{6.6}$$

Bei L^2-Produkten, die sich nicht über Ω erstrecken, hängen wir das Integrationsgebiet als Index an das Skalarprodukt $(\cdot, \cdot)$ an. Wir nutzen nun, dass die Testfunktion auf Γ_1

verschwindet, und setzen die Randbedingung (6.5c) in das Integral über Γ_2 ein:

$$\epsilon(\nabla u, \nabla v)+\epsilon(\beta u - g_2, v)_{\Gamma_2} + (b \cdot \nabla u, v) + (cu, v) = (f, v). \tag{6.7}$$

Dann bringen wir die Ausdrücke, in denen v, aber nicht u vorkommt, auf die rechte Seite der Gleichung und kürzen die linke Seite durch $a(u, v)$ sowie die rechte durch $(\widetilde{f}, v)$ ab. Die schwache Formulierung von (6.5) lautet damit:
Finde $u \in H^1(\Omega)$ mit $u = g_1$ auf Γ_1 und

$$a(u, v) = (\widetilde{f}, v) \qquad \forall v \in H^1(\Omega). \tag{6.8}$$

Gemäß der Herleitung der schwachen Formulierung ist jede klassische Lösung von (6.5) auch Lösung des schwachen Problems. Die Umkehrung gilt im Allgemeinen nicht; folgender Satz gibt jedoch ein hinreichendes Kriterium dafür.

Satz 6.3.
Seien c, f, β, g_1, g_2 stetige Funktionen, $b \in C^1(\bar{\Omega})$, Ω ein Lipschitzgebiet und sei eine Lösung $u \in C^2(\bar{\Omega})$ von (6.8) mit ausreichender Regularität bestimmt worden, welche zusätzlich der inhomogenen Dirichlet–Randbedingung $u = g_1$ auf Γ_1 genügt. Dann löst u auch das klassische Randwertproblem (6.5).

Beweis. Siehe [GR05] Satz 3.3. □

Bemerkung 6.4.
Der obige Satz besagt insbesondere, dass eine ausreichend glatte Lösung des schwachen Problems der Robin–Randbedingung genügt. Man zählt die Robin–Randbedingung daher zu den natürlichen Randbedingungen. Die Dirichlet–Randbedingungen ist hingegen vorausgesetzt worden, da sie nicht ohne Weiteres beim Übergang zur schwachen Formulierung berücksichtigt wird. In der schwachen Formulierung kann man die Dirichlet–Randbedingung sicherstellen, indem man den Ansatzraum auf Funktionen aus H^1 einschränkt, welche dieser Randbedingung genügen. Randbedingungen, welche im Ansatz- beziehungsweise im Testraum berücksichtigt werden, bezeichnet man auch als wesentliche Randbedingungen.

Der Einfachheit halber beschränken wir uns ab jetzt im Wesentlichen auf homogene Dirichlet–Randbedingungen. Die zugehörige schwache Formulierung lautet:
Finde $u \in V = H_0^1(\Omega)$ mit

$$\begin{aligned} a(u, v) &= f(v) \qquad \forall v \in V \quad \text{mit} \\ a(u, v) &= (\epsilon \nabla u, \nabla v) + (b \cdot \nabla u + cu, v) \qquad \text{und} \\ f(v) &= (f, v). \end{aligned} \tag{6.9}$$

Die Randbedingungen werden dabei im Ansatzraum berücksichtigt. Variationsprobleme vom Typ (6.9) haben wir bereits in Kapitel 3.1 diskutiert. Dort haben wir den Satz von Lax–Milgram als hinreichendes Kriterium für die eindeutige Existenz einer Lösung zur Verfügung gestellt. Unter geeigneten Voraussetzungen können wir diesen Satz auf die schwache Formulierung (6.9) anwenden:

Satz 6.5.
Sei Ω ein Lipschitzgebiet aus $\mathbb{R}^d$, $d \in \mathbb{N}$, $b, c \in L^\infty(\Omega)$, $f \in L^2(\Omega)$ und

$$c - \frac{1}{2} \nabla \cdot b \geq c_0 > 0 \qquad \text{auf } \Omega. \tag{6.10}$$

Dann besitzt die schwache Formulierung (6.9) eine eindeutige Lösung.

Beweis. Der Beweis erfolgt mit dem Satz von Lax–Milgram, siehe Satz 3.3. Seine Voraussetzungen sind erfüllt, denn:

(1.) H_0^1 ist ein Hilbertraum.

(2.) Das Funktional f ist nach Voraussetzung beschränkt und linear wegen der Linearität des L^2–Skalarproduktes.

(3.) Die Bilinearform $a(\cdot,\cdot)$ ist beschränkt, denn alle Terme von $a(\cdot,\cdot)$ lassen sich nach oben abschätzen:

$$\epsilon\,|(\nabla u, \nabla v)| \le \epsilon \|\nabla u\|_{L^2(\Omega)} \|\nabla v\|_{L^2(\Omega)} \le \|u\|_{H^1(\Omega)} \|v\|_{H^1(\Omega)}, \tag{6.11}$$

wobei die Cauchy–Schwarz–Ungleichung benutzt wurde.
Außerdem haben wir

$$\begin{aligned}
|(b\cdot\nabla u + cu, v)| &\le \|b\cdot\nabla u + cu\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)} \\
&\le (\|b\cdot\nabla u\|_{L^2(\Omega)} + \|cu\|_{L^2(\Omega)}) \|v\|_{L^2(\Omega)} \\
&\le (\|b\|_{L^\infty(\Omega)} \|u\|_{H^1(\Omega)} + \|c\|_{L^\infty(\Omega)} \|u\|_{L^2(\Omega)}) \|v\|_{L^2(\Omega)} \\
&\le (C\|u\|_{H^1(\Omega)} + C\|u\|_{L^2(\Omega)}) \|v\|_{L^2(\Omega)} \\
&\le C\|u\|_{H^1(\Omega)} \|v\|_{H^1(\Omega)}
\end{aligned} \tag{6.12}$$

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

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

$$\epsilon(\nabla u, \nabla u) \ge \epsilon\,|u|^2_{H^1(\Omega)}\,, \tag{6.13}$$

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

$$\begin{aligned}
(b\cdot\nabla u + cu, u) &= \int_\Omega u\, b\cdot\nabla u\, dx + \int_\Omega c\, u^2 dx \\
&= \int_\Omega \frac{1}{2}\, b\cdot\nabla(u^2) dx + \int_\Omega c\, u^2 dx \\
&= -\frac{1}{2}\int_\Omega (\nabla\cdot b) u^2\, dx + \int_\Omega c u^2 dx \\
&= \int_\Omega \underbrace{\left(c - \frac{1}{2}\nabla\cdot b\right)}_{\ge c_0} u^2 dx \\
&\ge c_0 \int_\Omega u^2 dx \\
&= c_0 \|u\|^2_{L^2(\Omega)}.
\end{aligned} \tag{6.14}$$

Dabei sind die Randintegrale gleich Null, da die Funktion u auf $\partial\Omega$ verschwindet.
Aus (6.13) und (6.14) ergibt sich

$$a(u,u) \ge C(\epsilon\,|u|^2_{H^1(\Omega)} + c_0\|u\|^2_{L^2(\Omega)}) = C\min\{\epsilon, c_0\}\|u\|^2_{H^1(\Omega)}. \tag{6.15}$$

Alle nötigen Voraussetzungen sind also erfüllt. □

Bemerkung 6.6.
Der Satz kann für allgemeinere Randbedingungen angewandt werden. Sind mit einer Funktion $g \in H^{1/2}(\partial\Omega)$ inhomogene Dirichlet–Randbedingungen vorgegeben, so setzt man mit Hilfe von Satz 8.8 aus [Wlo82] $g \in H^{1/2}(\partial\Omega)$ auf $\widetilde{g} \in H^1(\Omega)$ fort. Mit Hilfe der Fortsetzung $\widetilde{g}$ definiert man die Funktion $\widetilde{u} = u - \widetilde{g}$, setzt $u = \widetilde{u} + \widetilde{g}$ in die schwache Formulierung ein und bringt den Term mit $\widetilde{g}$ auf die rechte Seite. Das Resultat ist eine Bestimmungsgleichung der Form (6.9) mit homogenen Randbedingungen für $\widetilde{u}$. Für diese ist obiger Satz anwendbar, was die eindeutige Lösbarkeit des Problems mit inhomogenen Dirichlet–Randbedingungen liefert.

6.3 Das Standard–Galerkin–Verfahren

Um eine Näherungslösung der schwachen Formulierung (6.9) zu erhalten, verwenden wir Galerkin–Verfahren. Die Grundlagen hierzu haben wir in Kapitel 4 ausgearbeitet. Das Standard–Galerkin–Verfahren zur Lösung von (6.9) lautet:
Finde $u_h \in V_h$ mit

$$a(u_h, v_h) = f(v_h) \quad \forall v_h \in V_h^p. \tag{6.16}$$

Dabei sind a und f wie in der schwachen Formulierung. Für V_h^p wählen wir den Durchschnitt von $H_0^1(\Omega)$ mit einem polynomialen Finite–Element–Raum der Ordnung p auf dem Gebiet Ω, also P^p auf Dreieckselementen beziehungsweise Tetraedern oder die Räume Q^p auf Rechteckelementen beziehungsweise Hexaedern. Die Größe h steht hier für den maximalen Zellendurchmesser. Damit die Gebietszerlegung $\mathcal{T}_h$ zulässig im Sinne von Definition (5.3) gewählt werden kann, setzen wir Ω als polygonal beziehungsweise polyhedral berandet voraus. Nachfolgende Aussagen gelten gleichermaßen für die P^p- und Q^p-Elemente sowohl in zwei als auch in drei Dimensionen.
Da a koerzitiv ist, gibt es auch auf dem Unterraum V_h^p von H_0^1 eine eindeutige Lösung u_h. Es soll nun eine Fehlerabschätzung hergeleitet werden, welche darüber Auskunft gibt, wie gut u_h die exakte Lösung der schwachen Formulierung approximiert. Dazu benötigen wir insbesondere eine Interpolationsfehler–Abschätzug aus Abschnitt 5.4. Damit die Abschätzung benutzt werden kann, gehen wir davon aus, dass die Gebietszerlegung $\mathcal{T}_h$ quasi–uniform ist. Unter diesen Voraussetzugen gilt:

Satz 6.7.
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

$$c - \frac{1}{2}\nabla \cdot b \geq c_0 > 0 \qquad \text{auf } \Omega. \tag{6.17}$$

Sei ferner die Lösung u der schwachen Formulierung (6.9) ausreichend regulär, das heißt, $u \in H^p(\Omega) \cap H_0^1(\Omega)$. Dann gilt für die Lösung u_h des Standard–Galerkin–Verfahrens im Raum V_h^p die folgende Fehlerabschätzung

$$\|u - u_h\|_{H^1(\Omega)} \leq \frac{C}{\min\{c_0, \epsilon\}} h^p \, |u|_{H^{p+1}(\Omega)} \tag{6.18}$$

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

Beweis. Den Fehler zerlegen wir mit der Dreiecksungleichung in einen Interpolationsfehler- und Diskretisierungsfehler-Anteil:

$$\|u - u_h\|_{H^1(\Omega)} \leq \|u - u_I\|_{H^1(\Omega)} + \|u_I - u_h\|_{H^1(\Omega)}, \tag{6.19}$$

wobei u_I die Lagrange–Interpolierte von u ist. Den Interpolationsfehler schätzen wir mit Hilfe von Lemma 5.9 ab:

$$\|u - u_I\|_{H^1(\Omega)} \leq Ch^p \, |u|_{H^{p+1}(\Omega)} \, . \tag{6.20}$$

Wir benötigen nun noch eine Abschätzung für den Diskretisierungsfehler. Dazu gehen wir aus von

$$\|u_I - u_h\|^2_{H^1(\Omega)} \leq \frac{C}{\min\{c_0, \epsilon\}} a(u_I - u_h, u_I - u_h) = \frac{C}{\min\{c_0, \epsilon\}} a(u_I - u, u_I - u_h). \tag{6.21}$$

Die Gleichheit ergibt sich aus der Galerkin–Orthogonalität und die Ungleichung wegen der Koerzitivität von a, wobei der Vorfaktor aus (6.13) in Kombination mit (6.14) abgelesen werden kann. Die Terme der Bilinearform a schätzen wir getrennt voneinander nach oben ab: Wegen der Cauchy–Schwarz–Ungleichung und der Interpolationsfehler–Abschätzung (6.20) gelten

$$\begin{aligned} &\epsilon(\nabla(u_I - u), \nabla(u_I - u)) \\ &\qquad \leq \epsilon \, |u_I - u|_{H^1(\Omega)} \, |u_I - u_h|_{H^1(\Omega)} \\ &\qquad \leq \epsilon \, |u_I - u|_{H^1(\Omega)} \, \|u_I - u_h\|_{H^1(\Omega)} \\ &\qquad \leq \epsilon C h^p \, |u|_{H^{p+1}(\Omega)} \, \|u_I - u_h\|_{H^1(\Omega)} \end{aligned} \tag{6.22}$$

sowie die Abschätzung

$$\begin{aligned} &(b \cdot \nabla(u_I - u), u_I - u_h) \\ &\qquad \leq \|b \cdot \nabla(u_I - u)\|_{L^2(\Omega)} \|u_I - u_h\|_{H^1(\Omega)} \\ &\qquad \leq C \, |u_I - u|_{H^1(\Omega)} \, \|u_I - u_h\|_{H^1(\Omega)} \\ &\qquad \leq Ch^p \, |u|_{H^{p+1}(\Omega)} \, \|u_I - u_h\|_{H^1(\Omega)}. \end{aligned} \tag{6.23}$$

Den letzten Term beschränkt man durch

$$\begin{aligned} &(c(u_I - u), u_I - u_h) \\ &\qquad \leq C \, |u_I - u|_{H^0(\Omega)} \, \|u_I - u_h\|_{H^1(\Omega)} \\ &\qquad \leq Ch^{p+1} \, |u|_{H^{p+1}(\Omega)} \, \|u_I - u_h\|_{H^1(\Omega)} \end{aligned} \tag{6.24}$$

nach oben. Dividiert man in (6.21) durch $\|u_I - u_h\|_{H^1(\Omega)}$ und fasst alle Abschätzungen zusammen, so folgt die Behauptung. □

Gemäß obiger Fehlerabschätzung verschwindet der Fehler beim Übergang $h \to 0$. Gemessen am Interpolationsfehler geschieht dies mit der optimalen Konvergenzrate in h. Allerdings ist der Faktor auf der rechten Seite der Abschätzung proportional zu $1/\epsilon$. Im konvektions–dominanten Fall, in dem $0 < \epsilon \ll 1$ gilt, muss die Gitterweite h daher sehr klein gewählt werden, damit die im Satz gegebene Schranke brauchbar ist. Dies ist bei numerischen Rechnungen oft nicht möglich oder sinnvoll, da der Rechenaufwand zu groß wird. Rechnet man hingegen mit größeren Gitterweiten, so kann die Lösung im gesamten Lösungsgebiet unphysikalische Oszillationen aufweisen. Dies ist insbesondere bei Problemen mit Grenzschichten zu beobachten. In den nächsten Abschnitten werden wir modifizierte Galerkin–Verfahren vorstellen, welche bei dominanter Konvektion oftmals bessere Ergebnisse als das Standard–Galerkin–Verfahren liefern. Bei den modifizierten Galerkin–Verfahren ist es manchmal von

Vorteil, von einer Variante des Standard–Galerkin–Verfahrens auszugehen, bei welcher die homogene Randbedingung schwach behandelt wird. Wie Randbedingungen im schwachen Sinne behandelt werden können, diskutieren wir im Folgenden. Es sei aber erwähnt, dass oben geschilderte Probleme auch dort auftreten.

Bei der schwachen Behandlung der Randbedingungen lässt man im Approximationsraum V_h^p auch Funktionen zu, welche die Randbedingung nicht erfüllen. Für solche Funktionen lautet das Analogon zur Standard–Galerkin–Methode mit stark vorgegebenen Randbedingungen:
Finde $u \in V_h^p$ mit

$$\begin{aligned} &a(u_h, v_h) = f(v_h) \quad \forall v_h \in V_h^p, \qquad \text{wobei} \\ &a(u_h, v_h) = (\epsilon \nabla u_h, \nabla v_h) + (b \cdot \nabla u_h + c u_h, v_h) - \epsilon (n \cdot \nabla u_h, v_h)_\Gamma. \end{aligned} \tag{6.25}$$

Das letzte Integral bleibt hier stehen, da u auf dem Gebietsrand Γ ungleich Null sein darf. Dieses Galerkin–Verfahren besitzt den folgenden Mangel: Unter der Standard–Annahme $c - 1/2 \nabla \cdot b \geq c_0 > 0$ ist die Bilinerform a im Allgemeinen nicht mehr strikt koerzitiv und man hat keine eindeutige Lösbarkeit mehr.

Daher wird das Galerkin–Verfahren (6.25) durch die Hinzunahme dreier Terme in der Bilinearform a erweitert:

$$\begin{aligned} a_W(u_h, v_h) = \epsilon(\nabla u_h, \nabla v_h) + (b \cdot \nabla u_h + c u_h, v_h) - \epsilon (n \cdot \nabla u_h, v_h)_\Gamma \\ - \epsilon (u_h, n \cdot \nabla v_h)_\Gamma - (b \cdot n u_h, v_h)_{\Gamma_-} + \epsilon\gamma \sum_{E \subset \Gamma} \frac{1}{h_E} (u_h, v_h)_E. \end{aligned} \tag{6.26}$$

Dabei ist Γ_- der Einströmrand, an welchem $b \cdot n < 0$ gilt, γ ein neuer Parameter, E in zwei Dimensionen eine Kante der Gebietszerlegung und h_E die Länge dieser Kante (in drei Dimensionen gilt entsprechendes). Während der drittletzte Term für Symmetrie zwischen Ansatz- und Testfunktionen in den Randintegralen sorgt, sichern die beiden letzten Terme die Koerzitivität:

Lemma 6.8. Seien die Voraussetzungen von Satz 6.5 erfüllt und sei $\gamma > 0$ ausreichend groß. Dann gilt auf quasi–uniformen Gittern für die Bilinearform a_W aus (6.26) und alle $u_h \in V_h^p$ die Abschätzung

$$a_W(u_h, u_h) \geq \frac{1}{2} \left(\epsilon \, |u_h|^2_{H^1(\Omega)} + c_0 \|u_h\|^2_{L^2(\Omega)} + \| |b \cdot n|^{1/2} u_h \|^2_{L^2(\Gamma)} + \epsilon \sum_{E \subset \Gamma} \frac{1}{h_E} \|u_h\|^2_{L^2(E)} \right). \tag{6.27}$$

Da die Wurzel aus der rechten Seite dieser Abschätzung eine Norm auf V_h^p definiert, ist a_W strikt koerzitiv bezüglich dieser Norm.

Beweis. Wir behandeln zunächst einige Terme von a_W einzeln. Es gelten

$$\epsilon(\nabla u_h, \nabla u_h) \geq \epsilon \, |u_h|^2_{H^1(\Omega)}$$

sowie

$$\epsilon\gamma \sum_{E \subset \Gamma} \frac{1}{h_E} (u_h, u_h)_E = \epsilon\gamma \sum_{E \subset \Gamma} \frac{1}{h_E} \|u_h\|^2_{L^2(E)}.$$

Außerdem können wir abschätzen

$$\begin{aligned}(b\cdot\nabla u_h + cu_h, u_h) - (b\cdot nu_h, u_h)_{\Gamma_-} & \\ &= (c-\frac{1}{2}\nabla\cdot b, u_h^2) + \frac{1}{2}(b\cdot n, u_h^2)_\Gamma - (b\cdot nu_h, u_h)_{\Gamma_-} \\ &= (c-\frac{1}{2}\nabla\cdot b, u_h^2) + \frac{1}{2}(b\cdot n, u_h^2)_{\Gamma/\Gamma_-} + \frac{1}{2}(-b\cdot n, u_h^2)_{\Gamma_-} \\ &\geq c_0\|u_h\|^2_{L^2(\Omega)} + \frac{1}{2}\||b\cdot n|^{1/2}u_h\|^2_{L^2(\Gamma)}. \end{aligned} \tag{6.28}$$

Die Gleichungen folgen aus partieller Integration und Umschreiben der Randintegrale. Die Ungleichung folgt aus der Voraussetzung (6.10) sowie der Tatsache, dass $b\cdot n<0$ auf Γ_- gilt. Fassen wir alle bisherigen Aussagen zusammen, so ergibt sich:

$$\begin{aligned}a_W(u_h,u_h) \geq \epsilon\,|u_h|^2_{H^1(\Omega)} + c_0\|u_h\|^2_{L^2(\Omega)} + \frac{1}{2}\||b\cdot n|^{1/2}u_h\|^2_{L^2(\Gamma)} & \\ - 2\epsilon(n\cdot\nabla u_h, u_h)_\Gamma + \epsilon\gamma\sum_{E\subset\Gamma}\frac{1}{h_E}\|u_h\|^2_{L^2(E)}. & \end{aligned} \tag{6.29}$$

Den Minusterm schätzen wir noch beträgsmäßig nach oben ab:

$$2\epsilon\,|(n\cdot\nabla u_h, u_h)_E| \leq 2\epsilon h_E^{-1/2}\,|u_h|_{H^1(K)}\,\|u_h\|_{L^2(E)} \leq C(\frac{\epsilon}{2}\,|u_h|^2_{H^1(K)} + \frac{2\epsilon}{h_E}\|u_h\|_{L^2(E)}).$$

Die Abschätzungen folgen dabei aus der Cauchy–Schwarz–Ungleichung, der Spurungleichung (5.42) und der Relation $h_E\sim h_K$, welche wegen der Quasiuniformität des Gitters gilt. K ist dabei eine der Zellen, welche E als Kante beziehungsweise Seitenfläche besitzt. Summation über alle $E\subset\Gamma$ liefert zusammen mit der Wahl $\gamma\geq C(\frac{1}{2}+2)$ die gesuchte Abschätzung. Dass die Wurzel der rechten Seite von 6.27 eine Norm ist, rechnet man unter Verwendung der Normeigenschaften der einzelnen Terme direkt nach. □

Aufgrund der Koerzitivität liefert der Satz von Lax–Milgram die eindeutige Lösbarkeit des Standard–Galerkin–Verfahren mit schwach vorgegebenen Randbedingungen. Die Lösung wird im Allgemeinen nicht den vorgegebenen homogenen Randbedingungen gehorchen. Manchmal macht es allerdings Sinn, einen Fehler am Rand in Kauf zu nehmen, wenn die Näherung dafür in anderen Bereichen besser wird als diejenige des Galerkin–Verfahrens mit stark vorgegebenen Randbedingungen (siehe zum Beispiel die Arbeiten [FS95] und [BH07]). Darüber hinaus gelingen zuweilen theoretische Untersuchungen leichter mit schwach vorgegebenen Randbedingungen, wie zum Beispiel die Herleitung von Fehlerabschätzungen. Hierzu werden wir in den nächsten Abschnitten ein Beispiel sehen.

Bemerkung 6.9.
(1.) Eine klassische Lösung u der Konvektions–Diffusionsgleichung, $u\in H^1_0\cap C^2(\Omega)$, ist weiterhin Lösung des Standard–Galerkin–Verfahrens mit schwach vorgegebenen Randbedingungen, denn die künstlich hinzugenommenen Randterme verschwinden für u aus H^1_0.

(2.) In dem Paper [Nit72] von Nitsche sind erstmals schwach vorgegebene Randbedingungen diskutiert worden. Sie werden daher manchmal auch als Randbedingungen vom Nitsche–Typ bezeichnet.

(3.) Es gibt mehrere Varianten des Standard–Galerkin–Verfahrens mit schwach vorgegebenen Randbedingungen. Zum Beispiel ist es nicht zwingend erforderlich, die Randterme zu symmetrisieren. Einige weitere Varianten findet man in [Fai78].

6.4 Die Streamline–Diffusion–Methode

Wie im letzten Abschnitt erwähnt, sind die Ergebnisse des Standard–Galerkin–Verfahrens im konvektions-dominanten Fall oft nicht zufriedenstellend. Für diesen Fall benötigt man neue Methoden. Eine solche Methode ist die Streamline–Diffusion–Methode, auch streamline upwind Petrov–Galerkin–Methode genannt oder kurz SUPG–Methode. Die Idee dieser Methode besteht darin, die StandardFormulierung um Terme zu erweitern, welche künstlich für stärkere Diffusion sorgen. Numerische Rechnungen werden durch die künstliche Diffusion stabilisiert, wodurch unphysikalische Oszillationen der Näherungslösung verringert werden. Anders als zum Beispiel bei den Upwind–Verfahren [1] wird die künstliche Diffusion bei der Streamline–Diffusion–Methode nur in Richtung der Stromlinien aufgeprägt. Die künstliche Diffusion wirkt so in die Richtung, in welcher sie hauptsächlich gebraucht wird. Im Vergleich zu den Upwind–Verfahren, welche die künstliche Diffusion nicht derart gerichtet aufprägen, gelingt die Stabilisierung bereits mit geringerer künstlicher Diffusion. Dies ist geschickter, da man bei zu starker künstlicher Diffusion schlechte, verschmierte Lösung erhält. Die Streamline–Diffusion–Methode wurde von Brooks und Hughes in [HB79] entwickelt.

Wir gehen im Folgenden von den gleichen Voraussetzugen wie bei dem Standard–Galerkin–Verfahren mit stark vorgegebenen Randbedingungen aus, das heißt insbesondere: Das Gebiet $\Omega \subset \mathbb{R}^2$ sei polygonal berandet (oder polyhedral in drei Dimensionen), der Approximationsraum sei gegeben durch $V_h^p = P^p \cap H_0^1(\Omega)$ oder $V_h^p = Q^p \cap H_0^1(\Omega)$ und die Gebietszerlegung $\mathcal{T}_h$ sei quasi–uniform. Dann lautet die Streamline–Diffusion–Methode:
Finde $u_h \in V_h^p$, so dass für alle $v_h \in V_h^p$ gilt:

$$a_{SD}(u_h, v_h) = f_{SD}(v_h), \qquad \text{wobei} \tag{6.30a}$$

$$a_{SD}(u_h, v_h) = a(u_h, v_h) + \sum_{K \in \mathcal{T}_h} \delta_K \left(-\epsilon \Delta u_h + b \cdot \nabla u_h + c u_h, b \cdot \nabla v_h\right)_K, \tag{6.30b}$$

$$f_{SD}(v_h) = f(v_h) + \sum_{K \in \mathcal{T}_h} \delta_K \left(f, b \cdot \nabla v_h\right)_K \tag{6.30c}$$

mit $\delta_K \geq 0$. Das Standard–Galerkin–Verfahren ist hier durch einen Term der Form

$$\sum_{K \in \mathcal{T}_h} (\mathrm{Res}, b \cdot \nabla v_h) \tag{6.31}$$

ergänzt worden, wobei Res das Residuum der Konvektions–Diffusionsgleichung bezeichnet. Für eine ausreichend glatte Lösung $u \in H^2(\Omega) \cap H_0^1$ der schwachen Formulierung verschwindet das Residuum, so dass u ebenfalls (6.30a) erfüllt. Diese Eigenschaft bezeichnet man als Konsistenz. Aus der Konsistenz eines Verfahrens folgt unmittelbar seine Galerkin–Orthogonalität:

$$a_{SD}(u - u_h, v_h) = 0 \qquad \forall v_h \in V_h. \tag{6.32}$$

Dazu betrachtet man (6.30a) einmal mit der Lösung u sowie einmal mit u_h und subtrahiert beide Gleichungen voneinander.

Als nächstes soll mit Hilfe des Satzes von Lax–Milgram untersucht werden, unter welchen Voraussetzungen die Streamline–Diffusion–Methode eine eindeutige Lösung besitzt. Die

[1] Die Upwind–Verfahren können als Vorreiter der Streamline–Diffusion–Methode angesehen werden. Einige Arbeiten zu diesen Methoden sind zum Beispiel [CGMZ76], [Tab77] und [HHZM77].

kritischste Voraussetzung ist hierbei die Koerzitivität von a_{SD}. Um die Koerzitivität zu prüfen, führen wir die folgende Norm ein:

$$\|\!|u|\!\|_{SD} := \left(\epsilon \, |u|^2_{H^1(\Omega)} + \sum_{K \in \mathcal{T}_h} \delta_K \|b \cdot \nabla u\|^2_{L^2(K)} + c_0 \|u\|^2_{L^2(K)} \right)^{1/2}. \tag{6.33}$$

Mit dieser Norm zeigen wir nun:

Satz 6.10. Seien $c_K = \max_{x \in K} |c(x)|$ und

$$c - \frac{1}{2} \nabla \cdot b \geq w > 0 \qquad \text{auf } \Omega. \tag{6.34}$$

Der SD-Parameter δ_K erfülle

$$0 < \delta_K < \frac{1}{2} \min \left\{ \frac{c_0}{c_K^2}, \frac{h_K^2}{\epsilon \mu_{inv}^2} \right\} \tag{6.35}$$

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

$$a_{SD}(v_h, v_h) \geq \frac{1}{2} \|\!|v_h|\!\|^2_{SD} \qquad \forall v_h \in V_h. \tag{6.36}$$

Beweis. Wir formen zunächst um und schätzen dann ab:

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

Die Ungleichung ergibt sich hierbei aus derselben Rechnung wie in (6.14). Den letzten Term schätzen wir nun betragsmäßig nach oben ab, um ihn anschließend von den restlichen zu subtrahieren:

$$\begin{aligned}
&\left| \sum_{K \in \mathcal{T}_h} \delta_K \left(-\epsilon \Delta v_h + c v_h, b \cdot \nabla v_h \right)_K \right| \\
&\qquad \leq \sum_{K \in \mathcal{T}_h} \epsilon^2 \delta_K \|\Delta v_h\|^2_{L^2(K)} + \sum_{K \in \mathcal{T}_h} c_T^2 \delta_K \|v_h\|^2_{L^2(K)} + \frac{1}{2} \sum_{K \in \mathcal{T}_h} \|b \cdot \nabla v_h\|^2_{L^2(K)} \\
&\qquad \leq \frac{\epsilon}{2} \, |v_h|^2_{H^1(\Omega)} + \frac{c_0}{2} \|v_h\|^2_{L^2(\Omega)} + \frac{1}{2} \sum_{K \in \mathcal{T}_h} \|b \cdot \nabla v_h\|^2_{L^2(K)}.
\end{aligned} \tag{6.37}$$

Die erste Abschätzung folgt, da für jedes Skalarprodukt 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 (6.35) an δ_K und der lokal inversen Ungleichung (5.39). Schließlich erhalten wir aus (6.37) die Behauptung:

$$\begin{aligned}a_{SD}(v_h,v_h) \geq{}& \epsilon\,|v_h|^2_{H^1(\Omega)} + c_0\|v_h\|^2_{L^2(\Omega)} + \sum_{K\in\mathcal{T}_h}\delta_K\|b\cdot\nabla v_h\|^2_{L^2(K)} \\ &- \left|\sum_{K\in\mathcal{T}_h}\delta_K\left(-\epsilon\Delta v_h + cv_h, b\cdot\nabla v_h\right)_K\right| \\ \geq{}& \frac{1}{2}\left(\epsilon\,|v_h|^2_{H^1(\Omega)} + c_0\|v_h\|^2_{L^2(\Omega)} + \sum_{K\in\mathcal{T}_h}\|b\cdot\nabla v_h\|^2_{L^2(K)}\right). \qquad \square \qquad (6.38)\end{aligned}$$

Die restlichen Annahmen des Satz von Lax–Milgram prüft man hingegen leicht. So folgt zum Beispiel die Beschränktheit von f aus der Beschränktheit von b und der Cauchy–-Schwarz–Ungleichung. Damit liefert obiger Satz ein hinreichendes Kriterium für die eindeutige Lösbarkeit der Streamline–Diffusion–Methode. Nun stellt sich wieder Frage, wie gut diese Lösung die ursprüngliche Lösung der schwachen Formulierung (6.9) approximiert. Im nächsten Lemma schätzen wir dazu den Diskretisierungsfehler des Verfahrens ab:

Lemma 6.11. Seien die Voraussetzungen von Satz 6.10 erfüllt und sei $u \in H^{p+1}(\Omega)\cap H^1_0(\bar{\Omega})$ mit $p\geq 1$. Dann lässt sich der Diskretisierungfehler abschätzen durch

$$\|\!|u^I - u_h|\!\|_{SD} \leq C\left[\sum_{K\in\mathcal{T}_h}(\epsilon+\delta_K+\delta_K^{-1}h_K^2+h_K^2+h_K(1+\delta_K)+h_K^2(1+\delta_K))h_K^{2p}\,|u|^2_{H^{p+1}(K)}\right]^{1/2} \qquad (6.39)$$

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

Beweis. Der Beweis orientiert sich an den Ausführungen von [RST08] Seite 305. Wegen Satz 6.10 und der Galerkin–Orthogonalität (6.32) der Streamline–Diffusion–Methode können wir schreiben

$$\begin{aligned}\frac{1}{2}\|\!|u^I-u_h|\!\|_{SD} &\leq a_{SD}(u^I-u_h,u^I-u_h) \\ &= a_{SD}(u^I-u+u-u_h,u^I-u_h) \\ &= a_{SD}(u^I-u,u^I-u_h) + a_{SD}(u-u_h,u^I-u_h) \\ &= a_{SD}(u^I-u,u^I-u_h).\end{aligned}$$

Die Terme von a_{SD} werden nun einzeln abgeschätzt. Für den ersten Term in (6.30b) haben wir:

$$\begin{aligned}\epsilon(\nabla(u^I-u),\nabla(u^I-u_h)) &\leq \epsilon^{1/2}\left|u^I-u\right|_{H^1(\Omega)}\epsilon^{1/2}\left|u^I-u_h\right|_{H^1(\Omega)} \\ &\leq \epsilon^{1/2}\left|u^I-u\right|_{H^1(\Omega)}\|\!|u^I-u_h|\!\|_{SD} \\ &\leq C\epsilon^{1/2}h^p\left|u\right|_{H^{p+1}(\Omega)}\|\!|u^I-u_h|\!\|_{SD}.\end{aligned}$$

Zunächst wurde dabei die Cauchy–Schwarz–Ungleichung benutzt, dann die Definition der SD-Norm und schließlich die Interpolationsfehler–Abschätzung (5.31).
Um den zweiten und dritten Term nach oben zu beschränken, führen wir zunächst eine kurze Zwischenrechnung durch. Für zwei beliebige Funktionen g, h aus $H^1(\Omega)$ mit $h|_{\partial\Omega} = 0$ gilt unter Verwendung von partieller Integration:

$$\begin{aligned}(b \cdot \nabla g + cg, h) &= (\nabla \cdot (bg), h) - (g\nabla \cdot b, h) + (cg, h) \\ &= ((c - \nabla \cdot b)g, h) + (\nabla \cdot (bg), h) \\ &\underset{\text{p.I.}}{=} ((c - \nabla \cdot b)g, h) - (g, b \cdot \nabla h).\end{aligned}$$

Damit erhalten wir für den zweiten und dritten Term von a_{SD}:

$$\begin{aligned}&(b \cdot \nabla(u^I - u) + c(u^I - u), u^I - u_h) \\ &\quad = ([c - \nabla \cdot b][u - u^I], u^I - u_h) \; - \; (u^I - u, b \cdot \nabla[u^I - u_h]) \\ &\quad \leq C \left(\left(\sum_{K \in \mathcal{T}_h} \|u^I - u\|^2_{L^2(K)} \right)^{1/2} + \left(\sum_{K \in \mathcal{T}_h} \delta_K^{-1} \|u^I - u\|^2_{L^2(K)} \right)^{1/2} \right) |\!|\!| u^I - u_h |\!|\!|_{SD} \\ &\quad \leq Ch^p \left(\sum_{K \in \mathcal{T}_h} h_K^2 (1 + \delta_K^{-1}) \, |u|^2_{H^{p+1}(K)} \right)^{1/2} |\!|\!| u^I - u_h |\!|\!|_{SD}.\end{aligned}$$

Die erste Abschätzung folgt aus der Cauchy-Schwarz- und der Young-Ungleichung, die zweite aus der Interpolationsfehler–Abschätzung (5.31). Den lezten Term von a_{SD} beschränken wir schließlich wie folgt:

$$\begin{aligned}&\left| \sum_{K \in \mathcal{T}_h} \delta_K (-\epsilon \Delta[u^I - u_h] + b \cdot \nabla(u^I - u_h) + c(u^I - u_h), b \cdot \nabla(u^I - u_h))_K \right| \\ &\quad \leq \sum_{K \in \mathcal{T}_h} \delta_K^{1/2} \left(\epsilon \left| u^I - u_h \right|_{H^2(K)} + C \left| u^I - u_h \right|_{H^1(K)} + C \left| u^I - u_h \right|_{L^2(K)} \right) \\ &\qquad\qquad \times \delta_K^{1/2} \|b \cdot \nabla(u^I - u_h)\|^2_{L^2(K)} \\ &\quad \leq C \sum_{K \in \mathcal{T}_h} \delta_K^{1/2} (\epsilon h_K^{p-1} + h_K^p + h_K^{p+1}) \, |u|_{H^{p+1}(K)} \; \delta_K^{1/2} \|b \cdot \nabla(u^I - u_h)\|_{L^2(K)} \\ &\quad \leq C \left(\sum_{K \in \mathcal{T}_h} \delta_K (\epsilon h_K^{p-1} + h_K^p + h_K^{p+1})^2 \, |u|^2_{H^{p+1}(K)} \right)^{1/2} \left(\sum_{K \in \mathcal{T}_h} \delta_K \|b \cdot \nabla(u^I - u_h)\|^2_{L^2(K)} \right)^{1/2} \\ &\quad \leq C \left(\sum_{K \in \mathcal{T}_h} \delta_K h_K^{2p} (\epsilon h_K^{-1} + 1 + h_K)^2 \, |u|^2_{H^{p+1}(K)} \right)^{1/2} |\!|\!| u^I - u_h |\!|\!|_{SD} \\ &\quad \leq Ch^p \left(\sum_{K \in \mathcal{T}_h} \left(\epsilon + \delta_K + h_K(1 + \delta_K) + h_K(1 + \delta_K) + h_K^2(1 + \delta_K) \right) |u|^2_{H^{p+1}(K)} \right)^{1/2} \\ &\qquad\qquad \times |\!|\!| u^I - u_h |\!|\!|_{SD}.\end{aligned}$$

Die erste Abschätzung ergibt sich hierbei aus Kombination der Cauchy–Schwarz- und der Dreiecksungleichung, die zweite aus der Interpolationsfehler–Abschätzung (5.31), die

dritte aus der Tatsache, dass für $a_i, b_i \geq 0$ wegen der Young–Ungleichung $\sum_i a_i b_i \leq (\sum_i a_i^2)^{1/2}(\sum_i b_i^2)^{1/2}$ gilt, und die letzte aus $\delta_K \epsilon \leq C h_K^2$, was aus der Annahme (6.35) an δ_K folgt. Fassen wir die Abschätzungen der drei Terme von a_{SD} zusammen, so ist der Beweis vollständig. □

Um aus dem vorherigen Lemma die bestmögliche Konvergenzordnung abzuleiten, werden die Terme ϵ, δ_K, $\delta_K^{-1} h_K^2$, h_K^2, $h_K(1+\delta_K)$, $h_K^2(1+\delta_K)$ gegeneinander balanciert. Dazu setzen wir

$$\delta_K = \begin{cases} \delta_0 h_K, & \text{falls } \mathrm{Pe}_K > 1 \quad \text{(Konvektions–dominanter Fall)}, \\ \delta_1 h_K^2/\epsilon, & \text{falls } \mathrm{Pe}_K \leq 1 \quad \text{(Diffusions–dominanter Fall)}, \end{cases} \tag{6.40}$$

wobei δ_0, δ_1 zwei positive Konstanten sind und Pe_K die lokale Pécletzahl abkürzt, das heißt,

$$\mathrm{Pe}_K = \frac{\|b\|_{L^\infty(K)} h_K}{2\epsilon}. \tag{6.41}$$

Mit dieser Wahl des SD–Parameters gelangen wir zu folgender Fehlerabschätzung:

Satz 6.12 (**Globale Fehlerabschätzung der Streamline–Diffusion–Methode**)**.** Der SD–Parameter δ_K sei gemäß (6.40) gewählt und erfülle die Bedingung von Satz 6.10. Die Gitterweite h_K sei 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

$$|\!|\!| u - u_h |\!|\!|_{SD} \leq C(\epsilon^{1/2} + h^{1/2}) h^p \, |u|_{H^{p+1}(\Omega)} \tag{6.42}$$

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

Beweis. Den globalen Fehler zerlegen wir mit der Dreiecksungleichung in einen Interpolationsfehler- und einen Diskretisierungfehler-Anteil:

$$|\!|\!| u - u_h |\!|\!|_{SD} \leq |\!|\!| u^I - u_h |\!|\!|_{SD} + |\!|\!| u - u^I |\!|\!|_{SD}, \tag{6.43}$$

mit der Lagrange–Interpolation u_I von u. Den Diskretisierungsfehler können wir mit dem letzten Lemma beschränken. Wenn wir dabei die Wahl (6.40) für den SD–Parameters verwenden und alle Terme der Ordnung $\mathcal{O}(h^{p+\frac{1}{2}})$ zusammenfassen, folgt

$$|\!|\!| u^I - u_h |\!|\!|_{SD} \leq C(\epsilon^{1/2} + h^{1/2}) h^p \, |u|_{H^{p+1}(\Omega)} \, . \tag{6.44}$$

Der Interpolationsfehler wird wiederum durch Abschätzung (5.31) kontrolliert:

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

Beide Abschätzungen in (6.43) einsetzen liefert den Beweis. □

Bemerkung 6.13.

(i) Der wesentliche Vorteil dieser Fehlerabschätzung gegenüber derjenigen des Standard–Galerkin–Verfahrens ist, dass auf der rechten Seite keine Terme negativer Potenz in ϵ auftreten. Damit bleibt diese Abschätzung auch im konvektions-dominanten Fall brauchbar.

(ii) Die Interpolationsfehler–Abschätzung (5.31) gibt in der L^2-Norm eine optimale Konvergenzrate der Ordnung $\mathcal{O}(h^{p+1})$ vor. Die im Satz gegebene Fehlerabschätzung verfehlt diese Rate im konvektions– und im diffusions–dominanten Fall um den Faktor $h^{1/2}$. Für die H^1-Norm wäre eine optimale Konvergenzrate von der Ordnung $\mathcal{O}(h^p)$. Im diffusions–dominanten Fall wird diese erreicht, während im konvektions–dominanten Fall eine Ordnung von $\mathcal{O}(h^p\epsilon^{-1/2})$ realisiert wird.

(iii) Um die leichte Suboptimaltität der Streamline–Diffusion–Methode in der L^2–Norm zu beheben, wird beispielsweise in den Arbeiten [JSW87] und [ZR96] eine Erweiterung des Verfahrens vorgestellt. Hier wird künstliche Diffusion in kontrollierter Weise auch senkrecht zu den Stromlinien aufgeprägt. Für diese Methoden sind optimale Konvergenzraten bewiesen worden [Näv92], [Zho97]. Allerdings gelingt dies nur für spezielle Gitter und Lösungen mit höherer Regularität ($u \in H^5(\Omega)$).

(iv) Jede klassische Lösung der Konvektions–Diffusionsgleichung genügt einem Maximumsprinzip. Die Lösungen der Streamline–Diffusion–Methode besitzten diese Eigenschaft hingegen im Allgemeinen nicht. In der Folge beobachtet man bei numerischen Tests folgendes Verhalten: Obwohl es der Streamline–Diffusion–Methode gelingt, unphysikalische Oszillationen im Großteil des Lösungsgebiet zu unterdrücken, treten diese weiterhin in den Grenzschichten auf, also dort, wo sich die Lösung schnell ändert. Um dieses Problem zu beheben, sind Varianten der Streamline–Diffusion–Methode entwickelt worden, welche mittels so gennanter shock–capturing Terme die unerwünschte Oszillationen in den Grenzschichten unterdrücken (siehe zum Beispiel [TP86], [BE02], [JK07a] und [JK07b]). Die shock capturing Terme sind nichtlinear, was den Rechenaufwand vergrößert und zudem die Analysis erschwert. In der englischen Literatur sind diese Methoden auch unter dem Namen spurious oscillations at layers diminishing (SOLD) methods bekannt. In dieser Arbeit werden wir uns auf die Standardvariante der Streamline–Diffusion–Methode beschränken.

(v) Die Streamline–Diffusion Methode besitzt mit δ_K einen Parameter, der frei im durch Satz 6.10 festgelegten Intervall variiert werden kann, ohne dass sich die Fehlerabschätzung des letzten Satzes ändert. Dennoch wird die Güte numerischer Resultate durch die Wahl von δ_K beeinflusst. Es fragt sich also, wie δ_K optimalerweise gewählt werden sollte. Arbeiten in diese Richtung sind zum Beispiel [ST95] und [SE99]. Trotz der dort präsentierten Ergebnisse ist es im Allgemeinen unklar, welches δ_K optimal ist. Dies bleibt auch bei unseren Simulationen zu testen.

6.5 Die Kanten–Stabilisierung

Neben der Streamline–Diffusion–Methode verwenden wir in dieser Arbeit noch ein weiteres Stabilisierungsverfahren, die Kanten–Stabilisierung, im Englischen auch edge stabilization oder continuous interior penalty (CIP) stabilization method genannt. Sie besitzt gegenüber der Streamline–Diffusion–Methode den Vorteil, dass keine neuen unsymmetrischen Terme eingeführt werden und keine Ableitungen zweiter Ordnung berechnet werden müssen. Außerdem lässt sie die mass lumping Technik zu. Der Nachteil ist eine größere Anzahl

von Null verschiedener Einträge in der Systemmatrix. Der Grundgedanke der Kanten–Stabilisierung liegt darin, Unstetigkeiten in der ersten Ableitung der Näherungslösung zu bestrafen. Wird eine Lösung der schwachen Formulierung mit stetigen ersten Ableitungen genähert, so resultiert daraus ein stabilisierender Effekt. Die Idee der Kanten–Stabilisierung geht auf Douglas und Dupont [DD75] zurück und hat durch die Arbeiten von Burman und Hansbo neuerdings wieder an Bedeutung gewonnen [BH04], [BH06], [BFH06]. In diesem Abschnitt stellen wir die theoretischen Grundlagen der Kanten–Stabilisierungs–Methode vor (im Folgenden auch CIP-Methode genannt). Der Schwerpunkt liegt dabei darauf, die eindeutige Lösbarkeit der CIP-Methode zu untersuchen und eine globale Abschätzung für den Verfahrensfehler herzuleiten.

Wir gehen von einem polygonal oder in drei Dimensionen polyhedral berandeten Gebiet $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$ aus, zu welchem eine quasi–uniforme Gebietszerlegung gewählt wurde. Als Finite–Elemente–Räume werden die Räume $P^p(\Omega)$ und $Q^p(\Omega)$ benutzt. Beide werden durch V_h^p abgekürzt, wobei h für den maximalen Zellendurchmesser steht. Es hat sich gezeigt, siehe beispielsweise [BFH06] oder [BBJL07], dass die Analysis zur CIP-Methode einfacher mit schwach vorgegebenen Randbedingungen gelingt. Daher ist die Formulierung (6.25) mit der Bilinearform a_W aus (6.26) unser Ausgangspunkt, bei welcher die homogene Randbedingung nicht in den Finite–Element–Raum V_h^p integriert worden ist. Diese geht durch die Hinzunahme eines weiteren Stabilisierungsterms über in die CIP-Methode, welche lautet:
Finde $u_h \in V_h^p$, so dass für alle $v_h \in V_h^p$ gilt:

$$a_{\mathrm{CIP}}(u_h, v_h) = (f, v_h) \quad \text{mit} \tag{6.45a}$$

$$a_{\mathrm{CIP}}(u_h, v_h) = a_W(u_h, v_h) + j_h(u_h, v_h). \tag{6.45b}$$

Dabei ist j_h der Stabilisierungsterm der Methode. Dieser hat die Form

$$j_h(u_h, v_h) = \sum_{E \in \mathcal{E}_h} \tau h_E^2 \left(b_h \cdot [\nabla u_h]_E, b_h \cdot [\nabla v_h]_E\right)_E . \tag{6.46}$$

Die Summe erstreckt sich über die Menge $\mathcal{E}_h$ aller innerer Kanten (beziehungsweise Seitenflächen), τ ist ein neuer Parameter und $[w]_E$ ist der Sprung der Größe w über die Kante E in Richtung ihres Einheits-Normalenvektors n_E, das heißt:

$$[w]_E(x) = \lim_{s \to 0} \left(w(x + s n_E) - w(x - s n_E)\right) \quad \text{für } x \in E. \tag{6.47}$$

Da es zu jeder Kante zwei Normalenvektoren gibt, ist obige Definition nicht eindeutig. Die Normalenvektoren unterscheiden sich jedoch nur um ein Vorzeichen, so dass dies auch für die Definition des Sprungs gilt und sein Betrag im Stabilisierungsterm j_h eine wohldefinierte Größe ist. Für stückweise polynomialen Funktionen ist der Sprung im Punkt $x \in E$ einfach durch

$$[w]_E(x) = w_1(x) - w_2(x) \tag{6.48}$$

gegeben, wobei w_i jeweils die Einschränkungen von w auf eine der beiden Zellen ist, welche E als Kante besitzen. Der Stabilisierungsterm j_h ist symmetrisch und beinhaltet nur erste Ableitungen. Dies sind zwei Vorteile gegenüber der Streamline–Diffusion–Methode. Der Nachteil liegt darin, dass der Stabilisierungsterm Integrale über die inneren Kanten enthält. Über diese werden die Freiheitsgrade benachbarter Zellen miteinander verkoppelt (wie sehen wir in Abschnitt 7.1 genauer). Durch die Kopplung vergrößert sich die Anzahl der von Null verschiedenen Einträge in der Systemmatrix der Kanten–Stabilisierung, was den Rechenaufwand erhöht.

Die Größe b_h in (6.46) bezeichnet eine stückweise lineare Approximation von b bezüglich der Gebietszerlegung, also zum Beispiel die lineare Lagrange–Interpolierte. Wir setzen für die nachfolgende Analysis $b \in W^{1,\infty}(\Omega)$ voraus, so dass für b_h gilt:

$$\|b - b_h\|_{L^\infty(K)} \leq C h_K \|b\|_{W^{1,\infty}(K)}. \tag{6.49}$$

Für eine ausreichend reguläre Lösung $u \in H^{3/2+\epsilon}(\Omega),\ \epsilon > 0$, der schwachen Formulierung ist der Sprung von $[\nabla u]_E$ an allen inneren Kanten fast überall bezüglich des $(d-1)$-dimensionalen L^2–Maßes gleich Null. Der Stabilisierungterm verschwindet daher in diesem Fall. Zudem fallen die Randterme in a_W weg. Damit ist jede ausreichend reguläre Lösung u der schwachen Formulierung auch Lösung der CIP-Methode. Dies impliziert die Galerkin–Orthogonalität von u bezüglich der Lösung u_h von (6.45a) und a_{CIP}:

$$a_{\mathrm{CIP}}(u - u_h, v_h) = 0 \qquad \forall v_h \in V_h^p. \tag{6.50}$$

Bezüglich der Norm

$$|\!|\!| u |\!|\!|_{\mathrm{CIP}} = \left(\epsilon \|u\|^2_{H^1(\Omega)} + c_0 \|u\|^2_{L^2(\Omega)} + j_h(u,u) + \||b\cdot n|^{1/2} u\|^2_{L^2(\Gamma)} + \epsilon \sum_{E \subset \Gamma} \frac{1}{h_E} \|u\|^2_{L^2(E)} \right)^{1/2} \tag{6.51}$$

macht das folgende Lemma eine Aussage über die Koerzitivität von a_{CIP}:

Lemma 6.14.
Seien die Daten der schwachen Formulierung ausreichend regulär, das heißt, $b \in W^{1,\infty}(\Omega)$ sowie $c \in L^\infty(\Omega)$, $\gamma > 0$ groß genug und die Gebietszerlegung quasi–uniform. Außerdem gelte die Ungleichung

$$c - \frac{1}{2} \nabla b \geq c_0 > 0 \quad \text{auf } \Omega. \tag{6.52}$$

Dann gilt für alle $u_h \in V_h^p$ die Abschätzung

$$a_{\mathrm{CIP}}(u_h, u_h) \geq \frac{1}{2} |\!|\!| u_h |\!|\!|^2_{\mathrm{CIP}}. \tag{6.53}$$

Insbesondere ist dann die Bilinearform a_{CIP} strikt koerzitiv auf V_h^p.

Beweis. Der Beweis folgt direkt, indem man die Abschätzung aus Lemma 6.8 benutzt und auf der linken Seite der Abschätzung den Term $j_h(u_h, u_h) > 0$ hinzuaddiert sowie auf der rechten Seite die Hälfte davon. □

Unter den Voraussetzungen dieses Lemmas sowie der zusätzlichen Annahme $f \in L^2(\Omega)$ sind alle Voraussetzungen des Satz von Lax–Milgram erfüllt und wir haben ein hinreichendes Kriterium für die Existenz einer eindeutigen Lösung der CIP-Methode zur Hand. Als nächstes wollen wir eine Abschätzung des Verfahrensfehlers herleiten. Eine Schlüsselstelle im Beweis dieser Abschätzung ist die Existenz eines Interpolationsoperators wie er in nachfolgendem Lemma vorgestellt wird. Mit Hilfe dieses Operators werden wir eine beinahe optimale Fehlerabschätzung für den konvektiven Term angeben können.

Lemma 6.15.
Es gibt einen Interpolationsoperator $\pi^* : H^2(\mathcal{T}_h) \to V_h^p$, so dass für alle $v_h \in V_h^p$ und alle $K \in \mathcal{T}_h$ gilt

$$h_K \|b_h \cdot \nabla v_h - \pi^*(b_h \cdot \nabla v_h)\|^2_{L^2(K)} \leq C \sum_{E \in \mathcal{E}_h(K)} \int_E h_E^2 \left| b_h \cdot [\nabla v_h]_E \right|^2 ds \tag{6.54}$$

mit einer Konstante $C > 0$ und $\mathcal{E}_h(K) := \{E \in \mathcal{E}_h : \ E \cap K \neq \emptyset\}$.
$H^2(\mathcal{T}_h)$ bezeichnet dabei den Raum aller bezüglich der Gebietszerlegung $\mathcal{T}_h$ stückweisen H^2-Funktionen, also

$$H^2(\mathcal{T}_h) = \left\{u : \Omega \to \mathbb{R} : \ u|_K \in H^2(K) \ \forall K \in \mathcal{T}_h\right\}.$$

Beweis. Wir definieren π^* in den Knoten p_i der Triangulierung $\mathcal{T}_h$ durch

$$(\pi^* v)(p_i) := \frac{1}{m_i} \sum_{\{K : p_i \in K\}} v\,|_K\,(p_i), \tag{6.55}$$

wobei m_i die Anzahl der Zellen bezeichnet, welche den Knoten p_i enthalten. Da die Funktionen aus V_h^p durch ihre Werte in den Knoten eindeutig festgelegt sind, ist π^* wohldefiniert. Um Ungleichung (6.54) zu beweisen, benötigen wir mehrere Abschätzungen. Sei dazu

$$\Phi|_K := \Phi_K = (b_h \cdot \nabla v_h - \pi^*(b_h \cdot \nabla v_h))\,|_K. \tag{6.56}$$

Gemäß der Definition von π^* ist Φ_K ein Polynom mit $\Phi_K(p_j) = 0$ für alle Knoten von K, die nur zu Zelle K gehören. Aus diesen Eigenschaften ergibt sich für alle $K \in \mathcal{T}_h$ die Abschätzung:

$$\|\Phi_K\|_{L^2(K)} \leq C h_K^{1/2} \|\Phi_K\|_{L^2(\partial K)}. \tag{6.57}$$

Zum Beweis dieser Ungleichung transformiert beispielsweise [BFH06] auf das Referenzelement der Triangulierung zurück und verwendet dort die Normäquivalenz auf endlichdimensionalen Räumen sowie das Standard–Skalierungsargument aus [GR86] Seite 96. Die nächste benötigte Ungleichung verwendet die folgende skalierte l_1-Norm für $q_h \in V_h^p(E)$:

$$\|q_h\|_{l_1(E)} := h_E^{1/2} \sum_{\{j : p_j \in E\}} |q_h(p_j)|\,.$$

Wegen der Normäquivalenz auf $V_h^p(E)$ gibt es dann zwei Konstanten $C_1 > 0$ und $C_2 > 0$, so dass:

$$C_1 \|q_h\|_{L^2(E)} \leq \|q_h\|_{l_1(E)} \leq C_2 \|q_h\|_{L^2(E)} \qquad \forall q_h \in V_h^p(E),\ \forall E \in \mathcal{E}_h. \tag{6.58}$$

Um die nächste Abschätzung zu erhalten, setzen wir die Definition von π^* in Φ_K ein und nutzen die Stetigkeit von b_h

$$\Phi_K(p_j) = \frac{1}{m_j} \sum_{\{K' : p_j \in K'\}} b_h(p_j) \cdot (\nabla v_h|_K(p_j) - \nabla v_h|_{K'}(p_j))\,.$$

Damit folgt:

$$|\Phi_K(p_j)| \leq \frac{1}{m_j} \sum_{\{K' : p_j \in K'\}} \sum_{E' \in w(K,K')} |b_h(p_j) \cdot [\nabla v_h]_{E'}(p_j)|\,. \tag{6.59}$$

Hier wurde der Sprung $(\nabla v_h|_K(p_j) - \nabla v_h|_{K'}(p_j))$ abgeschätzt durch die Sprünge über die Menge der Kanten $w(K, K')$, welche zwischen K und K' liegen und welche p_j enthalten. Gibt es mehrere solcher Kantenmengen, so wählen wir eine beliebige davon aus (siehe Abbildung 6.2).
Die Summe in (6.59) soll nun über alle Kanten aus $\mathcal{E}_{h,j}$ geführt werden, auf welchen der Knoten p_j liegt. Wegen der Formregularität der Triangulierung gibt es nur endlich viele solcher Kanten und es gilt $h_E \sim h_K$ sowie $h_E \sim h_{E'}$. Daraus folgt

$$|\Phi_K(p_j)| \leq \sum_{E' \in \mathcal{E}_{h,j}} |b_h(p_j) \cdot [\nabla v_h]_{E'}(p_j)|\,,$$

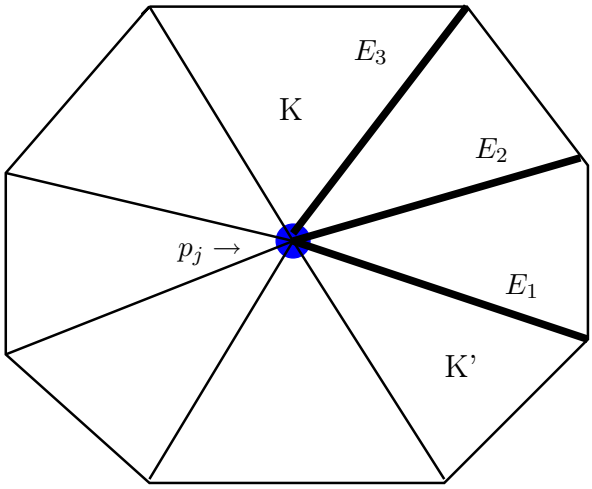

Abbildung 6.2: Veranschaulichung der Situation im Beweis von Lemma 6.15. Für $w(K, K')$ kann man beispielsweise die Kanten-Menge $\{E_1, E_2, E_3\}$ wählen.

und wegen Abschätzung (6.58) die letzte benötigte Ungleichung:

$$\|\Phi_K\|_{l_1(E)} \leq C \sum_{E' \in \mathcal{E}_h} \|b_h \cdot [\nabla v_h]_{E'}\|_{l_1(E')}. \tag{6.60}$$

Kombination aller Abschätzungen ergibt die Ungleichungs-Kette:

$$\begin{aligned}
h_K \|b_h \cdot \nabla v_h - \pi^*(b_h \cdot \nabla v_h)\|^2_{L^2(K)} &= h_K \|\Phi_K\|^2_{L^2(K)} \\
&\leq C h_K^2 \sum_{E \subset \partial K} \|\Phi_K\|^2_{L^2(E)} \\
&\leq C h_K^2 \sum_{E \subset \partial K} \|\Phi_K\|^2_{l_1(E)} \\
&\leq C h_K^2 \left(\sum_{E' \in \mathcal{E}_h(K)} \|b_h \cdot [\nabla v_h]_{E'}\|_{l_1(E')} \right)^2 \\
&\leq C \sum_{E' \in \mathcal{E}_h(K)} h_{E'}^2 \|b_h \cdot [\nabla v_h]_{E'}\|^2_{l_1(E')} \\
&\leq C \sum_{E' \in \mathcal{E}_h(K)} h_{E'}^2 \|b_h \cdot [\nabla v_h]_{E'}\|^2_{L^2(E')}.
\end{aligned}$$

Die Abschätzungen folgen der Reihe nach aus (6.57), (6.58), (6.60), $h_K \sim h_{E'}$ und wieder (6.58). Bei der vierten Abschätzung wurde zudem die Ungleichung $(a_1 + ... + a_n)^2 \leq C(a_1^2 + ... + a_n^2)$ benutzt. □

Bemerkung 6.16.
Gemäß [BFH06] Lemma 3.1 gilt auch die Umkehrung der Abschätzung in obigem Lemma, das heißt es gibt eine Konstante $\widetilde{C} > 0$ mit

$$\widetilde{C} \sum_{E \in \mathcal{E}_h(K)} \int_E h_E^2 \, |b_h \cdot [\nabla v_h]_E|^2 \, ds \leq h_K \|b_h \cdot \nabla v_h - \pi^*(b_h \cdot \nabla v_h)\|^2_{L^2(K)}.$$

Dieses Ergebnis besagt, dass die Interpolationsfehler-Abschätzung aus dem Lemma gemessen an der Ordnung von h_K nicht verbessert werden kann.

Der nächste Satz stellt die gesuchte Abschätzung des Verfahrensfehlers zur Verfügung:

Satz 6.17. (**Globale Fehlerabschätzung für die CIP-Methode**)
Seien die Daten der schwachen Formulierung genügend regulär, das heißt, $b \in W^{1,\infty}(\Omega)$, $c \in L^\infty(\Omega)$ sowie $f \in L^2(\Omega)$, sei $\gamma > 0$ ausreichend groß und die Gebietszerlegung quasi–uniform. Sei ferner eine ausreichend glatte Lösung der schwachen Formulierung vorgegeben, sprich: $u \in H^{p+1}(\Omega)$. Dann gilt für die Lösung u_h der CIP-Methode (6.45a) die folgende globale Fehlerabschätzung:

$$|\!|\!| u - u_h |\!|\!|_{\mathrm{CIP}} \leq C \left(\epsilon^{1/2} + h^{1/2} \right) h^p \|u\|_{H^{p+1}(\Omega)} \tag{6.61}$$

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

Beweis. Wir beschränken uns auf eine Beweisskizze. Der Fehler in der CIP–Norm wird mit der Dreiecksungleichung in einen Diskretisierungsfehler- und Interpolationsfehler–Anteil zerlegt:

$$|\!|\!| u - u_h |\!|\!|_{\mathrm{CIP}} \leq |\!|\!| u - u_\pi |\!|\!|_{\mathrm{CIP}} + |\!|\!| u_\pi - u_h |\!|\!|_{\mathrm{CIP}}. \tag{6.62}$$

Die Abschätzung des Diskretisierungs–Fehler gelingt wie folgt: Aus der Koerzitivität von a_{CIP} bezüglich $|\!|\!| \cdot |\!|\!|_{\mathrm{CIP}}$ und der Galerkin–Orthogonalität der Methode ergibt sich

$$\begin{aligned} \frac{1}{2} |\!|\!| u_h - u_\pi |\!|\!|_{\mathrm{CIP}}^2 &\leq a_{\mathrm{CIP}}(u_h - u_\pi, u_h - u_\pi) \\ &= a_{\mathrm{CIP}}(u - u_\pi, u_h - u_\pi). \end{aligned} \tag{6.63}$$

Dabei kürzt u_π die L^2–Interpolation von u in Ω aus Lemma 5.10 ab.
Division durch $|\!|\!| u_h - u_\pi |\!|\!|_{\mathrm{CIP}}/2$, Auseinanderziehen der Terme und Übergang zum Supremum liefert:

$$|\!|\!| u_h - u_\pi |\!|\!|_{\mathrm{CIP}} \leq 2 \sup_{w_h \in V_h^p} \frac{a_W(u - u_\pi, w_h)}{|\!|\!| w_h |\!|\!|_{\mathrm{CIP}}} + 2 \sup_{w_h \in V_h^p} \frac{j_h(u - u_\pi, w_h)}{|\!|\!| w_h |\!|\!|_{\mathrm{CIP}}}. \tag{6.64}$$

Für den zweiten Term gilt die folgende Ungleichungskette

$$\begin{aligned} |j_h(u - u_\pi, w_h)|^2 &\leq C j_h(u - u_\pi, u - u_\pi) j_h(w_h, w_h) \\ &\leq C \sum_{E \in \mathcal{E}_h} h_E^2 \|\nabla(u - u_\pi)\|_{L^2(E)}^2 |\!|\!| w_h |\!|\!|_{\mathrm{CIP}}^2 \\ &\leq C \sum_{K \in \mathcal{T}_h} h_K^2 (h_K^{-1} \|\nabla(u - u_\pi)\|_{L^2(K)}^2 + h_K \|\nabla(u - u_\pi)\|_{H^1(K)}^2) |\!|\!| w_h |\!|\!|_{\mathrm{CIP}}^2 \\ &\leq C \sum_{K \in \mathcal{T}_h} h_K^2 (h_K^{-2} \|u - u_\pi\|_{L^2(K)}^2) |\!|\!| w_h |\!|\!|_{\mathrm{CIP}}^2 \\ &\leq C h^{2p+1} \|u\|_{H^{p+1}(\Omega)}^2 |\!|\!| w_h |\!|\!|_{\mathrm{CIP}}^2. \end{aligned} \tag{6.65}$$

Die erste Ungleichung ergibt sich analog zur Cauchy–Schwarz–Ungleichung, die zweite aus der Beschränktheit von b und der Abschätzung für beliebiges $g \in L^2(\Omega)$

$$\sum_{E \in \mathcal{E}_h} \int_E [g]^2 ds \leq 2 \sum_{E \in \mathcal{E}_h} \int_E |g|^2 \, ds,$$

welche wegen der Dreiecksungleichung gilt. Die dritte Ungleichung in (6.65) folgt aus der Spurungleichung und der Tatsache, dass $h_E \sim h_K$ wegen der Quasiuniformität des Gitters gilt, die vierte durch mehrfachen Anwenden der inversen Ungleichung (5.39) und die fünfte aus der Interpolationsfehler–Abschätzung (5.33). Zieht man in obiger Abschätzung die

Wurzel und setzt in (6.64) ein, so erhält man eine Abschätzung für den zweiten Term mit der gewünschten Ordnung in h. Im ersten Term von (6.64) bringt man die Bilinearform a_W durch partielle Integration in die folgende Form

$$\begin{aligned} a_W(u-u_\pi, w_h) =& \epsilon(\nabla(u-u_\pi), \nabla w_h) + c((c-\nabla\cdot b)(u-u_\pi), w_h) \\ &+ (b\cdot n(u-u_\pi), w_h)_{\Gamma_+} - (u-u_\pi, b\cdot\nabla w_h) \\ &- \epsilon(n\cdot\nabla(u-u_\pi), w_h)_\Gamma - \epsilon(u-u_\pi, n\cdot\nabla w_h)_\Gamma \\ &+ \epsilon\gamma \sum_{E\subset\Gamma} \frac{1}{h_E}(u-u_\pi, w_h)_E. \end{aligned} \tag{6.66}$$

Dann schätzt man die Terme getrennt voneinander ab. Wir geben hierfür zwei Beispiele an, die anderen Abschätzungen verlaufen mit den gleichen Hilfsmitteln. Die Abschätzung des vierten Terms in (6.66) ist am aufwändigsten. Durch Nulladdition von $b_h \cdot w_h$ spaltet man den Term zunächst wie folgt auf:

$$(u-u_\pi, b\cdot\nabla w_h) = (u-u_\pi, (b-b_h)\cdot\nabla w_h) + (u-u_\pi, b_h\cdot\nabla w_h). \tag{6.67}$$

Den ersten Summanden schätzt man anschließend ab durch:

$$\begin{aligned} |(u-u_\pi, (b-b_h)\cdot\nabla w_h)| &\le C\|u-u_\pi\|_{L^2(\Omega)}\|b-b_h\|_{L^\infty(\Omega)}\,|w_h|_{H^1(\Omega)} \\ &\le C\sum_{K\in\mathcal{T}_h}\|u-u_\pi\|_{L^2(K)}\|b-b_h\|_{L^\infty(K)}\,|w_h|_{H^1(K)} \\ &\le C\sum_{K\in\mathcal{T}_h}\|u-u_\pi\|_{L^2(K)} h_K\,|w_h|_{H^1(K)} \\ &\le Ch^{p+1}\|u\|_{H^{p+1}(\Omega)}|\!|\!|w_h|\!|\!|_{\mathrm{CIP}}. \end{aligned}$$

Dabei wurden der Reihe nach die Cauchy–Schwarz–Ungleichung, die Dreiecksungleichung, die Voraussetzung (6.49) an b_h und die Interpolationsfehler–Abschätzung (5.33) mitsamt der Relation $h_K \le h\ \forall K \in \mathcal{T}_h$ benutzt. Eine Abschätzung für den zweiten Summanden ergibt sich wie folgt:

$$\begin{aligned} |(u-u_\pi, b_h\cdot\nabla w_h)| &= |(u-u_\pi, b_h\cdot\nabla w_h - \pi^*(b_h\cdot\nabla w_h)| \\ &\le C\sum_{K\in\mathcal{T}_h}\|u-u_\pi\|_{L^2(K)}\|b_h\cdot\nabla w_h - \pi^*(b_h\cdot\nabla w_h\|_{L^2(K))} \\ &\le C\sum_{K\in\mathcal{T}_h}\|u-u_\pi\|_{L^2(K)}\left(h_K^{-1}\sum_{E\in\mathcal{E}_h(K)}\int_E h_E^2\,|b_h\cdot[\nabla u_h]_E|^2\,ds\right)^{1/2} \\ &\le C\left(\sum_{K\in\mathcal{T}_h}\|u-u_\pi\|_{L^2(K)}\right)h^{1/2} j_h(u_h,u_h)^{1/2} \\ &\le Ch^{p+1/2}\|u\|_{H^{p+1}(\Omega)}|\!|\!|w_h|\!|\!|_{\mathrm{CIP}}. \end{aligned} \tag{6.68}$$

Die Gleichheit ergibt sich aus der L^2–Orthogonalität der globalen L^2–Projektion π_h bezüglich V_h^p. Die erste Ungleichung folgt aus der Cauchy–Schwarz- und der Dreiecksungleichung, die zweite Ungleichung aus Lemma 6.15, die dritte durch Vergleich mit der Definition von j_h und die letzte durch Anwenden der Interpolationsfehler–Abschätzung (5.33). Es sei erwähnt, dass dies die einzige Stelle im Beweis ist, an welcher Lemma 6.15 angewandt wird. Exemplarisch schätzen wir nun noch einen der Randterme ab. Dazu benötigen wir die folgende Ungleichung

$$h_E^{1/2}\|n\cdot\nabla w_h\|_{L^2(E)} \le C\left(|w_h|_{H^1(K)} + h_T\,|w_h|_{H^2(K)}\right) \le C\,|w_h|_{H^1(K)} \quad \forall E\subset\partial K, \tag{6.69}$$

welche aus der Spurungleichung (5.42) und der inversen Ungleichung (5.39) folgt. Den vorletzten Term in (6.66) beschränken wir nun nach oben durch:

$$\begin{aligned}|\epsilon(u-u_\pi, n\cdot\nabla w_h)_\Gamma| &\le C\sum_{E\subset\Gamma}\left(\frac{\epsilon}{h_E}\right)^{1/2}\|u-u_\pi\|_{L^2(E)}(\epsilon h_E)^{1/2}\|n\cdot\nabla w_h\|_{L^2(E)}\\ &\le C\left(\sum_{E\subset\Gamma}\frac{\epsilon}{h_E}\|u-u_\pi\|^2_{L^2(E)}\right)^{1/2}\left(\epsilon\sum_{E\subset\Gamma}\|w_h\|^2_{H^1(K)}\right)^{1/2}\\ &\le C\epsilon^{1/2}h^p\|u\|_{H^{p+1}(\Omega)}|\!|\!|w_h|\!|\!|_{\mathrm{CIP}}.\end{aligned}$$

Die erste Abschätzung folgt aus der Cauchy–Schwarz- und der Dreiecksungleichung sowie Erweiterung mit $h_E^{1/2}/h_E^{1/2}$, die zweite aus (6.69) sowie der Cauchy–Schwarz–Ungleichung für Summen, $\sum_i a_i b_i \le (\sum_i a_i^2)^{1/2}(\sum_i b_i^2)^{1/2}$ für a_i, b_i aus $\mathbb{R}_0^+$, und die dritte aus der Interpolationsfehler–Abschätzung (5.34). Einen ausführlicheren Beweis zur Abschätzung des Diskretisierungsfehler findet man in [RST08] von Seite 359 bis 361. Dort kann insbesondere nachgeschlagen werden, wie die restlichen Terme abgeschätzt werden. Es fehlt noch die Abschätzung des Interpolationsfehlers in (6.62). Dazu betrachten wir

$$\begin{aligned}|\!|\!|u-u_\pi|\!|\!|_{\mathrm{CIP}} &\le C\epsilon^{1/2}\,|u-u_\pi|_{H^1(\Omega)} + C\|u-u_\pi\|_{L^2(\Omega)} + j_h(u-u_\pi,u-u_\pi)^{1/2}\\ &\qquad + C\|u-u_\pi\|_{L^2(\Gamma)} + \left(\sum_{E\subset\Gamma}\frac{\epsilon}{h_E}\|u-u_\pi\|^2_{L^2(E)}\right)^{1/2}\\ &\le C\left(\epsilon^{1/2}+h^{1/2}\right)h^p\|u\|_{H^{p+1}(\Omega)}.\end{aligned}$$

Nach der elementaren ersten Abschätzung gelingt die zweite, indem man man die Integrale über das Gebiet mit Hilfe der Interpolationsfehler–Abschätzung (5.33) und das Randintegral mit (5.34) beschränkt. Den Term zu j_h haben wir bereits in (6.65) nach oben abgeschätzt. □

Bemerkung 6.18.
(i) Die Fehlerabschätzung sagt die gleichen Konvergenzrate voraus, wie wir sie für die Streamline–Diffusion–Methode erhalten haben: eine leicht subtopimale Konvergenzrate der Ordnung $\mathcal{O}(h^{p+1/2})$ in der L^2–Norm, eine leicht subtopimale Rate der Ordnung $\mathcal{O}(h^p\epsilon^{-1/2})$ in der H^1–Norm für den konvektions–dominanten Fall und eine optimale Rate der Ordnung $O(h^p)$ in der H^1–Norm für den diffusions-dominanten Fall.

(ii) Bei der Abschätzung des konvektiven Terms im Beweis bei (6.68) wird die L^2-Orthogonalität der globalen L^2-Projektion bezüglich V_h^p benutzt. Die L^2–Projektion darf als Interpolations–Operator genutzt werden, da die Randbedingungen schwach vorgegeben sind. Gibt man die Randbedingung hingegen stark vor, so liefert die L^2–Projektion Funktionen, die nicht mehr im Approximationsraum liegen, da sie im Allgemeinen die Randbedingungen nicht erhält. Demzufolge müsste ein anderer Interpolations–Operator verwandt werden. Nutzt man zum Beispiel die Lagrange–Interpolation, so geht die L^2–Orthogonalität verloren und obiger Beweis funktioniert nicht mehr. Dies erklärt wohl, wieso in der Literatur die Randbedingungen bei der CIP–Methode schwach vorgegeben werden.

(iii) In ihrer Standard–Formulierung leidet die CIP-Methode unter dem gleichen Problem wie die Streamline–Diffusion–Methode. Unphysikalische Oszillationen werden zwar im Großteil des Lösungsgebiets unterdrückt, doch in den Grenzschichten treten Instabilitäten auf, welche die Genauigkeit der Lösung beeinträchtigen. Diesem Problem begegnet

man mit shock capturing Termen, wie zum Beispiel dem folgenden aus [BH04]:

$$J_{sc}(u,v) = \sum_{K\in\mathcal{T}_h} \int_{\partial K} \Psi_K(u)\ \mathrm{sign}(t_{\partial K}\cdot\nabla(u|_K))\ \ t_{\partial K}\cdot\nabla(v|_K)ds, \tag{6.70}$$

wobei $t_{\partial K}$ der Tangentenvektor an den Rand ∂K der Zelle K ist und

$$\Psi_K(u) = \mathrm{diam}(K)(C_1\epsilon + C_2\mathrm{diam}(K)) \max_{E\subset\partial K} |[n_E\cdot\nabla u]_E| \tag{6.71}$$

mit zwei frei wählbaren Parametern C_1 und C_2. In [BH04] wird für Finite–Elemente erster Ordnung mit obigem shock capturing Term ein diskretes Maximumsprinzip gezeigt. Für stabilisierte Finite–Element–Methoden höherer Ordnung ist die Gültigkeit eines diskreten Maximumsprinzips allerdings noch nicht bewiesen worden [JK07a].

Kapitel 7

Numerische Ergebnisse für die Konvektions–Diffusionsgleichung

Die Konvektions–Diffusionsgleichung werden wir mit Hilfe des Standard–Galerkin–Verfahrens, der Streamline–Diffusion– und der CIP–Methode approximieren. Zunächst beschreiben wir dazu die Implementierung der CIP–Methode, welche im Rahmen dieser Arbeit durchgeführt worden ist. Zu den Kantenintegralen, die bei der CIP–Methode zu berechnen sind, geben wir exakte Werte an. Dann wird das Gitter spezifiziert, welches bei den Finite–Elemente–Methoden verwandt worden ist. Anschließend führen wir Rechnungen für ein Beispiel ohne Grenzschichten durch. Unter anderem wird dabei überprüft, ob die Stabilisierungsverfahren die von der Fehleranalysis vorausgesagten Konvergenzraten erreichen und wie beide Verfahren im Vergleich zueinander abschneiden. Danach wird auch ein Beispiel mit Grenzschichten betrachtet.

7.1 Implementierung der CIP–Methode

Die Konvektions–Diffusionsgleichung werden wir mit Hilfe von Finite–Elemente–Methoden lösen. Als Finite–Elemente–Code steht das Programmpaket MooNMD zur Verfügung. Das Standard–Galerkin–Verfahren und die Streamline–Diffusion–Methode sind dort schon implementiert, die CIP–Methode hingegen nur für den Spezialfall von Finiten–Elementen erster Ordnung. Im Rahmen dieser Arbeit ist die CIP–Methode auf Elemente höherer Ordnung erweitert worden.
Bei Standard–Galerkin–Verfahren und der Streamline–Diffusion–Methode stehen nur solche Basisfunktionen miteinander in Verbindung, welche gleichzeitig auf mindestens einer gemeinsamen Gitterzelle von Null verschieden sind. Mit anderen Worten, nur für solche Paare von Basisfunktionen gibt es von Null verschiedene Einträge in der Systemmatrix. Bei der CIP–Methode wird ein Stabilisierungsterm der folgenden Form benutzt:

$$j_h(u_h, v_h) = \sum_{E \in \mathcal{E}_h} \tau h_E^2 \left(b_h \cdot [\nabla u]_E, b_h \cdot [\nabla v]_E\right)_E \tag{7.1}$$

mit der Menge $\mathcal{E}_h$ aller inneren Kanten und dem Sprung $[\cdot]_E$ über die Kante E. Ist eine Funktion in Zelle K von Null verschieden, so ist der Sprung ihres Gradienten über die Kanten von K im Allgemeinen ungleich Null. Deshalb werden durch diesen Stabilisierungsterm alle Paare von Basisfunktionen verknüpft, welche jeweils zumindest in einer von zwei aneinandergrenzenden Zellen von Null verschieden sind. Um die neuen Verknüpfungen im Code zu berücksichtigen, muss die Struktur der Systemmatrix entsprechend erweitert werden.

Beispiel 7.1. Im Folgenden betrachten wir eine Situation, wie sie in Abbildung 7.1 für das P^1–Element gezeigt wird. Dabei geben wir explizit die Basisfunktionen an und berechnen das Kantenintegral über die innere Kante. Zum einem kann daran gesehen werden, wie die CIP–Methode zusätzliche Freiheitsgrade miteinander verknüpft; zum anderen kann die Korrektheit des Programms getestet werden.

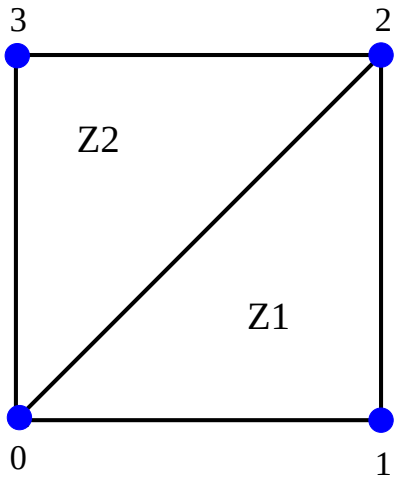

Abbildung 7.1: Knotenverteilung bei Beispiel 7.1. Die blauen Punkte entsprechen den Knoten. Für weitere Details siehe Text.

Um einen trivialen Fall zu vermeiden, nehmen wir die Randbedingungen als schwach vorgegeben an, so dass alle Basisfunktionen frei sind. Wie in Abschnitt 5.2 genauer erläutert worden ist, sind die Basisfunktionen des P^1–Elements dadurch definiert, dass sie jeweils in einem der Knoten (hervorgehoben durch die blauen Punkte) gleich 1 sind, während sie in allen anderen Knoten verschwinden. Eine explizite Formel für die Basisfunktionen kann schnell aus dem Konzept der baryzentrischen Koordinaten λ_i, $i = 1, 2, 3$, abgeleitet werden (vgl. Definition 5.4). Besitzt die Basisfunktion ϕ die Ecke $\mathbf{a}^i$ in Gitterzelle K als 1-Knoten, so ist sie auf K gegeben durch

$$\phi|_K(\lambda) = \lambda_i. \tag{7.2}$$

Auf allen Zellen, zu welchen der 1-Knoten nicht gehört, verschwindet die Basisfunktion hingegen. Wir wählen das Koordinatensystem im Folgenden so, dass das dargestellte Quadrat dem Einheitsquadrat $[0, 1]^2$ entspricht und der mit 0 indizierte Punkt im Ursprung liegt. Zudem seien in Zelle 1 die Eckpunkte gegeben durch $a_1 = (0, 0), a_2 = (1, 0), a_3 = (1, 1)$. Dann lauten die zugehörigen baryzentrischen Koordinaten auf Zelle 1:

$$\lambda_1 = 1 - x - y \quad \lambda_2 = x \quad \lambda_3 = y.$$

Entsprechend erhält man in Zelle 2 mit $a_1 = (0, 0), a_2 = (1, 1), a_3 = (0, 1)$ als Eckpunkten die baryzentrischen Koordinaten:

$$\lambda_1 = 1 - y \quad \lambda_2 = x \quad \lambda_3 = y - x.$$

Formel (7.2) liefert nun sofort den folgenden Satz von Basisfunktionen

$$\phi_0(x, y) = \begin{cases} 1 - x - y & \text{in Zelle 1} \\ 1 - y & \text{in Zelle 2} \end{cases}, \qquad \phi_1(x, y) = \begin{cases} x & \text{in Zelle 1} \\ 0 & \text{in Zelle 2} \end{cases},$$

$$\phi_2(x, y) = \begin{cases} y & \text{in Zelle 1} \\ x & \text{in Zelle 2} \end{cases}, \qquad \phi_3(x, y) = \begin{cases} 0 & \text{in Zelle 1} \\ y - x & \text{in Zelle 2} \end{cases},$$

wobei die Basisfunktion ϕ_i ihren 1-Knoten im Punkt i in Abbildung 7.1 hat. Als nächstes werten wir die Kantenintegrale über die innere Kante von Punkt 0 zu Punkt 2 aus. Der

Integrand ist als Produkt zweier Polynome ersten Grades im Allgemeinen ein Polynom zweiten Grades. Ein Polynom p zweiten Grades lässt sich in einer Dimension exakt durch die folgende Gaußquadraturformel integrieren:

$$\int_a^b p(s)ds = \frac{b-a}{2}(w_0 p(s_0) + w_1 p(s_1)) \tag{7.3}$$

mit den Gewichten $w_0 = w_1 = 1$ und den Quadraturpunkten

$$s_0 = -\frac{1}{\sqrt{3}}\frac{b-a}{2} + \frac{a+b}{2}, \quad s_1 = \frac{1}{\sqrt{3}}\frac{b-a}{2} + \frac{a+b}{2}.$$

Für die Kantenintegrale ergibt sich daraus die Formel

$$I = \sum_{j=1}^{2} \frac{h_E^3}{2} \left(b_{x,j}[\partial_x u_j] + b_{y,j}[\partial_y u_j]\right) \left(b_{x,j}[\partial_x v_j] + b_{y,j}[\partial_y v_j]\right) |_{s=s_i} \tag{7.4}$$

mit den Quadraturpunkten $s_0 \approx (0.2113, 0.2113)$, $s_1 \approx (0.7887, 0.7887)$ und der Kantenlänge $h_E = \sqrt{2}$. Um die Formeln benutzen zu können, benötigen wir noch die Sprünge der Basisfunktionen in den Quadraturpunkten. Die Sprünge der Basisfunktionen sind in beiden Quadraturpunkten gleich und lauten:

Nr. Basisfunktion	0	1	2	3
$[\partial_x u]$	-1	1	-1	1
$[\partial_y u]$	1	-1	1	-1

Setzen wir nun beispielsweise $b = b_h = (1, 2)$, so errechnet man mit Hilfe von Formel (7.4) die folgenden Werte für die Kantenintegrale:

Nr. Ansatzfunktion	Nr. Testfunktion	Integral
0	0	$\sqrt{2}$
0	1	$-\sqrt{2}$
0	2	$\sqrt{2}$
0	3	$-\sqrt{2}$
1	1	$\sqrt{2}$
1	2	$-\sqrt{2}$
1	3	$\sqrt{2}$
2	2	$\sqrt{2}$
2	3	$-\sqrt{2}$
3	3	$\sqrt{2}$

Durch die Kantenintegrale gibt es also eine Kopplung zwischen allen Basisfunktionen. Im Gegensatz zu dem Standard–Galerkin–Verfahren und der Streamline–Diffusion–Methode kommt also eine Kopplung zwischen Basisfunktion 1 und 3 hinzu. Ein zweites Testbeispiel mit stark vorgegebenen Dirichlet–Randbedingungen wird in Anhang A gegeben.

7.2 Wahl der Gebietszerlegung

Als Finite–Elemente verwenden wir in dieser Arbeit stetige Dreiecks- und Rechteckelemente. Diese sind in Kapitel 5 eingeführt worden. Zur vollständigen Definition der Finite–Elemte ist noch die Triangulierung des Lösungsgebiets Ω festzulegen. Das Lösungsgebiet

wird bei unseren Modellrechnungen immer das kleine Einheitsquadrat $[0,1]^2$ sein. Dieses zerlegen wir entweder in Quadrate oder in Dreiecke nach dem in Abbildung 7.2 gezeigten Schema. Auf der linken Seite der Abbildung ist jeweils die gröbste Gebietszerlegung gezeigt, auf welcher wir rechnen werden. Die gröbste Gebietszerlegung bezeichnen wir im Folgenden als Level–1–Gitter. Die rechte Seite der Abbildung zeigt die Gebietszerlegung nach einer Gitterverfeinerung. Gemäß diesem Schema wird das Lösungsgebiet im Finite–Elemente–Code immer feiner zerlegt, um Rechnungen größerer Genauigkeit zu erhalten. Das Gitter, welches aus dem Level–1–Gitter nach $n-1$ Verfeinerungsschritten hervorgeht, bezeichnen wir als Level–n–Gitter. Das Level–n–Gitter besitzt für Quadrate 4^n Gitterzellen und für Dreiecke $2 \cdot 4^n$. Der Rechenaufwand nimmt mit der Anzahl der Freiheitsgrade zu, welche der Anzahl der Knoten in den Finite–Element–Räumen entspricht. Die Finite–Elemente erster Ordnung auf Dreiecken sowie Quadraten besitzen auf dem Level–n–Gitter $(2^n+1)^2$ Knoten, die Elemente zweiter Ordnung $(2 \cdot 2^n + 1)^2$ und die Elemente dritter Ordnung $(3 \cdot 2^n + 1)^2$ Knoten.

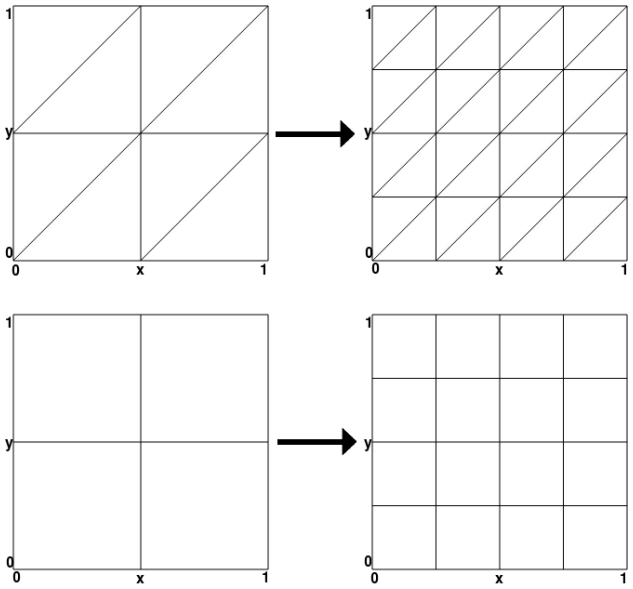

Abbildung 7.2: Zerlegung des kleinen Einheitsquadrates in Dreiecke oder Quadrate, wie sie im Finite–Element–Code benutzt wird. Links für das Level–1–Gitter, rechts auf dem Level–2–Gitter.

7.3 Rechnungen ohne Grenzschichten

In diesem Abschnitt betrachten wir das exponentielle Beispiel aus [BH04] mit der Lösung

$$u(x,y) = e^{-5(x-0.5)^2-15(y-0.5)^2} \tag{7.5}$$

auf dem Einheitsquadrat $\Omega = [0,1]^2$. Die Vorgaben auf Ω lauten $\epsilon = 10^{-6}$, $b = (1,0)^T$ und $c = 1$. Zu diesen Vorgaben wird die rechte Seite f der Konvektions-Diffusionsgleichung so angepasst, dass u die Gleichung löst. Dazu setzt man obige Vorgaben in (7.5) ein und führt die Ableitungen aus. Als Randwerte des Beispiels gibt man die Funktionswerte von u auf $\partial\Omega$ vor.
Die Lösung des obigen Beispiels ist in Abbildung 7.3 zu sehen. Wie man erkennt, ändert sich die Lösung über das Lösungsgebiet recht gleichmäßig ohne abrupte Änderungen in den Funktionswerten. Grenzschichten gibt es also keine.

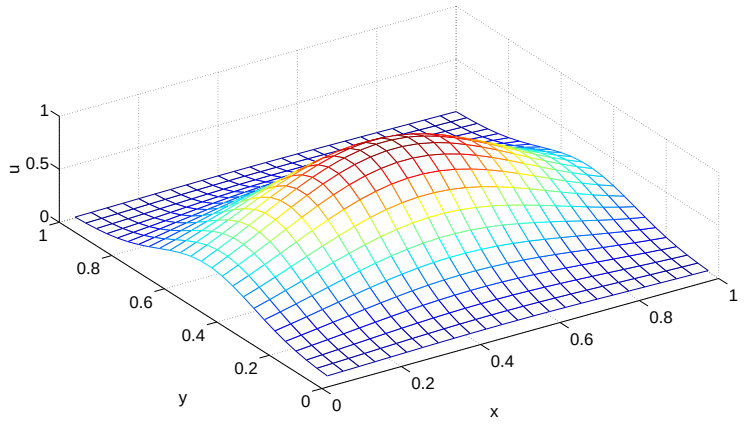

Abbildung 7.3: Lösung des exponentiellen Beispiels (7.5).

Im Folgenden soll untersucht werden, wie gut das exponentielle Beispiel durch das Standard–Galerkin–Verfahren, die Streamline–Diffusion–Methode und die CIP–Methode genähert wird. Bei der Streamline–Diffusion–Methode wird die Bilinearform a des Standard–Galerkin–Verfahrens erweitert um einen Term der Form

$$\sum_{K \in \mathcal{T}_h} \delta_K (\mathrm{Res}, (b \cdot \nabla) v_h)$$

mit dem Residuum Res der Konvektions–Diffusionsgleichung. Der Stabilisierungsparameter δ_K ist dabei gegeben durch

$$\delta_K = \begin{cases} \delta_0 h_K, & \text{falls } \mathrm{Pe}_K > 1 \qquad \text{(konvektions–dominanter Fall)}, \\ \delta_1 h_K^2 / \epsilon, & \text{falls } \mathrm{Pe}_K \leq 1 \qquad \text{(diffusions–dominanter Fall)}, \end{cases}$$

mit frei wählbaren globalen Parameter δ_0, δ_1 und der lokalen Pécletzahl

$$\mathrm{Pe}_K = \|b\|_{L^\infty(K)} h_K / (2\epsilon).$$

Unsere Rechnungen werden auf dem Level–2- bis Level–5–Gitter ausgeführt. Das Level–5–Gitter besitzt eine Gitterweite von 0.5^5. Somit ist für die Vorgaben an b und ϵ des exponentiellen Beispiels die lokale Pécletzahl stets größer als 1 und der konvektions–dominante Fall liegt vor. Demnach spielt nur der Parameter δ_0 eine Rolle. Bei der CIP–Methode wird das Standard–Galerkin–Verfahren erweitert durch den Term

$$\sum_{E \in \mathcal{E}_h} \tau h_E^2 \left(b_h \cdot [\nabla u_h]_E, b_h \cdot [\nabla v_h]_E \right)_E$$

mit dem frei wählbaren Stabilisierungsparameter τ (für die restlichen Bezeichnungen siehe Abschnitt 6.5).

Für beide Stabilisierungsverfahren sind im letzten Kapitel a–priori Fehlerabschätzungen für den Verfahrensfehler hergeleitet worden. Die Konvergenzraten, die dort für den L^2– und den H^1–Fehler vorausgesagt werden, sind in nachfolgender Tabelle aufgeführt. Daneben zeigt die Tabelle die an den Interpolationsfehler–Abschätzungen aus Abschnitt 5.4 gemessenen optimalen Raten.

Fehler	Konvergenzrate aus der Fehlerabschätzung	optimale Rate
L^2–Fehler	$h^{p+1/2}$	h^{p+1}
H^1–Fehler	$h^p \epsilon^{-1/2}$	h^p

Tabelle 7.1: Konvergenzraten für Finite–Elemente der Ordnung p für die Streamline–Diffusion– und die CIP–Methode. In der zweiten Spalte sind die durch die a-priori Fehlerabschätzungen aus Kapitel 6 vorausgesagten Raten gezeigt und in der dritten die an den Interpolationsfehler–Abschätzungen aus Abschnitt 5.4 gemessenen optimalen Raten.

Mit unseren Testrechnungen sollen im Wesentlichen nachfolgende Fragen untersucht werden:

- Zeigen die Stabilisierungsverfahren die von der Fehleranalysis aus Kapitel 6 vorausgesagten Konvergenzraten, das heißt bei einer Elementordnung von p eine Rate der Ordnung $\mathcal{O}(h^{p+1/2})$ bezüglich des L^2–Fehlers sowie eine Rate der Ordnung $\mathcal{O}(h^p \epsilon^{-1/2})$ bezüglich des H^1–Fehlers?
- Wie gut schneidet die CIP–Methode im Vergleich zum Standard–Galerkin–Verfahren und der Streamline–Diffusion–Methode ab?
- Wie hängen die Ergebnisse vom Gittertyp ab?[1]

Die Testrechnungen werden für Finite–Elemente erster, zweiter und dritter Ordnung auf dem Quadrat- und dem Dreiecksgitter durchgeführt. Wir zeigen die Werte der L^2– und H^1–Fehler mitsamt der zugehörigen Fehlerordnungen von Gitterlevel 2 bis 5. Aufgetragen sind die Fehler gegen τ, wobei τ entweder dem Stabilisierungsparameter der CIP–Methode entspricht oder dem Stabilisierungsparameter δ_0 der Streamline–Diffusion–Methode. Zur Lösung der linearen Gleichungssysteme setzen wir -wie auch bei allen folgenden Rechnungenden direkten Löser aus dem Paket UMFPACK [Dav04] ein.

Abbildung 7.4 zeigt die Ergebnisse für Finite–Elemente erster Ordnung auf dem Dreiecksgitter. Die schwarzen Kurven beziehen sich auf die Werte der CIP–Methode auf Gitterlevel 2, die roten auf Level 3, die grünen auf Level 4 und die blauen auf Level 5. Die orangefarbenen Kurven zeigen die Ergebnisse der Streamline–Diffusion–Methode auf Level 5 zum Vergleich. Gemäß der Abbildung genügen beide Stabilisierungsverfahren für nicht zu große Stabilisierungsparameter τ den von der Fehleranalysis vorausgesagten Konvergenzraten und sogar die an den Interpolationsfehler–Abschätzungen gemessenen optimalen Raten werden erreicht (vergleiche mit Tabelle 7.1). Die Konvergenzraten der CIP–Methode nimmt im dargestellten Bereich für zu große τ ab und fällt unter den von der Fehleranalysis vorausgesagten Wert. Das zugehörige τ wächst jedoch mit zunehmendem Gitterlevel. Auch wenn nicht dargestellt, bleiben für zu großes τ auch die Raten der Streamline–Diffusion–Methode hinter der Voraussage der Analysis zurück. Dass die Raten für größere τ–Werte für die dargestellten Gitterlevel verfehlt werden, stellt keinen Widerspruch zu den Fehlerabschätzungen der Analysis dar; letztere machen nur eine asymptotische Aussage für eine gegen Null strebende Gitterweite. Streng genommen kann man mit numerischen Testrechnungen nur feststellen, ob sich die Fehlerordnungen mit zunehmender Gitterweite so

[1] Bei gleichem Gitterlevel besitzen die Finiten–Elemente auf dem Quadratgitter ebenso viele Freiheitsgrade wie die Elemente auf dem Dreiecksgitter, siehe hierzu auch Abschnitt 7.2. Insofern handelt es sich also um einen fairen Vergleich.

ändert, dass beim Übergang $h \to 0$ die Fehlerordnungen der Analysis erreicht werden. Dies ist hier für beide Stabilisierungsverfahren der Fall. Im Vergleich liefern beide Stabilisierungsverfahren über ein großes τ–Intervall nahezu gleich große Fehlerwerte. Lediglich für große τ besitzt die CIP–Methode die größeren Fehler. Nach Tabelle 7.2 sind die Ergebnisse beider Stabilisierungsverfahren bei jeweils näherungsweise optimaler Wahl des Stabilisierungsparameters τ geringfügig besser als diejenigen des Standard–Galerkin–Verfahrens.
In [BH04] sind ebenfalls die Fehler und Fehlerordnungen der Finiten–Elemente erster Ordnung für das exponentielle Beispiel bestimmt worden. Gerechnet wird dort auf einem Dreiecksgitter, welches dem unsrigen bis auf die Gitterweite entspricht. Die Streamline–Diffusion– und die CIP–Methode werden auch dort betrachtet. Für beide Stabilisierungsverfahren sind unsere Fehlerwerte etwa vergleichbar mit denen von [BH04]. Rechnungen mit Elementen höherer Ordnung sind in [BH04] nicht durchgeführt worden.

τ	L^2	H^1	$\mathcal{O}(L^2)$	$\mathcal{O}(H^1)$
0.001	1.92e-4	5.94e-2	2.01	1.01
0.032	1.92e-4	5.95e-2	2.01	1.00
-	1.98e-4	6.12e-2	2.01	1.00

Tabelle 7.2: Fehler und Fehlerordnungen für das exponentielle Beispiel bei Finite–Elementen erster Ordnung auf dem Level–5–Dreiecksgitter. Die erste Zeile bezieht sich auf die CIP–Methode und die zweite auf die Streamline–Diffusion–Methode bei jeweils optimaler Wahl des Stabilisierungsparameters. Die dritte Zeile zeigt die Werte des Standard–Galerkin–Verfahrens.

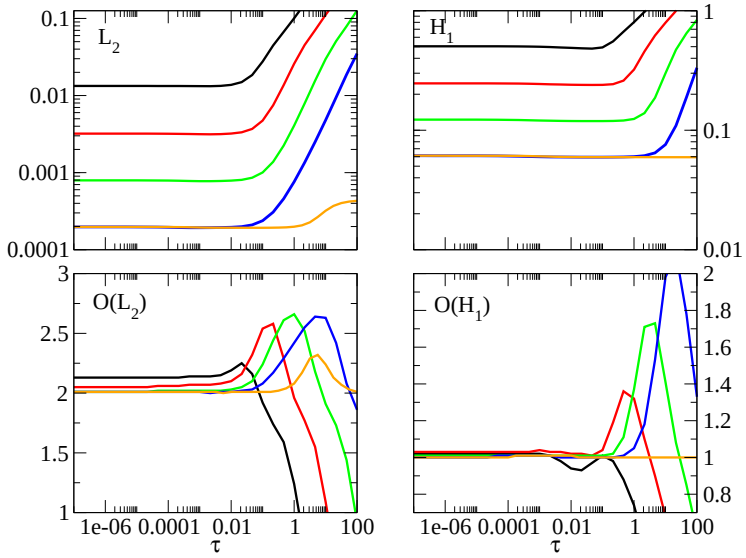

Abbildung 7.4: Fehler und Fehlerordnungen Finiter–Elemente erster Ordnung auf dem Dreiecksgitter aufgetragen gegen den Stabilisierungsparameter τ. In der oberen Reihe werden der L^2– und der H^1–Fehler gezeigt, in der unteren die zugehörigen Ordnungen. Die schwarzen, roten, grünen und blauen Kurven stellen die Ergebnisse der CIP–Methode dar und beziehen sich der Reihe nach auf Gitterlevel 2,3,4 und 5. Die orangefarbenen Kurven zeigt zum Vergleich die Resultate der Streamline–Diffusion–Methode auf dem Level–5–Gitter.

Für Finite–Elemente erster Ordnung auf dem Quadratgitter ergeben sich gemäß Abbildung 7.5 qualitativ ähnliche Ergebnisse wie für das Dreiecksgitter. Für ausreichend kleine τ liefern beide Stabilisierungsverfahren etwa gleich große Fehlerwerte und die Konvergenzraten der Interpolationsfehler–Abschätzungen werden erreicht. Wie Tabelle 7.3 zeigt, sind die Ergebnisse der Stabilisierungsverfahren bei jeweils optimaler Wahl von τ wiederum geringfügig besser als diejenigen des Standard–Galerkin–Verfahrens. Ein Vergleich mit Abbildung 7.4 oder Tabelle 7.2 zeigt, dass die Fehler auf dem Quadratgitter für alle Verfahren kleiner als auf dem Dreiecksgitter sind.

τ	L^2	H^1	$\mathcal{O}(L^2)$	$\mathcal{O}(H^1)$
0.001	1.11e-4	5.13e-2	2.00	1.00
0.001	1.10e-4	5.15e-2	2.00	1.00
-	1.11e-4	5.17e-2	2.01	1.00

Tabelle 7.3: Fehler und Fehlerordnungen für das exponentielle Beispiel bei Finite–Elementen erster Ordnung auf dem Level–5–Quadratgitter. Die erste Zeile bezieht sich auf die CIP–Methode und die zweite auf die Streamline–Diffusion–Methode bei jeweils optimaler Wahl des Stabilisierungsparameters. Die dritte Zeile zeigt die Werte des Standard–Galerkin–Verfahrens.

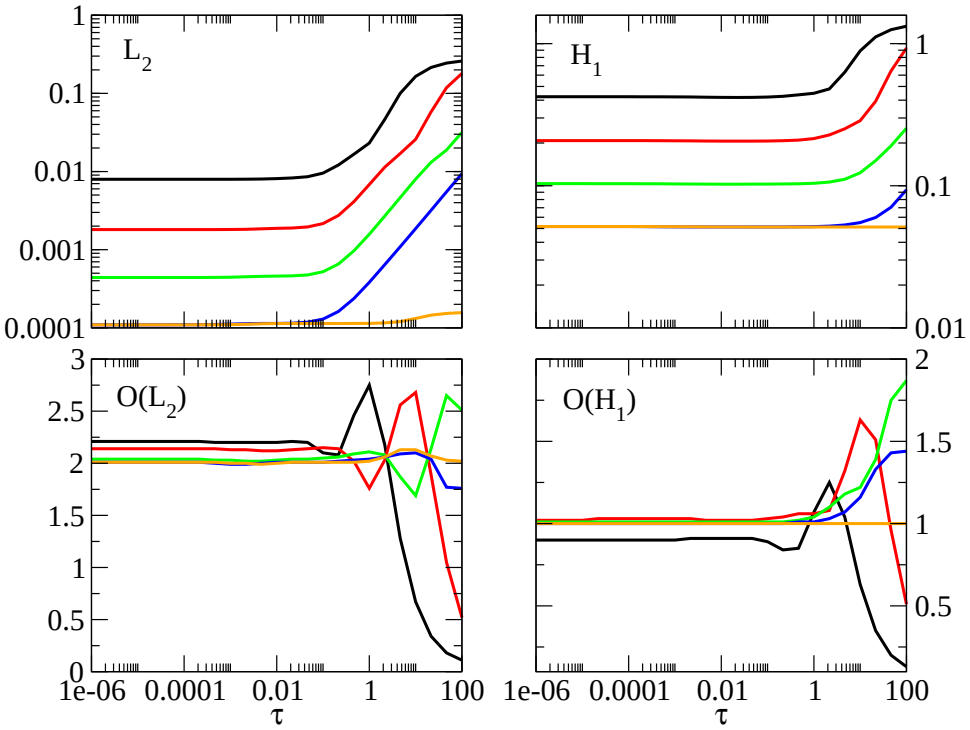

Abbildung 7.5: Zu sehen ist das gleiche wie in Abbildung 7.4 nur für Finite–Elemente erster Ordnung auf dem Quadratgitter.

Abbildung 7.6 zeigt die Ergebnisse für Finite–Elemente zweiter Ordnung auf dem Dreiecksgitter. Hier werden die von der Analysis vorausgesagten Konvergenzraten von der CIP–Methode in einem deutlich kleineren τ–Intervall als für Elemente erster Ordnung erreicht. Die maximal erreichte Ordnung und das τ-Intervall, in welchem die Voraussage der Fehleranalysis erfüllt werden, wachsen jedoch mit feiner werdendem Gitter. In dem τ–Bereich, in welchem die Fehlerordnungen am größten sind, sind die Fehler am kleinsten. Bei jeweils optimaler Wahl des Stabilisierungsparameters besitzt die Streamline–Diffusion–Methode etwas kleinere Fehlerwerte als die CIP–Methode, siehe hierzu Tabelle 7.4. Die Tabelle zeigt auch, dass die Fehler beider Stabilisierungsverfahren um mindestens einen Faktor drei kleiner als diejenigen des Standard–Galerkin–Verfahrens sind. Weiteren Tests zufolge wächst dieser Unterschied mit zunehmenden Gitterlevel.

τ	L^2	H^1	$\mathcal{O}(L^2)$	$\mathcal{O}(H^1)$
4.64e-3	3.44e-6	1.49e-3	2.96	1.99
0.1	2.59e-6	1.26e-3	2.98	2.03
-	1.56e-5	7.36e-3	2.07	1.07

Tabelle 7.4: Fehler und Fehlerordnungen für das exponentielle Beispiel bei Finite–Elementen zweiter Ordnung auf dem Level–5–Dreiecksgitter. Die erste Zeile bezieht sich auf die CIP–Methode und die zweite auf die Streamline–Diffusion–Methode bei jeweils optimaler Wahl des Stabilisierungsparameters. Die dritte Zeile zeigt die Werte des Standard–Galerkin–Verfahrens.

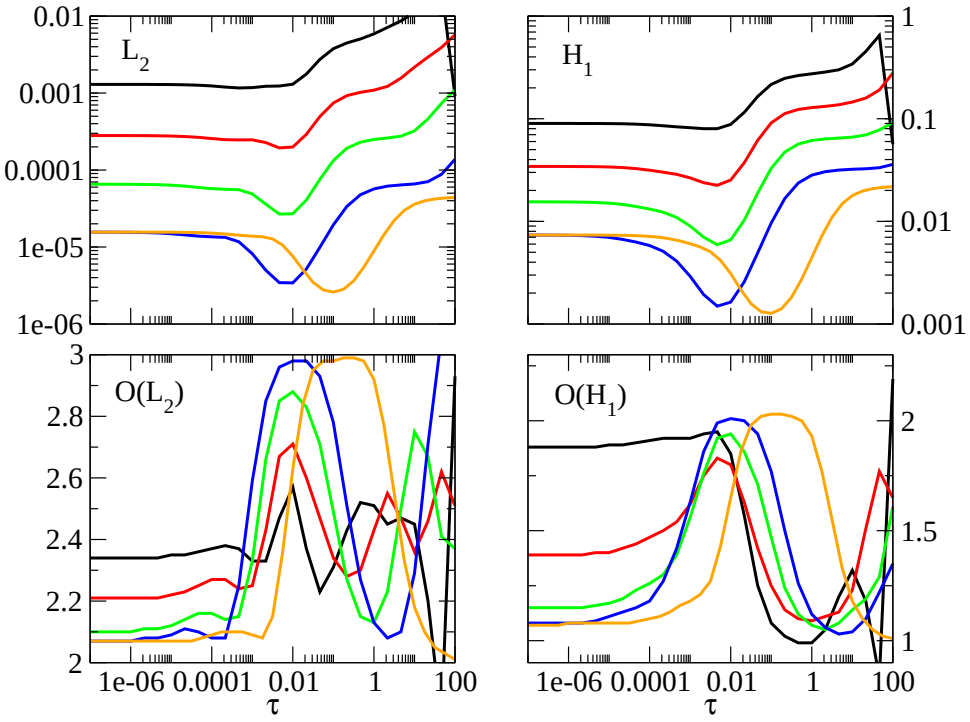

Abbildung 7.6: Zu sehen ist das gleiche wie in Abbildung 7.4 nur für Finite–Elemente zweiter Ordnung auf dem Dreiecksgitter.

Wie der Vergleich zwischen den Tabellen 7.4 und 7.5 zeigt, sind die Fehlerwerte der Stabilisierungsverfahren bei jeweils optimaler Wahl von τ auf dem Quadratgitter wiederum kleiner als auf dem Dreiecksgitter. Auch das τ–Intervall nahezu optimaler Stabilisierungsparameter ist größer als bei dem Dreiecksgitter, siehe hierzu Abbildung 7.7. Beide Verfahren besitzen etwa bei $\tau = 10$ die kleinsten Fehlerwerte. Bis zur zweiten Nachkommastelle sind die Fehler gleich. Das Standard–Galerkin–Verfahren liefert gemäß Tabelle 7.5 etwa vier Mal größere Fehlerwerte als die Stabilisierungsverfahren.

τ	L^2	H^1	$\mathcal{O}(L^2)$	$\mathcal{O}(H^1)$
10	2.08e-6	8.70e-4	2.98	2.01
10	2.08e-6	8.70e-4	2.98	2.01
-	8.70e-6	3.94e-3	2.11	1.12

Tabelle 7.5: Fehler und Fehlerordnungen für das exponentielle Beispiel bei Finite–Elementen zweiter Ordnung auf dem Level–5–Quadratgitter. Die erste Zeile bezieht sich auf die CIP–Methode und die zweite auf die Streamline–Diffusion–Methode bei jeweils optimaler Wahl des Stabilisierungsparameters. Die dritte Zeile zeigt die Werte des Standard–Galerkin–Verfahrens.

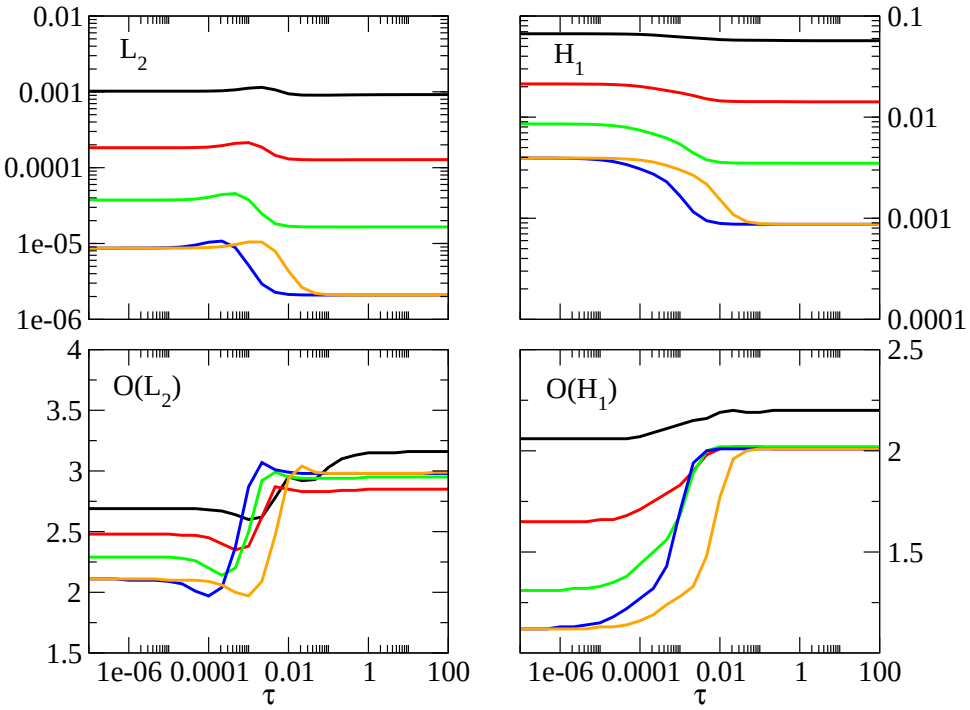

Abbildung 7.7: Zu sehen ist das gleiche wie in Abbildung 7.4 nur für Finite–Elemente zweiter Ordnung auf dem Quadratgitter.

Die Ergebnisse zu Finiten–Elementen dritter Ordnung auf dem Dreiecksgitter sind in Abbildung 7.8 zu sehen. Die an den Interpolationsfehler–Abschätzungen gemessenen optimalen Konvergenzraten werden durch die CIP–Methode auf dem Level–5–Gitter in dem τ–Intervall von 10^{-5} bis 10^{-3} erreicht. Selbiges gilt für die Streamline-Diffusion–Methode in einem etwa gleich großen Bereich. Auch die Fehlerwerte beider Verfahren sind ungefähr gleich groß. Im Vergleich zum Standard–Galerkin–Verfahren lassen sich gemäß Tabelle 7.6 durch die Stabilisierung die Fehler um nahezu eine Größenordnung verringern.

τ	L^2	H^1	$\mathcal{O}(L^2)$	$\mathcal{O}(H^1)$
0.001	1.63e-8	1.65e-5	4.02	3.01
0.031	1.58e-8	1.66e-5	4.01	3.01
-	1.60e-7	1.27e-4	3.18	2.20

Tabelle 7.6: Fehler und Fehlerordnungen für das exponentielle Beispiel bei Finite–Elementen dritter Ordnung auf dem Level–5–Dreiecksgitter. Die erste Zeile bezieht sich auf die CIP–Methode und die zweite auf die Streamline–Diffusion–Methode bei jeweils optimaler Wahl des Stabilisierungsparameters. Die dritte Zeile zeigt die Werte des Standard–Galerkin–Verfahrens.

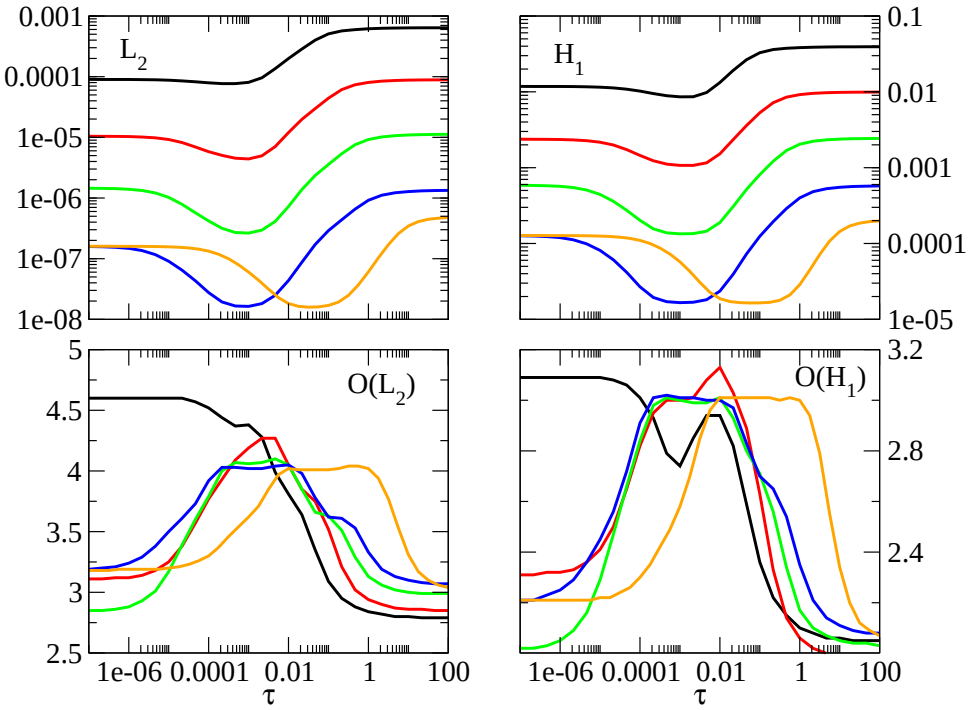

Abbildung 7.8: Zu sehen ist das gleiche wie in Abbildung 7.4 nur für Finite–Elemente dritter Ordnung auf dem Dreiecksgitter.

Abbildung 7.9 zeigt die Resultate für Finiten–Elemente dritter Ordnung auf dem Quadratgitter. Im Gegensatz zu den Elementen auf dem Dreiecksgitter können die Ergebnisse durch eine Stabilisierung kaum verbessert werden. Bei zu großen τ schneiden die Stabilisierungsverfahren sogar schlechter als das Standard–Galerkin–Verfahren ab. Die CIP–Methode reagiert hierbei sensibler auf zu groß gewählte τ. Erneut sind die Fehlerwerte aller Verfahren auf dem Quadratgitter kleiner als auf dem Dreiecksgitter.

τ	L^2	H^1	$\mathcal{O}(L^2)$	$\mathcal{O}(H^1)$
1.0e-6	1.202e-8	1.320e-5	3.97	3.05
1.0e-6	1.201e-8	1.321e-5	3.97	3.05
-	1.201e-8	1.321e-5	3.97	3.05

Tabelle 7.7: Fehler und Fehlerordnungen für das exponentielle Beispiel bei Finite–Elementen dritter Ordnung auf dem Level–5–Quadratgitter. Die erste Zeile bezieht sich auf die CIP–Methode und die zweite auf die Streamline–Diffusion–Methode bei jeweils optimaler Wahl des Stabilisierungsparameters. Die dritte Zeile zeigt die Werte des Standard–Galerkin–Verfahrens.

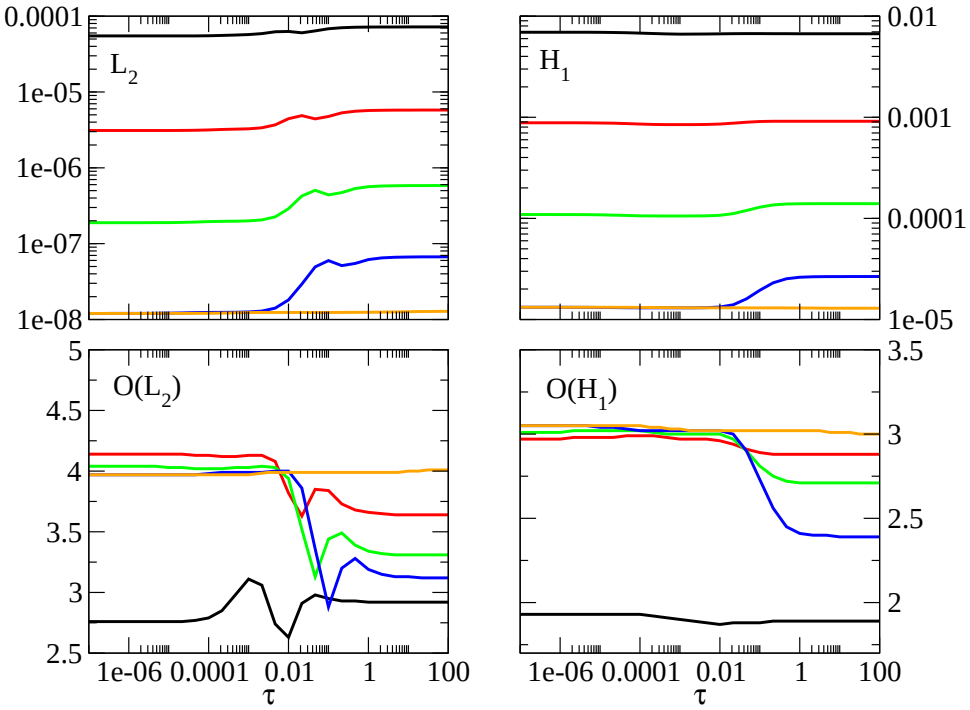

Abbildung 7.9: Zu sehen ist dasselbe wie in Abbildung 7.4 nur für Finite–Elemente dritter Ordnung auf dem Quadratgitter.

Obige Ergebnisse können wie folgt zusammengefasst werden: Bei allen betrachteten Fällen gibt es für beide Stabilisierungsverfahren auf dem Level–5–Gitter einen τ–Bereich, in welchem die von der Analysis vorausgesagten Konvergenzraten erreicht werden. Für die CIP–Methode haben wir uns zudem davon vergewissert, dass der Bereich mit feiner werdendem Gitter wächst. In manchem Fällen werden hierbei sogar die an der Interpolationsfehler–Abschätzung gemessenen optimalen Raten angenommen.
Die Streamline–Diffusion–Methode schneidet geringfügig besser als die CIP–Methode ab. Bei jeweils optimaler Wahl von τ sind die Fehler etwa gleich groß oder etwas kleiner bei der Streamline–Diffusion–Methode wie beispielsweise bei Finite–Elementen zweiter Ordnung auf dem Dreiecksgitter. Auch das Intervall beinahe optimaler τ ist bei der Streamline–Diffusion–Methode oft größer, wie Tabelle 7.8 zusammenfassend zeigt.
Abhängig vom Gittertyp und der Elementordnung können die Fehlerwerte durch eine Stabilisierung verringert werden. Relativ deutlich ist dies für die Finiten–Elemente zweiter Ordnung auf beiden Gittern und für Finite Elemente dritter Ordnung auf dem Dreiecksgitter. Es gibt aber auch Fälle, in denen das Standard–Galerkin–Verfahren nahezu gleich gute Ergebnisse wie die Stabilisierungverfahren liefert. Ein Beispiel hierfür sind Finite–Elemente dritter Ordnung auf dem Quadratgitter.
Die Fehler waren für alle Verfahren auf den Quadratgittern stets kleiner als auf den Dreiecksgittern. Auch der τ–Bereich nahezu optimaler Fehlerwerte wächst, wenn statt dem Dreiecksgitter das Quadratgitter verwandt wird.

Elementordnung	Gittertyp	$\tau_{\text{min,SD}}$	$\tau_{\text{max,SD}}$	$\tau_{\text{min,CIP}}$	$\tau_{\text{max,CIP}}$
1	Dreiecksgitter	0	2.1	0	0.046
1	Quadratgitter	0	46	0	0.046
2	Dreiecksgitter	0.05	0.18	3.53e-3	9.12e-3
2	Quadratgitter	0.046	$> 1e13$	4.6e-3	1e7
3	Dreiecksgitter	0.017	0.1	4.6e-4	2.1e-3
3	Quadratgitter	0	$> 1e13$	0	4.6e-3

Tabelle 7.8: τ–Bereiche für das exponentielle Beispiel, in denen die L^2– und H^1–Fehler um weniger als 10% von den Fehlerwerten bei einer optimalen Wahl von τ abweichen. Die dritte sowie vierte Spalte zeigen das τ–Intervall für die Streamline–Diffusion–Methode und die fünfte sowie sechste für die CIP–Methode. Betrachtet wird Gitterlevel 5.

Obige Ergebnisse sind qualitativ auch bei weiteren Beispielen beobachtet worden. Die Ergebnisse für ein anderes Beispiel aus [BH04] werden in Anhang B gezeigt. Auch bei diesem Beispiel sind unsere Ergebnisse für Elemente erster Ordnung auf dem Dreiecksgitter vergleichbar mit denen von [BH04].

Auch der Einfluss der Viskosität auf die Ergebnisse ist getestet worden. Verringert man ϵ von 10^{-6} weiter, ändern sich die Ergebnisse des Standard–Galerkin–Verfahrens und der beiden Stabilisierungsmethoden nur vernachlässigbar. Erhöht man ϵ, so bleiben die Ergebnisse zunächst gleich bis sie sich mit größer werdendem ϵ nach und nach verschlechtern. Bei großen ϵ befindet man sich im diffusions–dominanten Fall, in welchem üblicherweise auf eine Stabilisierung verzichtet werden kann.

7.4 Laufzeitmessungen

Im letztem Abschnitt haben wir unter anderem die Genauigkeit der Streamline–Diffusion– und der CIP–Methode miteinander verglichen. Gemessen an den Fehlerwerten hat die Streamline–Diffusion–Methode geringfügig besser abgeschnitten.

Ein weiteres wichtiges Vergleichskriterium ist die Laufzeit der Verfahren, das heißt die Zeit, welche die Verfahren benötigen, um die Systemmatrix zu assemblieren und das resultierende lineare Gleichungssystem zu lösen. Die Laufzeit untersuchen wir nun ebenfalls exemplarisch anhand des exponentiellen Beispiels. Für einen fairen Vergleich fixieren wir den Stabilisierungsparameter jeweils ungefähr auf den Wert, bei welchem die Fehlerwerte am kleinsten sind. Diese Werte sind auch den Tabellen aus dem letzten Abschnitt zu entnehmen. Den Stabilisierungsparameter δ_0 der Streamline–Diffusion–Methode kürzen wir durch τ ab.

Tabelle 7.9 zeigt Laufzeiten der Verfahren für Finite–Elemente erster Ordnung auf dem Quadrat– und dem Dreiecksgitter von Level vier bis sieben. Für einen ausreichend großen Gitterlevel benötigt die Streamline–Diffusion–Methode auf beiden Gittern weniger Zeit. Um den quantitativen Vergleich zwischen den Verfahren zu vereinfachen, gibt die Größe $\overline{\Delta t}$ den prozentuale Zuwachs in der Laufzeit an, welcher hinzunehmen ist, wenn statt der residualen Stabilisierung die CIP–Methode verwandt wird, das heißt:

$$\overline{\Delta t} = \frac{t_{\mathrm{CIP}} - t_{\mathrm{SD}}}{t_{\mathrm{SD}}},$$

wobei t_{SD} die Laufzeit der Streamline–Diffusion–Methode bezeichnet und t_{CIP} diejenige der CIP–Methode. Gemäß der Tabelle steigt der prozentuale Zuwachs mit zunehmendem Gitterlevel und beträgt auf dem Level–7–Dreiecksgitter bereits etwa 86% und auf dem Level–7–Quadratgitter etwa 204%.

Tabelle 7.9 erlaubt auch einen Vergleich zwischen den Laufzeiten auf dem Quadrat- und auf dem Dreiecksgitter. Gemessen an den Fehlerwerten waren die Quadratgitter in den vorausgehenden Abschnitten stets vorzuziehen. Gemäß der Tabelle sind die Rechnungen auf dem Quadratgitter bei der Streamline–Diffusion–Methode zudem schneller. Bei der CIP–Methode ist es aber umgekehrt, hier sind die Rechnungen auf dem Dreiecksgitter schneller.

Level	t_{SD} [s]	t_{CIP} [s]	$\overline{\Delta t}$	t_{SD} [s]	t_{CIP} [s]	$\overline{\Delta t}$
4	0.16	0.25	0.56	0.21	0.17	-0.19
5	0.40	0.75	0.87	0.57	0.31	-0.45
6	0.88	1.89	1.14	1.06	1.75	0.65
7	4.52	13.76	2.04	5.48	10.23	0.86

Tabelle 7.9: Laufzeit bei Finite–Elementen erster Ordnung für das exponentielle Beispiel. Die zweite bis vierte Spalte beziehen sich auf das Quadratgitter, die fünfte bis siebte auf das Dreiecksgitter. Für die Streamline–Diffusion–Methode ist auf dem Quadratgitter $\tau = 0.001$ und auf dem Dreiecksgitter $\tau = 0.032$. Für die CIP–Methode ist auf beiden Gittern $\tau = 0.001$.

Die Laufzeiten für Elemente zweiter Ordnung sind in Tabelle 7.10 zu sehen. Hier ist die

Laufzeit der Streamline–Diffusion–Methode auf allen Leveln und beiden Gittertypen geringer und der prozentuale Zuwachs $\overline{\Delta t}$ nimmt auf den höheren Level mit feiner werdendem Gitter zu. Im Vergleich zu den Elementen erster Ordnung ist der prozentuale Zuwachs ebenfalls gestiegen. Beide Stabilisierungsverfahren rechnen auf dem Dreieckgitter schneller als auf dem Quadratgitter. Ein Plausibilitätsargument hierfür ist, dass auf dem Quadratgitter mehr Freiheitsgrade in Verbindung miteinander stehen als auf dem Dreiecksgitter. So kann zum Beispiel bei der Streamline–Diffusion–Methode auf dem Quadratgitter ein Freiheitsgrad mit 25 Freiheitsgraden gekoppelt sein, während es auf dem Dreiecksgitter nur 19 mögliche Kopplungen sind. Jede der Kopplungen kann einen von Null verschiedenen Eintrag in der Systemmatrix erzeugen. Je größer die Anzahl der von Null verschiedenen Matrixeinträge ist, desto aufwendiger gestaltet sich im Allgemeinen die Lösung des zu Grunde liegenden Gleichungssystems.

Level	t_{SD} [s]	t_{CIP} [s]	$\overline{\Delta t}$	t_{SD} [s]	t_{CIP} [s]	$\overline{\Delta t}$
3	0.06	0.19	2.16	0.05	0.14	1.8
4	0.32	0.68	1.12	0.25	0.61	1.44
5	0.93	2.86	2.07	0.52	1.46	1.80
6	2.70	25.14	8.31	2.66	11.22	3.09

Tabelle 7.10: Laufzeit bei Finite–Elementen zweiter Ordnung für das exponentielle Beispiel. Die zweite bis vierte Spalte beziehen sich auf das Quadratgitter, die fünfte bis siebte auf das Dreiecksgitter. Für die Streamline–Diffusion–Methode ist auf dem Quadratgitter $\tau = 10$ und auf dem Dreiecksgitter $\tau = 0.1$. Für die CIP–Methode ist auf dem Quadratgitter $\tau = 10$ und auf dem Dreiecksgitter $\tau = 4.64 \cdot 10^{-3}$.

Nach Tabelle 7.11 ist auch für Elemente dritter Ordnung die Streamline–Diffusion–Methode gemessen an der Laufzeit vorzuziehen. Der prozentuale Zuwachs nimmt auf beiden Gittern mit dem Gitterlevel zu. Für Gitterlevel größer gleich vier ist der prozentuale Zuwachs im Vergleich zu den Elementen zweiter Ordnung weiter angestiegen. Auf den höheren Gitterleveln sind die Rechnungen bei beiden Stabilisierungsverfahren auf dem Dreiecksgitter schneller als auf dem Quadratgitter.

Level	t_{SD} [s]	t_{CIP} [s]	$\overline{\Delta t}$	t_{SD} [s]	t_{CIP} [s]	$\overline{\Delta t}$
3	0.18	0.52	1.88	0.13	0.37	1.84
4	0.41	1.87	3.56	0.36	1.13	2.13
5	1.30	16.83	11.9	1.11	6.19	4.57
6	7.31	164.20	21.4	6.50	51.30	6.89

Tabelle 7.11: Laufzeit bei Finite–Elementen dritter Ordnung für das exponentielle Beispiel. Die zweite bis vierte Spalte beziehen sich auf das Quadratgitter, die fünfte bis siebte auf das Dreiecksgitter. Für die Streamline–Diffusion–Methode ist auf dem Quadratgitter $\tau = 10^{-6}$ und auf dem Dreiecksgitter $\tau = 0.031$. Für die CIP–Methode ist auf dem Quadratgitter $\tau = 10^{-6}$ und auf dem Dreiecksgitter $\tau = 10^{-3}$.

Auf dem Level–6–Quadratgitter beträgt die Laufzeit der CIP–Methode 164.20 Sekunden und ist damit um einen Faktor 10 langsamer als die Streamline–Diffusion–Methode. Gemäß weiteren Tests werden von dieser Laufzeit 1.82 Sekunden für die Assemblierung der Terme

des Standard–Galerkin–Verfahrens und 3.4 Sekunden für die Assemblierung des Stabilisierungsterms verwandt, während die Lösung des linearen Gleichungssystems die restliche Zeit in Anspruch nimmt. Die CIP–Methode ist also vor allem deshalb langsamer als die Streamline–Diffusion–Methode, weil die Lösung des linearen Gleichungssystems mehr Zeit erfordert. Ursache hierfür ist das dichtere Besetzungsschema der Systemmatrix der CIP–Methode, welche durch die zusätzliche Kopplung der Freiheitsgrade über die Kantenintegrale verursacht wird (siehe hierzu auch Abschnitt 7.1).

Quantitativ kann das dichtere Besetzungsschema der CIP–Methode durch die Anzahl der von Null verschiedenen Einträge in der Systemmatrix erfasst werden. Diese sind in Tabelle 7.12 exemplarisch auf den Level–5–Gittern zu sehen. Für die CIP–Methode ist die maximale Anzahl $N_{\max,\mathrm{CIP}}$ der von Null verschiedenen Matrixeinträge gezeigt sowie die Anzahl $N_{\neq 0,\mathrm{CIP}}$ der Matrixelemente, welche speziell für das exponentielle Beispiel bei $\tau = 1$ ungleich Null sind. Gemäß weiteren Tests ändert sich $N_{\neq 0,\mathrm{CIP}}$ für andere $\tau > 0$ vernachlässigbar. Zum Vergleich enhält die Tabelle auch die maximale Anzahl $N_{\max,\mathrm{SD}}$ der von Null verschiedenen Matrixeinträge der Streamline–Diffusion–Methode. Die Anzahl der Matrixeinträge der Streamline–Diffusion–Methode, welche im vorliegenden Beispiel bei $\tau > 0$ ungleich Null sind, liegt nur geringfügig unterhalb von $N_{\max,\mathrm{SD}}$ und wird daher nicht dargestellt. Gemäß der Tabelle ist $N_{\neq 0,\mathrm{CIP}}$ stets größer als $N_{\max,\mathrm{SD}}$. Dies gilt für Elemente erster bis dritter Ordnung auf dem Dreiecks- sowie Quadratgitter. Zudem ist zu erkennen, dass bei den Rechnungen auf den Quadratgittern mehr von Null verschiedene Matrixeinträge auftreten als bei denjenigen auf den Dreiecksgittern.

Elementordnung	Gittertyp	$N_{\max,\mathrm{SD}}$	$N_{\neq 0,\mathrm{CIP}}$	$N_{\max,\mathrm{CIP}}$
1	Dreiecksgitter	2.8e5	4.38e5	5.16e5
1	Quadratgitter	3.60e5	5.94e5	8.28e5
2	Dreiecksgitter	1.85e6	3.29e6	4.00e6
2	Quadratgitter	2.57e6	5.11e6	7.64e6
3	Dreiecksgitter	6.19e6	1.20e7	1.48e7
3	Quadratgitter	9.09e6	1.98e7	3.06e7

Tabelle 7.12: Maximale Anzahl $N_{\max,\mathrm{SD}}$ der von Null verschiedenen Matrixeinträgen bei der Streamline–Diffusion–Methode, Anzahl $N_{\neq 0,\mathrm{CIP}}$ der von Null verschiedenen Matrixeinträgen bei der CIP–Methode speziell für das exponentielle Beispiel und maximale Anzahl $N_{\max,\mathrm{CIP}}$ der von Null verschiedenen Matrixeinträge der CIP–Methode. Betrachtet wird Gitterlevel 5. Für weitere Details siehe Text.

Als Hauptergebnis dieses Abschnitts halten wir fest, dass die Streamline–Diffusion–Methode deutlich schnellere Rechnungen als die CIP–Methode erlaubt. Die Unterschiede wachsen dabei mit dem Gitterlevel und der Elementordnung. Daneben hat sich gezeigt, dass die Rechnungen auf den Dreiecksgittern zumeist schneller sind als diejenigen auf dem Quadratgitter, da auf den Dreiecksgittern weniger Freiheitsgrade miteinander verkoppelt werden.

7.5 Rechnungen mit Grenzschichten

In Abschnitt 7.3 hat sich gezeigt, dass mit dem Standard–Galerkin–Verfahren auch im konvektions–dominanten Fall gute Ergebnisse erzielt werden können, wenn die Lösung keine Grenzschichten besitzt. Bei Lösungen mit Grenzschichten ist dies im Allgemeinen nicht der Fall, weil es in den Grenzschichten zu Oszillationen kommt, die sich über einen Großteil des Lösungsgebiets ausdehnen. Inwiefern sich die Ergebnisse durch die Streamline–Diffusion– und die CIP–Methode verbessern lassen, diskutieren wir in diesem Abschnitt. Es sei aber vorweggenommen, dass keine besonders guten Ergebnisse zu erwarten sind, weil wir bei unseren Rechnungen auf shock capturing Terme verzichten. Für eine Untersuchung mit shock capturing Termen sei auf [JK07a] und [JK07b] verwiesen.

Im Folgenden betrachten wir das Problem mit den Daten $\epsilon = 10^{-8}$, $b = (1,0)^T$, $c = 0$, $f = 1$ auf $\Omega = [0,1]^2$ und $u = 0$ auf $\partial\Omega$. Zu diesen Vorgaben erhält man die in Abbildung 7.10 gezeigte Lösung. Diese besitzt zwei parabolische Grenzschichten bei $y = 0$ und $y = 1$ und eine exponentielle bei $x = 1$.

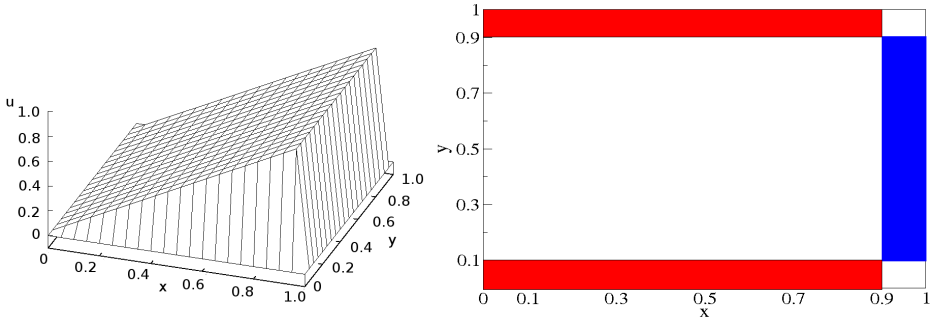

Abbildung 7.10: Links ist die Lösung des in diesem Abschnitt betrachteten Beispiels zu sehen, rechts die Bereiche, welche zur Definition der Kenngrößen L^1_{para} und L^1_{expo} benutzt werden. Der rot unterlegte Bereich entspricht Ω_{para} und der blau unterlegte Ω_{expo}.

Die Ergebnisse der CIP–Methode bei diesem Beispiel erklären wir exemplarisch anhand Finiter–Elemente erster Ordnung auf dem Level–5–Quadratgitter. Abbildung 7.5 zeigt hierzu die berechnete Lösung, wenn der Stabilisierungsparameter τ gleich 100 gewählt wird. An den parabolischen Grenzschichten treten Oszillationen auf, während die exponentielle Grenzschicht verschmiert ist. Das Verschmieren reicht auch in das Innere des Lösungsgebiets hinein, was bis ungefähr $x \geq 0.5$ zu erkennen ist. Das Verschmieren der Lösung ist auf eine zu starke Stabilisierung zurückzuführen. Daher verringern wir auf der rechten Seite von Abbildung 7.5 den Stabilisierungsparameter auf $\tau = 0.5$. Die exponentielle Grenzschicht ist nun nicht mehr verschmiert, dafür treten dort Oszillationen auf. Der Vorteil ist aber, dass die Oszillationen in der exponentiellen Grenzschicht nicht mehr soweit in das Innere des Lösungsgebiet hineinreichen wie das Verschmieren bei $\tau = 100$, so dass die die Lösung im Großteil des Lösungsgebiet einigermaßen gut approximiert wird.
In Abbildung 7.12 sehen wir die Lösung, wenn τ auf 0.01 verringert wird. Vor allem die Oszillationen in der exponentiellen Randschicht haben weiter zugenommen. Darüber hinaus haben sich die Oszillationen über den Großteil des Lösungsgebiets ausgebreitet. Bei

$\tau = 0.01$ ist die Stabilisierung also zu schwach, um die Lösung zu kontrollieren.

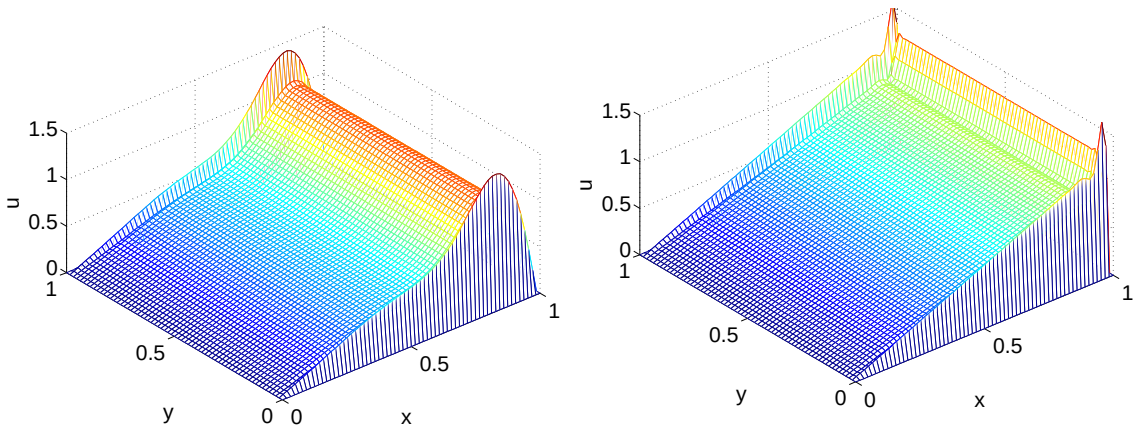

Abbildung 7.11: Links Lösung der CIP–Methode bei $\tau = 100$ und rechts bei $\tau = 0.5$.

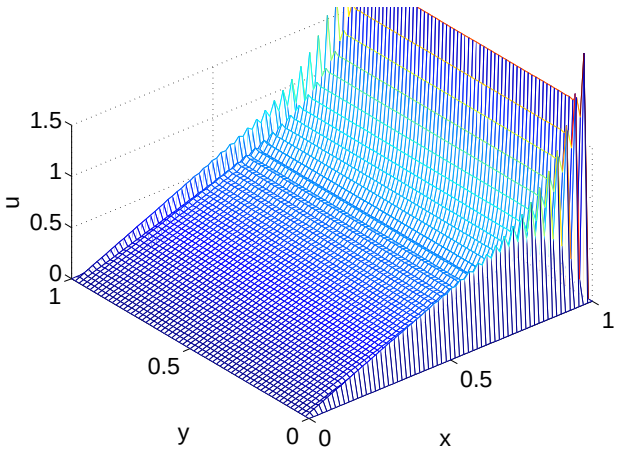

Abbildung 7.12: Lösung der CIP–Methode bei $\tau = 0.01$.

Als nächstes sollen obige Ergebnisse quantitativ anhand zweier Fehlergrößen beschrieben werden. Die erste davon ist

$$L^1_{\text{para}} = \frac{1}{N_{\text{para}}} \sum_{(x,y)\in\Omega_{\text{para}}} |u_h(x,y) - x| \tag{7.6}$$

wobei u_h der Näherungslösung entspricht, Ω_{para} der Menge aller Finite–Element Knoten (x_h, y_h) mit $1 > y_h \geq 0.9$ oder $0 < y_h \leq 0.1$ sowie $x_h \leq 0.9$ und N_{para} der Anzahl der Knoten in Ω_{para}. Die Menge Ω_{para} ist in Abbildung 7.10 als rot unterlegter Bereich zu sehen.
Die Bedeutung von L^1_{para} ergibt sich durch folgende Überlegung: Die exakte Lösung des Problems außerhalb der Grenzschichten lautet $u(x,y) = x$. Gemäß Abschnitt 6.1 besitzen die parabolischen Grenzschichten eine Breite der Ordnung $\mathcal{O}(\sqrt{\epsilon}\ln(1/\epsilon))$. Bei der Vorgabe $\epsilon = 10^{-8}$ ergibt dies etwa eine Breite von 0.001. Das Level–5–Quadratgitter besitzt eine Gitterweite von $(0.5)^5 = 0.03125$. Bei Finiten–Elementen nicht zu großer Ordnung liegen somit alle Knoten, welche nicht auf dem Gebietsrand liegen, außerhalb der parabolischen

Grenzschichten. Damit besteht die Menge Ω_{para} also aus Knoten, welche nahe der parabolischen Grenzschichten, aber noch nicht in den parabolischen Grenzschichten liegen. Folglich lautet die exakte Lösung $u(x, y) = x$ in Ω_{para} und die Fehlergröße L^1_{para} ist der lokale Fehler der Näherungslösung u_h in den parabolischen Grenzschichten.

Analog zu L^1_{para} definieren wir durch L^1_{expo} eine zweite Fehlergröße, welche den lokalen Fehler in der exponentiellen Grenzschicht misst:

$$L^1_{\text{expo}} = \frac{1}{N_{\text{expo}}} \sum_{(x,y)\in\Omega_{\text{expo}}} |u_h(x, y) - x| .$$

Hierbei bezeichnet Ω_{expo} die Menge aller Knoten (x_h, y_h) mit $1 > x_h \geq 0.9$ und $y_h \geq 0.1$ sowie $y_h \leq 0.9$. Die Größe N_{expo} gibt die Anzahl der Knoten in Ω_{expo} an. Die Menge Ω_{expo} ist in Abbildung 7.10 als blau unterlegter Bereich zu sehen. Es sei noch erwähnt, dass in der Arbeit [JK07a] andere Fehlergrößen verwandt werden, um die Fehler in den Grenzschichten zu beschreiben. Da bei unseren Rechnungen allerdings ohne shock capturing Terme gerechnet wird und daher deutlich größere Fehler als in [JK07a] auftreten sollten, sind die Fehlergrößen von dort in unserem Fall weniger sinnvoll.

Abbildung 7.13 zeigt in der oberen Zeile die Werte der Fehlergrößen für Finite–Elemente erster Ordnung auf dem Level–5–Quadratgitter. Sowohl L^1_{para} als auch L^1_{expo} werden etwa bei $\tau = 0.5$ minimal. Die Grenzschichten können demnach mit der CIP–Methode auf dem Level–5–Quadratgitter kaum besser als in Abbildung 7.5 auf der rechten Seite dargestellt genähert werden. Zum Vergleich sind in der Abbildung die Fehlerwerte der Streamline–Diffusion–Methode als orangefarbene Kurven aufgetragen. Auch hier liegt das Minimum der Fehler etwa bei bei $\tau = 0.5$. Die Fehlergröße L^1_{para} ist dort etwa so groß wie L^1_{para} der CIP–Methode, während L^1_{expo} etwa um 50% kleiner als die entsprechende Fehlergröße der CIP–Methode ist. Die Fehler sind allerdings auch bei der Streamline–Diffusion–Methode und einer optimalen Wahl von τ so groß, dass die Lösung in den Grenzschichten schlecht approximiert wird.
Wie die untere Zeile von Abbildung 7.13 zeigt, sind die Fehler beider Stabilisierungsverfahren für Elemente zweiter Ordnung auf demselben Gitter etwas kleiner. Davon abgesehen ist die Situation aber analog zu Elementen erster Ordnung.
Gemäß weiteren Tests verringern sich bei beiden Verfahren die Fehlergrößen L^1_{para} und L^1_{expo} mit der Gitterweite und das Aussehen der Näherungslösung nimmt immer mehr die Gestalt der exakten Lösung an. Um allerdings gute Lösungen in den Grenzschichten zu erhalten, muss die Gitterweite sehr klein gewählt werden, was einen hohen Rechenaufwand erfordert. Demnach sind beide Stabilisierungsverfahren für konvektions–dominante Problemen mit Grenzschichten wenig zu empfehlen, falls die Lösung auch in den Grenzschichten gut approximiert werden muss.

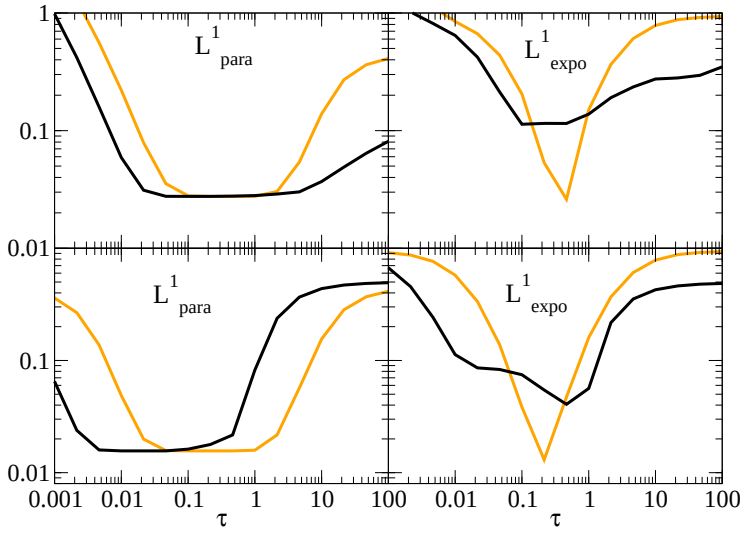

Abbildung 7.13: Lokale Fehlerwerte in den Grenzschichten der Streamline–Diffusion–Methode (orangefarbene Kurven) und der CIP–Methode (schwarze Kurven) im Auftrag gegen τ. Die Linke Spalte zeigt den Fehler L^1_{para} in den parabolischen Grenzschichten, die rechte den Fehler L^1_{expo} in der exponentiellen Grenzschicht. Die obere Zeile bezieht sich auf Elemente erster Ordnung, die untere auf Elemente zweiter Ordnung.

7.6 Zusammenfassung

Unsere Resultate für Probleme ohne Grenzschichten können wie folgt zusammengefasst werden: Die von der Analysis vorausgesagten Konvergenzraten werden durch die Streamline–Diffusion– und die CIP–Methode bei geeigneter Wahl der Stabilisierungsparameter erreicht. Das zugehörige Stabilisierungsparameter–Intervall wächst wie gewünscht mit feiner werdendem Gitter.
Abhängig von der Elementordnung und dem Gitterlevel sind die Ergebnisse der Stabilisierungsverfahren entweder besser oder ungefähr gleich den Resultaten des Standard–-Galerkin–Verfahrens.
Gemessen an den Fehlerwerten schneiden beide Stabilisierungsverfahren etwa gleich gut ab. Die Streamline–Diffusion–Methode erlaubt jedoch vor allem für feinere Gitterweiten und höhere Elementordnungen die deutlich schnelleren Rechnungen. Folglich ist der Streamline–Diffusion–Methode in der Regel der Vorzug zu geben.
Beide Stabilisierungsverfahren besitzen auf dem Quadratgitter stets kleinere Fehlerwerte als auf dem Dreiecksgitter. Die Rechnungen auf dem Dreiecksgitter sind jedoch außer bei Elementen erster Ordnung bei der Streamline–Diffusion–Methode schneller als auf dem Quadratgitter.
Auch ein Testbeispiel mit Grenzschichten ist betrachtet worden. Die Lösung des Standard–Galerkin–Verfahrens zeigt Oszillationen über das gesamte Lösungsgebiet. Diese können durch die Stabilisierungsverfahren im Großteil des Lösungsgebiets unterdrückt werden. Um jedoch eine gute Näherungen in den Grenzschichten zu erhalten, müsste das Gitter sehr fein gewählt werden. Dies ist oft nicht praktikabel. Insofern sind die Stabilisierungsverfahren ohne weitere Modifikation nicht für Probleme mit Grenzschichten zu empfehlen.

Kapitel 8

Die Oseen–Gleichung

Neben der Konvektions–Diffusionsgleichung untersuchen wir in dieser Arbeit die Oseen–Gleichung. Wie wir sehen werden, spielt die Oseen–Gleichung eine wichtige Rolle in der Fluiddynamik. In einem ersten Schritt überführen wir die Oseen–Gleichung in ein gemischtes Variationsproblem und untersuchen dessen Lösbarkeit. Zur Approximation einer Lösung stellen wir anschließend das Standard–Galerkin–Verfahren vor. Da das Standard–Galerkin–Verfahren in vielen Fällen instabil ist und unbrauchbare Ergebnisse liefert, führen wir mit der residualen und der Kanten–Stabilisierung verbesserte Verfahren ein. Für beide Verfahren untersuchen wir, wann eine eindeutige Lösung existiert, und leiten eine a–priori Abschätzung für den Verfahrensfehler ab.

8.1 Motivation und schwache Formulierung

Die Oseen–Gleichung auf einem offenen beschränkten Gebiet $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, lautet

$$-\nu\Delta u + (b \cdot \nabla u) + cu + \nabla p = f \qquad \text{in } \Omega, \tag{8.1a}$$

$$\nabla \cdot u = 0 \qquad \text{in } \Omega \tag{8.1b}$$

mit passenden Randbedingungen auf dem Gebietsrand $\partial\Omega$. Zu lösen ist die Gleichung nach u und p. Die Größen b, f und c sind die Daten des Problems. Gegenüber der Konvektions–Diffusionsgleichung ist hier u eine vektorwertige Funktion mit Bild in $\mathbb{R}^d$. Die Daten b und f sind ebenfalls vektorwertig mit Bild in $\mathbb{R}^d$, während ν, c und p skalare Größen sind. Ist $c \neq 0$, so bezeichnet man obige Gleichung auch manchmal als verallgemeinerte Oseen–Gleichung.

Die Oseen-Gleichung ist als lineare Näherung der Navier–Stokes–Gleichung eingeführt worden [Ose10]. In diesem Fall ist u das Geschwindigkeitsfeld eines inkompressiblen Fluids, p das Druckfeld, ν die Viskosität des Fluids und $c = 0$. Daneben hat die Oseen–Gleichung in der numerischen Mathematik an Bedeutung gewonnen, da sie unter anderem bei der iterativen Lösung der Navier–Stokes–Gleichung zu lösen ist. Hier entspricht b der Näherungslösung von u aus dem letzten Iterationsschritt. Handelt es sich zusätzlich um ein instationäres Problem bei dem eine Zeitdiskretisierung durchgeführt wird, so tritt auch der Term cu mit c ungleich Null in der Oseen–Gleichung auf. Weiterhin ist die Oseen–Gleichung als Hilfsproblem bei Trubulenzmodellen von Interesse [LR06].

In der Oseen–Gleichung tritt nur die erste Ableitung des Druckfeldes auf. Damit kann der Druck aus der Gleichung nur bis auf eine Konstante bestimmt werden. Um dennoch unter geeigneten Bedingungen die Eindeutigkeit einer Lösung zu erhalten, ist der Druck

zu normieren. Eine Normierung gelingt zum Beispiel durch die Forderung, dass p einen Mittelwert von 0 auf Ω besitzt:

$$\int_\Omega p(x)dx = 0. \tag{8.2}$$

Der Einfachheit halber betrachten wir die Oseen–Gleichung im Wesentlichen mit homogenen Dirichlet–Randwerten:

$$u = 0 \qquad \text{auf } \partial\Omega.$$

Für eine Behandlung der Oseen-Gleichung mit analytischen Methoden sei auf [Bat00] verwiesen. Wir beschränken uns hingegen auf die schwache Lösungstheorie der Gleichung. Für die schwache Formulierung setzen wir $b \in [W^{1,\infty}]^d$, $d = 2, 3$, $\nu > 0$, $c \in L^\infty(\Omega)$, $c \geq c_{\min} \geq 0$, $f \in [L^2(\Omega)]^d$ und $(\nabla \cdot b) = 0$ in Ω voraus. Der Übergang zur schwachen Formulierung gelingt mit den folgenden Schritten:

- Multipliziere die Gleichung (8.1a) mit einer beliebige Testfunktion $v \in H_0^1(\Omega)$ sowie (8.1b) mit einer beliebigen Testfunktion $q \in L^2(\Omega)$.
- Integriere beide Gleichungen über Ω.
- Wende in (8.1a) partielle Integration an, um die Ableitungen zweiter Ordnung im Diffusionsterm sowie die Ableitung im Druckterm auf die Testfunktion v_h abzuwälzen. Damit geht (8.1a) über in:

$$\nu(\nabla u, \nabla v) - \nu(n\nabla u, v)_{\partial\Omega} + ((b\cdot\nabla)u, v) + (cu, v) - (p, \nabla\cdot v) + (p, n\cdot v)_{\partial\Omega} = (f, v). \tag{8.3}$$

- Da die Testfunktion v auf dem Gebietsrand verschwindet, setze die beiden Randintegrale in obiger Gleichung auf Null.

Damit lautet die schwache Formulierung:
Finde $(u, p) \in H_0^1(\Omega) \times L_*^2(\Omega)$, so dass für alle $(v, q) \in H_0^1(\Omega) \times L_*^2(\Omega)$ gilt:

$$\begin{aligned} a(u, v) + b(v, p) &= (f, v), \\ b(u, q) &= 0 \qquad \text{mit} \\ a(u, v) &= \nu(\nabla u, \nabla v) + ((b \cdot \nabla)u, v) + (cu, v) \quad \text{und} \\ b(u, q) &= -(\nabla \cdot u, q). \end{aligned} \tag{8.4}$$

Der Raum für den Druck ist dabei modifiziert worden zu

$$L_*^2(\Omega) = \left\{ q \in L^2(\Omega) : \int_\Omega q(x)dx = 0 \right\}, \tag{8.5}$$

so dass die Normierung aus (8.2) gilt. Gemäß obiger Herleitung löst eine klassische Lösung auch die schwache Formulierung. Die Umkehrung gilt hingegen im Allgemeinen nicht.
Die schwache Formulierung der Oseen–Gleichung entspricht einem gemischten Variationsproblem. Gemischte Variationsprobleme des obigen Typs haben wir in Abschnitt 3.2 kennen gelernt (vergleiche mit Gleichung (3.9)). Mit der dort eingeführten Theorie beweisen wir nun die eindeutige Lösbarkeit von (8.4). Dazu benötigen wir noch folgendes Resultat für den Divergenz-Operator div mit $\mathrm{div}(u) = \nabla \cdot u$ in der bisherigen Schreibweise.

Lemma 8.1.
Der Divergenz-Operator ist ein Isomorphismus von $L_*^2(\Omega)$ auf das orthogonale Komplement $Z^\perp$ von

$$Z = \{v \in [H_0^1(\Omega)]^d : \nabla \cdot v = 0\} = \{v \in [H_0^1(\Omega)]^d : b(v, q) = 0 \ \forall q \in L_*^2\}.$$

Beweis. Siehe [GR86] Korollar 2.4. Zentrales Hilfsmittel ist der Satz vom abgeschlossenen Bild. □

Damit können wir wie gewünscht die eindeutige Lösbarkeit der schwachen Formulierung beweisen:

Satz 8.2.
Seien die Voraussetzungen an die Daten b, c, f wie oben. Dann besitzt die schwache Formulierung der Oseen–Gleichung eine eindeutige Lösung $(u, p) \in [H_0^1(\Omega)]^d \times L_*^2(\Omega)$.

Beweis. Die Behauptung folgt aus Satz 3.10. Die Voraussetzungen des Satzes sind erfüllt, denn:

(1.) Die Bilinearformen a und b sind stetig, was man schnell mit Hilfe der Cauchy–Schwarz–Ungleichung und der Beschränktheit von b und c nachrechnet.

(2.) a ist Z-elliptisch, das heißt, es gilt:

$$a(v, v) \geq C\|v\|_{H^1(\Omega)} \qquad \forall v \in Z \tag{8.6}$$

mit einer Konstanten $C > 0$. Sei dazu $v \in [H_0^1(\Omega)]^d$ beliebig. Durch Anwenden der partiellen Integration erhält man:

$$((b \cdot \nabla)v, v) = -\frac{1}{2}(v, (\nabla \cdot b)v).$$

Da die rechte Seite der Gleichung wegen der Voraussetzung $\nabla \cdot b = 0$ verschwindet, können wir schreiben:

$$\begin{aligned} a(v, v) &= \nu(\nabla v, \nabla v) + ((b \cdot \nabla)v, v) + (cv, v) \\ &= \nu\,|v|_{H^1(\Omega)}^2 + c\|v\|_{L^2}^2 \\ &\geq C\|v\|_{H^1(\Omega)}, \end{aligned}$$

wobei die letzte Ungleichung aus der Friedrichs–Ungleichung (2.5) folgt. Weil $v \in H_0^1(\Omega)$ beliebig war, ist a insbesondere Z-elliptisch.

(3.) b genügt der Babuška–Brezzi–Bedingung:

$$\inf_{0 \neq q \in L_*^2} \sup_{0 \neq v \in [H_0^1(\Omega)]^d} \frac{b(v, q)}{|v|_{H^1(\Omega)}\,\|q\|_{L^2(\Omega)}} \geq \widetilde{C} \tag{8.7}$$

mit einer Konstante $\widetilde{C} > 0$: Sei $q \in L_*^2$ beliebig vorgegeben. Gemäß Lemma 8.1 existiert zu diesem q ein $w \in Z^\perp$ mit

$$\nabla \cdot w = q \qquad \text{und} \qquad |w|_{H^1(\Omega)} \leq C\|q\|_{L^2(\Omega)}.$$

mit einer Konstanten C. Die Ungleichung gilt hierbei wegen der Stetigkeit des Divergenz–Operators. Mit dieser Wahl von w ergibt sich:

$$\sup_{0 \neq v \in [H^1(\Omega)]^d} \frac{b(v, q)}{|v|_{H^1(\Omega)}} \geq \frac{b(w, q)}{|w|_{H^1(\Omega)}} = \frac{(\nabla \cdot w, q)}{|w|_{H^1(\Omega)}} = \frac{\|q\|_{L^2(\Omega)}^2}{|w|_{H^1(\Omega)}} \geq \frac{1}{C}\|q\|_{L^2(\Omega)}.$$

Damit erfüllt b die Babuška–Brezzi–Bedingung und alle Voraussetzungen von Satz 3.10 sind nachgewiesen. □

Bemerkung 8.3.
Die Randbedingungen sind bisher als homogen angenommen worden. Nun betrachten wir den Fall inhomogener Dirichlet–Randbedingungen

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

mit $g \in H^{1/2}(\Omega)$. Es stellt sich die Frage, wann eine eindeutige Lösung bei inhomogenen Randbedingungen existiert. Ein notwendiges Kriterium ist, dass die Randfunktion g kompatibel mit der Divergenzfreiheit des Geschwindigkeitsfelds ist. Inwiefern die Kompatiblität gegeben sein muss, zeigt folgende Rechnung unter Verwendung von partieller Integration im zweiten Schritt:

$$0 = \int_\Omega \nabla \cdot u(x)dx = \int_{\partial\Omega} (u \cdot n)(s)ds = \int_{\partial\Omega} (g \cdot n)(s)ds.$$

Ist diese Kompatibilitäts-Bedingung erfüllt, gibt es gemäß [GR86] Lemma 2.2 ein Vektorfeld $u_0 \in H^1(\Omega)$ mit

$$\nabla \cdot u_0 = g \quad \text{in } \Omega \text{ und } u_0 = g \quad \text{auf } \partial\Omega.$$

Definiere nun die Funktion $\widetilde{u} = u - u_0$. Einsetzen von $u = \widetilde{u} + u_0$ in die schwache Formulierung und Ausnutzen der Linearität von a ergibt das folgende Variationsproblem:
Finde $(\widetilde{u}, p) \in H_0^1(\Omega) \times L_*^2(\Omega)$, so dass für alle $(v, q) \in H_0^1(\Omega) \times L_*^2(\Omega)$ gilt:

$$\begin{aligned} a(\widetilde{u}, v) + b(v, p) &= (f, v) - a(u_0, v), \\ b(\widetilde{u}, q) &= 0. \end{aligned}$$

Die gesuchte Lösung hat hier wieder homogene Dirichlet–Randbedingungen zu erfüllen. Satz 8.2 ist also anwendbar und liefert die eindeutige Existenz einer Lösung $(\widetilde{u}, p)$. Rückrechnen auf (u, p) ergibt das gewünschte Lösbarkeitskriterium. Wir fassen zusammen: Wenn die Randfunktion g mit $g \in H^{1/2}(\partial\Omega)$ gegeben ist und g der Kompatibilitäts-Bedingung $\int_{\partial\Omega}(n \cdot g)(s)ds = 0$ genügt, so besitzt das Problem mit inhomogenen Dirichlet–Randwerten g eine eindeutige Lösung.

8.2 Das Standard–Galerkin–Verfahren

Die Lösung der schwachen Formulierung werden wir mit gemischten Galerkin–Verfahren vom Typ (4.15) approximieren. Das Standard–Galerkin–Verfahren zur Lösung von (8.4) lautet:
Finde $(u_h, p_h) \in W_h^{r,s}$, so dass für alle $(v_h, p_h) \in W_h^{r,s}$ gilt:

$$a(u_h, v_h) + b(v_h, p_h) = (f, v_h), \tag{8.8a}$$

$$b(u_h, q_h) = 0. \tag{8.8b}$$

Nach Konstruktion ist eine Lösung der schwachen Formulierung auch Lösung des obigen diskreten Problems. Dies ergibt analog zu (6.32) die Galerkin–Orthogonalität des Verfahrens.
Als Approximationsraum $W_h^{r,s}$ verwenden wir $[V_h^r]^d \times Q_h^s = [\mathcal{V}_h^r]^d \cap [H_0^1(\Omega)]^d \times \mathcal{Q}_h^s \cap L_*^2(\Omega)$, wobei $\mathcal{V}_h^r$ und $\mathcal{Q}_h^s$ geeignet gewählte Finite–Elemente–Räume sind. In zwei Dimensionen wählen wir stetige Finite–Elemente auf Dreiecken oder auf Rechtecken (in drei Dimensionen analog Finite–Elemente auf Tetraedern oder Hexaedern). Die Definitionen und Eigenschaften dieser Finiten–Elemente können in Kapitel 5 nachgeschlagen werden. Damit eine zulässige Gebietszerlegung möglich ist, setzen wir Ω als polygonal beziehungsweise polyhedral berandet voraus.

Der Einfachheit halber beschränken wir uns auf Finite–Elemente–Räume $\mathcal{V}_h^r$, $\mathcal{Q}_h^s$ auf derselben Gebietszerlegung $\mathcal{T}_h$. Sind die Ordnungen $r, s \in \mathbb{N}$ der Räume gleich, spricht man von equal order Elementen. Neben equal order Elementen werden auch Raumpaare mit verschiedener Ordnung vorkommen. Vorteilhaft ist dabei die Wahl $r = s + 1$, so dass die Geschwindigkeit mit Elementen höherer Ordnung approximiert wird. Den Grund hierfür sehen wir gleich. Elementpaare mit $r = s + 1$ bezeichnet man als Taylor–Hood–Elemente. Für eine Übersicht über weitere gebräuchliche Elementpaare siehe [GS00] Abschnitt 3.12.2.

Im letzten Abschnitt haben wir bewiesen, dass die schwache Formulierung eine eindeutige Lösung besitzt. Zum Beweis ist Satz 3.10 benutzt worden. Da der Approximationsraum $W_h^{r,s}$ nach Konstruktion ein Unterraum von $[H_0^1(\Omega)]^d \times L_*^2(\Omega)$ ist, sind für das Standard–Galerkin–Verfahren alle Voraussetzungen des Satzes bis auf die diskrete Babuška–Brezzi–Bedingung

$$\inf_{0 \neq q_h \in Q_h^s} \sup_{0 \neq v_h \in [V_h^r]^d} \frac{b(v_h, q_h)}{|v_h|_{H^1(\Omega)} \, \|q_h\|_{L^2(\Omega)}} \geq \beta > 0 \tag{8.9}$$

erfüllt. Während die diskrete Babuška–Brezzi–Bedingung für equal order Elemente im Allgemeinen verletzt ist, gilt sie gemäß [BF91] Abschnitt VI.6 für die Taylor–Hood–Elemente. [1] Im Fall von Taylor–Hood–Elementen kann also Satz 3.10 angewandt werden und liefert die eindeutige Lösbarkeit des Standard–Galerkin–Verfahrens. Im Fall von equal order Elementen erhält man hingegen üblicherweise keine eindeutige Lösung. Darüber hinaus resultieren für equal order Elemente die folgenden beiden Probleme aus dem Verletzt sein der diskreten Babuška–Brezzi–Bedingung: Weder sinnvolle Fehlerabschätzungen sind möglich, noch erhält man aus numerischen Rechnungen brauchbare Ergebnisse.

Da bei den Taylor–Hood–Elementen die diskrete Babuška–Brezzi–Bedingung erfüllt ist, erhält man zumindest im Spezialfall $b = 0$ brauchbare Fehlerabschätzungen (siehe zum Beispiel [GR86] Theorem 4.3 oder [BP79]). Im konvektions–dominanten Fall mit $\|b\|_{L^\infty(\Omega)} \gg \nu$ liefert die Analysis jedoch auch hier unzufriedenstellende Vorhersagen. In der Folge sind auch die numerischen Ergebnisse verbesserungswürdig. Um die Analysis sowie die numerischen Ergebnisse im konvektions–dominanten Fall zu verbessern, greift man auf Stabilisierungsverfahren zurück. Die Stabilisierungsverfahren, welche wir im Folgenden kennen lernen werden, sind ebenfalls auf equal order Elemente anwendbar. Dort sorgen sie zusätzlich dafür, dass die diskrete Babuška–Brezzi–Bedingung in einer modifizierten Form erfüllt ist. In der nächsten Bemerkung treffen wir einige Vorbereitungen für die Stabilisierungsverfahren.

Bemerkung 8.4.
(i) Addiert man Gleichung (8.8b) zu (8.8a), so gelangt man zu der folgenden äquivalenten Formulierung:
Finde $(u_h, p_h) \in W_h^{r,s}$, so dass für alle $(v_h, p_h) \in W_h^{r,s}$ gilt:

$$a_{\mathrm{SG}}(u_h, p_h; v_h, q_h) = (f, v_h) \quad \text{mit} \tag{8.10}$$

$$a_{\mathrm{SG}}(u_h, p_h; v_h, q_h) := \nu(\nabla u_h, \nabla v_h) + ((b \cdot \nabla)u_h, v_h) + (cu_h, v_h) - (\nabla \cdot u_h, q_h). \tag{8.11}$$

Diese Darstellung ist etwas kompakter als (8.8) und wird daher manchmal vorgezogen.

[1] Dies ist insofern plausibel, da der Druckraum bei Taylor–Hood–Elementen kleiner als der Geschwindigkeitsraum ist, wodurch das Infimum in (8.9) einen größeren Wert besitzt.

(ii) In (8.8) sind die Randbedingungen im Ansatzraum $[V_h^r]^d = [\mathcal{V}_h^r]^d \cap [H_0^1(\Omega)]^d$ berücksichtigt worden. Man sagt, die Randbedingungen sind stark vorgegeben. Daneben gibt es das Standard–Galerkin–Verfahren mit schwach vorgegebenen Randbedingungen. Hier werden die Randbedingungen nicht durch die Wahl des Ansatzraums erzwungen. Diese Verfahren erleichtern manchmal die Analysis und werden im Folgenden als Ausgangspunkt für eines der Stabilisierungsverfahren gewählt. Ohne weitere Stabilisierung hat man jedoch die gleichen Problemen wie bei dem Verfahren mit stark vorgegebenen Randbedingungen. Das Standard–Galerkin–Verfahren mit schwach vorgegebenen Randbedingungen lautet:
Finde $(u_h, p_h) \in [\mathcal{V}_h^r]^d \times Q_h^s$, so dass für alle $(v_h, q_h) \in [\mathcal{V}_h^r]^d \times Q_h^s$ gilt:

$$\begin{aligned} a_W(u_h, v_h) + b_W(v_h, p_h) &= (f, v_h), \\ b_W(u_h, q_h) &= 0 \qquad \text{oder kompakter} \end{aligned}$$

$$a_W(u_h, v_h) + b_W(v_h, p_h) + b_W(u_h, q_h) = (f, v_h), \tag{8.12}$$

wobei die Bilinearformen gegeben sind durch

$$\begin{aligned} a_W(u_h, v_h) &= a(u_h, v_h) - \nu(n\nabla u_h, v_h)_{\partial\Omega} - \nu(u_h, n\nabla v_h)_{\partial\Omega} \\ &\quad - ((b\cdot n)u_h, v_h)_{\partial\Omega_-} + \gamma(\nu/\widetilde{h}\, u_h, v_h)_{\partial\Omega} \\ &\quad + \gamma(\max\{|b|, \nu/\widetilde{h}\}u_h \cdot n, v_h \cdot n)_{\partial\Omega}, \\ b_W(v_h, p_h) &= b(v_h, p_h) + (v_h \cdot n, p_h)_{\partial\Omega}. \end{aligned} \tag{8.13}$$

Dabei ist n der äußere Normalenvektor des Gebietsrandes $\partial\Omega$, $\gamma > 0$ ein neuer Parameter und $\partial\Omega_-$ der Einströmrand, an welchem $b \cdot n < 0$ gilt. Ferner ist $\widetilde{h}$ auf Ω definiert durch $\widetilde{h}|_K = h_K$ mit dem Durchmesser h_K der Gitterzelle K. Im Vergleich zum Standard–Galerkin–Verfahren sind die letzten vier Terme in a_W künstlich ergänzt worden, die ersten drei davon in Analogie zur schwachen Formulierung der Konvektions–Diffusionsgleichung. Die Terme $\nu(n\nabla u_h, v_h)_{\partial\Omega}$ und $(p_h, v_h \cdot n)_{\partial\Omega}$ kommen hingegen aus der partiellen Integration, siehe Gleichung (8.3). Da die künstlich hinzugenommenen Terme für eine Lösung $(u, p) \in H_0^1(\Omega) \times L_*^2(\Omega)$ der schwachen Formulierung verschwinden, ist (u, p) ebenfalls Lösung von (8.12) und das Verfahren bleibt Galerkin–orthogonal, das heißt, für alle $(v_h, q_h) \in [\mathcal{V}_h^r]^d \times Q_h^s$ gilt:

$$a_W(u - u_h, v_h) + b_W(v_h, p - p_h) + b_W(u - u_h, q_h) = (f, v_h). \tag{8.14}$$

Für weitere Details siehe [Nit72] oder [FS95]

8.3 Die Residuale Stabilisierung

In diesem Abschnitt führen wir ein Stabilisierungsverfahren ein, mit welchem die Ergebnisse des Standard–Galerkin–Verfahrens verbessert werden können: die residuale Stabilisierung. Man spricht von residualer Stabilisierung, weil die Bilinearformen des Standard–Galerkin–Verfahrens um Terme der Form (Res, w) mit einer geeigneten Größe w und dem Residuum Res von (8.1a) oder (8.1b) erweitert werden. Dies hat unmittelbar den Vorteil, dass eine Lösung der schwachen Formulierung auch Lösung des stabilisierten Galerkin–Verfahrens ist, woraus die Galerkin–Orthogonalität der residualen Stabilisierung folgt.

Im Folgenden betrachten wir die klassische residuale Stabilisierung. Diese lautet:
Finde $(u_h, p_h) \in W_h^{r,s}$ mit

$$a_{\mathrm{RB}}(u_h, p_h; v_h, q_h) = f_{\mathrm{RB}}(v_h, q_h) \qquad \forall (v_h, q_h) \in W_h^r. \tag{8.15}$$

Dabei ist der Approximationsraum $W_h^{r,s}$ wie für das Standard–Galerkin–Verfahren mit stark vorgegebenen Randbedingungen gewählt worden, also

$$W_h^{r,s} = [V_h^r]^d \times Q_h^s = [\mathcal{V}_h^r]^d \cap [H_0^1(\Omega)]^d \times \mathcal{Q}_h^s \cap L_*^2(\Omega).$$

Des Weiteren ist a_{SG} wie in Bemerkung 8.4 Teil (i) und die Stabilisierungsterme sind gegeben durch:

$$\begin{aligned} a_{\mathrm{RB}}(u_h, p_h; v_h, q_h) = a_{\mathrm{SG}}(u_h, p_h; v_h, q_h) + \underbrace{\sum_{K\in\mathcal{T}_h} \gamma_K(\nabla\cdot u_h, \nabla\cdot v_h)_K}_{\text{Grad-Div}} \\ + \underbrace{\sum_{K\in\mathcal{T}_h} \delta_K(-\nu\Delta u_h + (b\cdot\nabla)u_h + cu_h + \nabla p_h,\ (b\cdot\nabla)v_h + \nabla q_h)_K}_{\text{SUPG+PSPG}}, \\ f_{\mathrm{RB}}(v_h, q_h) = f(v_h, p_h) \quad + \quad \overbrace{\sum_{K\in\mathcal{T}_h} \delta_K(f, (b\cdot\nabla)v_h + \nabla q_h)_K}^{} \end{aligned} \tag{8.16}$$

mit den zellenweise definierten Stabilisierungsparametern δ_K und γ_K. Drei Stabilisierungsterme sind hier zum Standard–Galerkin–Verfahren hinzugenommen worden:
(1.) der streamline upwind Petrov–Galerkin (SUPG) Term der Form

$$\sum_{K\in\mathcal{T}_h} \delta_K(\mathrm{Res}, (b\cdot\nabla)v_h).$$

Dieser Term ist nützlich bei konvektions–dominanten Problemen, insbesondere wenn die Lösung Grenzschichten aufweist. Die SUPG–Stabilisierung haben wir bereits für die Konvektions–Diffusionsgleichung in Abschnitt 6.4 diskutiert.

(2.) der pressure stabilization Petrov–Galerkin (PSPG) Term der Form

$$\sum_{K\in\mathcal{T}_h} \delta_K(\mathrm{Res}, \nabla q_h).$$

Dieser Term ist vor allem bei equal order Elementen nützlich, wo er eine modifizierte Variante der diskreten Babuška–Brezzi–Bedingung sichert. Die Idee der Druck–Stabilisierung geht zurück auf die Arbeiten [JS86] und [HFB86].

(3.) der Grad-Div Term der Form

$$\sum_{K\in\mathcal{T}_h} \gamma_K(\nabla\cdot u_h, \nabla\cdot v_h)_K.$$

Wird das Standard–Galerkin–Verfahren verwandt, so führt die Divergenz-Nebenbedingung auf die Gleichung

$$(\nabla\cdot u_h, q_h) = 0 \quad \forall q \in Q_h^s.$$

Die daraus berechnete Näherungslösung u_h wird in der Regel nicht divergenzfrei sein. Insbesondere wenn der Druckraum, gegen den getestet wird, klein ist, sind größere Abweichungen davon zu erwarten. Um die Abweichung $|\nabla\cdot u_h|$ zu verringern, wird die Divergenz-Nebenbedingung über den Grad-Div Term gegen Basisfunktionen aus dem Geschwindigkeitsraum getestet. Dies kann insbesondere bei Taylor–Hood–Elementen effektiv sein, bei welchen der Geschwindigkeitsraum größer als der Druckraum ist. Numerische Rechnungen

haben zudem gezeigt, dass die Grad-Div Stabilisierung im konvektions–dominanten Fall die Robustheit des Verfahrens verbessert [FF92], [TL91].

Als nächstes soll ein hinreichendes Kriterium für die eindeutige Lösbarkeit des stabilisierten Galerkin–Verfahrens formuliert werden. Dazu prüfen wir, wann a_{RB} strikt koerzitiv ist und wenden den Satz von Lax–Milgram an (Satz 3.3). Die Gebietszerlegung $\mathcal{T}_h$ setzen wir als quasiuniform voraus, damit die Hilfsmittel aus Abschnitt 5.4 benutzt werden können. Zudem führen wir eine gitterabhängige Norm ein:

$$\begin{aligned} |\!|\!|(v_h,q_h)|\!|\!|_{\mathrm{RB}}^2 &= |[v_h]|^2 + \alpha\|q_h\|_{L^2(\Omega)}^2 \quad \text{mit} \\ |[(v_h,q_h)]|^2 &= \|c^{1/2}v_h\|_{L^2(\Omega)}^2 + \nu\|\nabla v_h\|_{L^2(\Omega)}^2 \\ &\quad + \sum_{K\in\mathcal{T}_h}\gamma_K\|\nabla\cdot v_h\|_{L^2(K)}^2 + \sum_{K\in\mathcal{T}_h}\delta_K\|(b\cdot\nabla)v_h + q_h\|_{L^2(K)}^2. \end{aligned} \tag{8.17}$$

Die Größe $\alpha > 0$ ist von den Daten des Problems abhängig und wird im Laufe des Abschnitts spezifiziert. Nachfolgende Ergebnisse gelten sowohl für Taylor–Hood als auch für equal order Elemente. Mit Hilfe obiger Norm erhalten wir die folgende Aussage:

Lemma 8.5. Die Stabilisierungsparameter δ_K und γ_K seien zellenweise so gewählt, dass gilt

$$0 < \delta_K \le \frac{1}{2}\min\left\{\frac{h_K^2}{\mu_{\mathrm{inv}}^2 r^4\nu}, \frac{1}{\|c\|_{L^\infty(K)}}\right\}, \quad 0 \le \gamma_K. \tag{8.18}$$

Dann ist a_{RB} strikt koerzitiv bezüglich der Norm $|[\cdot]|$:

$$a_{\mathrm{RB}}(v_h,q_h;v_h,q_h) \ge \frac{1}{2}\,|[(v_h,q_h)]|^2 \quad \forall (v_h,q_h)\in W_h^{r,s}.$$

Beweis. Für den Beweis führen wir die folgenden Abkürzungen ein

$$A^2 := \|c^{1/2}v_h\|_{L^2(\Omega)}^2 + \nu\|\nabla v_h\|_{L^2(\Omega)}^2, \quad B^2 := \|q_h\|_{L^2(\Omega)}^2 \tag{8.19}$$

und

$$\begin{aligned} X^2 &:= \sum_{K\in\mathcal{T}_h}\delta_K\|(b\cdot\nabla)v_h + q_h\|_{L^2(K)}^2, \\ Y^2 &:= \sum_{K\in\mathcal{T}_h}\delta_K\|-\nu\Delta v_h + cv_h\|_{L^2(K)}^2, \\ Z^2 &:= \sum_{K\in\mathcal{T}_h}\gamma_K\|\nabla\cdot v_h\|_{L^2(K)}^2. \end{aligned}$$

Damit können wir schreiben $|[(v_h,q_h)]| = A^2 + X^2 + Z^2$. Wegen der Definition von a_{RB}, partieller Integration ähnlich zu (6.14) im konvektiven Term und der Relation $\nabla\cdot b = 0$ ergibt sich:

$$a_{\mathrm{RB}}(v_h,q_h;v_h,q_h) \ge A^2 + X^2 + Z^2 - XY.$$

Unter Verwendung der inversen Ungleichung (5.40) und der Annahme (8.18) an die Stabilisierungsparameter rechnet man die Relation $Y \le A$ nach. Anwenden der Young–Ungleichung liefert schließlich die Behauptung:

$$a_{\mathrm{RB}}(v_h,q_h;v_h,q_h) \ge \frac{1}{2}\left(A^2 + X^2 + Z^2\right). \quad \square$$

Sind die Stabilisierungsparameter passend gewählt, liefert der Satz von Lax–Milgram die eindeutige Existenz einer Lösung (die restlichen Voraussetzungen des Satzes prüft man leicht). Dies kann immer durch ausreichend kleine Wahl von δ_K erreicht werden, auch dann wenn die Konstante μ_{inv} aus der inversen Abschätzung nicht bekannt sein sollte. Unser nächstes Ziel ist die Angabe einer a–priori Fehlerabschätzung. Dazu geben wir im folgenden Satz eine inf–sup Abschätzung an, welche im Gegensatz zu Lemma 8.5 ebenfalls Kontrolle über die L^2-Norm des Drucks gibt. Bei den nachfolgenden Beweisen folgen wir [LR06]. In [LR06] wird von etwas allgemeineren Voraussetzungen ausgegangen. Die Viskosität ν wird dort als ortsabhängige Größe behandelt. Dieser Fall tritt in der Praxis zuweilen auf, beispielsweise dann, wenn ein Fluid mit temperaturabhängiger Viskosität untersucht wird und die Temperatur im Lösungsgebiet variiert. Bei unseren Modellrechnungen wird die Viskosität aber eine konstante Größe sein. Folglich beschränken wir uns auch bei der Analysis auf konstantes ν.

Satz 8.6.
Die Stabilisierungsparameter δ_K und γ_K seien zellenweise so gewählt, dass gilt

$$0 < \widetilde{C}\frac{h_K^2}{r^2} \leq \delta_K \leq \frac{1}{2}\min\left\{\frac{h_K^2}{\mu_{\text{inv}}^2 r^4 \nu}, \frac{1}{\|c\|_{L^\infty(K)}}\right\}, \quad 0 \leq \delta_K \|b\|_{L^\infty(K)}^2 \leq \gamma_K \tag{8.20}$$

mit einer Konstanten $\widetilde{C} > 0$. Außerdem sei der Faktor α in der Definition der RB-Norm gegeben durch

$$\alpha^{-1/2} \sim \sqrt{\gamma} + \frac{1}{\widetilde{C}} + C_F\|c\|_{L^\infty(\Omega)} + \frac{C_F\|b\|_{L^\infty(\Omega)}}{\sqrt{\nu + c_{\min}C_F^2}} + \max_{K\in\mathcal{T}_h}\frac{h_K\|b\|_{L^\infty(\Omega)}}{\nu}, \tag{8.21}$$

wobei C_F die Friedrichskonstante aus Lemma 2.19 ist und die Abkürzung $\gamma = \max_{K\in\mathcal{T}_h}\gamma_k$ verwandt wurde. Dann erfüllt die Bilinearform a_{RB} die diskrete Babuška–Brezzi–Bedingung:

$$\inf_{0\neq(u_h,p_h)\in W_h^{r,s}} \sup_{0\neq(v_h,q_h)\in W_h^{r,s}} \frac{a_{\text{RB}}(u_h,p_h;v_h,q_h)}{|\!|\!|(u_h,p_h)|\!|\!|_{\text{RB}}\,|\!|\!|(v_h,q_h)|\!|\!|_{\text{RB}}} \geq \beta \tag{8.22}$$

mit einer positiven Konstanten β, die unabhängig vom Polynomgrad r, der Gitterweite h und der Viskosität ν ist.

Beweis. Sei ein beliebiges $(u_h,p_h) \in W_h^{r,s}$ vorgegeben. Gemäß Verführths Trick - siehe [FS91] oder [TV96] - existiert ein $z \in V_h^r$ mit

$$\nabla\cdot z = -p_h \quad \text{und} \quad \|z\|_{H^1(\Omega)} \leq C\|p_h\|_{L^2(\Omega)}.$$

Zu z definiert man durch $z_h = i_h^r z$ die Scott–Zhang–Interpolation aus Lemma 5.13. Nun rechnet man unter Verwendung der Fehlerabschätzung für die Scott-Zhang–Interpolation nach, dass bei der Wahl von α, δ_K und γ_K wie im Satz die folgende Ungleichung gilt:

$$a_{\text{RB}}(u_h,p_h;z_h,0) \geq \beta|\!|\!|(u_h,p_h)|\!|\!|_{\text{RB}}|\!|\!|(z_h,0)|\!|\!|_{\text{RB}}.$$

Da $(u_h,p_h) \in W_h^{r,s}$ beliebig gewählt ist, folgt daraus die Behauptung. Die Rechnung, welche auf obige Ungleichung führt, benutzt die inverse Ungleichung (5.40) und weitere Standard–Ungleichungen. Für mehr Details bezüglich dieser Abschätzung siehe [LR06] Lemma 2.2. □

Bemerkung 8.7.
(i) Um aus obigem Satz Vorhersagen für numerische Rechnungen abzuleiten, muss sichergestellt werden, dass die Stabilisierungsparameter δ_K und γ_K zellenweise in dem durch (8.20)

vorgegebenen Intervall liegen. Die Konstante $\widetilde{C}$ und die Konstante μ_{inv} aus der inversen Abschätzung sind aber nicht immer bekannt. Demnach ist auch nicht klar, ob überhaupt eine geeignete Wahl von δ_K möglich ist oder ob die untere Schranke nicht größer als die obere in (8.20) ist. Wie wir sehen werden, kann jedoch oft durch numerische Tests eine geeignete Wahl für die Stabilisierungsparaemter getroffen werden.

(ii) Im Fall der Taylor–Hood–Elemente gilt der obige Satz auch, wenn auf die untere Schranke für δ_K verzichtet wird. Dann ist jedoch die Konstante β in der Stabilitätsabschätzung durch die Konstante aus (8.9) zu ersetzen, welche abhängig von der Elementordnung ist [BBJL07].

Zum Beweis der a–priori Fehlerabschätzung benötigen wir das folgende technische Lemma, welches mit Hilfe der üblichen Standard–Abschätzungen nachgerechnet werden kann:

Lemma 8.8.
Seien die Voraussetzungen von Satz 8.6 gegeben. Dann gilt für alle $(u,p) \in W$ mit $u \in H^2(\mathcal{T}_h)$ und $(v,q) \in W_h^{r,s}$

$$a_{\text{RB}}(u,p;v_h,q_h) \leq C\Theta(u,p)|\!|\!|(v_h,q_h)|\!|\!|_{\text{RB}}, \qquad \text{wobei}$$

$$\Theta(u,p) := |[(u,p)]| + \left(\sum_{K\in\mathcal{T}_h} \delta_K^{-1}\|u\|^2_{L^2(K)}\right)^{1/2} + \left(\sum_{K\in\mathcal{T}_h} \delta_K\|-\nu\Delta u + cu\|^2_{L^2(K)}\right)^{1/2}$$
$$+ \left(\sum_{K\in\mathcal{T}_h} 3\left(\nu+\gamma_K+\frac{c_{\min}h_K^2}{\mu_{\min}^2 r^4}\right)^{-1}\|p\|^2_{L^2(K)}\right)^{1/2}. \tag{8.23}$$

Beweis. Siehe [LR06] Lemma 2.3. □

Damit sind wir in der Lage die gewünschte Fehlerabschätzung anzugeben:

Satz 8.9.
Sei $(u,p) \in [H^k(\Omega)]^d \times H^l(\Omega)$ mit $k,l > \frac{d}{2}$ die Lösung der schwachen Formulierung und $(u_h,p_h) \in V_h^r \times Q_h^s$ die aus (8.15) bestimmte Näherungslösung. Dann gilt für eine Wahl der Stabilisierungsparameter wie in Satz 8.6 die folgende Fehlerabschätzung:

$$|\!|\!|(u-u_h,p-p_h)|\!|\!|^2_{\text{RB}} \leq C\sum_{K\in\mathcal{T}_h}\left(M_K^u\frac{h_K^{2(r_u-1)}}{r^{2(k-1)}}\|u\|^2_{H^{r_u}(K)} + M_K^p\frac{h_K^{2(r_p-1)}}{r^{2(l-1)}}\|u\|^2_{H^{r_p}(K)}\right) \tag{8.24}$$

mit $r_u = \min\{k, r+1\}$, $r_p = \min\{l, s+1\}$ und den Größen

$$M_K^u = \frac{h_K^2}{r^2\delta_K} + \delta_K\left(\frac{\|c\|^2_{L^\infty(K)}h_K^2}{r^2} + \|b\|^2_{L^\infty(K)} + \frac{r^2\nu^2}{h_K^2}\right) + \gamma_K + \nu + \frac{\|c\|_{L^\infty(K)}h_K^2}{r^2},$$
$$M_K^p = \delta_K + \frac{h_K^2}{s^2\max\{\nu,\gamma_K\}}. \tag{8.25}$$

Beweis. Für den Beweis führen wir zunächst einige Abkürzungen ein. Den Verfahrensfehler bezeichnen wir mit $E_h := (u-u_h, p-p_h)$, den Diskretisierungsfehler mit $e_h := (u_I - u_h, p_I - p_h)$ und den Interpolationsfehler mit $e_I := (u-u_I, p-p_I)$. Dabei sind u_I und p_I geeignete Interpolationen von u und p, also zum Beispiel die Scott–Zhang–Interpolation für u und die Lagrange–Interpolation für p. Mit der Dreiecksungleichung spalten wir den Verfahrensfehler auf in den Diskretisierungs- und den Interpolationsfehler:

$$|\!|\!|E_h|\!|\!|_{\text{RB}} \leq |\!|\!|e_h|\!|\!|_{\text{RB}} + |\!|\!|e_I|\!|\!|_{\text{RB}}.$$

Den Interpolationsfehler schätzen wir ab durch:

$$\begin{aligned}\|\!|e_I|\!\|_{\mathrm{RB}} &= \left(|[e_I]|^2 + \alpha\|p - p_I\|^2_{L^2(\Omega)}\right)^{1/2} \\ &\leq |[e_I]| + \sqrt{\alpha}\|p - p_I\|_{L^2(\Omega)} \\ &\leq C\Theta(e_I).\end{aligned}$$

Die erste Ungleichung folgt aus der Relation $(a^2 + b^2)^{1/2} \leq a + b$ für $a, b \geq 0$ und die zweite aus der Definition von α in (8.21) sowie von Θ in (8.23). Den Diskretisierungsfehler beschränken wir nach oben durch:

$$\begin{aligned}\|\!|e_h|\!\|_{\mathrm{RB}} &\leq C\frac{a_{\mathrm{RB}}(e_h; v_h, q_h)}{\|\!|(v_h, q_h)|\!\|_{\mathrm{RB}}} \\ &= C\frac{a_{\mathrm{RB}}(e_I; v_h, q_h)}{\|\!|(v_h, q_h)|\!\|_{\mathrm{RB}}} \\ &\leq C\Theta(e_I).\end{aligned}$$

In der ersten Ungleichung ist (v_h, q_h) geeignet aus $W_h^{r,s}$ gewählt worden, so dass die Abschätzung gilt. Eine solche Wahl ist möglich, da gemäß Satz 8.6 die Bilinearform a_{RB} der diskreten Babuška–Brezzi–Bedingung genügt. Die Gleichheit ergibt sich aus der Galerkin–Orthogonalität des Verfahrens und die letzte Ungleichung aus Lemma 8.8. Als Zwischenergebnis erhalten wir damit:

$$\|\!|E_h|\!\|_{\mathrm{RB}} \leq C\Theta(e_I)$$

mit einer Konstanten $C > 0$. Daraus ergibt sich die gewünschte Abschätzung, indem man die Interpolationsfehler–Abschätzung für die Lagrange- und die Scott–Zhang–Interpolation aus Lemma 5.8 und 5.13 benutzt. □

Durch eine geschickte Wahl der Stabilisierungsparameter δ_K und γ_K lässt sich die a–priori Abschätzung des obigen Satzes verfeinern. Wir betrachten zunächst die equal order Elemente:

Satz 8.10.
Es gelten die Voraussetzungen des letzten Satzes mit $s = r$ und $l = k$. Durch die Wahl

$$\begin{aligned}\delta_K &\sim \frac{h_K^2}{rh_K\|b\|_{L^\infty(K)} + \|c\|_{L^\infty(K)}h_K^2 + r^4\nu}, \\ \gamma_K &\sim \frac{h_K\|b\|_{L^\infty(K)}}{r} + \frac{\|c\|_{L^\infty(K)}h_K^2}{r^2} + r^2\nu\end{aligned} \tag{8.26}$$

erhält man die Fehlerabschätzung:

$$\|\!|(u - u_h, p - p_h)|\!\|^2_{\mathrm{RB}} \leq C\sum_{K\in\mathcal{T}_h} M_K\frac{h_K^{2(\widetilde{k}-1)}}{r^{2(k-1)}}\left(\|u\|^2_{H^k(K)} + \|p\|^2_{H^k(K)}\right). \tag{8.27}$$

Dabei ist $\widetilde{k} = \min\{k, r+1\}$ und

$$M_K = \nu r^2 + \frac{h_K\|b\|_{L^\infty(K)}}{r} + \frac{\|c\|_{L^\infty(K)}h_K^2}{r^2}. \tag{8.28}$$

Beweis. Einsetzen der Stabilisierungsparameter in (8.24). □

Bemerkung 8.11.
Aus obigem Satz folgen bei hinreichender Regularität der Lösung im konvektions–dominanten Fall mit $\nu \leq \|b\|_{L^\infty(\Omega)} h$ die Fehlerabschätzungen:

$$\begin{aligned}
\|u - u_h\|_{L^2(\Omega)} &\leq C \frac{h^{r+1/2}}{r^{k-2}} \left(\|u\|_{H^{r+1}(\Omega)} + \|p\|_{H^{r+1}(\Omega)} \right), \\
\|p - p_h\|_{L^2(\Omega)} &\leq C \frac{h^{r+1}}{r^{k-2}} \nu^{-1/2} \left(\|u\|_{H^{r+1}(\Omega)} + \|p\|_{H^{r+1}(\Omega)} \right), \\
\|u - u_h\|_{H^1(\Omega)} &\leq C \frac{h^{r}}{r^{k-2}} \left(\|u\|_{H^{r+1}(\Omega)} + \|p\|_{H^{r+1}(\Omega)} \right).
\end{aligned} \tag{8.29}$$

Hinreichend regulär ist dabei $(u, p) \in [H^k(\Omega)]^d \times H^k(\Omega)$ mit $k \geq r + 1$. Unter denselben Regularitätsbedingungen erhält man im diffusions–dominanten Fall mit $\nu \geq \|b\|_{L^\infty(\Omega)} h$:

$$\begin{aligned}
\|u - u_h\|_{L^2(\Omega)} &\leq C \frac{h^{r}}{r^{k-2}} \nu^{1/2} \left(\|u\|_{H^{r}(\Omega)} + \|p\|_{H^{r+1}(\Omega)} \right), \\
\|p - p_h\|_{L^2(\Omega)} &\leq C \frac{h^{r}}{r^{k-2}} \nu \left(\|u\|_{H^{r+1}(\Omega)} + \|p\|_{H^{r+1}(\Omega)} \right), \\
\|u - u_h\|_{H^1(\Omega)} &\leq C \frac{h^{r}}{r^{k-2}} \left(\|u\|_{H^{r+1}(\Omega)} + \|p\|_{H^{r+1}(\Omega)} \right).
\end{aligned} \tag{8.30}$$

(ii) Gemäß der Wahl der Stabilisierungsparameter in (8.26) nimmt der Einfluss der Grad-Div Stabilisierung mit wachsender Elementordnung r zu, während der Einfluss der SUPG/PSPG Stabilisierung abnimmt.

Im Fall von Taylor–Hood–Elemente mit $s = r - 1$ ist eine andere Wahl von δ_K und γ_K günstig:

Satz 8.12.
Es gelten die Voraussetzungen von Satz 8.9 mit $s = r - 1$ und $l = k - 1$. Durch die Wahl

$$\delta_K \sim \frac{h_K^2}{r^2(\nu + 1)} \quad \text{und} \quad \gamma_K \sim \nu \tag{8.31}$$

erhält man die Fehlerabschätzung:

$$|\!|\!| (u - u_h, p - p_h) |\!|\!|_{\mathrm{RB}}^2 \leq C \sum_{K \in \mathcal{T}_h} \frac{h_K^{2\widetilde{k}}}{r^{2(k-1)}} \left((\nu + 1) \|u\|_{H^k(K)}^2 + \frac{1}{\nu + 1} \|p\|_{H^{k-1}(K)}^2 \right) \tag{8.32}$$

mit $\widetilde{k} = \min\{k - 1, r\}$.

Gemäß der Wahl der Stabilisierungsparameter in (8.31) verringert sich der Einfluss der SUPG/PSPG Stabilisierung, wenn die Elementordnung größer wird. Abschließend seien einige Modifikationen der klassischen residualen Stabilisierung erwähnt.

Bemerkung 8.13.
(i) Wir haben die SUPG und PSPG Stabilisierungsterme der Einfachheit halber mit demselben Parameter δ_K versehen. Stattdessen können beide Terme auch mit verschiedenen Parameter angesetzt werden. Insbesondere bei den Taylor–Hood–Elementen ist dies oft vorteilhaft. Dort hat sich gezeigt, dass die PSPG Stabilisierung zumeist überflüssig ist, siehe [MLR09]. Setzt man den PSPG Parameter auf Null, so verringert sich der numerische Aufwand, während die Ergebnisse etwa gleich gut bleiben.

(ii) Weitere residuale Stabilisierungsverfahren sind die Galerkin least-squares (GLS) Methode [FF92] und die algebraic subgrid-scale (ASGS) Methode [Cod00].

8.4 Die Kanten–Stabilisierung

Neben der klassischen residualen Stabilisierung werden wir die Kanten–Stabilisierung betrachten, auch continuous interior penalty (CIP) Methode genannt. Die Idee der CIP–Methode besteht darin, die Sprünge in den Ableitungen von Druck und Geschwindigkeit über den Rand der Gitterzellen zu bestrafen. Gemäß [BFH06] bietet die CIP–Methode die folgenden Vorteile gegenüber der residualen Stabilisierung:

- Es werden keine neuen unsymmetrischen Terme eingeführt.
- Eine unphysikalische Kopplung zwischen Druck und Geschwindigkeit wird vermieden.
- Die Methode besitzt größere Kompatibilität mit anderen numerischen Verfahren: Die mass lumping Technik bleibt anwendbar und bei zeitabhängigen Problemen hat man eine größere Auswahl an Zeitschrittverfahren.

Demgegenüber steht der Nachteil, dass die Systemmatrix der CIP–Methode eine größere Anzahl von Null verschiedener Einträge besitzt. Die CIP–Methode lautet:

Finde $(u_h, p_h) \in W_h^r$, so dass für alle $(v_h, q_h) \in W_h^r$ gilt:

$$a_{\mathrm{CIP}}(u_h, p_h; v_h, q_h) = (f, v_h) \tag{8.33}$$

mit der Bilinearform

$$a_{\mathrm{CIP}}(u_h, p_h; v_h, q_h) = a_W(u_h, v_h) + b_W(v_h, p_h) - b_W(u_h, q_h) + j_u(u_h, v_h) + j_p(p_h, q_h)$$

und dem Approximationsraum

$$W_h^r = [V_h^r]^d \times Q_h^r = [\mathcal{V}_h^r]^d \times \mathcal{Q}_h^r \cap L_*^2(\Omega).$$

Die Standard–Galerkin–Formulierung (8.12) mit schwachen Randbedingungen ist hier um folgende Stabilisierungsterme erweitert worden:

$$\begin{aligned} j_u(u_h, v_h) = &\sum_{K \in \mathcal{T}_h} \tau \zeta(\mathrm{Re}_K) h_K^2 \int_{\partial K} \|b \cdot n\|_{L^\infty(\partial K)} [\nabla u_h \cdot n] \cdot [\nabla v_h \cdot n] ds \\ &+ \sum_{K \in \mathcal{T}_h} \tau \zeta(\mathrm{Re}_K) h_K^2 \|b\|_{L^\infty(K)} \int_{\partial K} [\nabla \cdot u_h][\nabla \cdot v_h] ds, \end{aligned} \tag{8.34}$$

$$j_p(p_h, q_h) = \sum_{K \in \mathcal{T}_h} \tau \min\left\{ \frac{1}{\|b\|_{L^\infty(K)}}, \frac{h_K}{\nu} \right\} h_K^2 \int_{\partial K} [\nabla p_h] \cdot [\nabla q_h] ds \tag{8.35}$$

mit dem Stabilisierungs–Parameter $\tau > 0$, der Jacobimatrix ∇u von u, dem äußeren Normalenvektor n des Randes ∂K von Zelle K, ihrem Durchmesser h_K und den Abkürzungen:

$$\zeta(x) = \min\{1, x\}, \quad \mathrm{Re}_K = \frac{\|b\|_{L^\infty(K)} h_K}{\nu}. \tag{8.36}$$

Re_K ist die Reynoldszahl in Zelle K. Die Integrale erstrecken sich nur über Kanten (beziehungsweise Seitenflächen), welche *nicht* auf dem Gebietsrand liegen. Ferner kennzeichnet der Ausdruck $[w]$ den Sprung der Größe w über den Zellenrand ∂K in Richtung seines äußeren Normalenvektors n , das heißt:

$$[w]_E(x) = \lim_{s \to 0} (w(x + sn) - w(x - sn)) \quad \text{für } x \in E. \tag{8.37}$$

Für stückweise polynomiale Funktionen ist der Sprung im Punkt $x \in \partial K$ einfach durch

$$[w](x) = w_1(x) - w_2(x) \tag{8.38}$$

gegeben, wobei w_i jeweils die Einschränkung von w auf eine der beiden Zellen ist, zu welchen der Punkt x gehört. Ähnlich wie bei der residualen Stabilisierung erfüllen die Stabilisierungsterme folgende Aufgaben: Der erste Term in (8.34) stabilisiert das Verfahren im konvektions–dominanten Fall, der zweite stärkt die Divergenz–Nebenbedingung und der Stabilisierungsterm in (8.35) sichert eine modifizierte diskrete Babuška–Brezzi–Bedingung. Im Folgenden untersuchen wir die eindeutige Lösbarkeit der CIP–Methode und geben eine a–priori Fehlerabschätzung an. Dabei orientieren wir uns an [BFH06]. Im Vergleich zur Analysis der residualen Stabilisierung wird in [BFH06] von leicht modifizierten Voraussetzungen ausgegangen. Der Einfachheit halber werden nur equal order Elemente behandelt und die Größe c in der Oseen–Gleichung wird als Konstante echt größer Null angenommen. Die Annahme an die Gebietszerlegung $\mathcal{T}_h$ ist hingegen allgemeiner. Während wir bei der residualen Stabilisierung eine quasiuniforme Gebietszerlegung angenommen haben, wird diese Eigenschaft in [BFH06] nur noch lokal gefordert. Lokale Quasiuniformität kann nützlich sein, wenn man das Gitter an kritischen Stellen im Lösungsgebiet, wie beispielsweise an Singularitäten, lokal verfeinern möchte. Bei späteren Simulationen werden wir dies jedoch nicht tun, so dass wir für die Analysis eine "global" quasiuniforme Gebietszerlegung voraussetzen können.

Weitere Vereinfachungen in der Analysis von [BFH06] sind:
(1.) Die Randbedingungen werden schwach vorgegeben. Dies hat den Vorteil, dass die globale L^2–Projektion aus Lemma 5.10 verwandt werden kann. Die globale L^2–Projektion erhält im Allgemeinen keine homogenen Randbedingungen und ist somit für eine Analysis mit stark vorgegeben Randbedingungen ungeeignet.

(2.) Alle Stabilisierungsterme sind mit demselben Stabilisierungsparameter τ angesetzt worden. Darüber hinaus wird der gleiche Wert dem Parameter γ zugewiesen, welcher in der Bilinearform a_W für schwach vorgegebene Randbedingungen auftritt. Stattdessen könnten die Parameter auch verschieden voneinander gewählt werden. Darauf werden wir später nochmals eingehen, wenn die numerischen Resultate diskutiert werden. Es soll insbesondere geprüft werden, inwieweit die Ergebnisse durch Verwendung unterschiedlicher Parameter verbessert werden können.

(3.) Auf eine Angabe der Elementordnung in der a–priori Fehlerabschätzung wird verzichtet.

Im ersten Schritt zum Beweis der eindeutigen Lösbarkeit der CIP–Methode zeigen wir die Koerzitivität der Bilinearform $a_W + j_u$ gegenüber einer geeigneten Norm. Eine solche ist

$$\begin{aligned} |\!|\!| v_h |\!|\!|_{\mathrm{CIP}}^2 = c\|v_h\|_{L^2(\Omega)}^2 + \nu\|\nabla v_h\|_{L^2(\Omega)}^2 + j_u(v_h, v_h) + \| \left|b \cdot n\right|^{1/2} v_h\|_{L^2(\partial\Omega)}^2 \\ + \|\tau^{1/2}(\nu/\widetilde{h})^{1/2} v_h\|_{L^2(\partial\Omega)}^2 + \|\tau^{1/2} \max\{|b|, \nu/\widetilde{h}\}^{1/2} n \cdot v_h\|_{L^2(\partial\Omega)}^2, \end{aligned} \tag{8.39}$$

wobei $\widetilde{h} : \bar{\Omega} \to \mathbb{R}^+$ definiert ist durch $\widetilde{h}|_K := h_K$.

Lemma 8.14.
Es gibt eine Konstante $C > 0$, welche nur vom Gebiet Ω und dem Stabilisierungsparameter τ abhängt, so dass für alle $v_h \in [V_h^r]^d$ gilt:

$$a_W(v_h, v_h) + j_u(v_h, v_h) \geq C |\!|\!| v_h |\!|\!|_{\mathrm{CIP}}^2. \tag{8.40}$$

Beweis. Aufgrund der Definition der Bilinearform a_W gilt:

$$\begin{aligned} a_W(v_h, v_h) + j_u(v_h, v_h) \geq c\|v_h\|^2_{L^2(\Omega)} + \nu\|\nabla v_h\|^2_{L^2(\Omega)} + j_u(v_h, v_h) + \| \, |b \cdot n|^{1/2} \, v_h\|^2_{L^2(\partial\Omega)} \\ + \|\tau^{1/2}(\nu/\widetilde{h})^{1/2} v_h\|^2_{L^2(\partial\Omega)} + \|\tau^{1/2} \max\{|b|, \nu/\widetilde{h}\}^{1/2} n \cdot v_h\|^2_{L^2(\partial\Omega)} \\ - 4\nu(n\nabla v_h, v_h)_{\partial\Omega}, \end{aligned} \tag{8.41}$$

wobei der Term $(b \cdot \nabla v_h, v_h) - ((b \cdot n)v_h, v_h)_{\partial\Omega_-}$ analog zu (6.28) unter Verwendung der Annahme $\nabla \cdot b = 0$ abgeschätzt worden ist. Mit der Cauchy–Schwarz– und der Spurungleichung (5.41) beschränkt man nun den letzten Term betragsmäßig nach oben. Subtraktion dieser oberen Schranke von den restlichen Termen liefert die Behauptung. □

Im zweiten Schritt zum Beweis der eindeutigen Lösbarkeit der CIP–Methode zeigen wir, dass die Bilinearform b_W eine modifizierte diskrete Babuška–Brezzi–Bedingung erfüllt. Als Vorbereitung benötigen wir das folgende Resultat zur Stabilität der L^2–Projektion:

Lemma 8.15. (**Stabilität der L^2–Projektion**)
Seien die Konstante η sowie die Größenregularitätskonstante ξ der Gebietszerlegung aus Definition 5.7 ausreichend klein und sei $\phi \in V_h^1$ mit $\phi > 0$ auf Ω und

$$|\nabla\phi(x)| \leq \eta h_K^{-1}\phi(x) \qquad \forall x \in K, \quad \forall K \in \mathcal{T}_h. \tag{8.42}$$

Dann besitzt die globale L^2–Projektion $\pi_h : L^2(\Omega) \to V_h^r$ die folgenden Stabilitätseigenschaften:

$$\|\phi\pi_h u\|_{L^2(\Omega)} \leq C\|\phi u\|_{L^2(\Omega)} \quad \forall u \in L^2(\Omega), \tag{8.43}$$

$$\|\phi\nabla\pi_h u\|_{L^2(\Omega)} \leq C\|\phi\nabla u\|_{L^2(\Omega)} \quad \forall u \in H^1(\Omega) \tag{8.44}$$

mit einer Konstanten $C > 0$.

Beweis. Siehe [Bom04] Lemma 2.2 und Lemma 2.4.

Mit obigem Lemma können wir zeigen:

Lemma 8.16.
Unter den Voraussetzungen dieses Abschnitts erfüllt die Bilinearform b_W die folgende diskrete Babuška–Brezzi–Bedingung:

$$\inf_{0=q_h \in Q^r_{h,0}} \sup_{v_h \in [V_h^r]^d} \frac{|b_W(v_h, q_h)|}{\|q_h\|_{L^2(\Omega)}\|v_h\|_{H^1(\Omega)}} \geq C. \tag{8.45}$$

Dabei ist $C > 0$ eine Konstante und der Raum $Q^r_{h,0}$ ist definiert durch

$$Q^r_{h,0} = \{q_h \in Q_h^r : \; j_p(q_h, q_h) = 0\}. \tag{8.46}$$

Beweis.
Sei $q_h \in Q^r_{h,0}$ vorgegeben. Nach Lemma 8.1 gibt es zu q_h ein $v_q \in [H_0^1]^d$ mit

$$\nabla \cdot v_q = q_h \quad \text{und} \quad \|v_q\|_{H^1(\Omega)} \leq C\|q_h\|_{L^2(\Omega)}. \tag{8.47}$$

Im Folgenden zeigen wir, dass für die L^2–Projektion $\pi_h v_q$ von v_q gilt:

$$|b_W(v_q, \pi_h q_h)| \leq C\|q_h\|_{L^2(\Omega)}\|\pi_h v_q\|_{H^1(\Omega)}.$$

Da q_h beliebig vorgegeben ist, folgt daraus die Behauptung.
Mit Hilfe von partieller Integration ergibt sich:

$$\begin{aligned}\|q_h\|^2_{L^2(\Omega)} &= (\nabla \cdot v_q, q_h) \\ &= (\nabla \cdot v_q - \nabla \cdot \pi_h v_q, q_h) + (\nabla \cdot \pi_h v_q, q_h) \\ &= (v_q - \pi_h v_q, \nabla q_h) - (n \cdot \pi_h v_q, q_h)_{\partial\Omega} + (\nabla \cdot \pi_h v_q, q_h) \\ &= (v_q - \pi_h v_q, \nabla q_h) - b_W(\pi_h v_q, q_h). \end{aligned} \tag{8.48}$$

Da $q_h \in Q^r_{h,0}$ eine stetige erste Ableitung in Ω besitzt, gehört ∇q zu $[V^r_h]^d$ und der erste Term auf der rechten Seite von (8.48) verschwindet wegen der L^2-Orthogonalität von π_h bezüglich $[V^r_h]^d$. Somit ergibt sich

$$|b_W(\pi_h v_q, q_h)| = \|q_h\|^2_{L^2(\Omega)}.$$

Die rechte Seite schätzen wir nun nach unten ab durch:

$$\begin{aligned}\|q_h\|^2_{L^2(\Omega)} &\geq C\|q_h\|_{L^2(\Omega)}\|v_q\|_{H^1(\Omega)} \\ &\geq C\|q_h\|_{L^2(\Omega)}\|\pi_h v_q\|_{H^1(\Omega)}.\end{aligned}$$

Die erste Ungleichung folgt aus (8.47) und die zweite aus der H^1-Stabilität von π_h (siehe (8.43) mit der Wahl $\phi = 1$). □

Nach obigen Vorbereitungen können wir folgendes Resultat formulieren:

Satz 8.17.
Die CIP–Methode (8.33) besitzt eine eindeutige Lösung.

Beweis. Die Aussage folgt aus der Koerzitvität von $a_W + j_u$ und Satz 3.5, der wegen Lemma 8.16 in Richtung (i) nach (iii) angewandt werden darf. Für mehr Details siehe [BFH06] Theorem 3.5. □

Als nächstes soll eine a–priori Fehlerabschätzung für die CIP–Methode angegeben werden. Im Folgenden stellen wir dazu die nötigen Hilfsmittel zur Verfügung.

Lemma 8.18. (**Modifizierte Galerkin–Orthogonalität**)
Sei die Lösung (u, p) der schwachen Formulierung ausreichend regulär, das heißt, $(u, p) \in [H^{3/2+\epsilon}(\Omega)]^d \times L^2_0(\Omega)$ mit $\epsilon > 0$ und sei $(u_h, p_h) \in W^r_h$ die Lösung von (8.33). Dann gilt für alle $(v_h, q_h) \in W^r_h$:

$$a_W(u - u_h, v_h) + b_W(v_h, p - p_h) - b_W(u - u_h, q_h) + j_u(u - u_h, v_h) + j_p(p_h, q_h) = 0. \tag{8.49}$$

Beweis.
Aufgrund der vorausgesetzten Regularität der Lösung besitzt ∇u auf jedem Zellenrand eine wohldefinierte Spur. Damit verschwindet der Sprung $[\nabla u]$ über jeden Zellenrand und es gilt $j_u(u, v_h) = 0$. Aus der letzten Gleichung und der Galerkin–Orthogonalität des Standard–Galerkin–Verfahrens (8.14) mit schwach vorgegebenen Randbedingungen ergibt sich die Behauptung. □

Bemerkung 8.19.
Besitzt das Druckfeld p der Lösung die gleiche Regularität wie die Geschwindigkeit u, so erhält man mit der gleichen Argumentation wie im Beweis die Aussage $j_p(p, q_h) = 0$. In diesem Fall ist eine Lösung (u, p) der schwachen Formulierung auch Lösung der CIP–Methode und man hat die übliche Galerkin–Orthogonalität:

$$a_{\mathrm{CIP}}(u - u_h, p - p_h; v_h, q_h) = 0 \qquad \forall (v_h, q_h) \in W^r_h.$$

Eine Schlüsselstelle im Beweis der a–priori Fehlerabschätzungen ist die Existenz eines Interpolationsoperators, wie er im nächsten Lemma definiert wird.

Lemma 8.20.
Es gibt einen Interpolationsoperator $\pi^* : [H^2(\mathcal{T}_h)]^d \to [V_h^r]^d,\ d = 2, 3$, so dass für alle $v_h \in [V_h^r]^d$ und alle $K \in \mathcal{T}_h$ gilt

$$h\|b_h \cdot \nabla v_h - \pi^*(b_h \cdot \nabla v_h)\|^2_{L^2(\Omega)} \leq C \sum_{K \in \mathcal{T}_h(K)} \int_{\partial K} h_K^2 \, |b_h \cdot n| \, |[\nabla v_h \cdot n]|^2 \, ds \tag{8.50}$$

mit einer Konstante $C > 0$. $H^2(\mathcal{T}_h)$ bezeichnet hierbei den Raum aller bezüglich der Gebietszerlegung $\mathcal{T}_h$ stückweisen H^2-Funktionen, also

$$H^2(\mathcal{T}_h) = \left\{ u : \Omega \to \mathbb{R} : \ u|_K \in H^2(K) \ \forall K \in \mathcal{T}_h \right\}.$$

Beweis. Die Konstruktion von π^* und der Beweis verlaufen analog zu Lemma 6.15, siehe auch [BFH06] Lemma 3.1.

Das folgende Lemma erweitert die Interpolationsfehler–Abschätzungen für die globale L^2–Projektion.

Lemma 8.21.
Sei die Größenregularitätskonstante ξ ausreichend nahe 1. Dann gilt für alle $u \in [H^k(\Omega)]^d$ mit $k \geq 2$ die Interpolationsfehler–Abschätzung bezüglich der globalen L^2–Projektion $\pi_h : [L^2(\Omega)]^d \to [V_h^r]^d$:

$$|\!|\!| u - \pi_h u |\!|\!|^2_{\mathrm{CIP}} \leq C \sum_{K \in \mathcal{T}_h} \left(c h_K^{2r_u} + \max\{\nu, \|b\|_{L^\infty(K)} h_K\} h_K^{2(r_u - 1)} \right) \|u\|^2_{H^{r_u}(K)} \tag{8.51}$$

mit $r_u = \min\{k, r+1\}$ und einer Konstanten $C > 0$.
Außerdem gilt für alle $p \in H^l(\Omega)$ mit $l \geq 1$ die Interpolationsfehler–Abschätzung bezüglich der globalen L^2–Projektion $\pi_h : L^2(\Omega) \to Q_h^r$:

$$\begin{aligned} \|\widetilde{h}^{1/2} \phi_\pi (p - \pi_h p)\|^2_{L^2(\partial\Omega)} + \|\phi_\pi (p - \pi_h p)\|^2_{L^2(\Omega)} + j(\pi_\pi p, \pi_\pi p) \\ \leq C \sum_{K \in \mathcal{T}_h} \min\{\|b\|^2_{L^\infty(K)} h_K, h_K^2/\nu\} h_K^{2r_p - 1} \|p\|^2_{H^{s_p}(K)} \end{aligned} \tag{8.52}$$

mit $r_p = \min\{l, r+1\}$ und einer Konstanten $C > 0$. Die Größe ϕ_π in der Abschätzung ist zellenweise definiert durch die lineare L^2–Projektion von $\phi|_K = \nu^{-1/2} \min\{\mathrm{Re}_K^{-1/2}, 1\}$ auf die jeweilige Zelle.

Beweis. Wir betrachten zunächst die Abschätzung (8.51) für die Geschwindigkeit. In [BFH06] Lemma 4.6 wird der Interpolationsfehler zunächst aufgespalten in

$$|\!|\!| u - \pi_h u |\!|\!|^2_{\mathrm{CIP}} \leq |\!|\!| I_h u - \pi_h u |\!|\!|^2_{\mathrm{CIP}} + |\!|\!| I_h u - u |\!|\!|^2_{\mathrm{CIP}}, \tag{8.53}$$

wobei I_h der Lagrange-Interpolationsoperator ist. Anschließend wird die Stabilität der L^2–Projektion aus Lemma 8.15, die Spurungleichung (5.41), die inverse Ungleichung (5.40) und die Interpolationsfehler–Abschätzung für die Lagrange–Interpolation aus Lemma 5.8 angewandt, um die gewünschte Abschätzung zu erhalten. Die Abschätzung für den Druck in (8.52) erfolgt gemäß [BFH06] mit ähnlichen Argumenten, wenn statt dem Lagrange–Interpolationsoperator der Clément–Interpolationsoperator $\mathcal{C}_h : L^2(\Omega) \to Q_h^r$ aus Lemma 5.12 benutzt wird. Hier wird von der Clément–Interpolation Gebrauch gemacht, da

$p \in H^1(\Omega)$ im Allgemeinen zu geringe Regularität besitzt, um die Abschätzungen für die Lagrange–Interpolation anwenden zu können. Die Abschätzung des dritten Terms in (8.52) ist am aufwendigsten und wird hier exemplarisch durchgeführt. Es gilt

$$\begin{aligned}
j_p(\pi_h p, \pi_h p) &= \sum_{K\in\mathcal{T}_K} \zeta(\mathrm{Re}_K)\frac{h_K^2}{\|b\|_{L^\infty(K)}}\|[\nabla\pi_h p - \mathcal{C}_h\nabla p]\|^2_{L^2(\partial K)} \\
&\leq C\sum_{K\in\mathcal{T}_K} \zeta(\mathrm{Re}_K)\frac{h_K}{\|b\|_{L^\infty(K)}}\|\nabla\pi_h p - \mathcal{C}_h\nabla p\|^2_{L^2(K)} \\
&= C\sum_{K\in\mathcal{T}_K} \zeta(\mathrm{Re}_K)\frac{h_K}{\|b\|_{L^\infty(K)}}\|\nabla\pi_h p - \nabla\mathcal{C}_h p + \nabla\mathcal{C}_h p - \nabla p + \nabla p - \mathcal{C}_h\nabla p\|^2_{L^2(K)} \\
&\leq C\sum_{K\in\mathcal{T}_K} \zeta(\mathrm{Re}_K)\frac{h_K}{\|b\|_{L^\infty(K)}} \\
&\qquad \Big(\|\nabla\pi_h(p-\mathcal{C}_h p)\|_{L^2(K)} + \|\nabla\mathcal{C}_h p - \nabla p\|_{L^2(K)} + \|\nabla p - \mathcal{C}_h\nabla p\|^2_{L^2(K)}\Big) \\
&\leq C\sum_{K\in\mathcal{T}_h} \min\{\|b\|^2_{L^\infty(K)}h_K, h_K^2/\nu\}h_K^{2r_p-1}\|p\|^2_{H^{s_p}(K)}. \qquad (8.54)
\end{aligned}$$

Die erste Gleichheit ergibt sich aus der Relation $[\mathcal{C}_h\nabla p] = 0$, welche gilt, da der Clément–Interpolationsoperators von $L^2(\Omega)$ in einen Finite–Element–Raum stetiger Funktionen abbildet, siehe Abschnitt 5.4. Die erste Ungleichung folgt durch Benutzen der Abschätzung

$$\sum_K \int_{\partial K}[g]^2 ds \leq 2\sum_K \int_{\partial K}|g|^2\, ds$$

und anschließendes Anwenden der Spurungleichung (5.41) sowie der inversen Ungleichung (5.40). Die zweite Gleichheit in (8.54) folgt durch Nulladdition und die zweite Ungleichung aus der Dreiecksungleichung sowie der Relation $\pi_h(\mathcal{C}_h p) = \mathcal{C}_h p$, welche gilt, da die L^2–Projektion nach Konstruktion der Identität auf Q_h^r entspricht. Für die letzte Ungleichung in (8.54) ist schließlich die Interpolationsfehler–Abschätzung für die Clément–Interpolation aus Lemma 5.12 verwandt worden. □

Bemerkung 8.22.
Die Aufspaltung in (8.53) mit der Lagrange–Interpolation in [BFH06] wird durchgeführt, da die Gebietszerlegung $\mathcal{T}_h$ dort allgemeiner ist. In [BFH06] wird die Quasiuniformität von $\mathcal{T}_h$ lediglich lokal gefordert. In diesem Fall ist die Interpolationsfehler–Abschätzung aus Lemma 5.10 für die L^2–Projektion nicht mehr gültig. Die Abschätzung für die Lagrange–Interpolation ist aber weiterhin erlaubt und wird daher stattdessen benutzt. Da unsere Gebietszerlegungen immer quasiuniform sein werden, können wir die Interpolationsfehler–Abschätzung für die L^2–Projektion noch verwenden, was die Aufspaltung (8.53) überflüssig macht. Damit sollte die Abschätzung für die Geschwindigkeit für unsere Zwecke deutlich einfacher als in [BFH06] ausfallen. Die Abschätzung des Drucks lässt sich hingegen nicht entscheidend vereinfachen, da für die L^2–Projektion keine Interpolationsfehler–Abschätzungen zu beliebigen Funktionen aus $H^1(\Omega)$ möglich sind.

Mit dem letzten Lemma sind alle Vorbereitungen zum Beweis der gesuchten Fehlerabschätzung getroffen.

Satz 8.23. (**Fehlerabschätzung der CIP–Methode**)
Sei $(u,p) \in [H^k(\Omega)]^d \times H^l(\Omega)$ mit $k \geq 2$, $l \geq 1$, die Lösung der schwachen Formulierung

(8.4) und $(u_h, p_h) \in W_h^r$ die aus (8.33) bestimmte Näherungslösung. Dann gelten unter den Voraussetzungen von Lemma 8.21 die folgenden globale Fehlerabschätzungen:

$$\begin{aligned}
\|\!|u - u_h|\!\|_{\mathrm{CIP}} \leq C & \left(\sum_{K \in \mathcal{T}_h} \left(c h_K^{2r_u} + \max\{\|b\|_{L^\infty(K)} h_K, \nu\} h_K^{2(r_u-1)} \right) |u|^2_{H^{r_u}(K)} \right)^{1/2} \\
& + C \max_{K \in \mathcal{T}} \{ c^{-1/2} \, |b|_{W^{1,\infty}(K)} \, h_K^{r_u} \} \|u\|_{H^{r_u}(\Omega)} \\
& + C \left(\sum_{K \in \mathcal{T}_h} \min\{\|b\|^{-1}_{L^\infty(K)}, h_K/\nu\} h_K^{2s_p - 1} \|p\|^2_{H^{r_p}(K)} \right)^{1/2}
\end{aligned} \tag{8.55}$$

sowie

$$\|p - p_h\|_{L^2(\Omega)} \leq C \left(c^{1/2} + \max_{K \in \mathcal{T}_h} \{\|b\|_{L^\infty(K)} h_K, \nu\}^{1/2} + c^{-1/2} \|b\|_{L^\infty(\Omega)} \right) \mathrm{KR}_u \tag{8.56}$$

mit der Konvergenzrate KR_u von u in (8.55), $r_u = \min\{k, r+1\}$, $r_p = \min\{l, r+1\}$ und einer Konstanten $C > 0$.

Beweis. Siehe [BFH06] Theorem 4.9 und 4.12. Die Beweise sind länglich und erfordern einige technische Zwischenrechnungen. Um zumindest einen Eindruck zu gewinnen, wie aus den bisher präsentierten Resultaten eine Fehlerabschätzung abgeleitet werden kann, führen wir die Rechnung exemplarisch für den konvektiven Term bei der Abschätzung des Geschwindigkeitsfehlers durch. Dies ist zugleich eine der Schlüsselstellen im Beweis. Zunächst zerlegen wir den Fehler $\|\!|u - u_h|\!\|_{\mathrm{CIP}}$ soweit, dass der konvektive Term einzeln von den anderen Beiträgen zum Fehler behandelt werden kann. Mit der Dreiecksungleichung schätzen wir nach oben ab

$$\|u - u_h\|_{\mathrm{CIP}} \leq \|\!|e^\pi|\!\|_{\mathrm{CIP}} \|\!|e_h|\!\|_{\mathrm{CIP}} \tag{8.57}$$

mit den Bezeichnungen

$$u - u_h = \underbrace{u - \pi_h}_{e^\pi} + \underbrace{\pi_h u - u_h}_{e_h}.$$

Der erste Term in (8.57) wird bereits durch Lemma 8.21 kontrolliert. Der zweite muss noch abgeschätzt werden und wird mit Hilfe der Koerzitivität aus Lemma 8.14 und der modifizierten Galerkin–Orthogonalität aus Lemma 8.18 wie folgt zerlegt:

$$\begin{aligned}
C \|\!|e_h|\!\|^2_{\mathrm{CIP}} + j_p(p - p_h, p - p_h) &\leq a_W(e_h, e_h) + j_u(e_h, e_h) \\
&\leq a_W(e^\pi, e_h) + b_W(e_h, \pi_h(p - p_h)) - b_W(e^\pi, \pi_h(p - p_h)) \\
&\quad + j_u(e^\pi, e_h) - j_p(\pi_h(p - p_h), p - p_h)
\end{aligned} \tag{8.58}$$

mit

$$\begin{aligned}
a_W(e^\pi, e_h) = c(e^\pi, e_h) &+ 2\nu(\nabla e^\pi, \nabla e_h) - ((b \cdot n) e^\pi, e_h)_{\partial\Omega_-} + \tau(\nu/\widetilde{h}\, e^\pi, e_h)_{\partial\Omega} \\
&+ \tau(\max\{|b|, \nu/\widetilde{h}\} e^\pi \cdot n, e_h \cdot n)_{\partial\Omega} - 2\nu(\nabla e^\pi \cdot n, e_h)_{\partial\Omega} \\
&- 2\nu(e^\pi, \nabla e_h \cdot n)_{\partial\Omega} + (\nabla e^\pi \cdot b, e_h).
\end{aligned}$$

Die ersten fünf Terme von $a_W(e^\pi, e_h)$ sind symmetrisch und können direkt mit Hilfe der Cauchy–Schwarz–Ungleichung und der Interpolationsfehler–Abschätzung aus Lemma 8.21 wie gewünscht nach oben beschränkt werden. Die hinteren drei sind unsymmetrisch und

erfordern mehr Aufwand. Der letzte Term ist der Beitrag des konvektiven Terms. Dieser soll nun abgeschätzt werden. Mit partieller Integration folgt

$$(b \cdot \nabla e^\pi, e_h) \leq |(e^\pi, (b \cdot n) e_h)| + |(e^\pi, \nabla e_h \cdot b)| \,.$$

Der erste Term wird wie oben durch die Cauchy–Schwarz–Ungleichung und Lemma 8.21 abgeschätzt. Den zweiten zerlegen wir nochmals durch Nulladdition mit der stückweise linearen Interpolation b_h von b und der Dreiecksungleichung:

$$|(e^\pi, b \cdot \nabla e_h)| \leq |(e^\pi, \nabla e_h \cdot (b - b_h))| + |(e^\pi, \nabla e_h \cdot b_h)| \,. \tag{8.59}$$

Die beiden Terme auf der rechten Seite beschränken wir einzeln. Der erste lässt sich abschätzen durch:

$$\begin{aligned}
|(e^\pi, \nabla e_h \cdot (b - b_h))| &\leq C \sum_{K \in \mathcal{T}_h} h_K \, |b_h|_{W^{1,\infty}(K)} \, \|e^\pi\|_{L^2(K)} \|\nabla e_h\|_{L^2(K)} \\
&\leq C \sum_{K \in \mathcal{T}_h} c^{-1/2} \, |b_h|_{W^{1,\infty}(K)} \, \|e^\pi\|_{L^2(K)} \|c^{1/2} e_h\|_{L^2(K)} \\
&\leq C \max_{K \in \mathcal{T}_h} \{ c^{-1/2} \, |b_h|_{W^{1,\infty}(K)} \, h_K^{r_u} \} \|u\|_{H^{r_u}(\Omega)} |\!|\!| e_h |\!|\!|_{\mathrm{CIP}}.
\end{aligned}$$

Die erste Ungleichung folgt aus der Cauchy–Schwarz–Ungleichung und weil b in $W^{1,\infty}(\Omega)$ liegt, die zweite aus der inversen Ungleichung (5.40) sowie Einsmultiplikation mit $c^{1/2}$ und die dritte aus der Interpolationsfehler–Abschätzung für die L^2–Projektion aus Lemma 5.10. Der zweite Term in (8.59) wird nun noch wie folgt beschränkt

$$\begin{aligned}
|(e^\pi, \nabla e_h \cdot b_h)| &= |(e^\pi, \nabla e_h \cdot b_h - \pi^*(\nabla e_h \cdot b_h))| \\
&\leq C h^{-1/2} \|e^\pi\|_{L^2(\Omega)} \|h^{1/2} \nabla e_h \cdot b_h - \pi^*(\nabla e_h \cdot b_h)\|_{L^2(\Omega)} \\
&\leq C h^{-1/2} \|e^\pi\|_{L^2(\Omega)} \left(\sum_{K \in \mathcal{T}_h} \int_{\partial K} h_K^2 \|b \cdot n\|_{L^\infty(\partial K)} \, |[\nabla e_h \cdot n]|^2 \, ds \right)^{1/2} \\
&\leq C \left(\sum_{K \in \mathcal{T}_h} \|b\|_{L^\infty(K)} h_K^{2r_u - 1} \|u\|^2_{H^{r_u}(K)} \right)^{1/2} \\
&\qquad \left(\sum_{K \in \mathcal{T}_h} \int_{\partial K} h_K^2 \|b \cdot n\|_{L^\infty(\partial K)} \, |[\nabla e_h \cdot n]|^2 \, ds \right)^{1/2} \\
&\leq C \left(\sum_{K \in \mathcal{T}_h} \|b\|_{L^\infty(K)} h_K^{2r_u - 1} \|u\|^2_{H^{r_u}(K)} \right)^{1/2} |\!|\!| e_h |\!|\!|_{\mathrm{CIP}}.
\end{aligned} \tag{8.60}$$

Die erste Gleichung gilt wegen der Orthogonalität der L^2–Projektion und die Ungleichungen der Reihe nach wegen der Cauchy–Schwarz–Ungleichung, der Interpolationsfehler–Abschätzung für π^* aus Lemma 8.20 und der Interpolationsfehler–Abschätzung für π_h aus Lemma 5.10. Die letzte Ungleichung überprüft man mit der folgenden Zwischenrechnung (siehe auch Lemma 4.8 in [BFH06]): Es bezeichne A_1 die Menge aller Gitterzellen von $\mathcal{T}_h$

auf denen $\|b\|_{L^\infty(\Omega)} h_K \geq \nu$ ist und A_2 die diejenige mit $\|b\|_{L^\infty(\Omega)} h_K < \nu$. Dann gilt:

$$\begin{aligned}
&\sum_{K\in\mathcal{T}_h} \int_{\partial K} h_K^2 \|b\cdot n\|_{L^\infty(\partial K)} \left|[\nabla e_h \cdot n]\right|^2 ds \\
&\quad \leq \sum_{K\in A_1} \int_{\partial K} h_K^2 \|b\cdot n\|_{L^\infty(\partial K)} \left|[\nabla e_h \cdot n]\right|^2 ds + \sum_{K\in A_2} h_K \nu \int_{\partial K} \left|[\nabla e_h \cdot n]\right|^2 ds \\
&\quad \leq \sum_{K\in\mathcal{T}_h} \int_{\partial K} h_K^2 \zeta(\mathrm{Re}_K) \|b\cdot n\|_{L^\infty(\partial K)} \left|[\nabla e_h \cdot n]\right|^2 ds + C\|\nu^{1/2}\nabla e_h\|^2_{L^2(\Omega)} \\
&\quad \leq C\left(j_u(e_h,e_h) + \|\nu^{1/2}\nabla e_h\|^2_{L^2(\Omega)}\right) \\
&\quad \leq C|\!|\!| e_h |\!|\!|_{\mathrm{CIP}}.
\end{aligned}$$

Dabei ist für die erste Ungleichung die Relation $\|b\|_{L^\infty(\Omega)} h_K < \nu$ für alle $K \in A_2$ benutzt worden. Bei der zweiten Ungleichung wurde im ersten Summand verwandt, dass $\zeta(\mathrm{Re}_K) = 1$ für alle $K \in A_2$ ist, und die Summe wurde über alle $K \in \mathcal{T}_h$ erweitert. Der zweite Summand wurde mit Hilfe der Spurungleichung (5.41) und der inversen Ungleichung (5.40) nach oben beschränkt. Die beiden letzten Ungleichungen gelten direkt nach Definition von j_u und der CIP-Norm $|\!|\!| \cdot |\!|\!|_{\mathrm{CIP}}$. Damit ist die Fehlerabschätzung für den konvektiven Term vollständig. □

Bemerkung 8.24.
(i) Aus obigem Satz folgen bei ausreichender Regularität der Lösung im konvektions–dominanten Fall wegen $\nu \leq \|b\|_{L^\infty(\Omega)} h$ die Fehlerabschätzungen:

$$\begin{aligned}
\|u-u_h\|_{L^2(\Omega)} &\leq Ch^{r+1/2}\left(\|u\|_{H^{r+1}(\Omega)} + \|p\|_{H^{r+1}(\Omega)}\right), \\
\|p-p_h\|_{L^2(\Omega)} &\leq Ch^{r+1/2}\left(\|u\|_{H^{r+1}(\Omega)} + \|p\|_{H^{r+1}(\Omega)}\right), \\
\|u-u_h\|_{H^1(\Omega)} &\leq Ch^{r+1/2}\nu^{-1/2}\left(\|u\|_{H^{r+1}(\Omega)} + \|p\|_{H^{r+1}(\Omega)}\right).
\end{aligned} \tag{8.61}$$

Ausreichend regulär meint dabei $(u,p) \in [H^k(\Omega)]^d \times H^k(\Omega)$ mit $k \geq r+1$. Unter denselben Regularitätsbedingungen ergibt sich im diffusions–dominanten Fall mit $\nu \geq \|b\|_{L^\infty(\Omega)} h$:

$$\begin{aligned}
\|u-u_h\|_{L^2(\Omega)} &\leq Ch^r\left(\nu^{1/2}\|u\|_{H^r(\Omega)} + h^{1/2}\|p\|_{H^{r+1}(\Omega)}\right), \\
\|p-p_h\|_{L^2(\Omega)} &\leq C\nu^{1/2}h^r\left(\nu^{1/2}\|u\|_{H^{r+1}(\Omega)} + h^{1/2}\|p\|_{H^{r+1}(\Omega)}\right), \\
\|u-u_h\|_{H^1(\Omega)} &\leq Ch^r\left(\|u\|_{H^{r+1}(\Omega)} + \|p\|_{H^{r+1}(\Omega)}\right).
\end{aligned} \tag{8.62}$$

Kapitel 9

Numerische Ergebnisse zur Oseen–Gleichung

In diesem Abschnitt werden die numerischen Ergebnisse zur Oseen–Gleichung vorgestellt. Betrachtet werden die residuale Stabilisierung und die CIP–Methode. Wir beginnen mit einem Überblick über die Implementierung der Stablisierungsverfahren. Dann vergleichen wir die Fehlerwerte und die Fehlerordnungen der residualen Stabilisierung sowie der CIP–Methode miteinander und diskutieren den Vorschlag aus [BBJL07] für die Stabilisierungsparameter der CIP–Methode. Speziell für die CIP–Methode untersuchen wir dann, ob die Fehler durch eine geeignetere Wahl der Stabilisierungsparameter verringert werden können. Gerechnet wird bei obigen Simulationen mit equal order Elementen. Neben equal order Elementen sind auch Taylor–Hood–Elemente von Interesse. Zu diesen geben wir als nächstes Ergebnisse für die CIP–Methode an. Schließlich vergleichen wir die residuale Stabilisierung mit der CIP–Methode anhand ihrer Laufzeiten.

9.1 Grundlagen zu den Simulationen

Das Standard–Galerkin–Verfahren zur Lösung der Oseen–Gleichung lautet: Finde $(u_h, p_h) \in W_h^{r,s}$, so dass für alle $(v_h, p_h) \in W_h^{r,s}$ gilt:

$$a_{\mathrm{SG}}(u_h, p_h; v_h, q_h) = (f, v_h) \quad \text{mit} \tag{9.1}$$

$$a_{\mathrm{SG}}(u_h, p_h; v_h, q_h) := \nu(\nabla u_h, \nabla v_h) + ((b \cdot \nabla)u_h, v_h) + (cu_h, v_h) - (\nabla \cdot u_h, q_h). \tag{9.2}$$

Bei der residualen Stabilisierung erweitern wir das Standard–Galerkin–Verfahren um die Terme

$$\sum_{K \in \mathcal{T}_h} \delta_K(\mathrm{Res}, (b \cdot \nabla)v_h + \nabla q_h) + \sum_{K \in \mathcal{T}_h} \gamma_K(\nabla \cdot u_h, \nabla \cdot v_h).$$

Im Folgenden werden wir uns bei der residualen Stabilisierung auf equal order Elemente beschränken (für Ergebnisse zu Taylor–Hood–Elementen siehe beispielsweise [MLR09]). Für die equal order Elemente setzen wir die Stabilisierungsparameter δ_K, γ_K motiviert durch Satz 8.10 zellenweise an durch:

$$\delta_K = \delta \frac{h_K^2}{r h_K \|b\|_{L^\infty(K)} + \|c\|_{L^\infty(K)} h_K^2 + r^4 \nu},$$

$$\gamma_K = \gamma \frac{h_K \|b\|_{L^\infty(K)}}{r} + \frac{\|c\|_{L^\infty(K)} h_K^2}{r^2} + r^2 \nu. \tag{9.3}$$

Die Größen δ und γ sind globale Parameter, welche bei unseren Simulationen variiert werden.

Als nächstes führen wir unseren Ansatz für die CIP–Methode ein. Ansatz (8.33), welchen wir in Abschnitt 8.4 betrachtet haben, wird hierbei leicht modifiziert. Zum einen werden wir im Code die Randbedingungen stark vorgeben, was die Implementierung des Verfahrens erleichtert. Zum anderen ergänzen wir das Standard–Galerkin–Verfahren (9.2) um die Terme:

$$\begin{aligned}
&\sum_{E\in\mathcal{E}_h}\int_E \tau_1[\nabla u_h\cdot n_E][\nabla v_h\cdot n_E]ds\\
+&\sum_{E\in\mathcal{E}_h}\int_E \tau_2[\nabla\cdot u_h][\nabla\cdot v_h]ds\\
+&\sum_{E\in\mathcal{E}_h}\int_E \tau_3[\nabla p_h\cdot n_E][\nabla q_h\cdot n_E]ds,
\end{aligned}$$

mit der Menge $\mathcal{E}_h$ aller inneren Kanten der Gebietszerlegung, dem Normalenvektor n_E an die Kante E und der Jacobimatrix ∇u von u. Gemäß [BBJL07] bleiben für diesen Ansatz unsere Resultate aus Abschnitt 8.4 gültig. Darüber hinaus bietet der Ansatz den Vorteil, dass in [BBJL07] eine Empfehlung für die Stabilisierungsparameter τ_i gegeben wird. Diese lautet:

$$\begin{aligned}
\tau_1 &= \tau_{\text{grad}}\|b\cdot n_E\|_{L^\infty(E)}\frac{h_E^2}{r^\alpha},\\
\tau_2 &= \tau_{\text{div}}\|b\|_{L^\infty(E)}\frac{h_E^2}{r^\alpha},\\
\tau_3 &= \tau_p\min(1,\text{Re}_E)\frac{h_E^2}{\|b\|_{L^\infty(E)}r^\alpha},\\
\text{mit } \text{Re}_E &= \frac{\|b\|_{L^\infty(E)}h_E}{\nu r^{1/2}},\ \alpha=\frac{7}{2}\ \text{ und } \tau_{\text{grad}}=\tau_{\text{div}}=\tau_p=1.
\end{aligned}\tag{9.4}$$

Während nach der Empfehlung von [BBJL07] die Parameter τ_{grad}, τ_{div} und τ_p auf 1 zu fixieren sind, werden wir auch andere Werte zulassen.

Bei der Implementierung der CIP–Methode ist die Systemmatrix des Standard–Galerkin–Verfahrens wegen der hinzukommenden Stabilisierungsterme zu erweitern. Für ein zweidimensionales Problem ist die Systemmatrix des Standard–Galerkin–Verfahrens von der Form:

$$\begin{pmatrix} A_{11} & A_{12} & B_1\\ A_{21} & A_{22} & B_2\\ B_1^T & B_2^T & C\end{pmatrix} = \begin{pmatrix} A & 0 & B_1\\ 0 & A & B_2\\ B_1^T & B_2^T & 0\end{pmatrix}.\tag{9.5}$$

Die Matrixblöcke A_{ij}, B_i und C werden bei dieser Schreibweise nach dem Raum unterschieden, aus welchem die jeweilige Ansatz- und Testfunktion stammt. Die folgende Tabelle zeigt das zugehörige Schema, wobei die beiden Geschwindigkeitskomponenten mit x und y gekennzeichnet sind:

Matrixblock	Testfunktion	Ansatzfunktion
A_{11}	v_h^x	u_h^x
A_{12}	v_h^x	u_h^y
A_{21}	v_h^y	u_h^x
A_{22}	v_h^y	u_h^y
B_1	v_h^x	p_h
B_2	v_h^y	p_h
B_1^T	q_h	u_h^x
B_2^T	q_h	u_h^y
C	q_h	p_h

Durch die Stabilisierungsterme der CIP–Methode ist die Matrix (9.5) durch eine Matrix des Typs

$$\begin{pmatrix} \bar{A}_{11} & \bar{A}_{12} & 0 \\ \bar{A}_{21} & \bar{A}_{22} & 0 \\ 0 & 0 & \bar{C} \end{pmatrix} \tag{9.6}$$

zu erweitern. Die Systemmatrix der CIP–Methode ist damit von der Form

$$\begin{pmatrix} A_{11} & A_{12} & B_1 \\ A_{21} & A_{22} & B_2 \\ B_1^T & B_2^T & C \end{pmatrix}. \tag{9.7}$$

Insbesondere werden also Ansatz- und Testfunktionen des Druckfeldes miteinander gekoppelt. Eine zusätzliche Kopplung zwischen Druck und Geschwindigkeit gibt es jedoch entgegen der Erweiterung bei der residualen Stabilisierung nicht.
Da die Stabilisierungsterme Kantenintegrale enthalten, besitzen die Matrixblöcke in (9.6) ein dichteres Besetzungsschema als die Blöcke der Systemmatrix des Standard–Galerkin–Verfahrens. Für jeden Block einzeln gesehen ist dies analog zur Konvektions–Diffusionsgleichung (vergleiche mit Abschnitt 7.1). Um die CIP–Methode für die Oseen–Gleichung zu erhalten, wird dieses Schema für jeden der Matrixblöcke in (9.6) benutzt.

Bei den nachfolgenden Rechnungen wählen wir Finite–Element–Räume auf Quadraten oder auf Dreiecken. Das Lösungsgebiet ist das Einheitsquadrat $[0,1]^2$, welches nach dem in Abbildung 7.2 gezeigten Schema trianguliert wird. Die Gebietszerlegung ist für den Druck- und Geschwindigkeitsraum identisch. Zur Lösung des linearen Gleichungssystems setzen wir den direkten Löser aus dem Paket UMFPACK [Dav04] ein. Der Einfachheit halber beschränken wir uns bei allen Beispielen auf die Vorgabe $c = 0$ in Ω in der Oseen–Gleichung. Im Vergleich zur Situation in Abschnitt 7.2 gibt es hier mit u_x, u_y und p drei unbekannte Größen statt nur einer. Die Anzahl der Freiheitsgrade ist daher bei gleicher Elementordnung und gleichem Gittertyp dreimal so groß wie diejenige aus Abschnitt 7.2.
Zu Testzwecken sind die Stabilisierungsverfahren auch auf Beispiele angewandt worden, bei welchen die Lösung im Ansatzraum des Verfahrens liegt. Erwartungsgemäß haben beide Verfahren die Lösung bis auf Rundungsfehler exakt berechnet. Getestet wurde dies für beide Gittertypen bis Finite–Elemente vierter Ordnung.
Die Ergebnisse der residualen Stabilisierung konnten zudem mit den Resultaten aus anderen Arbeiten verglichen werden. So sind etwa zu Beispiel 8.0.1 aus [Roe07] vergleichbare

Fehler- und Ordnungswerte berechnet worden. Für die CIP–Methode sind bisher nur die numerischen Resultate in [BFH06] zur Oseen–Gleichung veröffentlicht worden. Gerechnet wird dort allerdings für ein dreidimensionales Lösungsgebiet, welches nicht im Rahmen dieser Arbeit in MooNMD implementiert worden ist. Ein direkter Vergleich mit Literaturwerten wurde daher für die CIP–Methode nicht durchgeführt.

9.2 Das sincos–Beispiel

Als erstes untersuchen wir das sincos-Beispiel mit der folgenden Lösung auf $\Omega = [0,1]^2$:

$$\begin{aligned} u_x &= \sin(\pi x), \\ u_y &= -\pi \cos(\pi y), \\ p &= \sin(\pi x)\cos(\pi y). \end{aligned} \tag{9.8}$$

Als Vorgaben wählen wir $b = u$, $c = 0$ und $\nu = 10^{-6}$ auf Ω. Zu diesen Vorgaben werden die rechte Seite f der Oseen–Gleichung und die Randwerte so angepasst, dass (9.8) die Oseen–Gleichung löst. Dies gelingt leicht, indem man die Werte der obigen Lösung als inhomogenen Dirichlet–Randwerte auf $\partial\Omega$ vorgibt und f durch Einsetzen von (9.8) in die Oseen–Gleichung (8.1) festlegt.
Die Lösung des sincos-Beispiels ist in Abbildung 9.2 zu sehen.

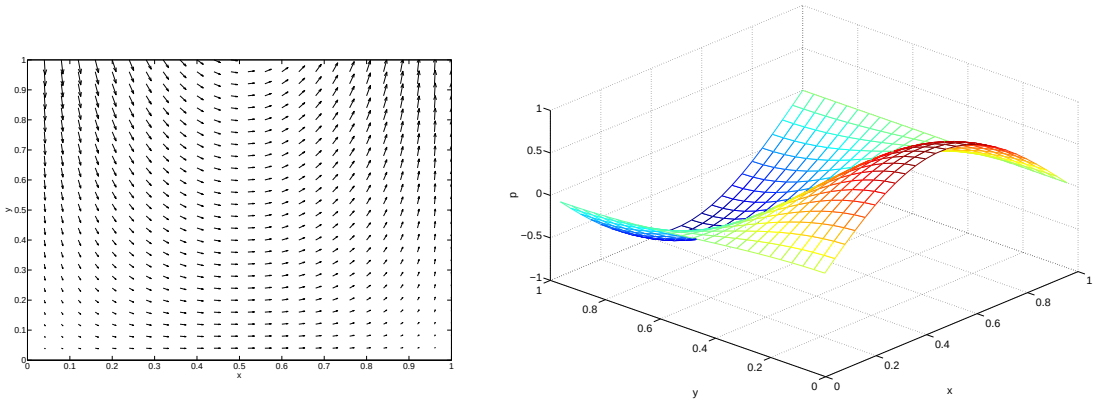

Abbildung 9.1: Links Vektorplot für das Geschwindigkeitsfeld des sincos–Beispiels, rechts das zugehörige Druckfeld.

Die folgenden Seiten zeigen für equal order Elemente die Fehler und Fehlerordnungen der CIP–Methode im Vergleich zur residualen Stabilisierung. Der Einfachheit halber fixieren wir dabei die Stabilisierungsparameter τ_{grad}, τ_{div}, τ_p der CIP–Methode auf einen gemeinsamen Wert τ. Selbiges tun wir mit den Stabilisierungsparametern δ und γ der residualen Stabilisierung. Für die residuale Stabilisierung ist bekannt, dass die Ergebnisse durch eine unterschiedliche Wahl von δ und γ (zumindest leicht) verbessert werden können, siehe beispielsweise [LR06]. Dennoch kann es insbesondere aus praktischer Sicht attraktiv sein, die Anzahl der Parameter, deren optimaler Wert vorab nicht bekannt ist, zu reduzieren.
Gerechnet wird auf dem Dreiecks- und Quadratgitter für Finite–Elemente erster bis dritter Ordnung. Für erste und zweite Ordnung sind die Rechnungen bis zum Level–5–Gitter ausgeführt worden sowie für Elemente dritter Ordnung bis zum Level–4–Gitter. Die Fehlerordnungen werden mit den Voraussagen der globalen Fehlerabschätzungen aus Kapitel

8 verglichen. Dazu ist zwischen dem konvektions– und dem diffusions–dominanten Fall zu unterscheiden. Der konvektions–dominante Fall wird durch die Ungleichung $\nu \leq \|b\|_{L^\infty(\Omega)} h$ angezeigt. Da unsere Rechnungen maximal bis zum Level–5–Gitter durchgeführt werden, ist die Gitterweite h stets größer als 0.03. Wegen der Vorgabe $\nu = 10^{-6}$ und der groben Abschätzung $\|b\|_{L^\infty(\Omega)} < 10$ liegt bei allen Rechnungen der konvektions–dominanten Fall vor. Für diesen Fall sind die Raten der residualen Stabilisierung aus (8.29) zu entnehmen und die Raten der CIP–Methode aus (8.61). Für beide Verfahren werden die gleichen Fehlerordnungen prognostiziert. Diese sind in der folgenden Tabelle in Abhängigkeit von der Elementordnung r zusammengefasst. Daneben enthält die Tabelle die an den Interpolationsfehler–Abschätzungen aus Abschnitt 5.4 gemessenen optimalen Konvergenzraten.

Fehler	Konvergenzrate aus der Fehlerabschätzung	optimale Rate
L^2–Fehler von u	$h^{r+1/2}$	h^{r+1}
H^1–Fehler von u	h^r	h^r
L^2–Fehler von p	$h^{r+1/2}$	h^{r+1}
H^1–Fehler von p	-	h^r

Tabelle 9.1: Konvergenzraten für Finite–Elemente der Ordnung r für die residuale Stabilisierung und die CIP–Methode. In Spalte zwei sind die durch die a–priori Fehlerabschätzungen aus Kapitel 8 vorausgesagten Raten gezeigt, in Spalte drei die an den Interpolationsfehler–Abschätzungen aus Abschnitt 5.4 gemessenen optimalen Raten. Für den H^1–Fehler des Druckfeldes macht die Analysis keine Aussage.

Bei dem Vergleich unserer Resultate mit den Voraussagen der Fehlerabschätzungen ist zu beachten, dass letztere streng genommen nur eine asymptotische Aussage für eine gegen Null strebende Gitterweite machen. Da wir nicht für beliebig kleine Gitterweiten rechnen können, kann es im Folgenden also durchaus vorkommen, dass die vorausgesagten Raten verfehlt werden.

Bei der Diskussion unserer Ergebnisse sollen vor allem die folgenden Fragen erörtert werden:

- Liefert die neu implementierte CIP–Methode die aus der Fehleranalysis vorausgesagten Konvergenzraten?
- Wie gut ist die Empfehlung $\tau = 1$?
- Wie gut ist die CIP–Methode im Vergleich zur residualen Stabilisierung?
- Inwiefern unterscheiden sich die Ergebnisse zwischen dem Dreiecks- und dem Quadratgitter?

Wir beginnen mit den Ergebnissen zu den Elementen erster Ordnung auf dem Dreiecksgitter. Wie Abbildung 9.2 zeigt, nehmen die Fehler der CIP–Methode erwartungsgemäß bei jeder Gitterverfeinerung ab. Die Abnahme der Fehler erfolgt außer bei zu großem τ mit den Konvergenzraten, die oberhalb der Raten aus Tabelle 9.1 liegen. Das Minimum des L^2– und H^1–Fehlers im Geschwindigkeits- sowie im Druckfeld wird gleichzeitig bei etwa $\tau = 0.017$ erreicht und dies unabhängig von der Gitterweite. Tabelle 9.2 zeigt zum Vergleich die Fehlerwerte bei der Wahl $\tau = 1$, welche in [BBJL07] empfohlen wird. Die Fehler bei

$\tau = 1$ sind größer als bei $\tau = 0.017$, liegen aber zumindest in der gleichen Größenordnung. Weiterhin ist zu erkennen, dass eine Stabilisierung benötigt wird, um gute Resultate zu erhalten. Dies zeigt sich besonders in den Fehlern des Druckfeldes. Für kleine Stabilisierungsparameter wachsen die Fehler im Druckfeld mit abnehmendem τ weiter an. Dies ist auch außerhalb des dargestellten Parameterbereichs der Fall. Reduziert man beispielsweise τ auf einen Wert von 10^{-7}, so beträgt der H^1–Fehler im Druckfeld bereits ungefähr 17. Die relativ großen Fehler im Druckfeld sind darauf zurückzuführen, dass die diskreten Babuška–Brezzi–Bedingung bei den equal order Elementen verletzt ist.

Die orangefarbene Kurve zeigt zum Vergleich die Werte der residualen Stabilisierung für das Level–5–Gitter. Hier liefert die Wahl $\tau = 0.215$ in allen Fehlern die optimalen Werte. Diese sind gemäß Tabelle 9.2 etwa gleich groß wie die Fehlerwerte der CIP–Methode.
Wünschenswert ist neben möglichst kleinen Fehlern bei einer optimalen Wahl des Stabilisierungsparameters, dass die Fehler möglichst langsam anwachsen, wenn der Parameter von seinem optimalen Wert entfernt wird. Auch in diesem Kriterium schneiden beide Stabilisierungsverfahren etwa gleich gut ab. Während die residuale Stabilisierung besser für große τ ist, liefert die CIP–Methode bei kleinen τ die besseren Ergebnisse.

τ	$L^2(u)$	$H^1(u)$	$L^2(p)$	$H^1(p)$	$\mathcal{O}(L^2(u))$	$\mathcal{O}(H^1(u))$	$\mathcal{O}(L^2(p))$	$\mathcal{O}(H^1(p))$
1.0	1.51e-3	1.03e-1	1.09e-3	6.23e-2	2.55	1.26	2.81	1.28
1.77e-2	3.39e-4	9.07e-2	3.19e-4	5.50e-2	2.00	1.00	2.03	1.02
0.215	4.13e-4	9.06e-2	3.39e-4	5.48e-2	2.00	1.00	2.01	1.01

Tabelle 9.2: Fehler und Ordnungen für Finite–Elemente erster Ordnung auf dem Level–5–Dreiecksgitter. Die beiden oberen Zeilen beziehen sich auf die CIP–Methode und die untere auf die residuale Stabilisierung.

Abbildung 9.3 zeigt die Resultate für Finite–Elemente erster Ordnung auf dem Quadratgitter. Qualitativ gleicht die Situation derjenigen von Finiten–Elemente auf dem Dreiecksgitter. Insbesondere liefert die Wahl $\tau = 1$ gemäß Tabelle 9.3 Fehlerwerte in der gleichen Größenordnung wie die optimale Wahl $\tau \approx 0.001$. Quantitativ gibt es hingegen Unterschiede: Die Fehlerwerte auf dem Quadratgitter sind für beide Stabilisierungsverfahren kleiner und der Bereich optimaler τ fällt sichtbar größer aus.

τ	$L^2(u)$	$H^1(u)$	$L^2(p)$	$H^1(p)$	$\mathcal{O}(L^2(u))$	$\mathcal{O}(H^1(u))$	$\mathcal{O}(L^2(p))$	$\mathcal{O}(H^1(p))$
1.0	4.85e-4	6.65e-2	6.00e-4	3.89e-2	2.48	1.05	2.71	1.26
1.0e-3	3.21e-4	6.51e-2	3.04e-4	3.14e-2	2.00	1.00	2.00	1.00
1.0e-3	3.22e-4	6.51e-2	3.04e-4	3.14e-2	2.00	1.00	2.00	1.00

Tabelle 9.3: Fehler und Ordnungen für Finite–Elemente erster Ordnung auf dem Level–5–Quadratgitter. Die beiden oberen Zeilen beziehen sich auf die CIP–Methode und die untere auf die residuale Stabilisierung.

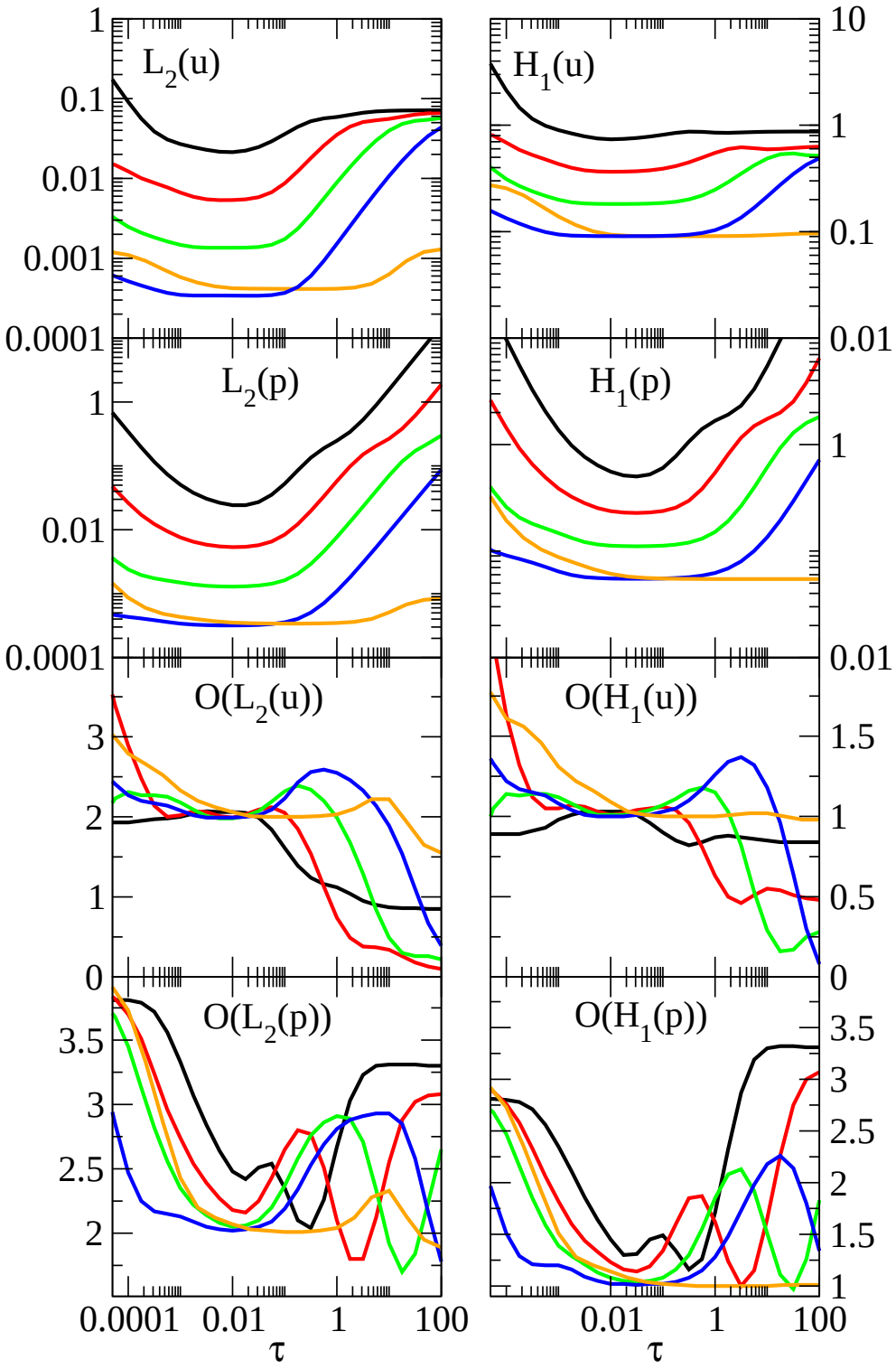

Abbildung 9.2: Fehler und Fehlerordnungen für Finite–Elemente erster Ordnung auf dem Dreiecksgitter gegen den Stabilisierungsparameter τ. Die Graphen in den ersten beiden Reihen zeigen die Fehlerwerte im Geschwindigkeitsfeld u und im Druckfeld p in der L^2– beziehungsweise H^1–Norm. Die Graphen in den beiden unteren Reihen geben die zugehörigen Fehlerordnungen an. Die orangefarbenen Linien beziehen sich auf die residuale Stabilisierung auf dem Level–5–Gitter. Die restlichen Kurven sind Resultate der CIP–Methode. Für die schwarzen, roten, grünen und blauen Linien wurde der Reihe nach auf Level 2,3,4 und 5 gerechnet.

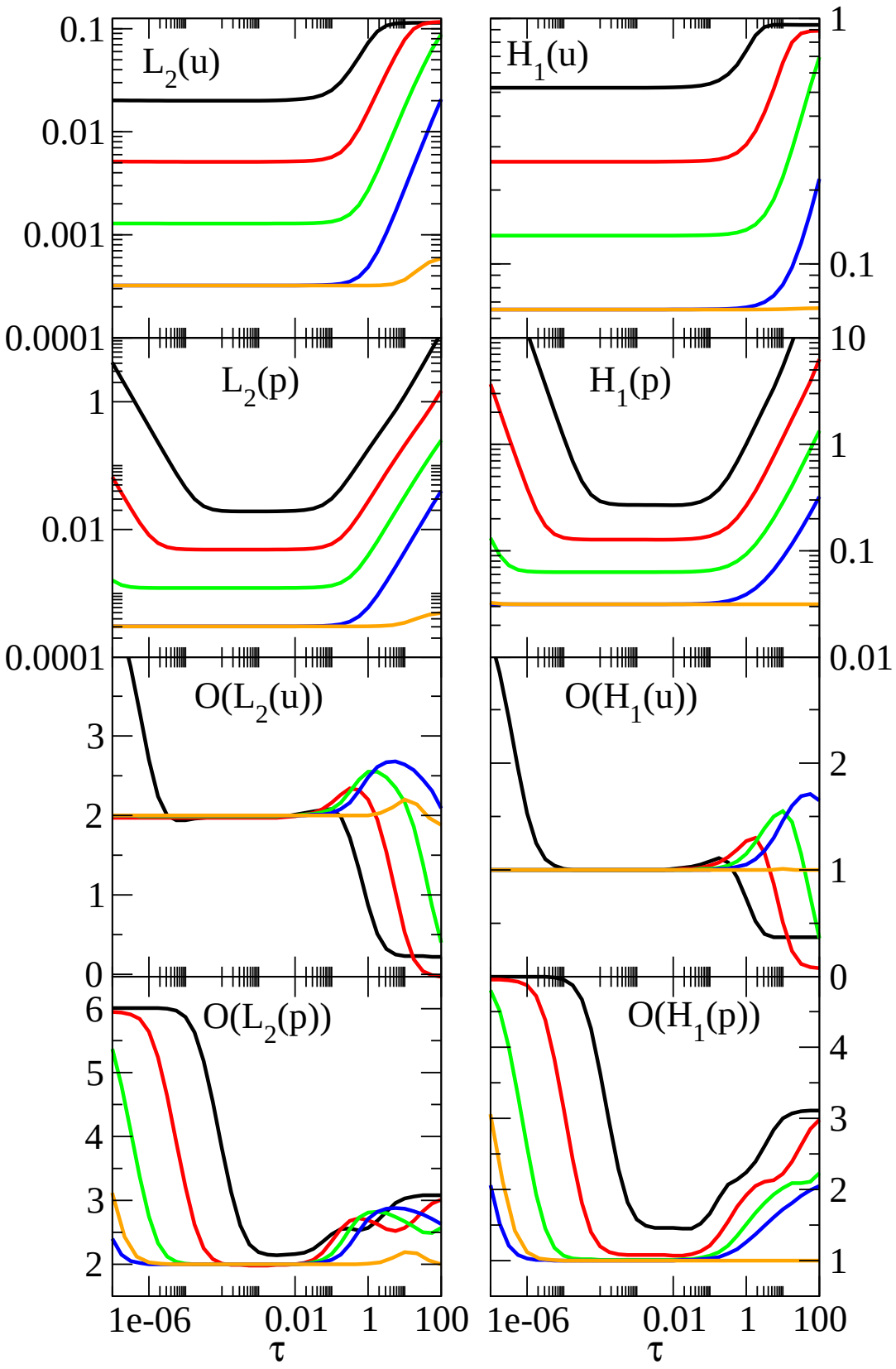

Abbildung 9.3: Fehler und Fehlerordnungen für Finite–Elemente erster Ordnung auf dem Quadratgitter. Für weitere Details siehe die Beschreibung von Abbildung 9.2.

Abbildung 9.4 zeigt die Resultate für Finite–Elemente zweiter Ordnung auf dem Dreiecksgitter. Auf Level 5 erreicht die CIP–Methode fast im gesamten Parameterbereich die von der Analysis vorausgesagten Raten mit Ausnahme der Rate des H^1–Geschwindigkeitsfehlers, welcher nur in einem kleinen Intervall um $\tau = 1$ größer als die vorausgesagte Ordnung 2 ist. Im Großteil des dargestellten Parameterbereichs steigt die Konvergenzrate des H^1–Fehlers allerdings mit dem Gitterlevel, so dass der H^1–Fehler vermutlich die vorausgesagte Rate auf einem größeren Level erreicht. Zudem besitzt die CIP–Methode sehr gute Raten im Druckfeld. Alle dargestellten Fehlerwerte der CIP–Methode werden etwa bei $\tau = 0.177$ gleichzeitig minimal. Die Fehlerwerte bei $\tau = 1$ besitzen gemäß Tabelle 9.4 dieselbe Größenordnung wie die Werte für eine optimale Wahl von τ. Die residuale Stabilisierung weist gegenüber der CIP–Methode bei einer optimalen Wahl von τ die kleineren Fehlerwerte im Druckfeld auf, während sie im Geschwindigkeitsfeld die größeren Fehler liefert. Ferner ist der Parameterbereich kleiner Fehler bei der residualen Stabilisierung größer als derjenige der CIP–Methode. Dies liegt hauptsächlich daran, dass die CIP–Methode schlechter bei großen τ abschneidet. Dafür besitzt die CIP–Methode auf Level 5 fast überall die besseren Konvergenzraten. Besonders deutlich ist dies an den Raten des Druckfeldes zu erkennen, welche bei der CIP–Methode für ausreichend große τ ungefähr um 1 besser sind.

τ	$L^2(u)$	$H^1(u)$	$L^2(p)$	$H^1(p)$	$\mathcal{O}(L^2(u))$	$\mathcal{O}(H^1(u))$	$\mathcal{O}(L^2(p))$	$\mathcal{O}(H^1(p))$
1.0	4.24e-6	8.17e-4	1.01e-5	3.49e-3	3.25	2.13	4.01	3.01
0.177	2.13e-6	7.79e-4	2.93e-6	1.12e-3	3.18	2.03	3.91	2.84
1.0	4.27e-6	1.15e-3	9.99e-7	5.35e-4	2.77	1.85	3.04	2.00

Tabelle 9.4: Fehler und Ordnungen für Finite–Elemente zweiter Ordnung auf dem Level–5–Dreiecksgitter. Die beiden oberen Zeilen beziehen sich auf die CIP–Methode und die untere auf die residuale Stabilisierung.

Die Fehler für das Quadratgitter sind gemäß Abbildung 9.5 bei beiden Stabilisierungsverfahren kleiner als auf dem Dreiecksgitter. Diese Aussage ist für alle abgebildeten τ und alle Fehlerwerte gültig. Ferner gibt es bei der CIP–Methode ein relativ großes τ–Intervall, in welchem die Fehler fast optimal sind und in welchem die Konvergenzrate des H^1–Fehlers die Voraussage der Fehlerabschätzung erreicht. Zu diesem Bereich gehört auch die Wahl $\tau = 1$. Wie Tabelle 9.5 zeigt, schneiden beide Stabilisierungsverfahren bei einer jeweils optimalen Wahl des Stabilisierungsparameters etwa gleich gut ab. Die CIP–Methode ist dabei leicht vorzuziehen, da sie einen vergleichsweise großen Bereich kleiner Fehler im Geschwindigkeitsfeld besitzt. Außerdem sei bemerkt, dass die CIP–Methode auch auf dem Level–5–Quadratgitter für beinahe alle τ bessere Konvergenzraten als die residuale Stabilisierung besitzt.

τ	$L^2(u)$	$H^1(u)$	$L^2(p)$	$H^1(p)$	$\mathcal{O}(L^2(u))$	$\mathcal{O}(H^1(u))$	$\mathcal{O}(L^2(p))$	$\mathcal{O}(H^1(p))$
1.0	1.07e-6	4.17e-4	7.01e-7	2.32e-4	3.06	2.01	3.70	2.50
1.77	1.01e-6	4.15e-4	5.25e-7	2.11e-4	3.03	2.00	3.16	2.10
1.0	1.20e-6	4.63e-4	5.01e-7	2.11e-4	2.98	1.99	3.01	2.01

Tabelle 9.5: Fehler und Ordnungen für Finite–Elemente zweiter Ordnung auf dem Level–5–Quadratgitter. Die beiden oberen Zeilen beziehen sich auf die CIP–Methode und die untere auf die residuale Stabilisierung.

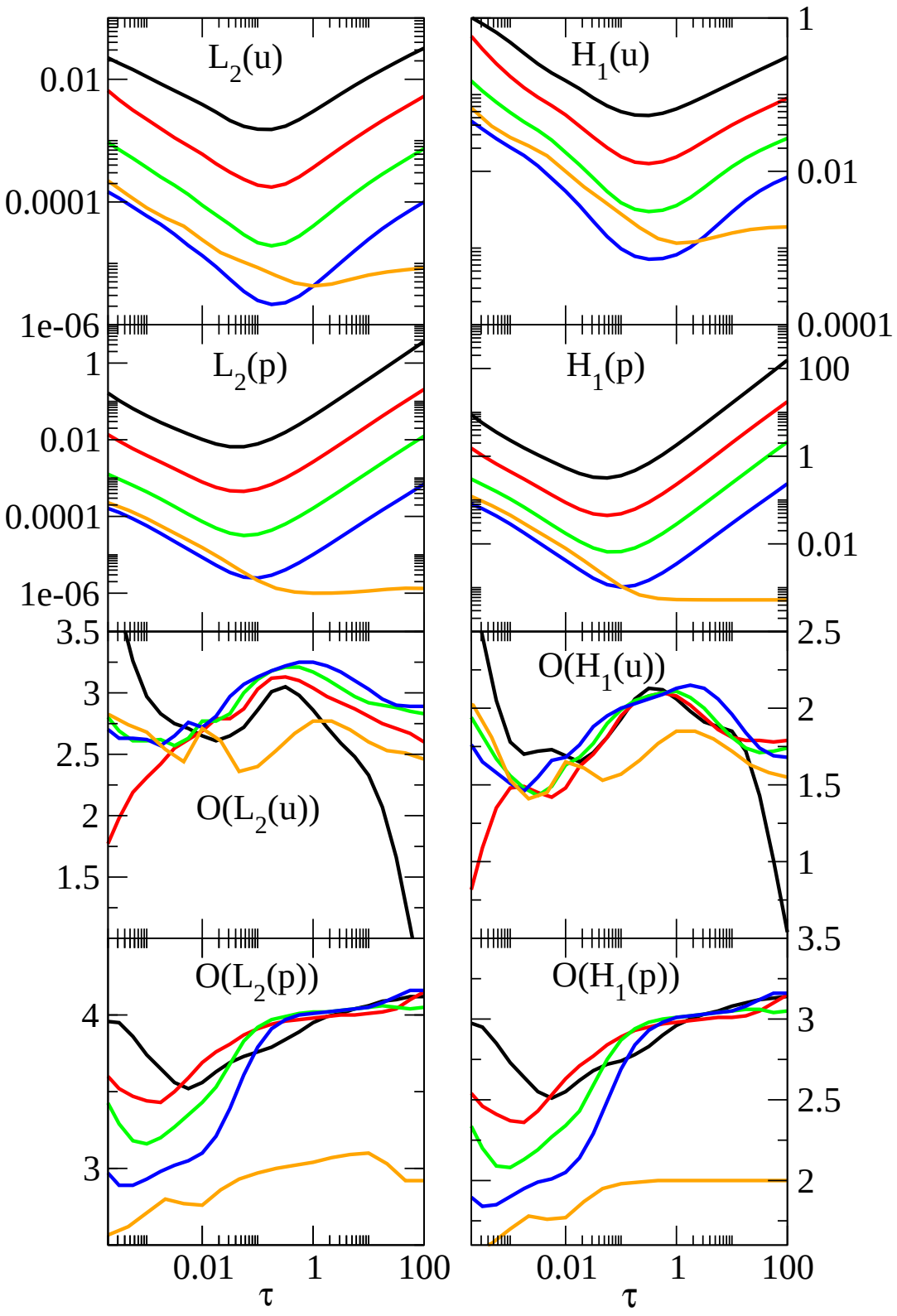

Abbildung 9.4: Fehler und Fehlerordnungen für Finite–Elemente zweiter Ordnung auf dem Dreiecksgitter. Für weitere Details siehe die Beschreibung von Abbildung 9.2.

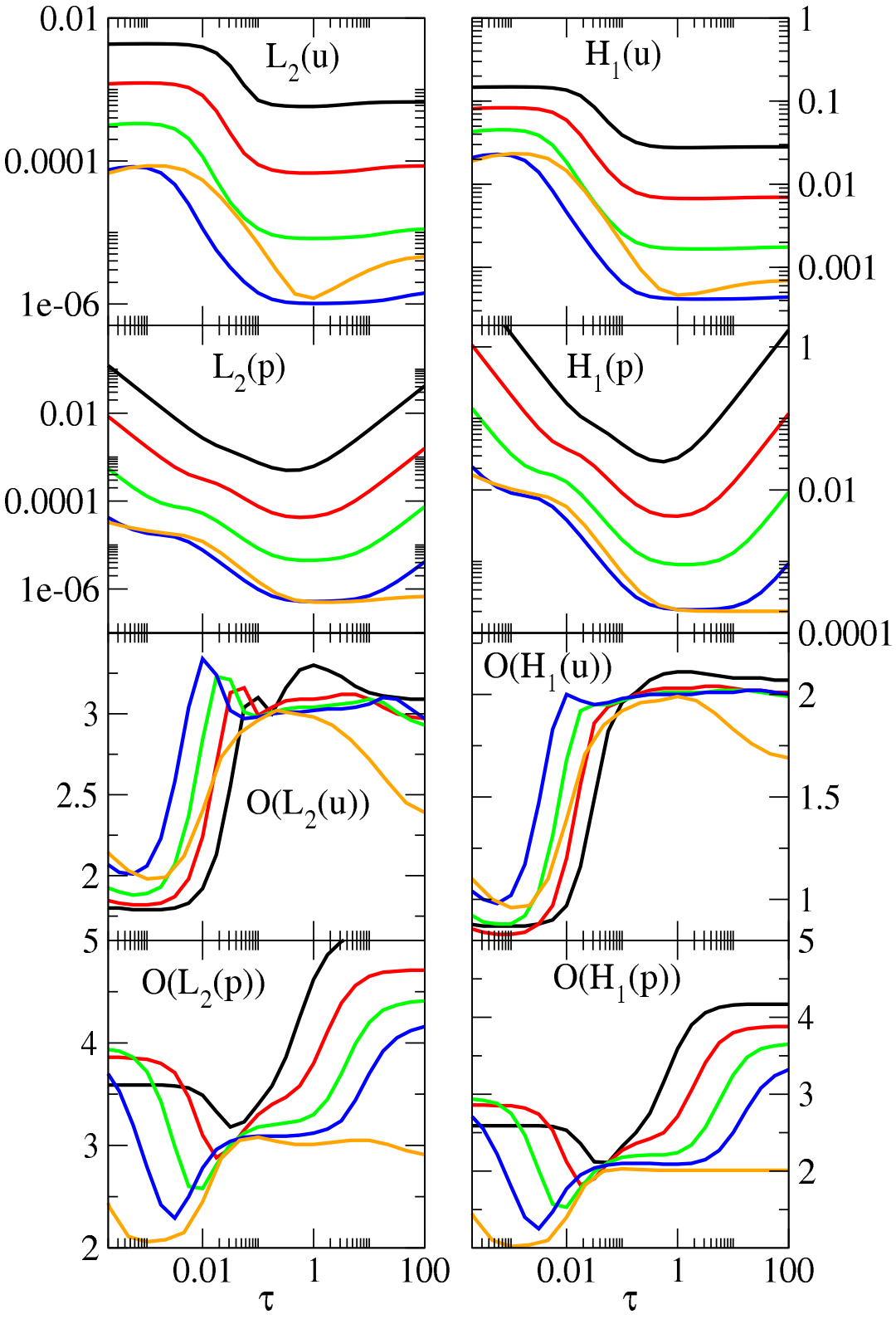

Abbildung 9.5: Fehler und Fehlerordnungen für Finite–Elemente zweiter Ordnung auf dem Quadratgitter. Für weitere Details siehe die Beschreibung von Abbildung 9.2.

Die Ergebnisse für Finite–Elemente dritter Ordnung auf dem Dreiecksgitter sind in Abbildung 9.6 zu sehen. Die CIP–Methode erreicht die aus der Theorie vorausgesagten Fehlerordnungen in einem Bereich etwa um $\tau = 0.031$. Zudem werden dort alle Fehler etwa gleichzeitig minimal. Bei $\tau = 1$ sind die Fehler schon merklich größer; nach Tabelle 9.6 liegt dort der L^2–Fehler der Geschwindigkeit mehr als eine Größenordnung über dem Fehler bei $\tau = 0.031$. Dennoch erreicht man auch für $\tau = 1$ bereits auf dem Level–4–Gitter beinahe die vorausgesagten Konvergenzraten, wobei die Raten mit feiner werdendem Gitter anwachsen.
Bei jeweils optimaler Wahl von τ besitzt die residuale Stabilisierung etwas kleinere Fehlerwerte als die CIP–Methode. Außerdem bleiben die Fehler der residualen Stabilisierung in einem deutlich größeren τ–Intervall kleiner als die Fehler der CIP–Methode. Die residuale Stabilisierung ist hier also vorzuziehen.

τ	$L^2(u)$	$H^1(u)$	$L^2(p)$	$H^1(p)$	$\mathcal{O}(L^2(u))$	$\mathcal{O}(H^1(u))$	$\mathcal{O}(L^2(p))$	$\mathcal{O}(H^1(p))$
1.0	1.03e-6	1.70e-4	4.66e-7	1.19e-4	3.98	2.91	4.20	3.27
3.16e-2	8.18e-8	3.09e-5	6.77e-8	2.93e-5	4.07	2.98	4.29	3.19
0.56	7.89e-8	2.48e-5	7.57e-8	2.57e-5	4.17	3.05	4.02	3.00
0.1	7.35e-8	2.76e-5	6.52e-8	2.76e-5	4.05	3.01	4.03	3.01

Tabelle 9.6: Fehler und Ordnungen für Finite–Elemente dritter Ordnung auf dem Level–5–Dreiecksgitter. Die beiden oberen Zeilen beziehen sich auf die CIP–Methode und die beiden unteren auf die residuale Stabilisierung.

Abbildung 9.7 zeigt die Resultate für Finite–Elemente dritter Ordnung auf dem Quadratgitter. Hier sind die Fehlerwerte stets kleiner als diejenigen auf dem Dreiecksgitter. Darüber hinaus fällt der Bereich größer aus, in welchem die Fehler fast optimal sind.
Auch auf dem Quadratgitter werden die vorausgesagten Konvergenzordnungen in dem Bereich kleiner Fehler durch die CIP–Methode erreicht. Zudem wächst dieser Bereich mit dem Gitterlevel. Bei der Wahl $\tau = 1$ liegen die Fehler wieder etwa um eine Größenordnung über den Fehlern, welche bei der optimalen Wahl $\tau = 0.0056$ erreicht werden, siehe Tabelle 9.7. Die Konvergenzordnungen für $\tau = 1$ wachsen mit feiner werdendem Gitter und entsprechen auf Level 4 beinahe den Vorhersagen der Fehleranalysis.
Die residuale Stabilisierung liefert über einen weiten Bereich von τ–Werten kleinere Fehler als die CIP–Methode. Außerdem besitzt die residuale Stabilisierung bei optimaler Wahl von τ die kleineren Fehlerwerte. Wie auf den Dreieckselementen ist die residualen Stabilisierung also der CIP–Methode vorzuziehen.

τ	$L^2(u)$	$H^1(u)$	$L^2(p)$	$H^1(p)$	$\mathcal{O}(L^2(u))$	$\mathcal{O}(H^1(u))$	$\mathcal{O}(L^2(p))$	$\mathcal{O}(H^1(p))$
1.0	6.87e-7	8.97e-5	2.87e-7	6.09e-5	3.86	2.89	4.01	3.10
5.62e-3	6.01e-8	1.70e-5	2.51e-8	8.06e-6	4.00	2.99	4.13	3.23
4.64e-2	5.90e-8	1.69e-5	2.57e-8	7.89e-6	4.01	2.99	4.00	3.00
46.4	6.78e-8	1.50e-5	3.04e-8	6.64e-6	4.17	3.09	4.01	3.01

Tabelle 9.7: Fehler und Ordnungen für Finite–Elemente dritter Ordnung auf dem Level–5–Quadratgitter. Die beiden oberen Zeilen beziehen sich auf die CIP–Methode und die beiden unteren auf die residuale Stabilisierung.

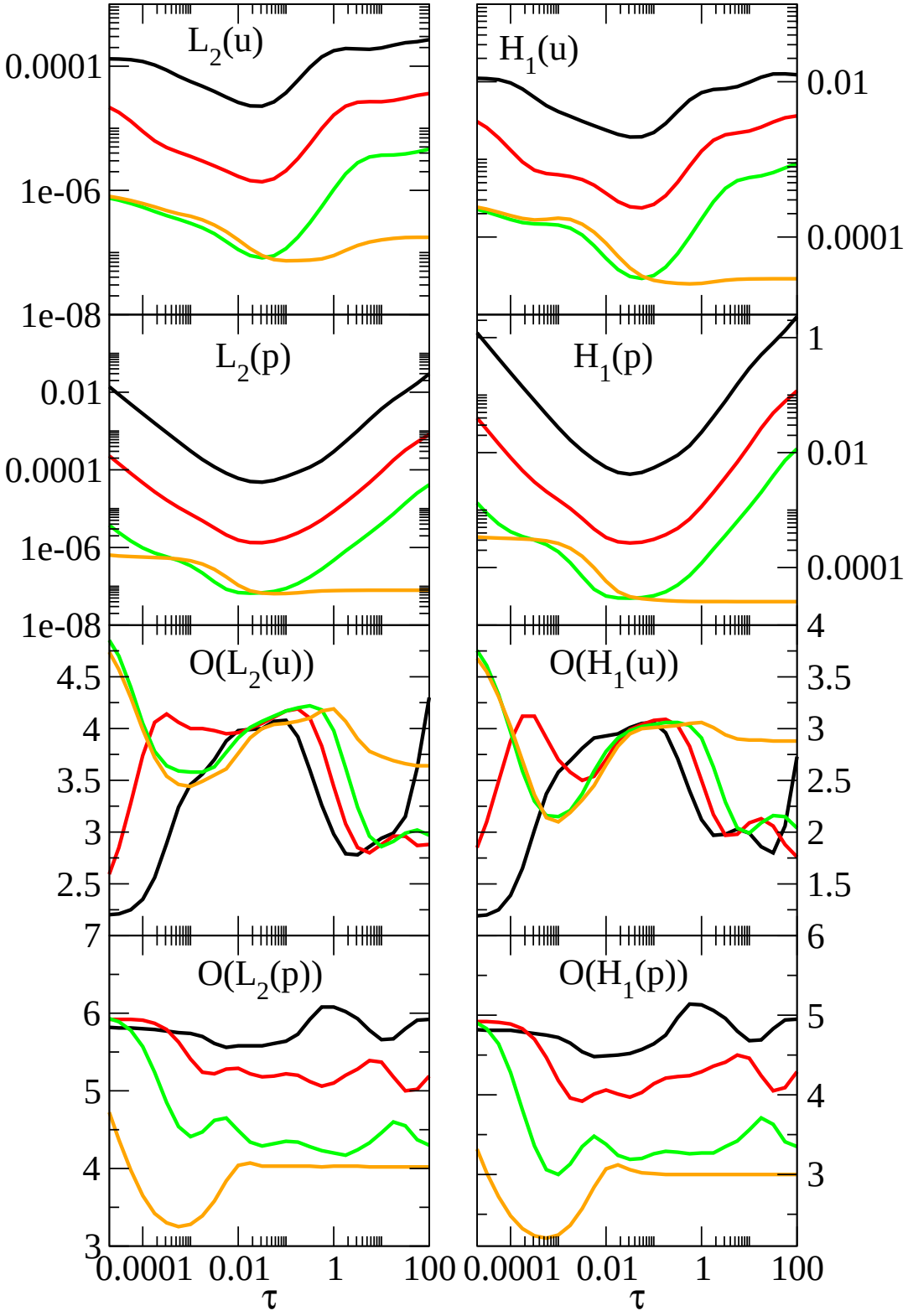

Abbildung 9.6: Fehler und Fehlerordnungen für Finite–Elemente dritter Ordnung auf dem Dreiecksgitter. Für weitere Details siehe die Beschreibung von Abbildung 9.2.

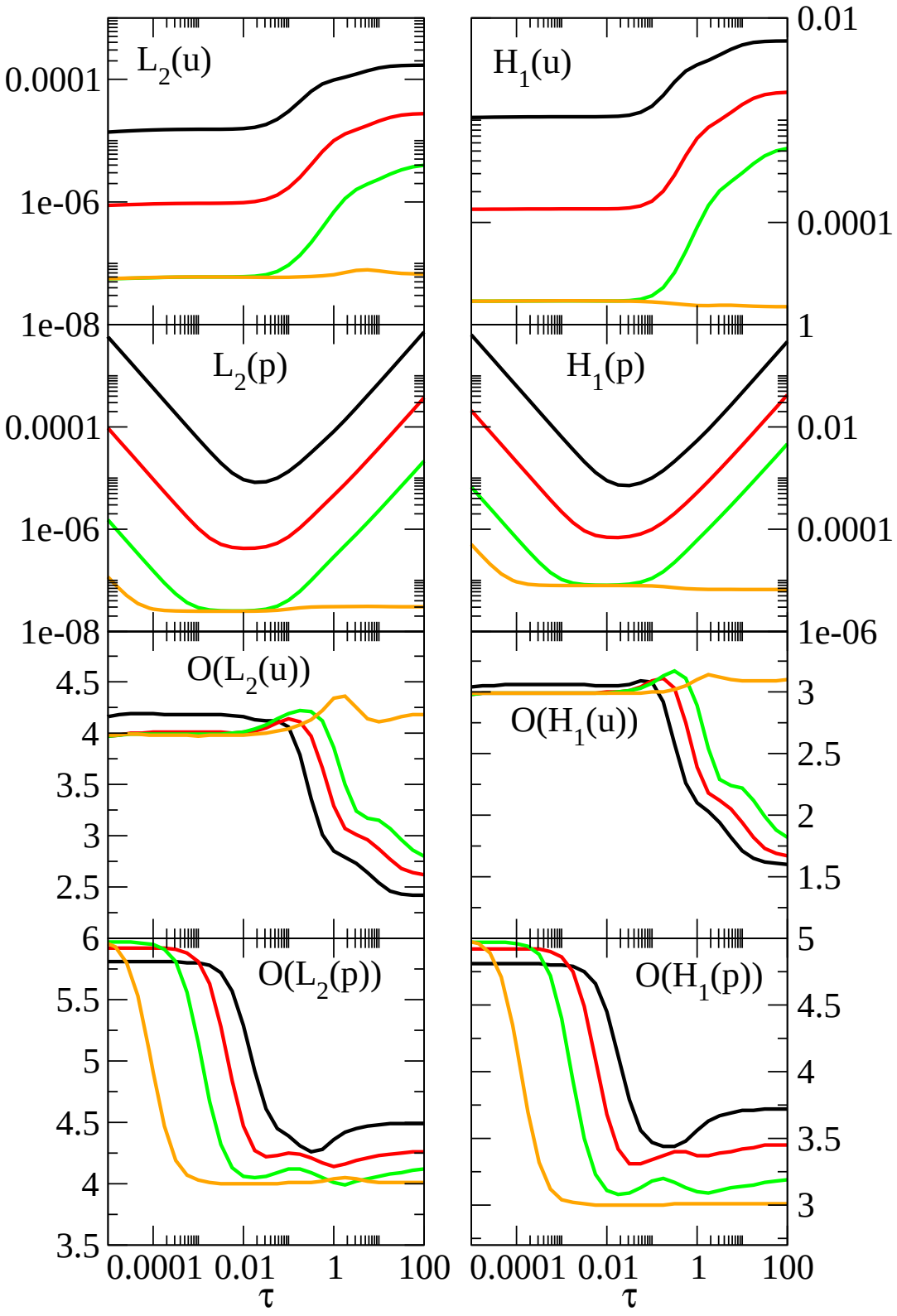

Abbildung 9.7: Fehler und Fehlerordnungen für Finite–Elemente dritter Ordnung auf dem Quadratgitter. Für weitere Details siehe die Beschreibung von Abbildung 9.2.

Bevor wir das nächste Beispiel betrachten, fassen wir unsere bisherigen Ergebnisse kurz zusammen. Die CIP–Methode erreicht bei allen betrachteten Fällen in einem geeignet gewählten τ–Intervall die vorhergesagten Konvergenzraten aus Tabelle 9.1. In dem entsprechenden Intervall werden zugleich die kleinsten Fehlerwerte erreicht. Im Vergleich zu der optimalen Wahl des Stabilisierungsparameter τ liefert die Wahl $\tau = 1$ bei Elementen erster und zweiter Ordnung Fehler in der gleichen Größenordnung. Insbesondere für Elemente zweiter Ordnung ist die Wahl $\tau = 1$ beinahe optimal. Bei Elementen dritter Ordnung sind die Unterschiede hingegen größer. Hier liegen die Fehler bei $\tau = 1$ etwa um eine Größenordnung über den Fehlern bei einer optimalen Wahl von τ. Die aus der Fehleranalysis vorausgesagten Ordnungen werden bei $\tau = 1$ jedoch in allen Fällen näherungsweise erreicht oder sogar übertroffen. Die CIP–Methode liefert im Vergleich zur residualen Stabilisierung für Finite–Elemente erster und zweiter Ordnung etwa gleich gute Ergebnisse. Für Finite–Elemente dritter Ordnung schneidet die residuale Stabilisierung auf dem Level–4–Gitter jedoch besser ab, vor allem weil der Bereich kleiner Fehler deutlich größer für die residuale Stabilisierung ausfällt. Die optimale Wahl von τ ist bei beiden Stabilisierungsverfahren davon abhängig, ob auf einem Quadrat- oder Dreiecksgitter gerechnet wird, und davon, welche Elementordnung betrachtet wird. Die optimalen τ–Werte können dabei von Fall zu Fall um mehrere Größenordnungen auseinanderliegen. Ferner sei noch bemerkt, dass die residuale Stabilisierung bei großen Stabilisierungsparametern stets kleinere Fehler im Druckfeld als die CIP–Methode aufweist. Das Quadratgitter ist in allen Fällen dem Dreiecksgitter vorzuziehen: Die Fehler sind dort für beide Stabilisierungsverfahren und für alle Gitterlevel kleiner und das τ–Intervall guter Ergebnisse größer.

9.3 Das polynomiale Beispiel

Die gleichen Rechnungen wie im letzten Abschnitt führen wir nun für ein anderes Beispiel durch. Unter anderem stellt sich die Frage, welche der Erkenntnisse für das sincos–Beispiel auf andere Beispiele übertragbar sind. Das Beispiel besitzt die folgende polynomiale Lösung

$$\begin{aligned} u_x &= 2x^2(1-x)^2y(1-y)(1-2y), \\ u_y &= -2y^2(1-y)^2x(1-x)(1-2x), \\ p &= x^3 + y^3 - 0.5. \end{aligned} \tag{9.9}$$

Als Vorgaben wählen wir $b = u$, $c = 0$ und $\nu = 10^{-6}$. Zu diesen Vorgaben bestimmen wir die rechte Seite der Oseen–Gleichung sowie die Randbedingungen analog zum sincos–Beispiel, so dass (9.9) die Oseen–Gleichung (8.1) löst.
Das Geschwindigkeits- und Druckfeld der Lösung ist in den Abbildungen 9.8 und 9.9 zu sehen. Wie dort zu erkennen ist, zeichnet sich das polynomiale Beispiel unter anderem dadurch aus, dass das Druckfeld im Großteil des Lösungsgebietes betragsmäßig deutlich größer als das Geschwindigkeitsfeld ist. Mit anderen Worten: Das Geschwindigkeitsfeld wird durch das Druckfeld dominiert.
Da die Lösung des Beispiels ein Polynom ist, liegt die Lösung für Finite–Elemente ausreichend großer Elementordnung im Ansatzraum und sollte demnach bis auf Rundungsfehler exakt berechnet werden können. Eine ausreichend große Ordnung für den Druckraum wird durch Finite–Elemente dritter Ordnung erreicht, eine ausreichend große Ordnung im Geschwindigkeitsraum entweder durch Elemente vierter Ordnung auf dem Quadratgitter oder durch Elemente siebter Ordnung auf dem Dreiecksgitter.

Bei unseren Rechnungen betrachten wir im Wesentlichen Gitter bis Level 5, welche eine Gitterweite von $(0.5)^5$ besitzen. Man vergewissert sich wie für das sincos–Beispiel, dass bei obigen Vorgaben die Relation $\nu \geq h\|b\|_{L^\infty(\Omega)}$ erfüllt ist. Daher liegt der konvektionsdominante Fall vor und die von der Analysis vorausgesagten Konvergenzraten können erneut Tabelle 9.1 entnommen werden.

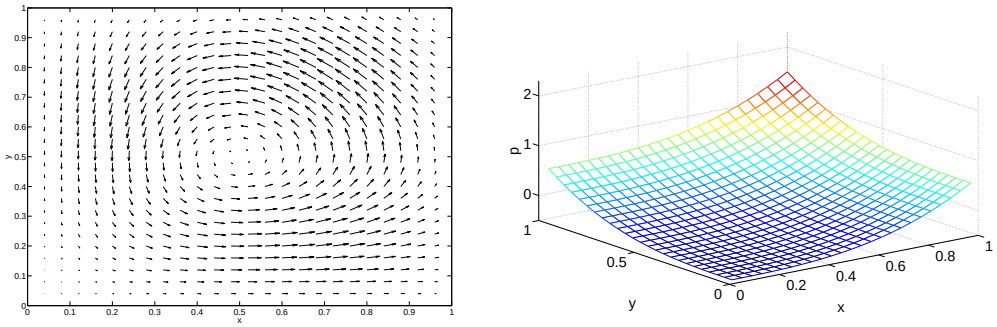

Abbildung 9.8: Links Vektorplot für das Geschwindigkeitsfeld des polynomialen Beispiels, rechts das zugehörige Druckfeld.

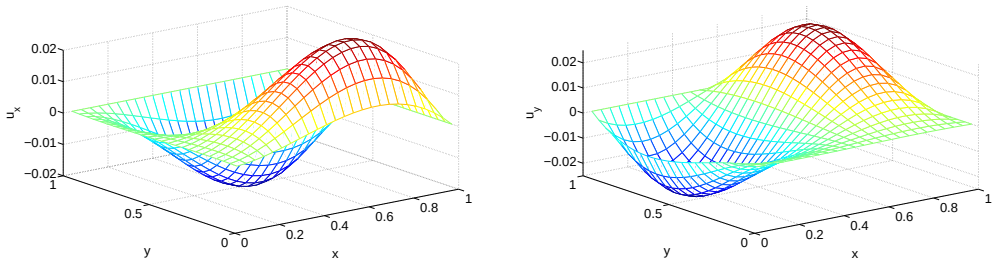

Abbildung 9.9: Links die x–Komponente des Geschwindigkeitsfeld, rechts die y–Komponente.

Abbildung 9.10 zeigt die Fehler und Fehlerordnungen für Finite–Elemente erster Ordnung auf Dreiecken. Im Vergleich zum sincos–Beispiel werden hier die Fehler erst bei einem kleinen Stabilisierungsparameter τ gut. Für ausreichend kleine τ erreichen sowohl die CIP–Methode als auch die residuale Stabilisierung die aus der Analysis vorausgesagten Fehlerordnungen. Bei der Wahl $\tau = 1$, wie sie in [BBJL07] für die CIP–Methode vorgeschlagen wird, sind die Fehler jedoch deutlich größer und auch die Ordnung ist schlecht, siehe hierzu insbesondere Tabelle 9.8. Leider werden in [BBJL07] keine numerischen Ergebnisse präsentiert, so dass nicht klar wird, für welche Fälle der Parametervorschlag getestet worden ist. In dem vorliegenden Fall ist der Vorschlag jedenfalls weit von dem optimalen Wert entfernt.
Für kleine τ besitzen die residuale Stabilisierung und die CIP–Methode beinahe identische Fehlerwerte. Mit zunehmendem τ wachsen die Fehler der residualen Stabilisierung aber langsamer. Somit ist die residuale Stabilisierung hier leicht vorzuziehen.
Abbildung 9.10 vermittelt den Eindruck, dass die Fehlerwerte für ausreichend kleine Stabilisierungsparameter beim Übergang $\tau \to 0$ konstant bleiben. Tatsächlich ist dies noch für viel kleinere τ–Werte als in der Abbildung dargestellt der Fall. Dennoch darf nicht ganz auf eine Stabilisierung verzichtet werden. So erhält man mit dem Standard–Galerkin–Verfahren, welches einer Wahl $\tau = 0$ entspricht, keine sinnvollen Ergebnisse.

τ	$L^2(u)$	$H^1(u)$	$L^2(p)$	$H^1(p)$	$\mathcal{O}(L^2(u))$	$\mathcal{O}(H^1(u))$	$\mathcal{O}(L^2(p))$	$\mathcal{O}(H^1(p))$
1.0	4.87e-1	3.28e+1	9.38e-4	3.15e-2	0.70	-0.33	1.21	1.31
1.0e-8	1.71e-5	3.14e-3	5.76e-5	2.23e-2	2.71	1.04	2.29	1.11
1.0e-8	1.70e-5	3.14e-3	5.58e-5	2.21e-2	2.71	1.04	2.05	1.02

Tabelle 9.8: Fehler und Ordnungen für Finite–Elemente erster Ordnung auf dem Level–5–Dreiecksgitter. Die beiden oberen Zeilen beziehen sich auf die CIP–Methode und die untere auf die residuale Stabilisierung.

Die Ergebnisse auf dem Quadratgitter für Elemente erster Ordnung ähneln gemäß Abbildung 9.11 qualitativ den Resultaten auf dem Dreiecksgitter. Die Fehlerwerte für optimal gewähltes τ sind hier etwas besser als diejenigen des Dreiecksgitters, siehe hierzu Tabelle 9.9.

τ	$L^2(u)$	$H^1(u)$	$L^2(p)$	$H^1(p)$	$\mathcal{O}(L^2(u))$	$\mathcal{O}(H^1(u))$	$\mathcal{O}(L^2(p))$	$\mathcal{O}(H^1(p))$
1.0	7.24e-1	4.48e+1	2.14e-3	3.79e-2	1.19	0.27	0.79	1.13
1.0e-8	8.14e-6	2.14e-3	5.27e-5	2.29e-2	2.35	1.00	2.08	1.02
1.0e-8	8.15e-6	2.14e-3	4.46e-5	2.21e-2	2.35	1.00	2.02	1.00

Tabelle 9.9: Fehler und Ordnungen für Finite–Elemente erster Ordnung auf dem Level–5–Quadratgitter. Die beiden oberen Zeilen beziehen sich auf die CIP–Methode und die untere auf die residuale Stabilisierung.

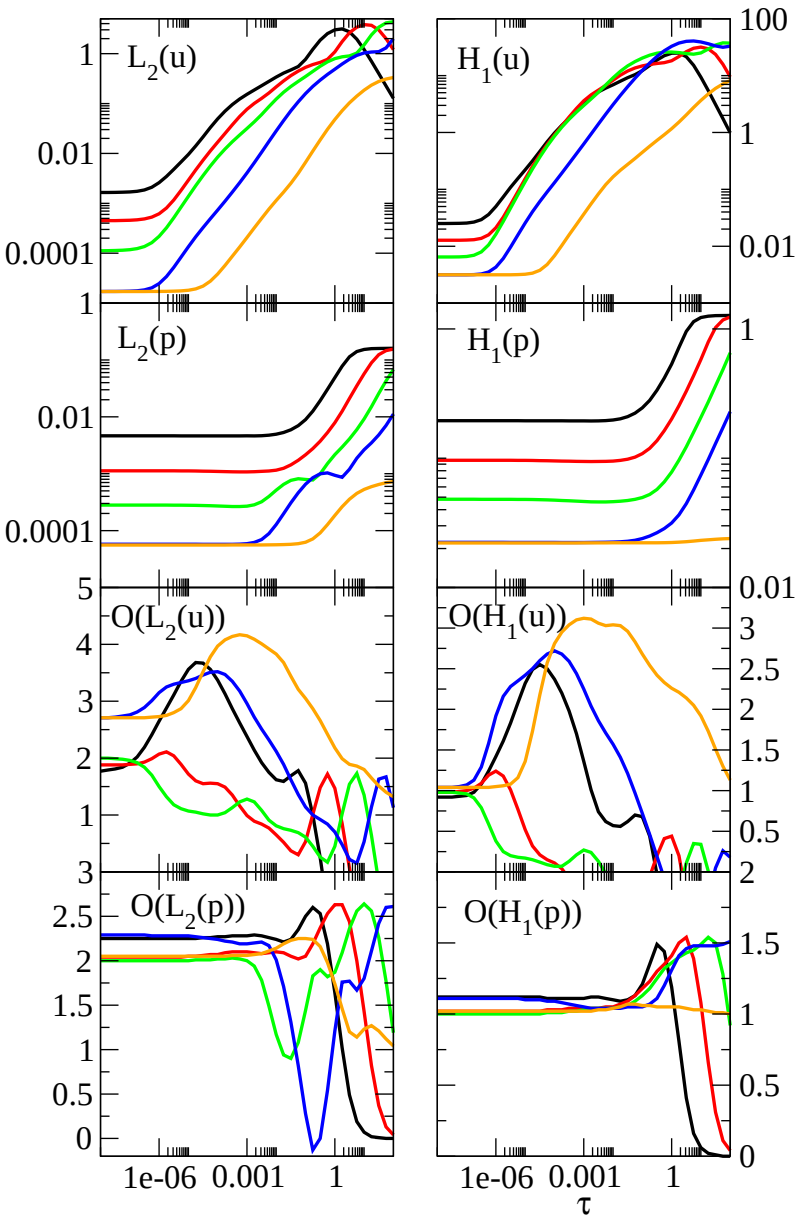

Abbildung 9.10: Fehler und Fehlerordnungen für Finite–Elemente erster Ordnung auf dem Dreiecksgitter. Für weitere Details siehe die Beschreibung von Abbildung 9.2.

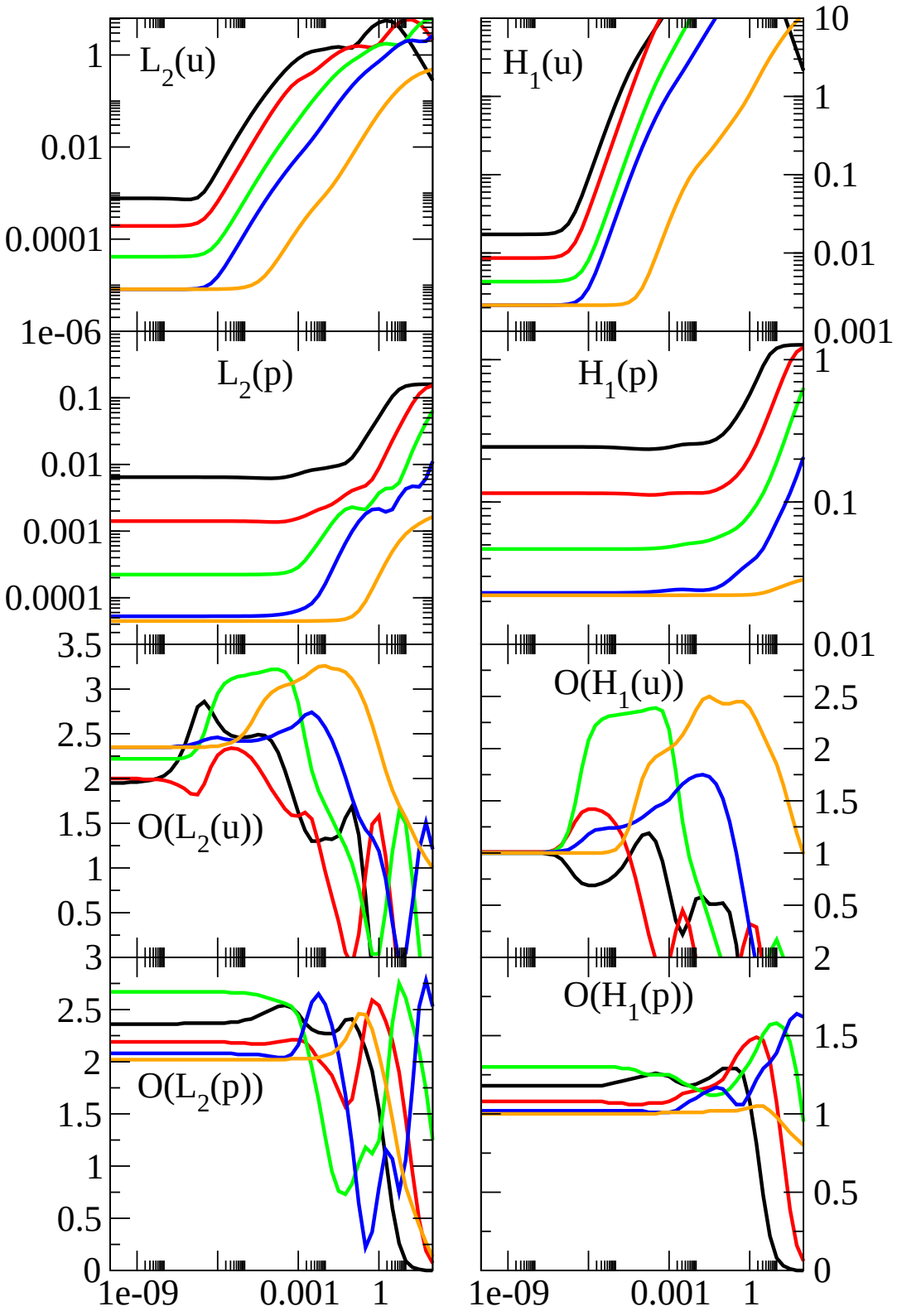

Abbildung 9.11: Fehler und Fehlerordnungen für Finite–Elemente erster Ordnung auf dem Quadratgitter. Für weitere Details siehe die Beschreibung von Abbildung 9.2.

Wie Abbildung 9.12 für das Dreiecksgitter zeigt, unterscheidet sich die Situation bei Finiten–Elementen zweiter Ordnung deutlich von derjenigen bei Elementen erster Ordnung. Die guten Fehlerwerte werden hier gerade bei relativ großen τ–Werten erreicht. Bei der CIP–Methode nehmen der L^2–Fehler und der H^1–Fehler ihr Minimum versetzt an, nämlich ersterer bei $\tau \approx 10$ und letzterer bei $\tau \approx 56$. Die Wahl $\tau = 1$ ist hiervon nicht weit entfernt und ergibt Fehlerwerte in der gleich Größenordnung, siehe hierzu auch Tabelle 9.10. Zudem ist zu beobachten, dass die Minima der Fehlerwerte im Geschwindigkeitsfeld mit feiner werdendem Gitter zu größeren τ verschoben werden. Für das Druckfeld wächst der Bereich fast optimaler Stabilisierungsparameter mit zunehmender Gitterverfeinerung, so dass für ein geeignet gewähltes konstantes τ auf allen Gittern fast optimale Fehlerwerte erreicht werden können. Die durch die Fehleranalysis vorausgesagten Ordnungen werden für das Druckfeld in einem großen τ–Intervall auf allen Gitterleveln erreicht und im Geschwindigkeitsfeld für genügend große τ in der Abbildung. Im Vergleich zur residualen Stabilisierung schneidet die CIP–Methode diesmal etwas besser ab. Die Druckfehler sind bei jeweils optimaler Wahl von τ zwar für beide Verfahren fast identisch, die CIP–Methode besitzt jedoch die kleineren Fehler im Geschwindigkeitsfeld.

τ	$L^2(u)$	$H^1(u)$	$L^2(p)$	$H^1(p)$	$\mathcal{O}(L^2(u))$	$\mathcal{O}(H^1(u))$	$\mathcal{O}(L^2(p))$	$\mathcal{O}(H^1(p))$
1.0	1.64e-5	6.79e-3	1.63e-7	7.74e-5	2.45	1.29	2.99	2.01
10	6.34e-6	1.53e-3	1.66e-7	7.74e-5	4.52	1.02	3.19	2.00
56.2	2.53e-5	4.87e-4	2.09e-7	7.75e-5	4.74	3.53	4.24	2.01
1.77	2.17e-5	4.48e-3	1.66e-7	7.75e-5	3.42	2.79	3.03	2.01

Tabelle 9.10: Fehler und Ordnungen für Finite–Elemente zweiter Ordnung auf dem Level–5–Dreiecksgitter. Die ersten drei Zeilen beziehen sich auf die CIP–Methode und die letzte auf die residuale Stabilisierung.

Die Ergebnisse für Finite–Elemente zweiter Ordnung auf dem Quadratgitter sind in Abbildung 9.13 zu sehen. Erneut sind die Fehler auf dem Quadratgitter kleiner als die Fehler auf dem Dreiecksgitter. Die besten Fehlerwerte werden wie für das Dreiecksgitter für ausreichend große τ angenommen. In diesem Bereich nehmen der L^2– und H^1–Druckfehler der CIP–Methode ihre an den Interpolationsfehler–Abschätzungen gemessenen optimalen Werte an. Im Gegensatz hierzu liegt die Konvergenzrate der Geschwindigkeitsfehler bei großen τ unterhalb der Voraussage der Fehlerabschätzung. Die vorausgesagte Rate wird nur für die kleineren τ–Werte erreicht, bei welchen die Fehler im Geschwindigkeitsfeld größer sind. Ein solches Verhalten ist bei der residuale Stabilisierung nicht zu beobachten. Im τ–Intervall kleiner Fehler sind hier auch die Konvergenzraten gut. Wiederum für die CIP–Methode spricht, dass die Fehler im Geschwindigkeitsfeld für große τ um beinahe zwei Größenordnungen unterhalb der Fehler liegen, welche von der residualen Stabilisierung bei einer optimalen Wahl von τ erreicht werden, siehe hierzu Tabelle 9.11.
Für die CIP–Methode scheint eine Wahl für τ optimal zu sein, welche oberhalb des dargestellten τ–Intervalls liegt. Für noch größere τ kommt es aber zu numerischen Problemen. Die Zeit zur Lösung des linearen Gleichungssystem verdreifacht sich etwa, wenn τ von 100 auf 1000 erhöht wird. Zudem kommt es bei Variation von τ zu starken Schwankungen in den Konvergenzraten der Geschwindigkeitsfehler. Hierbei werden die Raten teilweise negativ, das heißt, der Fehler wird von Level 4 zu Level 5 wieder größer. Daher verzichten wir hier auf die Angabe einer optimalen Wahl von τ. Auf obige Probleme kommen wir in Abschnitt 9.5 zurück.

τ	$L^2(u)$	$H^1(u)$	$L^2(p)$	$H^1(p)$	$\mathcal{O}(L^2(u))$	$\mathcal{O}(H^1(u))$	$\mathcal{O}(L^2(p))$	$\mathcal{O}(H^1(p))$
1.0	1.75e-5	7.30e-3	1.85e-7	7.74e-5	2.56	1.42	2.99	2.01
393	4.92e-7	9.74e-5	1.872-07	7.76e-5	2.23	0.94	2.99	2.00
3.16	2.45e-5	3.73e-3	1.66e-7	7.73e-5	3.36	2.99	3.04	2.01
1.77	2.17e-5	4.48e-3	1.66e-7	7.75e-5	3.42	2.79	3.03	2.01

Tabelle 9.11: Fehler und Ordnungen für Finite–Elemente zweiter Ordnung auf dem Level–5–Quadratgitter. Die ersten beiden Zeilen beziehen sich auf die CIP–Methode und die beiden letzten auf die residuale Stabilisierung.

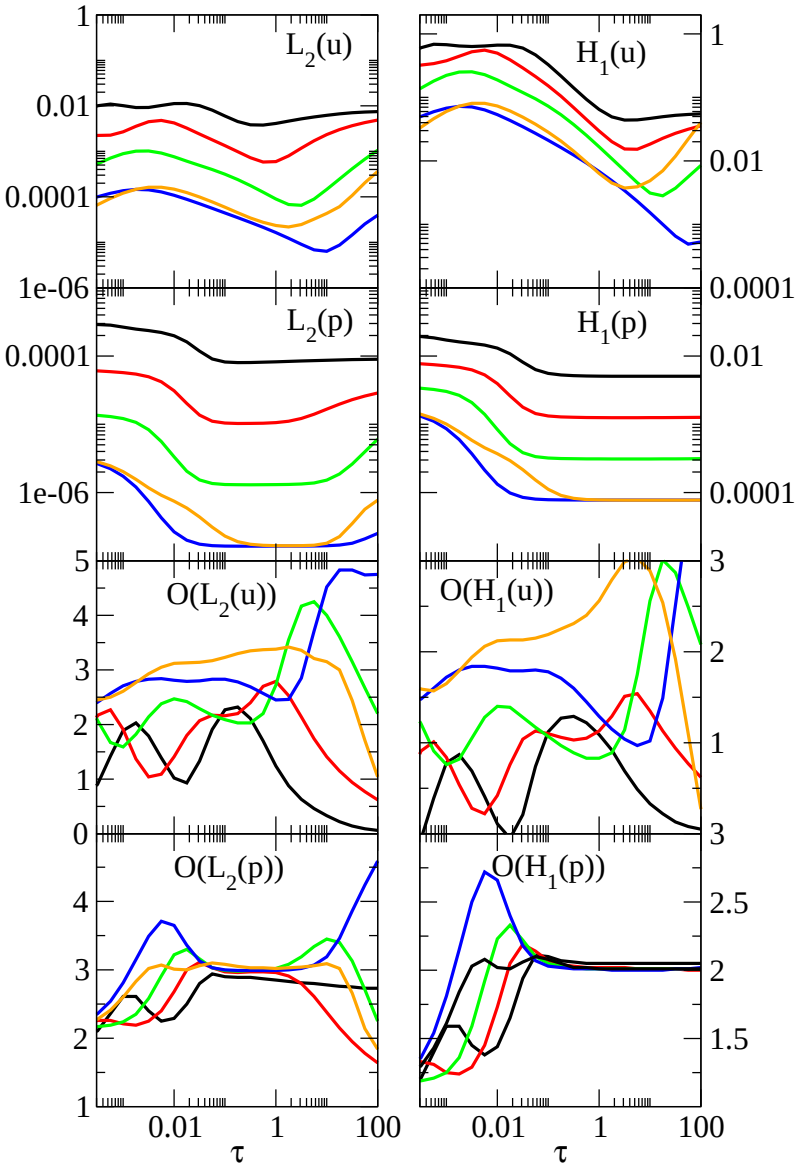

Abbildung 9.12: Fehler und Fehlerordnungen für Finite–Elemente zweiter Ordnung auf dem Dreiecksgitter. Für weitere Details siehe die Beschreibung von Abbildung 9.2.

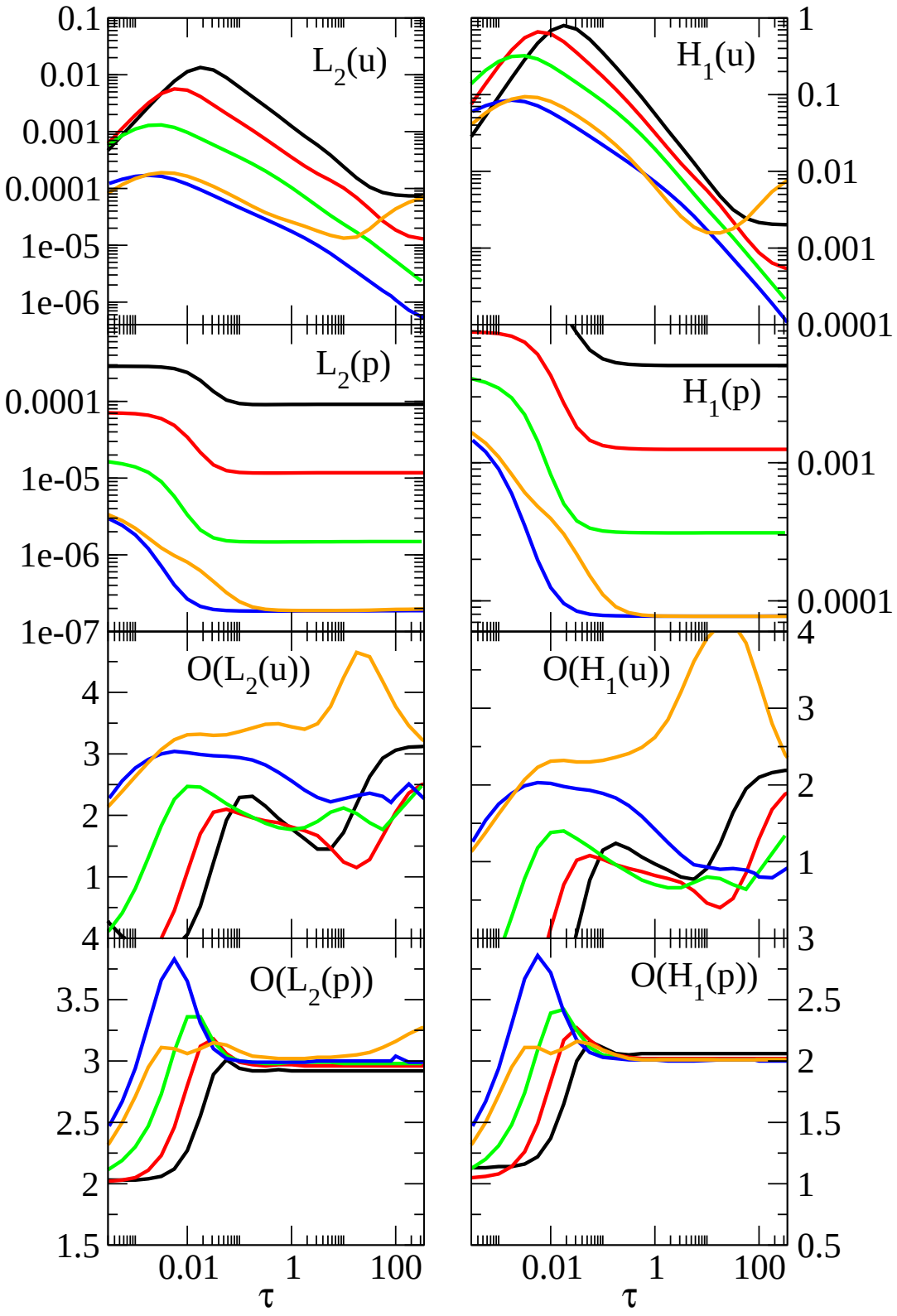

Abbildung 9.13: Fehler und Fehlerordnungen für Finite–Elemente zweiter Ordnung auf dem Quadratgitter. Für weitere Details siehe die Beschreibung von Abbildung 9.2.

Abbildung 9.14 zeigt die Ergebnisse für Finite–Elemente dritter Ordnung auf dem Dreiecksgitter. Die von der Analysis vorausgesagten Konvergenzraten werden von der CIP–Methode auf dem Level–4–Gitter außer für zu große τ erreicht und werden teilweise sogar deutlich überboten. Bei zu großen τ kommt es hingegen ebenso wie bei der residualen Stabilisierung zu numerischen Problemen. Die Fehler der CIP–Methode werden bei $\tau = 0.0177$ etwa gleichzeitig minimal. Die Fehlerwerte sind nach Tabelle 9.12 um eine Größenordnung kleiner als die Werte bei der Wahl $\tau = 1$. Die vorgegebenen Konvergenzraten werden jedoch auch bei $\tau = 1$ realisiert.
Die Fehler der residualen Stabilisierung sind etwa vergleichbar mit denen der CIP–Methode. Während die residuale Stabilisierung gemäß Tabelle 9.12 die kleineren Werte im H^1–Fehler des Drucks besitzt, besitzt die CIP–Methode die kleineren Werte in den anderen Fehlerwerten. Auch der Bereich beinahe optimaler Stabilisierungsparameter τ ist in beiden Fällen ungefähr gleich groß.
Das Druckfeld der Lösung besitzt die Form $x^3 + y^3 - 0.5$. Als Polynom dritten Grades liegt das Druckfeld somit im Ansatzraum für den Druck. Dennoch erkennt man im Druckfeld auf den niedrigeren Leveln noch relativ große Fehler. Das Druckfeld wird aber in Abhängigkeit von dem Geschwindigkeitsfeld berechnet, dessen Lösung nicht im Ansatzraum des Verfahrens liegt. Die Fehler, die im Geschwindigkeitsfeld gemacht werden, übertragen sich somit teilweise auf das Druckfeld.

τ	$L^2(u)$	$H^1(u)$	$L^2(p)$	$H^1(p)$	$\mathcal{O}(L^2(u))$	$\mathcal{O}(H^1(u))$	$\mathcal{O}(L^2(p))$	$\mathcal{O}(H^1(p))$
1.0	4.38e-8	1.01e-5	4.45e-10	1.24e-7	5.50	3.66	4.24	2.54
1.77e-2	9.59e-9	4.16e-6	1.73e-11	5.73e-9	4.41	3.16	4.81	3.71
1.77	2.36e-8	4.66e-6	3.38e-11	1.25e-9	4.62	3.37	5.78	3.91

Tabelle 9.12: Fehler und Ordnungen für Finite–Elemente dritter Ordnung auf dem Level–5–Dreiecksgitter. Die ersten beiden Zeilen beziehen sich auf die CIP–Methode und die beiden letzten auf die residuale Stabilisierung.

Die Resultate für Finite–Elemente dritter Ordnung auf Quadraten sind in Abbildung 9.15 zu sehen. Die Fehler beider Stabilisierungsverfahren sind kleiner als bei den Rechnungen auf dem Dreiecksgitter. Die CIP–Methode erreicht die von der Analysis vorausgesagten Konvergenzraten in einem relativ großen τ–Intervall, wobei die Ordnung insbesondere im L^2–Fehler der Geschwindigkeit deutlich größer ausfällt. Die Wahl $\tau = 1$ liefert ebenfalls die vorausgesagten Konvergenzraten. Die Fehlerwerte im Druckfeld sind jedoch gemäß Tabelle 9.13 bei $\tau \approx 0.0177$ um mindestens eine Größenordnung kleiner.
Beide Stabilisierungsverfahren berechnen die Lösung recht genau. Die Fehlerwerte bei einer jeweils optimalen Wahl des Stabilisierungsparameters sind bei der residualen Stabilisierung etwas kleiner. Auffällig sind die Oszillationen des L^2–Druckfehlers mit τ, wenn die residuale Stabilisierung verwandt wird. Da die Fehler hier besonders klein sind, handelt es sich vermutlich um Rundungsfehler.

Wie oben für die Dreiecksgitter plausibel gemacht, beobachten wir auch hier relativ große Fehler im Druckfeld auf den niedrigeren Leveln, da das Geschwindigkeitsfeld noch nicht im Ansatzraum der Verfahren liegt. Bei Finite–Elementen vierter Ordnung auf dem Quadratgitter liegt die Lösung jedoch im Ansatzraum der Verfahren. Erwartungsgemäß sind hier beide Methode bis auf Rundungsfehler exakt. So besitzen alle Fehler bereits auf dem Level–2–Gitter Werte kleiner gleich 10^{-12}.

τ	$L^2(u)$	$H^1(u)$	$L^2(p)$	$H^1(p)$	$\mathcal{O}(L^2(u))$	$\mathcal{O}(H^1(u))$	$\mathcal{O}(L^2(p))$	$\mathcal{O}(H^1(p))$
1.0	5.24e-9	1.21e-6	1.13e-10	2.38e-8	6.81	4.08	4.68	2.85
1.77e-2	9.56e-10	5.53e-7	3.33e-12	9.51e-10	5.45	3.10	5.75	3.71
0.01	9.09e-10	5.37e-7	1.24e-12	3.76e-10	4.57	3.02	5.86	3.62

Tabelle 9.13: Fehler und Ordnungen für Finite–Elemente dritter Ordnung auf dem Level–5–Quadratgitter. Die ersten beiden Zeilen beziehen sich auf die CIP–Methode und die letzte auf die residuale Stabilisierung.

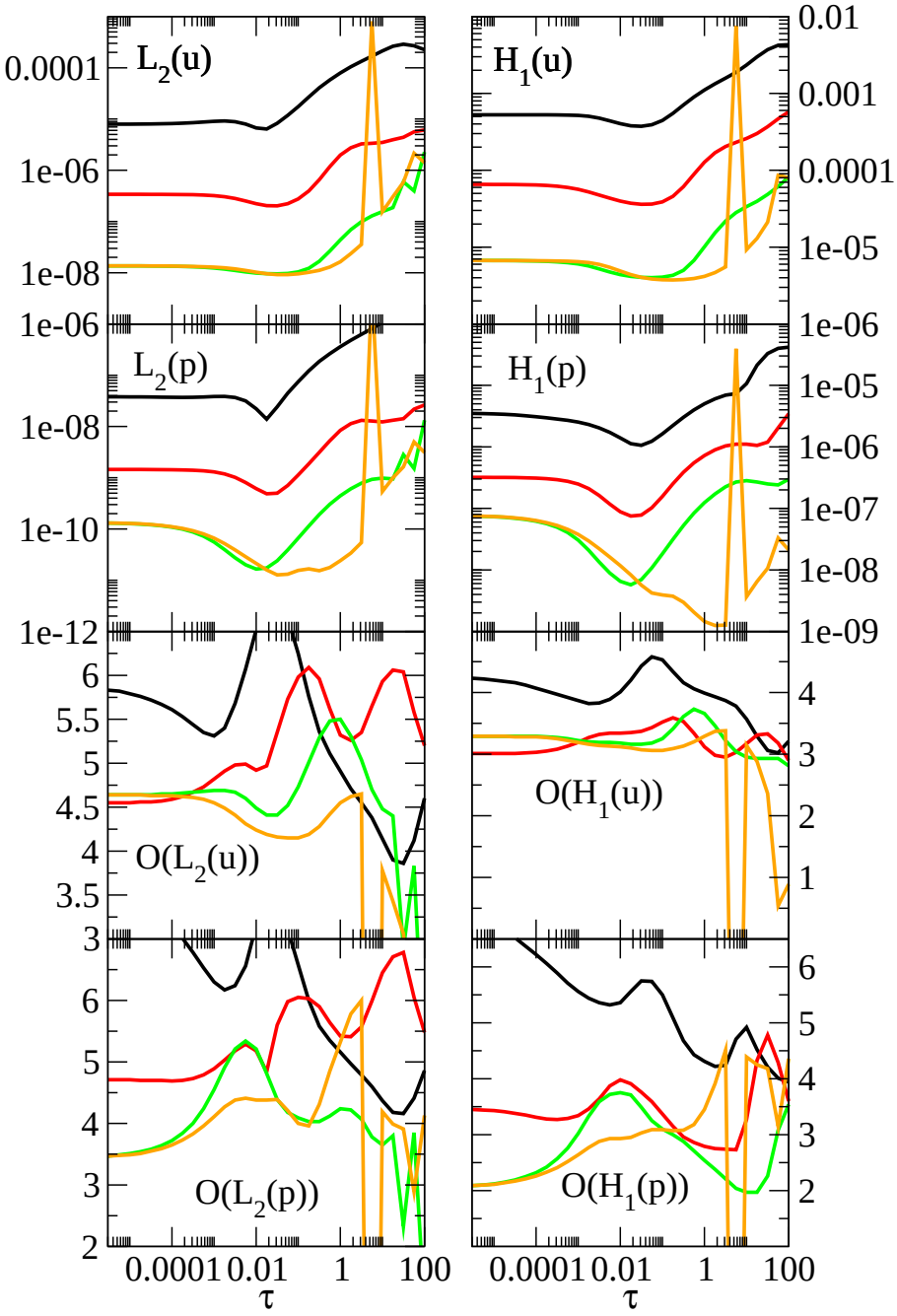

Abbildung 9.14: Fehler und Fehlerordnungen für Finite–Elemente dritter Ordnung auf dem Dreiecksgitter. Für weitere Details siehe die Beschreibung von Abbildung 9.2.

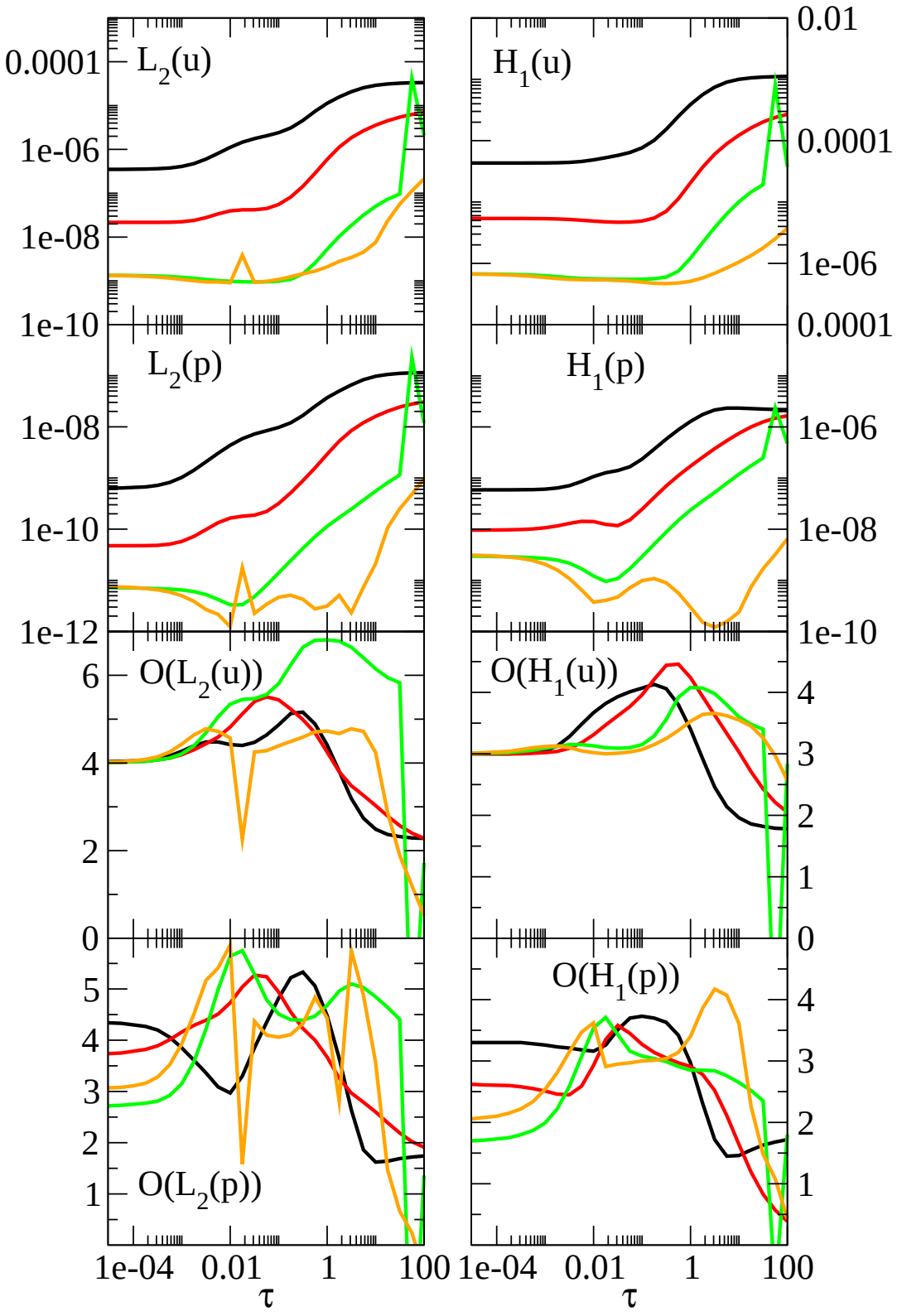

Abbildung 9.15: Fehler und Fehlerordnungen für Finite–Elemente dritter Ordnung auf dem Quadratgitter. Für weitere Details siehe die Beschreibung von Abbildung 9.2.

Insgesamt können die Ergebnisse für das polynomiale Beispiel wie folgt zusammengefasst werden:
Bei allen Rechnungen gibt es für die CIP–Methode ein τ–Intervall, in welchem auf einem ausreichend feinem Gitter die von der Fehleranalysis vorausgesagten Konvergenzraten erreicht werden.
Die Wahl $\tau = 1$ führt zu teilweise deutlich größeren Fehlerwerten als die jeweils optimale Wahl des Stabilisierungsparameters. Bei Finite–Elementen erster und dritter Ordnung können durch eine geeignetere Wahl von τ die Fehlerwerte um mindestens eine Größenordnung verringert werden. Eine geeignete Wahl für Elemente erster Ordnung ist unabhängig von dem Gittertyp und der Gitterweite $\tau \leq 10^{-8}$. Damit liegt in diesem Fall der Bereich beinahe optimaler Fehlerwerte, um mehrere Größenordnungen von dem Parametervorschlag $\tau = 1$ entfernt.
Insgesamt schneidet die CIP–Methode etwa genauso gut wie die residuale Stabilisierung ab. Welche der Verfahren vorzuziehen ist, hängt von der Elementordnung und dem Gittertyp ab. Beispielweise verdient die CIP–Methode bei Elementen zweiter Ordnung den Vorrang, weil sie die kleineren Fehler im Geschwindigkeitsfeld besitzt, während die residuale Stabilisierung auf dem Level-3-Quadratgitter besser abschneidet, da sie die kleineren Fehlerwerte im Druckfeld liefert.
Bei allen Rechnungen sind die Fehlerwerte auf dem Quadratgitter kleiner als auf den Dreiecksgitter. Zudem ist das τ–Intervall fast optimaler Fehlerwerte größer für das Quadratgitter.

Neben dem polynomialen und dem sincos–Beispiel ist im Rahmen dieser Arbeit noch ein drittes Beispiel untersucht worden. Dabei handelt es sich um ein Beispiel mit einer exponentielles Lösung für das Geschwindigkeitsfeld und einem Druckfeld, welches in ganz Ω gleich Null ist. Die Ergebnisse zu diesem Beispiel werden in Anhang C gezeigt. Beide Stabilisierungsverfahren erreichen hier die an den Interpolationsfehler–Abschätzungen gemessenen optimalen Konvergenzraten in einem großen τ–Intervall. Weiterhin fällt auf, dass die residuale Stabilisierung das Druckfeld besser als die CIP–Methode approximiert. Ansonsten sind die Ergebnisse vergleichbar mit denen des sincos–Beispiels.

Bei den bisherigen Rechnungen war die Viskosität stets auf $\nu = 10^{-6}$ fixiert. In Anhang D wird geprüft, wie sich die Resultate mit ν ändern.

9.4 Optimale Wahl der Stabilisierungsparameter

In den letzten Abschnitten sind die Stabilisierungsparameter τ_{grad}, τ_{div}, τ_p bei der CIP–Methode immer gleich gewählt worden. Selbiges gilt für die Parameter δ und γ bei der residualen Stabilisierung. Für die residualen Stabilisierung ist bekannt, dass teilweise bessere Ergebnisse erzielt werden können, wenn die Parameter verschieden voneinander gewählt werden, siehe zum Beispiel [LR06]. In diesem Abschnitt prüfen wir, ob dies auch für die CIP–Methode möglich ist. Es sei nochmal daran erinnert, dass in [BBJL07] die Wahl $\tau_{\text{grad}} = \tau_{\text{div}} = \tau_p = 1$ vorgeschlagen wird. Dieser Vorschlag impliziert, dass keine deutlichen Verbesserungen durch eine unterschiedliche Wahl der Stabilisierungsparameter erzielt werden sollten. Die Viskosität fixieren wir weiterhin auf $\nu = 10^{-6}$.

Als erstes betrachten wir das sincos–Beispiel mit Elementen zweiter Ordnung auf dem Level–4–Quadratgitter. Gemäß Abbildung 9.5 nehmen hier alle Fehler etwa gleichzeitig bei der Wahl $\tau = 1$ ihren minimalen Wert an. Auf diesen Wert fixieren wir τ_p. Die anderen beiden Stabilisierungsparameter τ_{grad} und τ_{div} werden nun jeweils von 0.01 bis 100 varriiert.

Die Fehlerwerte hierzu sind in Abbildung 9.16 zu sehen. Wie man erkennt, wachsen die Fehler im Geschwindigkeitsfeld, wenn für τ_{div} größere Werte als 1 gewählt werden. Entfernt man τ_{grad} zu sehr von 1, wachsen die Fehler im Geschwindigkeitsfeld ebenfalls. Außerdem nehmen die Fehler im Druckfeld zu, wenn τ_{grad} zu groß wird. Nur τ_{div} kann kleiner als 1 gewählt werden, ohne dass einer der Fehler wächst. Allerdings verbessern sich die Ergebnisse nur geringfügig für eine kleinere Wahl als 1. Somit ist bei der Vorgabe $\tau_p = 1$ die Wahl $\tau_{\text{grad}} = \tau_{\text{div}} = 1$ nahezu optimal. Obige Rechnungen sind ebenfalls mit $\tau_p = 10$ und $\tau_p = 0.1$ durchgeführt worden. Auch hier ergab sich qualitativ dasselbe Bild. Somit ist die Wahl $\tau_p = \tau_{\text{grad}} = \tau_{\text{div}} = 1$ hier als beinahe optimal anzusehen.

Stichprobenartig ist obiger Test ebenfalls für Finite–Elemente erster Ordnung auf dem Level–3–Dreiecksgitter durchgeführt worden. Für den Fall, dass alle Stabilisierungsparameter auf einen gemeinsamen Wert τ fixiert werden, ergeben sich ungefähr bei $\tau = 0.01$ die kleinsten Fehler. Diesen Wert geben wir für τ_p vor. Nach Abbildung 9.17 gibt es einen relativ großen Bereich für τ_{div}- und τ_{grad}-Werte, in denen die Fehlerwerte beinahe optimal sind. Wie oben bereits beobachtet, kann auch hier der Parameter τ_{div} verringert werden ohne, dass sich die Ergebnisse verschlechtern, wenn τ_p und τ_{grad} auf 0.01 fixiert werden.

Auch für das polynomiale Beispiel sind die Parameter variiert worden. Dabei wurde nicht beobachtet, dass durch eine unterschiedliche Wahl der Stabilisierungsparameter die Ergebnisse erwähnenswert verbessert werden könnten. Abbildung 9.18 zeigt exemplarisch die Ergebnisse für Finite–Elemente erster Ordnung auf dem Level–3–Quadratgitter. Nach der gleichen Vorgehensweise wie oben wählen wir hier τ_p zu 10^{-8} und variieren τ_{div} und τ_{grad}. Gezeigt ist diesmal nur der L^2–Fehler, da die anderen Fehlerwerte bis zur vierten Nachkommastelle konstant geblieben sind. Zudem sind die Änderungen im L^2–Fehler sehr gering. Auch wenn die Ergebnisse durch eine kleinere Wahl von τ_{div} und τ_{grad} verbessert werden können, ist die Wahl $\tau_{\text{grad}} = \tau_{\text{div}} = 1$ nahezu optimal.

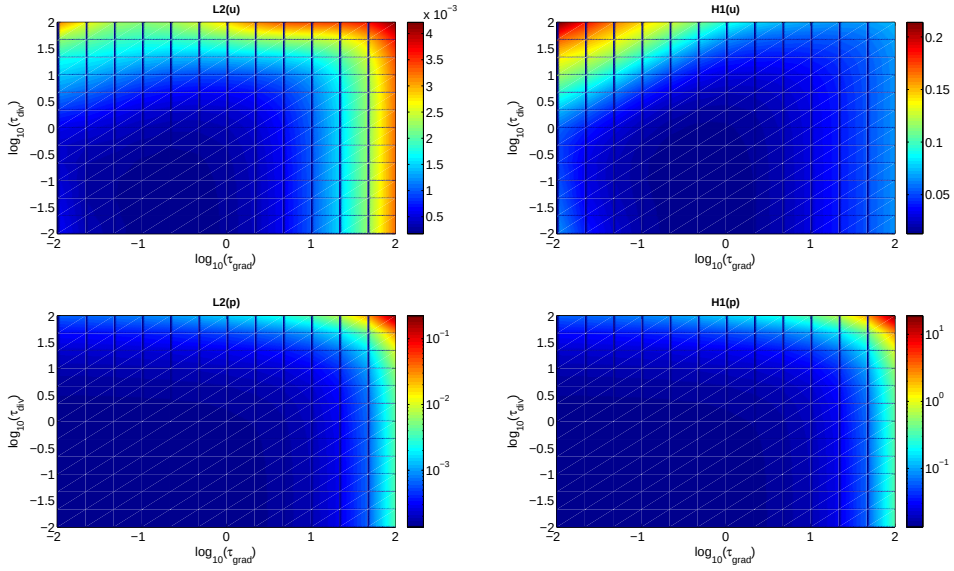

Abbildung 9.16: Fehlerwerte der CIP–Methode bei $\tau_p = 1$ und Variation von τ_{grad} und τ_{div}. Betrachtet wird das sincos–Beispiel mit Finite–Elementen zweiter Ordnung auf dem Level–4–Quadratgitter. Für weitere Details siehe Text.

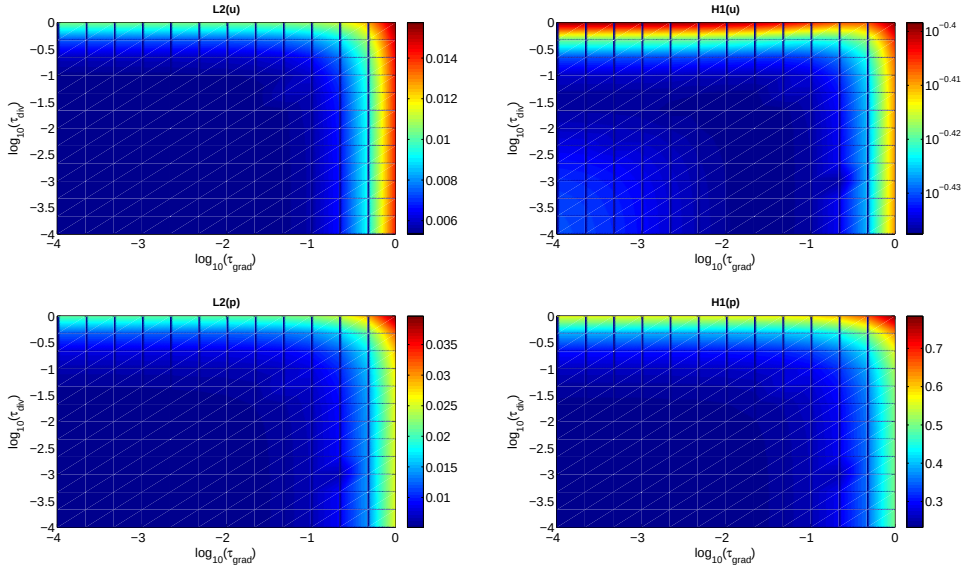

Abbildung 9.17: Fehlerwerte der CIP–Methode bei $\tau_p = 0.01$ und Variation von τ_{grad} und τ_{div}. Betrachtet wird das sincos–Beispiel mit Finite–Elementen erster Ordnung auf dem Level–3–Dreiecksgitter. Für weitere Details siehe Text.

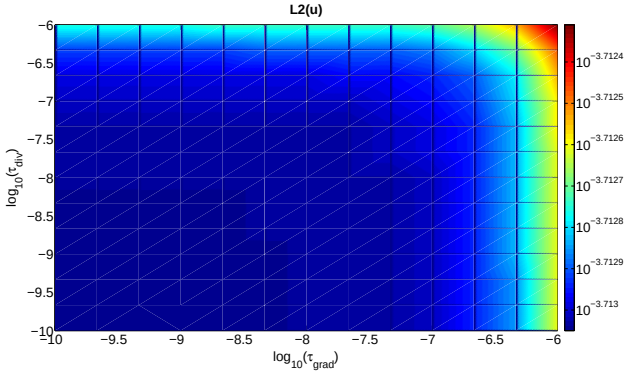

Abbildung 9.18: Fehlerwerte der CIP–Methode bei $\tau_p = 0.01$ und Variation von τ_{grad} und τ_{div}. Betrachtet wird das polynomiale Beispiel mit Finite–Elementen erster Ordnung auf dem Level–3–Quadratgitter. Für weitere Details siehe Text.

Bei obigen Tests konnte τ_{div} stets kleiner als die anderen Parameter gewählt werden, ohne dass sich die Ergebnisse verschlechtert haben. Dies wirft die Frage auf, ob die Divergenzstabilisierung überhaupt gebraucht wird. Möglicherweise könnte der Stabilisierungsterm, welcher die Sprünge des Gradienten der Geschwindigkeit bestraft, schon zur Stabilisierung im Geschwindigkeitsfeld ausreichen.

Wir betrachten zunächst wieder das sincos–Beispiel mit Elementen zweiter Ordnung auf dem Level–5–Quadratgitter. Die blaue Kurven in Abbildung 9.19 beziehen sich auf die CIP–Methode mit der Wahl $\tau_p = \tau_{\text{grad}} = \tau_{\text{div}} = \tau$ und die schwarzen auf $\tau_{\text{div}} = 0$ sowie $\tau_p = \tau_{\text{grad}} = \tau$. Die orangefarbenen Linien zeigen zum Vergleich die Ergebnisse der residualen Stabilisierung. Tatsächlich sind die Fehler der CIP–Methode ohne Divergenzsta-

bilisierung genauso gut wie die der vollen Stabilisierung, wenn jeweils ein optimaler Wert von τ gewählt wird. Für kleinere τ sind die Geschwindigkeitsfehler der vollen Stabilisierung kleiner, während bei größeren τ das Verfahren mit $\tau_{\text{div}} = 0$ die kleineren Fehlerwerte im Druckfeld liefert. Auch die Fehlerordnungen beider Verfahren sind etwa gleich gut. In dem vorliegenden Fall kann also tatsächlich auf die Divergenzstabilisierung verzichtet werden.

In Abbildung 9.20 ist statt dem Quadratgitter ein Dreiecksgitter gewählt worden. Bei jeweils optimaler Wahl von τ schneidet die CIP–Methode gemessen an den Fehlerwerten ohne Divergenzstabilisierung hier sogar etwas besser ab, da bei ihr die Fehler im Druckfeld um ungefähr 20 Prozent kleiner sind. Die Fehlerordnungen sind dafür aber etwas besser bei der vollen Stabilisierung.

Als nächstes betrachten wir das sincos–Beispiel auf dem Level–4–Quadratgitter für Elemente dritter Ordnung. Wie Abbildung 9.21 zeigt, wachsen hier die Fehler im gesamten dargestellten τ-Bereich im Vergleich zur vollen Stabilisierung, wenn τ_{div} gleich Null gewählt wird. In diesem Fall wird die Divergenzstabilisierung also benötigt. Auch wenn nicht dargestellt, kann ein ähnliches Verhalten für das polynomiale Beispiel auf dem Level–5–Dreiecksgitter mit Finiten Elementen zweiter Ordnung beobachtet werden.

Obige Ergebnisse können wie folgt zusammengefasst werden: Durch eine unterschiedliche Wahl der Stabilisierungsparameter konnten die Fehlerwerte in keinem Fall deutlich, also zum Beispiel um eine Größenordnung, verbessert werden. Während eine Wahl von $\tau_{\text{grad}} \neq \tau_p$ oft zu schlechteren Ergebnissen führt, kann τ_{div} oft kleiner als die anderen Stabilisierungsparameter gewählt werden. In einigen Fällen kann sogar auf die Divergenzstabilisierung verzichtet werden, ohne dass die Fehler anwachsen, und zuweilen erzielt man damit sogar bessere Resultate. Es gibt aber auch Fälle, bei denen es sinnvoll ist, die Divergenzstabilisierung zu verwenden, wie beispielsweise für das sincos–Beispiel bei Finite–Elementen zweiter Ordnung auf dem Dreiecksgitter.

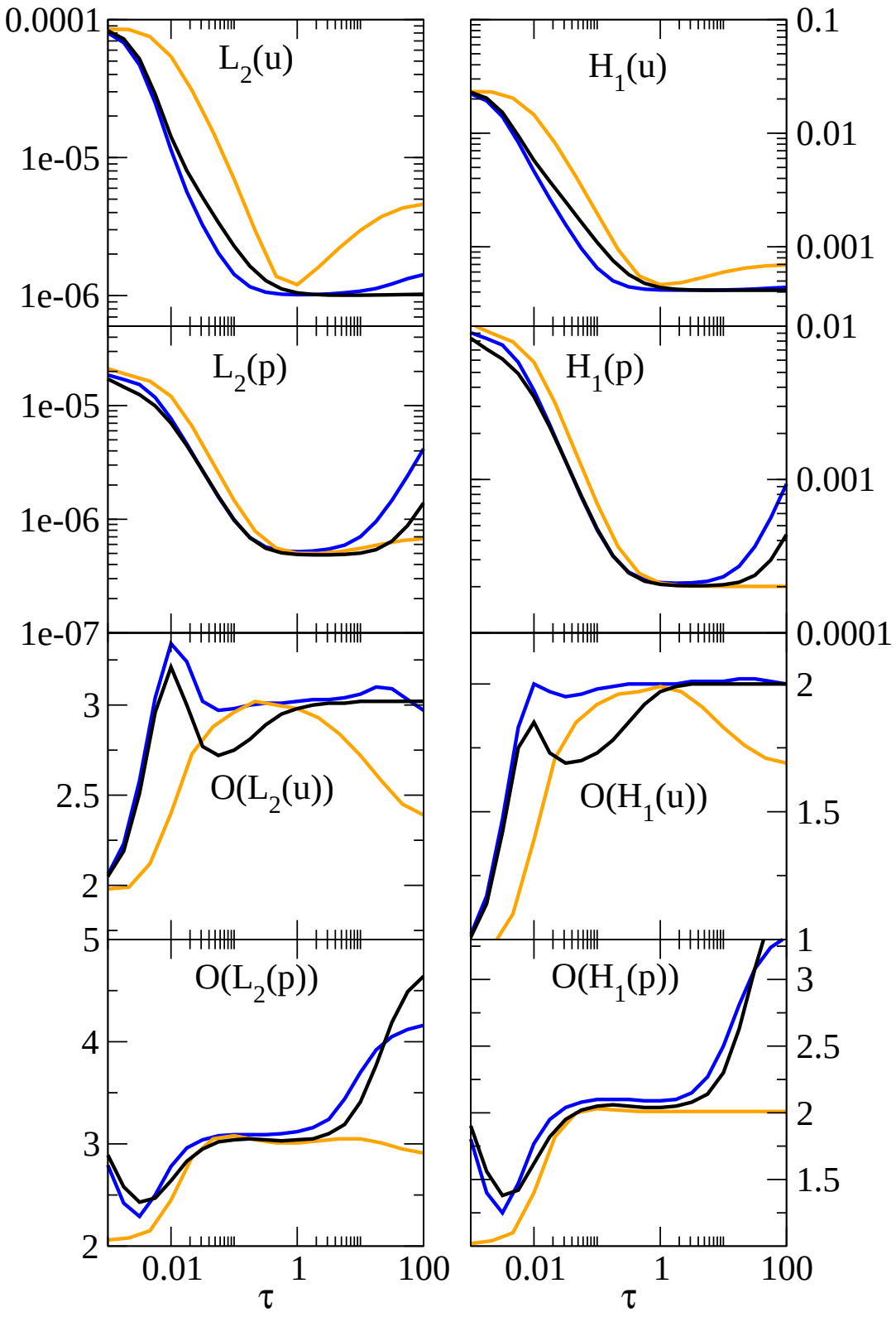

Abbildung 9.19: Fehler und Fehlerordnungen der Stabilisierungsverfahren bei Finiten–Elementen zweiter Ordnung für das sincos–Beispiel auf dem Level–5–Quadratgitter. Die blauen Kurven zeigen die Werte für die CIP–Methode mit der Wahl $\tau = \tau_{\text{grad}} = \tau_{\text{div}} = 1$, die schwarzen mit $\tau = \tau_{\text{grad}} = \tau_p$ sowie $\tau_{\text{div}} = 0$ und die orangefarbenen für residuale Stabilisierung mit $\tau = \delta = \gamma$.

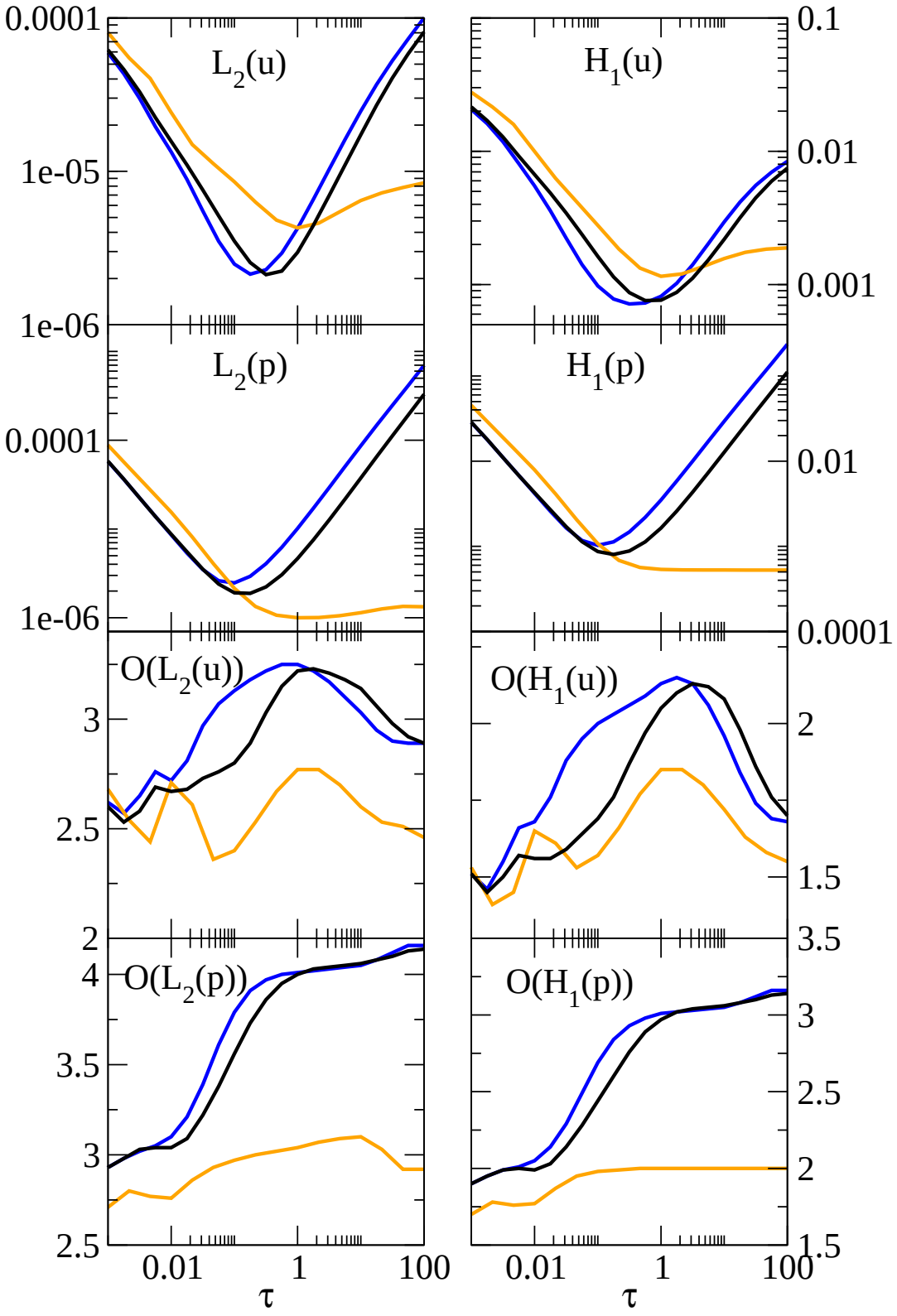

Abbildung 9.20: Fehler und Fehlerordnungen der Stabilisierungsverfahren bei Finiten–Elementen zweiter Ordnung für das sincos–Beispiel auf dem Level–5–Dreiecksgitter. Die blauen Kurven zeigen die Werte für die CIP–Methode mit der Wahl $\tau = \tau_{\text{grad}} = \tau_{\text{div}} = 1$, die schwarzen mit $\tau = \tau_{\text{grad}} = \tau_p$ sowie $\tau_{\text{div}} = 0$ und die orangefarbenen für residuale Stabilisierung mit $\tau = \delta = \gamma$.

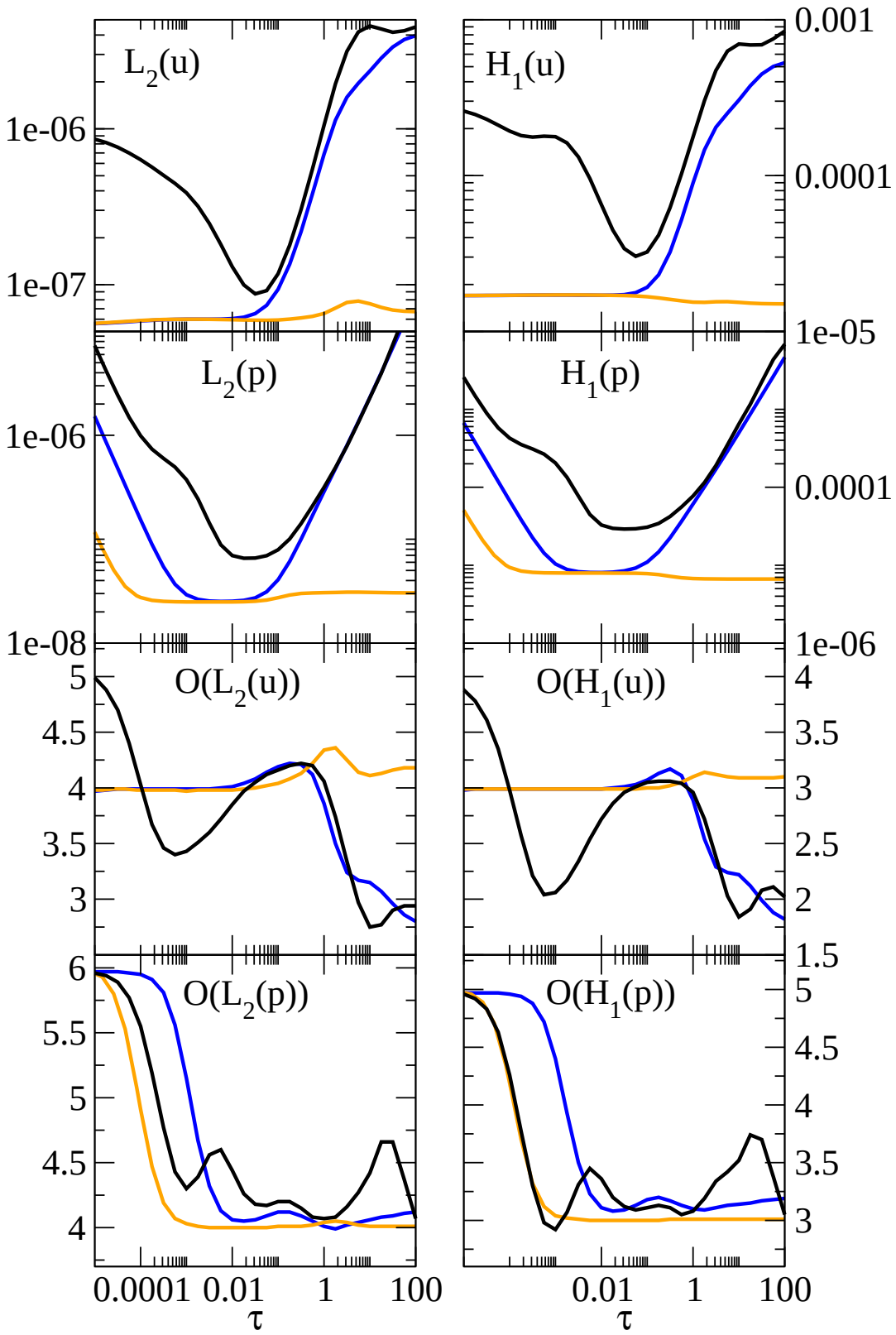

Abbildung 9.21: Fehler und Fehlerordnungen der Stabilisierungsverfahren bei Finiten–Elementen dritter Ordnung für das sincos–Beispiel auf dem Level–4–Quadratgitter. Die blauen Kurven zeigen die Werte für die CIP–Methode mit der Wahl $\tau = \tau_{\text{grad}} = \tau_{\text{div}} = 1$, die schwarzen mit $\tau = \tau_{\text{grad}} = \tau_p$ sowie $\tau_{\text{div}} = 0$ und die orangefarbenen für residuale Stabilisierung mit $\tau = \delta = \gamma$.

9.5 Taylor–Hood–Elemente

Bei Taylor–Hood–Elementen handelt es sich um Finite–Elemente, bei denen das Geschwindigkeitsfeld mit Elementen höherer Ordnung als das Druckfeld approximiert wird. In Abschnitt 8.3 haben wir die Taylor–Hood–Elemente für die residuale Stabilisierung theoretisch untersucht. Für den Fall, dass die Elementordnung in der Geschwindigkeit um eins größer als die Ordnung im Druckfeld ist, wurde dort eine a–priori Fehlerabschätzung hergeleitet. Im Folgenden beschränken wir uns auf das polynomiale und das sincos–Beispiel mit $\nu = 10^{-6}$. In beiden Fällen liegt der konvektions–dominante Fall vor. Für den konvektions–dominanten Fall sind von der Analysis in Satz 8.12 die in Tabelle 9.14 zusammengefassten Fehlerordnungen vorausgesagt worden.

Fehler	Konvergenzrate aus der Fehlerabschätzung	optimale Rate
L^2–Fehler von u	h^r	h^{r+1}
H^1–Fehler von u	$h^{r-1/2}$	h^r
L^2–Fehler von p	h^r	h^r
H^1–Fehler von p	-	h^{r-1}

Tabelle 9.14: Konvergenzraten der Taylor–Hood–Elemente der Ordnung r im Geschwindigkeitsfeld und der Ordnung $r-1$ im Druckfeld für die residuale Stabilisierung, wenn der konvektions–dominante Fall vorliegt. In Spalte zwei sind die von der a–priori Fehlerabschätzung (8.32) vorausgesagten Raten gezeigt, in Spalte drei die an den Interpolationsfehler–Abschätzungen aus Abschnitt 5.4 gemessenen optimalen Raten.

Taylor–Hood–Elemente sind für die residuale Stabilisierung bereits numerisch untersucht worden und haben sich als konkurrenzfähig zu den equal order Elementen erwiesen, siehe beispielsweise [MLR09]. Für die CIP–Methode findet man hingegen weder analytische noch numerische Untersuchungen zu den Taylor–Hood–Elementen. In diesem Abschnitt betrachten wir nun die Taylor–Hood–Elemente auch für die CIP–Methode. Abkürzend sprechen wir im Folgenden von 21–Taylor–Hood–Elementen, wenn die Elementordnung im Geschwindigkeitsfeld gleich zwei ist und im Druckfeld gleich eins (analog dazu führen wir auch die Bezeichnung 32– und 43–Taylor–Hood–Element ein).

Abbildung 9.22 zeigt unter anderem die Fehler und Fehlerordnungen der 21–Taylor–Hood–Elemente für das sincos–Beispiel auf dem Level–4–Quadratgitter. Aufgetragen werden die Werte gegen den Stabilisierungsparameter τ. Zunächst betrachten wir die roten und magentafarbenen Kurven. Die magentafarbenen Kurven beziehen sich auf die Resultate für die 21–Taylor–Hood–Elemente mit $\tau_p = \tau_{\text{grad}} = \tau_{\text{div}} = \tau$ und die roten auf die 21–Taylor–Hood–Elemente mit $\tau_{\text{grad}} = \tau_{\text{div}} = \tau$ sowie $\tau_p = 0$. In allen Fehlerwerten schneidet die Taylor–Hood Methode ohne Druckstabilisierung, also mit $\tau_p = 0$, deutlich besser ab. Vor allem gilt dies für große τ. Auch bei allen weiteren Tests war die Wahl $\tau_p = 0$ stets vorzuziehen. Wie wir in Abschnitt 8.4 erwähnt haben, sichert der Druckstabilisierungsterm eine modifizierte diskrete Babuška–Brezzi–Bedingung. Diese ist aber für die Taylor–Hood–Elemente gemäß Abschnitt 8.1 ohnehin schon erfüllt. Es ist also durchaus plausibel, dass hier auf die Druckstabilisierung verzichtet werden kann.

Die 21–Taylor–Hood–Elemente mit $\tau_{\text{grad}} = \tau_{\text{div}} = \tau$ und $\tau_p = 0$ sollen nun mit den equal order Elementen verglichen werden. Bei den equal order Elementen weisen wir allen Stabilisierungsparameter den gemeinsamen Wert τ zu. Die grünen Kurven beziehen sich auf

die equal order Elemente erster Ordnung, die blauen auf die equal order Elemente zweiter Ordnung. Vor allem die Geschwindigkeitsfehler der Taylor–Hood–Elemente sind kleiner als diejenigen der Elemente erster Ordnung. In ihrem optimalen Bereich erreichen die Fehler sogar dasselbe Niveau wie die Fehler der Elemente zweiter Ordnung. Die im Vergleich zu den Elementen zweiter Ordnung kleinere Wahl des Druckraums bei den 21–Taylor–Hood–Elementen macht sich in den Geschwindigkeitsfehlern also kaum bemerkbar. Der kleinere Druckraum macht sich allerdings in den Druckfehlern bemerkbar: Die equal order Elemente zweiter Ordnung besitzen bei jeweils optimaler Wahl von τ die deutlich kleineren Fehler im Druckfeld, siehe hierzu auch Tabelle 9.15.

Die 21–Taylor–Hood–Elemente mit $\tau_{\text{grad}} = \tau_{\text{div}} = \tau$ und $\tau_p = 0$ erreichen für eine optimale Wahl von τ gemäß Tabelle 9.15 eine Ordnung von 2.78 bzw. 1.95 im L^2– bzw. H^1–Fehler der Geschwindigkeit und eine Ordnung von 2 bzw. 1 im L^2– bzw. H^1–Fehler des Druckfeldes. Diese Ordnungen liegen über den Voraussagen für die Taylor–Hood–Elemente der residualen Stabilisierung, welche in Tabelle 9.14 aufgeführt sind. Folglich könnten auch für die Taylor–Hood–Elemente der CIP–Methode ähnliche Fehlerabschätzungen gelten, wie sie für die Taylor–Hood–Elemente der residualen Stabilisierung bewiesen worden sind.

Als nächstes vergleichen wir die roten mit den schwarzen Kurven in Abbildung 9.22. Die schwarzen beziehen sich ebenfalls auf die 21–Taylor–Hood–Elemente, allerdings für solche mit $\tau_{\text{grad}} = \tau$ und $\tau_{\text{div}} = \tau_p = 0$. Auf die Divergenzstabilisierung wird also verzichtet oder, äquivalent dazu, stabilisiert wird nur über die Sprünge im Gradienten der Geschwindigkeit. Im letzten Abschnitt haben wir die Divergenzstabilisierung für die equal order Elemente diskutiert. Im vorliegenden Fall kann bei großen τ auf die Divergenzstabilisierung verzichtet werden, während sie bei kleinen τ benötigt wird.

τ	$L^2(u)$	$H^1(u)$	$L^2(p)$	$H^1(p)$	$\mathcal{O}(L^2(u))$	$\mathcal{O}(H^1(u))$	$\mathcal{O}(L^2(p))$	$\mathcal{O}(H^1(p))$
3.16e-5	1.28e-3	1.30e-1	1.21e-3	6.31e-2	1.99	1.00	2.00	1.02
100	1.04e-5	1.75e-3	2.54e-4	6.29e-2	2.78	1.95	2.00	1.00
1.0	8.22e-6	1.67e-3	4.49e-6	9.04e-4	3.04	2.01	3.30	2.24

Tabelle 9.15: Fehler und Fehlerordnungen für das sincos–Beispiel auf dem Level–4–Quadratgitter. Die erste Zeile bezieht sich auf die equal order Elemente erster Ordnung, die zweite auf die 21–Taylor–Hood–Elemente und die letzte auf die equal order Elemente zweiter Ordnung. Für die Bedeutung von τ siehe Text.

Abbildung 9.23 zeigt die Ergebnisse für 32–Taylor–Hood–Elemente. Betrachtet wird erneut das sincos–Beispiel auf dem Level–4–Quadratgitter. Die roten Kurven beziehen sich auf die 32–Taylor–Hood–Elemente mit $\tau_{\text{grad}} = \tau_{\text{div}} = \tau$ und $\tau_p = 0$, die blauen auf die equal order Elemente zweiter Ordnung und die grünen auf die equal order Elemente dritter Ordnung jeweils mit $\tau_p = \tau_{\text{grad}} = \tau_{\text{div}} = \tau$.
Die 32–Taylor–Hood–Elemente besitzen die kleinsten Fehler bei den großen τ–Werten in der Abbildung. Die Fehlerordnungen sind dort ebenfalls am besten und erreichen außer im H^1–Fehler der Geschwindigkeit die Raten, welche für die Taylor–Hood–Elemente der residualen Stabilisierung bewiesen worden sind. Im H^1–Fehler fehlt hingegen etwa noch eine halbe Ordnung. Wie Tabelle 9.16 zeigt, schließt der Bereich kleiner Fehler noch weitaus größere τ als dargestellt der Fall mit ein.

Bei jeweils optimaler Wahl der Stabilisierungsparameter sind die Fehlerwerte der 32–Taylor–Hood–Elemente im Geschwindigkeitsfeld kleiner und im Druckfeld ungefähr gleich

der Werte der equal order Elemente zweiter Ordnung. Die Fehlerwerte der equal order Elemente dritter Ordnung sind bei jeweils optimaler Wahl von τ allesamt kleiner als diejenigen der 32–Taylor–Hood–Elemente. Auch in den Geschwindigkeitsfehlern erkennt man einen deutlichen Unterschied, obwohl hier die Ansatzräume beider Verfahren gleich groß sind. Die relativ großen Fehlerwerte im Geschwindigkeitsfeld werden möglicherweise durch Rückkopplung mit dem Druckfeld hervorgerufen, in welchem die Fehler vergleichbar mit den equal order Elementen zweiter Ordnung sind.

Die schwarzen Kurven in Abbildung 9.23 kennzeichnen die 32–Taylor–Hood–Elemente ohne Divergenzstabilisierung, das heißt, mit der Wahl $\tau_{\text{grad}} = \tau$ und $\tau_{\text{div}} = \tau_p = 0$. Wie man erkennt, wachsen alle Fehler beinahe um eine Größenordnung, wenn auf die Divergenzstabilisierung verzichtet wird. Auch für das polynomiale Beispiel, welches als nächstes betrachtet wird, sind die Fehler mit Divergenzstabilisierung stets kleiner gleich den Fehlern ohne Divergenzstabilisierung. Die Wahl $\tau_{\text{div}} = 0$ scheint also keine Verbesserung für die Taylor–Hood–Elemente zu bringen und wird daher nicht weiter diskutiert.

τ	$L^2(u)$	$H^1(u)$	$L^2(p)$	$H^1(p)$	$\mathcal{O}(L^2(u))$	$\mathcal{O}(H^1(u))$	$\mathcal{O}(L^2(p))$	$\mathcal{O}(H^1(p))$
1.0	8.22e-6	1.67e-3	4.49e-6	9.04e-4	3.04	2.01	3.30	2.24
100	1.75e-6	2.19e-4	3.83e-6	8.08e-4	2.89	1.98	2.97	2.02
2.0e+4	1.68e-6	1.54e-4	3.88e-6	8.08e-4	2.86	2.04	2.97	2.02
1.9e+5	1.65e-6	1.56e-4	3.87e-6	8.08e-4	2.92	2.02	2.97	2.02
5.62e-3	6.01e-8	1.70e-5	2.51e-8	8.06e-6	4.00	2.99	4.13	3.23

Tabelle 9.16: Fehler und Fehlerordnungen für das sincos–Beispiel auf dem Level–4–Quadratgitter. Die erste Zeile bezieht sich auf die equal order Elementen zweiter Ordnung, die zweite bis vierte auf die 32–Taylor–Hood–Elemente und die letzte auf die equal order Elemente dritter Ordnung. Für die Bedeutung von τ siehe Text.

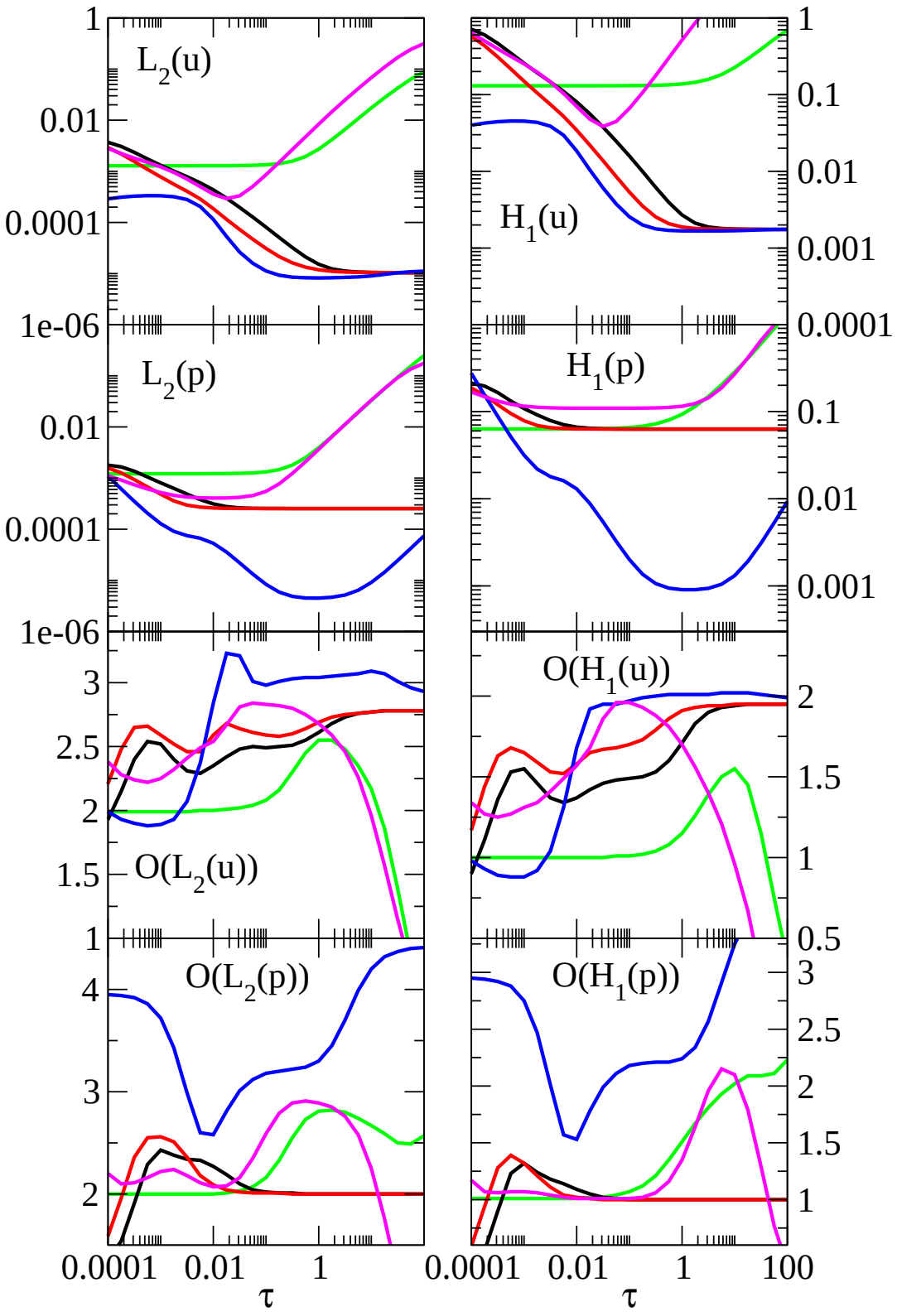

Abbildung 9.22: Fehler und Fehlerordnungen der CIP–Methode auf dem Level–4–Quadratgitter für das sincos–Beispiel. Die grünen Kurven kennzeichnen die Werte der equal order Elemente erster Ordnung, die blauen diejenigen der equal order Elemente zweiter Ordnung. Die restlichen Kurven zeigen die Werte der 21–Taylor–Hood–Elemente, wobei sich die Ergebnisse voneinander durch die Wahl für die Stabilisierungsparameter τ_{grad}, τ_{div}, τ_p unterscheiden. Für weitere Details siehe Text.

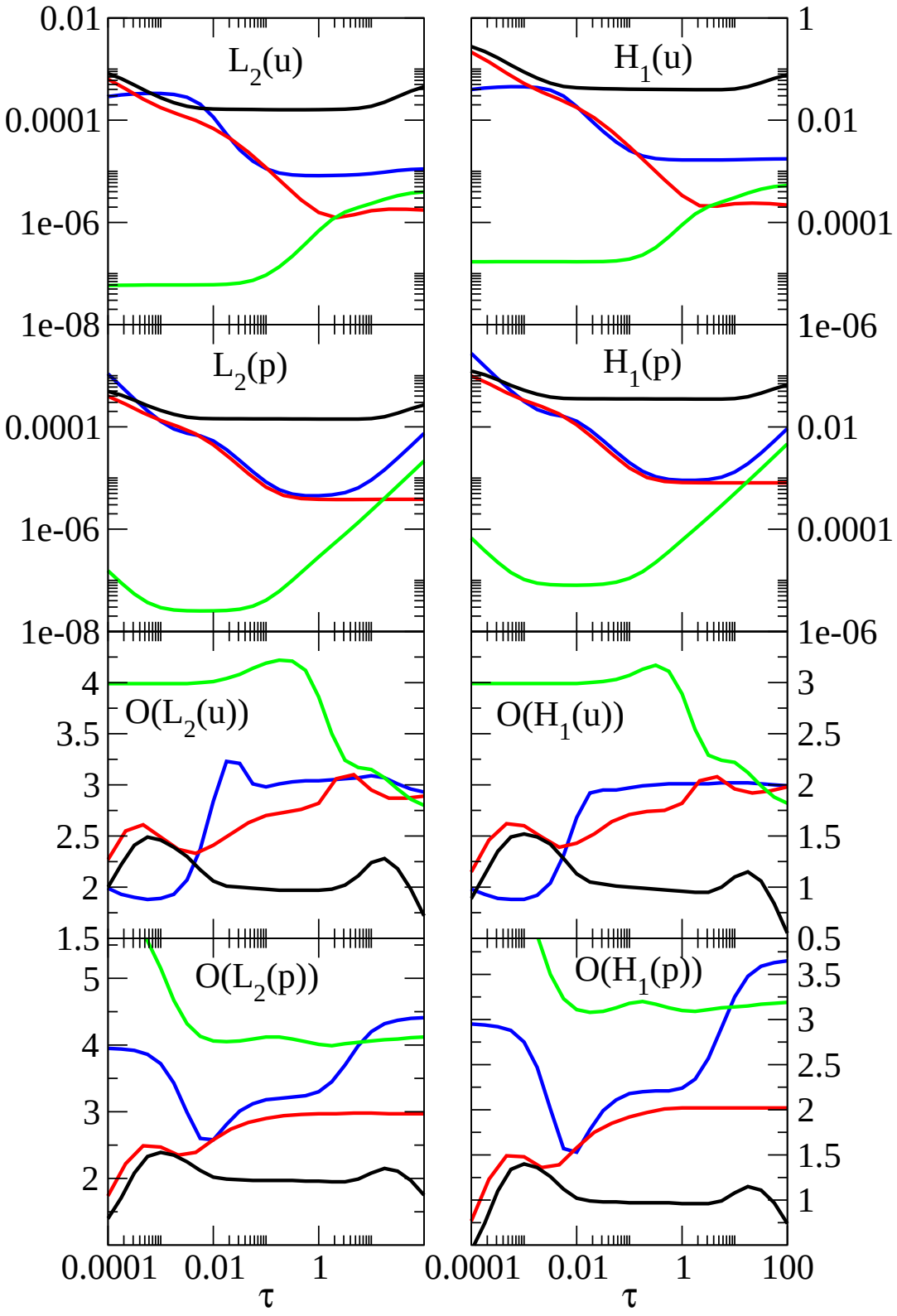

Abbildung 9.23: Fehler und Fehlerordnungen der CIP–Methode auf dem Level–4–Quadratgitter für das sincos–Beispiel. Die blauen Kurven kennzeichnen die Werte der equal order Elemente zweiter Ordnung, die grünen diejenigen der equal order Elemente dritter Ordnung. Die restlichen Kurven zeigen die Werte der 32–Taylor–Hood–Elemente, wobei sich die Ergebnisse voneinander durch die Wahl für die Stabilisierungsparameter τ_{grad}, τ_{div}, τ_p unterscheiden. Für weitere Details siehe Text.

In Abbildung 9.24 sind die Resultate der 21–Taylor–Hood–Elemente für das polynomiale Beispiel auf dem Level–4–Quadratgitter zu sehen. Es sei daran erinnert, dass es für diese Konfiguration bei großen τ numerische Probleme mit den equal order Elementen zweiter Ordnung gibt. Näheres hierzu ist in Abschnitt 9.3 besprochen worden. Ein Vergleich zwischen den equal order Elementen zweiter Ordnung und den 21–Taylor–Hood–Elementen ist daher nur bedingt möglich.
Die roten Kurven in der Abbildung beziehen sich auf die 21–Taylor–Hood–Elemente mit $\tau_{\text{grad}} = \tau_{\text{div}} = \tau$ und $\tau_p = 0$. Die kleinsten Fehlerwerte werden bei relativ großen τ erreicht. Die Fehlerordnungen entsprechen dort sogar den an den Interpolationsfehlern gemessenen optimalen Konvergenzraten aus Tabelle 9.14.

Bei großen τ kommt es bei den 21–Taylor–Hood–Elemente mit $\tau_p = 0$ zu keinen vergleichbaren numerischen Problemen wie bei den equal order Elementen mit $\tau_p = \tau$ zweiter Ordnung. Somit könnten die numerischen Problemen für große τ bei den Elementen zweiter Ordnung durch einen zu großen Druckstabilisierungsterm verursacht werden. Gemäß weiteren numerischen Tests ist dies aber nicht der Fall. So kommt es bei den equal order Elementen zweiter Ordnung auch zu Problemen, wenn τ_p auf den Wert 1 fixiert wird, während $\tau = \tau_{\text{grad}} = \tau_{\text{div}}$ vergrößert werden.

Da es bei den equal order Elementen erster Ordnung zu keinen numerischen Problemen kommt, sollen diese mit den 21–Taylor–Hood–Elementen verglichen werden. Bei jeweils optimaler Wahl der Stabilisierungsparameter ergibt sich hier ein ähnliches Bild wie für das sincos–Beispiel. Gemäß Tabelle 9.17 sind die Fehlerwerte für die Taylor–Hood–Elemente im Druckfeld nur wenig kleiner, wohingegen die Fehler im Geschwindigkeitsfeld deutlich geringer ausfallen.

τ	$L^2(u)$	$H^1(u)$	$L^2(p)$	$H^1(p)$	$\mathcal{O}(L^2(u))$	$\mathcal{O}(H^1(u))$	$\mathcal{O}(L^2(p))$	$\mathcal{O}(H^1(p))$
1.0e-8	4.15e-5	4.30e-3	2.22e-4	4.66e-2	2.22	1.00	2.67	1.30
1.0e+6	3.36e-7	6.96e-5	1.78e-4	4.41e-2	3.01	2.00	2.00	1.00

Tabelle 9.17: Fehler und Fehlerordnungen für das polynomiale Beispiel auf dem Level–4–Quadratgitter. Die erste Zeile bezieht sich auf die equal order Elemente erster Ordnung und die zweite auf die 21–Taylor–Hood–Elemente. Für die Bedeutung von τ siehe Text.

Abbildung 9.25 zeigt die Ergebnisse der 32–Taylor–Hood–Elemente mit $\tau_p = 0$ und $\tau_{\text{grad}} = \tau_{\text{div}} = \tau$ für das polynomiale Beispiel, gekennzeichnet durch die roten Kurven. Die blauen Kurven stellen die Resultate der equal order Elemente zweiter Ordnung dar und die grünen diejenigen der equal order Elemente dritter Ordnung. Wie zuvor ist für die equal order Elemente der Stabilisierungsparameter τ gleich τ_p, τ_{grad} und τ_{div}. Gerechnet wird auf dem Level–4–Quadratgitter.
Die Kurven sind in der Abbildung nicht über das gesamte dargestellte τ–Intervall verfolgt worden; für die equal order Elemente zweiter Ordnung aufgrund der in Abschnitt 9.3 geschilderten numerischen Probleme und für die anderen Elemente, weil in dem ausgelassenen τ-Bereich die Fehlerwerte bereits deutlich größer sind als bei einer optimalen Wahl von τ und weil zudem die Fehler weiter anzuwachsen scheinen.

Die Geschwindigkeitsfehler der 32–Taylor–Hood–Elemente verringern sich für große τ beinahe auf die Fehler der equal order Elementen dritter Ordnung bei optimaler Wahl von τ, siehe hierzu auch Tabelle 9.18. Die Ergebnisse im Druckfeld sind hingegen eher vergleich-

bar mit den Resultaten der equal order Elemente zweiter Ordnung. Im Bereich nahezu optimal gewählter τ erreichen die 32–Taylor–Hood–Elemente im Druckfeld sicher und im Geschwindigkeitsfeld beinahe die an der Interpolationsfehler–Abschätzung gemessenen optimalen Raten. Insbesondere liegen die Raten damit oberhalb der für die 32–Taylor–Hood–Elemente der residualen Stabilisierung vorausgesagten Konvergenzordnungen in Tabelle 9.14.

τ	$L^2(u)$	$H^1(u)$	$L^2(p)$	$H^1(p)$	$\mathcal{O}(L^2(u))$	$\mathcal{O}(H^1(u))$	$\mathcal{O}(L^2(p))$	$\mathcal{O}(H^1(p))$
4.13e+5	4.05e-9	8.37e-7	1.46e-6	3.12e-4	3.96	2.94	2.98	2.02
3.16e-2	9.44e-10	5.50e-7	4.78e-12	1.08e-9	5.47	3.09	5.29	3.43

Tabelle 9.18: Fehler und Fehlerordnungen für das polynomiale Beispiel auf dem Level–4–Quadratgitter. Die erste Zeile bezieht sich auf die 32–Taylor–Hood–Elemente und die zweite auf die equal order Elemente dritter Ordnung. Für die Bedeutung von τ siehe Text.

Rechnungen sind bei dem polynomialen Beispiel auch für die 43–Taylor–Hood–Elemente auf dem Quadratgitter durchgeführt worden. In diesem Fall liegt die Lösung im Ansatzraum und die Ergebnisse sind wie gewünscht bereits auf dem Level–0–Gitter bis auf Rundungsfehler exakt.
Als Hauptergebnis dieses Abschnitts halten wir fest, dass die Taylor–Hood–Elemente der CIP–Methode bei optimaler Wahl des Stabilisierungsparameters τ bei fast allen Beispielen bereits auf dem Level–4–Gitter die für die residuale Stabilisierung vorausgesagten Konvergenzraten erreicht haben. Dies legt die Vermutung nahe, dass solche Abschätzungen auch für die CIP–Methode gelten.

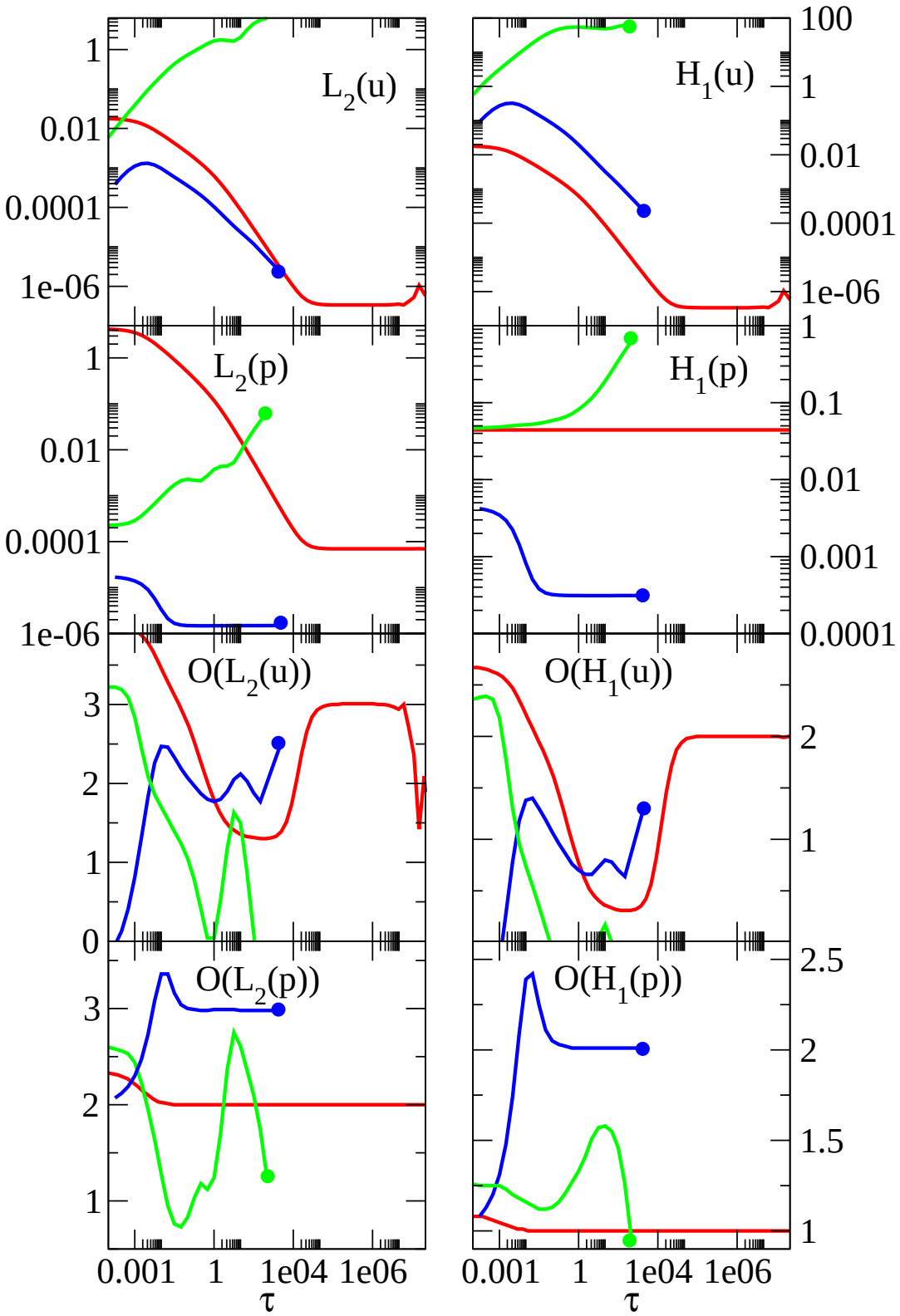

Abbildung 9.24: Fehler und Fehlerordnungen der CIP–Methode auf dem Level–4–Quadratgitter für das polynomiale Beispiel. Die roten Kurven kennzeichnen die Werte der 21-Taylor–Hood–Elemente, die grünen Kurven diejenigen der equal order Elemente erster Ordnung und die blauen die Werte der equal order Elemente zweiter Ordnung. Für weitere Details siehe Text.

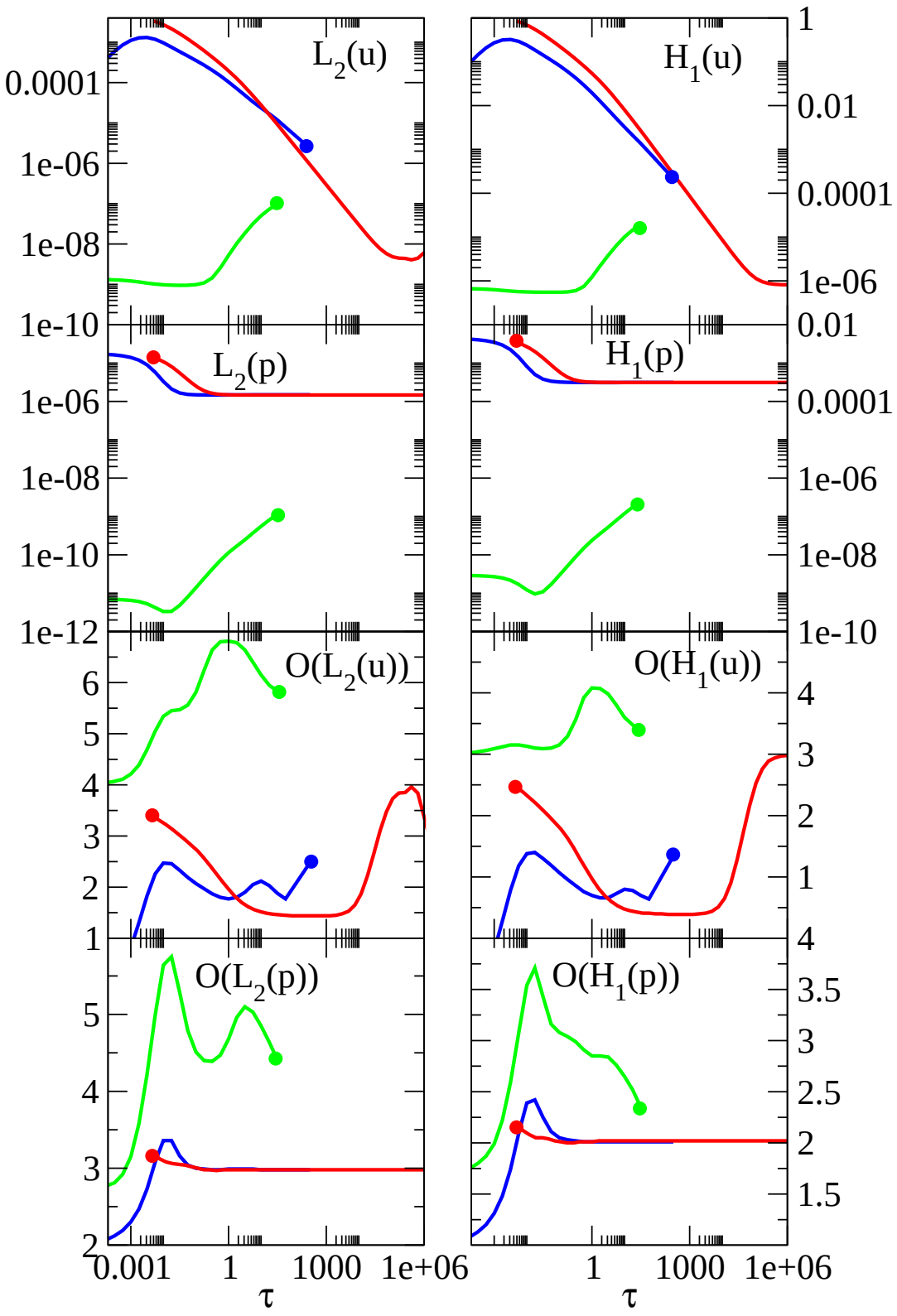

Abbildung 9.25: Fehler und Fehlerordnungen der CIP–Methode auf dem Level–4–Quadratgitter für das polynomiale Beispiel. Die roten Kurven kennzeichnen die Werte der 32–Taylor–Hood–Elemente, die blauen Kurven diejenigen der equal order Elemente zweiter Ordnung und die grünen die Werte der equal order Elemente dritter Ordnung. Für weitere Details siehe Text.

9.6 Laufzeitmessungen

In Abschnitt 9.2 und 9.3 haben wir die Genauigkeit der residualen Stabilisierung mit derjenigen der CIP–Methode verglichen. Gemessen an den Fehlerwerten in den L^2– sowie H^1–Normen im Druck- und Geschwindigkeitsfeld hat die residuale Stabilisierung geringfügig besser abgeschnitten.

Ein weiteres wichtiges Vergleichskriterium ist die Laufzeit der Verfahren, das heißt die Zeit, welche die Verfahren benötigen, um die Systemmatrix zu assemblieren und das resultierende lineare Gleichungssystem zu lösen. Die Laufzeit untersuchen wir in diesem Abschnitt exemplarisch für das sincos–Beispiel. Für einen fairen Vergleich setzen wir bei der residualen Stabilisierung $\delta = \gamma = \tau$ sowie bei der CIP–Methode $\tau_{\text{grad}} = \tau_{\text{div}} = \tau_p = \tau$ und fixieren τ jeweils ungefähr auf den Wert, bei welchem die Fehlerwerte am kleinsten sind. Diese Werte sind den Tabellen von Abschnitt 9.2 zu entnehmen.

Tabelle 9.19 zeigt die Laufzeiten der Verfahren für Finite–Elemente erster Ordnung auf dem Quadrat– und dem Dreiecksgitter von Level zwei bis fünf. Die residuale Stabilisierung braucht auf beiden Gittern weniger Zeit. Um den quantitativen Vergleich zwischen den Verfahren zu vereinfachen, gibt die Größe $\overline{\Delta t}$ den prozentuale Zuwachs in der Laufzeit an, welcher hinzunehmen ist, wenn statt der residualen Stabilisierung die CIP–Methode verwandt wird, das heißt:

$$\overline{\Delta t} = \frac{t_{\text{CIP}} - t_{\text{RB}}}{t_{\text{RB}}},$$

wobei t_{RB} die Laufzeit der residualen Stabilisierung bezeichnet und t_{CIP} diejenige der CIP–Methode. Gemäß der Tabelle beträgt der prozentuale Zuwachs auf dem Dreiecksgitter mindestens 40%. Auf dem Quadratgitter ist die residuale Stabilisierung sogar um über 200% schneller, wobei die Unterschiede mit zunehmenden Gitterlevel etwas abnehmen.

Tabelle 9.19 erlaubt auch einen Vergleich zwischen den Laufzeiten auf dem Quadrat- und dem Dreiecksgitter. Gemessen an den Fehlerwerten waren die Quadratgitter in den vorausgehenden Abschnitten stets vorzuziehen. Nach der Tabelle sind die Rechnungen auf dem Quadratgitter bei der residualen Stabilisierung zudem schneller. Bei der CIP–Methode ist es aber umgekehrt, hier sind die Rechnungen auf dem Dreiecksgitter schneller.

Level	t_{RB} [s]	t_{CIP} [s]	$\overline{\Delta t}$	t_{RB} [s]	t_{CIP} [s]	$\overline{\Delta t}$
2	0.04	0.15	2.75	0.08	0.14	0.75
3	0.15	0.54	2.6	0.33	0.55	0.66
4	0.50	1.68	2.36	1.05	1.48	0.40
5	2.01	6.06	2.01	3.07	5.97	0.94

Tabelle 9.19: Laufzeit in Sekunden bei den Finite–Elementen erster Ordnung für das sincos–Beispiel. Die zweite bis vierte Spalte beziehen sich auf das Quadratgitter, die fünfte bis siebte auf das Dreiecksgitter. Für die residualen Stabilisierung ist auf dem Quadratgitter $\tau = 0.215$ und auf dem Dreiecksgitter $\tau = 0.001$. Für die CIP–Methode ist auf dem Quadratgitter $\tau = 0.0177$ und auf dem Dreiecksgitter $\tau = 0.001$.

Die Laufzeiten für Elemente zweiter Ordnung sind in Tabelle 9.20 zu sehen. Auch hier ist die Laufzeit der residualen Stabilisierung auf allen Leveln und beiden Gittertypen geringer.

Erneut ist der prozentuale Zuwachs $\overline{\Delta t}$ auf dem Quadratgitter größer. Im Vergleich zu den Elementen erster Ordnung hat aber der prozentuale Zuwachs auf dem Quadratgitter abgenommen, während er auf dem Dreiecksgitter zugenommen hat. Die CIP–Methode rechnet auf dem Dreiecksgitter schneller als auf dem Quadratgitter, bei der residualen Stabilisierung gilt dies auf den höheren Gitterleveln.

Level	t_{RB} [s]	t_{CIP} [s]	$\overline{\Delta t}$	t_{RB} [s]	t_{CIP} [s]	$\overline{\Delta t}$
2	0.20	0.61	2.05	0.25	0.68	1.72
3	0.64	1.96	2.06	0.84	1.95	1.32
4	3.96	11.53	1.91	2.88	7.07	1.46
5	37.59	92.79	1.47	18.80	44.09	1.35

Tabelle 9.20: Laufzeit in Sekunden bei den Finite–Elementen zweiter Ordnung für das sincos–Beispiel. Die zweite bis vierte Spalte beziehen sich auf das Quadratgitter, die fünfte bis siebte auf das Dreiecksgitter. Für die residuale Stabilisierung ist auf dem Quadratgitter $\tau = 1.0$ und auf dem Dreiecksgitter $\tau = 1.0$. Für die CIP–Methode ist auf dem Quadratgitter $\tau = 1.77$ und auf dem Dreiecksgitter $\tau = 0.177$.

Gemäß Tabelle 9.21 ist auch für Elemente dritter Ordnung die residuale Stabilisierung gemessen an der Laufzeit vorzuziehen. Der prozentuale Zuwachs auf den Quadratgittern hat im Vergleich zu den Elementen zweiter Ordnung abgenommen, so dass $\overline{\Delta t}$ auf den höheren Gitterleveln hier nun auf dem Dreiecksgitter größer als auf dem Quadratgitter ist. Bei beiden Stabilisierungsverfahren kann auf dem Dreiecksgitter schneller als auf dem Quadratgitter gerechnet werden.

Level	t_{RB} [s]	t_{CIP} [s]	$\overline{\Delta t}$	t_{RB} [s]	t_{CIP} [s]	$\overline{\Delta t}$
2	0.98	1.61	0.64	0.66	0.75	0.13
3	4.94	9.06	0.83	2.05	4.13	1.01
4	33.39	67.49	1.02	10.24	25.54	1.49

Tabelle 9.21: Laufzeit in Sekunden bei den Finite–Elementen dritter Ordnung für das sincos–Beispiel. Die zweite bis vierte Spalte beziehen sich auf das Quadratgitter, die fünfte bis siebte auf das Dreiecksgitter. Für die residualen Stabilisierung ist auf dem Quadratgitter $\tau = 46.4$ und auf dem Dreiecksgitter $\tau = 0.1$. Für die CIP–Methode ist auf dem Quadratgitter $\tau = 0.046$ und auf dem Dreiecksgitter $\tau = 0.03162$.

Bei Elementen zweiter Ordnung auf dem Level–5–Quadratgitter und bei Elementen dritter Ordnung auf dem Level–4–Quadratgitter beträgt die Laufzeit der CIP–Methode bereits über eine Minute. Weitere Tests haben gezeigt, dass auf diesen Gittern bei beiden Stabilisierungsverfahren mehr als 90% der Laufzeit zur Lösung des linearen Gleichungssystems verwandt werden. Demnach ist dort die längere Laufzeit der CIP–Methode vor allem darauf zurückzuführen, dass sich die Lösung des Gleichungssystems aufwändiger gestaltet. Grund hierfür ist das dichtere Besetzungsschema der Systemmatrix, welche bei der CIP–Methode durch die zusätzliche Kopplung der Freiheitsgrade über die Kantenintegrale verursacht wird.

Kapitel 10

Zusammenfassung und Ausblick

Betrachtet worden sind stabilisierte Finite–Element–Methoden für die Konvektions–Diffusions– und die Oseen–Gleichung. Bei der Konvektions–Diffusionsgleichung sind das Standard–Galerkin–Verfahren, die Streamline–Diffusion–Methode und die CIP–Methode (auch Kanten–Stabilisierung genannt) untersucht worden. Nachdem die theoretischen Grundlagen der Methoden gelegt worden sind, haben wir numerische Simulationen durchgeführt, deren Ergebnisse wie folgt zusammengefasst werden können:

Bei Problemen ohne Grenzschichten haben sich beide Stabilisierungsverfahren als geeignet erwiesen. Getestet worden sind unter anderem das exponentielle und das Tangens–Hyperbolicus–Beispiel aus [BH04]. Gerechnet wurde dabei auf einem Dreiecks– und einem Quadratgitter mit Elementen erster, zweiter und dritter Ordnung. Auch der Gitterlevel wurde variiert. Bei unseren numerischen Tests werden die von der Analysis vorausgesagten Konvergenzraten in der L^2–Norm und der H^1–Seminorm von beiden Stabilisierungsverfahren erreicht. Hierzu ist der Stabilisierungsparameter τ der Verfahren passend zu wählen. In dem zugehörigen τ–Bereich sind in der Regel zugleich die Fehlerwerte am kleinsten.
Was eine passende Wahl für τ ist, hängt von der Elementordnung ab. Dazu kann das exponentielle Beispiel auf dem Quadratgitter betrachtet werden. Für Elemente erster Ordnung erhält man nahezu optimale Fehlerwerte, wenn τ kleiner gleich 0.01 gewählt wird. Für Elemente zweiter Ordnung sollte τ hingegen größer gleich 0.01 gewählt werden. Auch der Gittertyp hat einen Einfluss auf die optimale Wahl von τ. Dazu kann erneut das exponentielle Beispiel mit Elementen zweiter Ordnung betrachtet werden. Auf dem Quadratgitter werden die kleinsten Fehler für τ größer gleich 0.01 erreicht, während auf dem Dreiecksgitter eine Wahl von $\tau = 4.64 \cdot 10^{-3}$ optimal ist.

Gemessen an den L^2– und H^1–Fehlerwerten schneiden beide Stabilisierungsverfahren etwa gleich gut ab. Die Streamline–Diffusion–Methode erlaubt jedoch vor allem für feinere Gitterweiten und höhere Elementordnungen die deutlich schnelleren Rechnungen. So war beispielsweise die Laufzeit der Streamline–Diffusion–Methode mit Elementen dritter Ordnung auf dem Level–5–Quadratgitter um einen Faktor 10 kleiner als diejenige der CIP–Methode. Gemäß weiteren Tests wird dabei mehr als 95% der Laufzeit zur Lösung des linearen Gleichungssystems verwandt. Demnach ist die längere Laufzeit der CIP–Methode vor allem darauf zurückzuführen, dass sich die Lösung des Gleichungssystems aufwändiger gestaltet. Grund hierfür ist das dichtere Besetzungsschema der Systemmatrix, welche bei der CIP–Methode durch die zusätzliche Kopplung der Freiheitsgrade über die Kantenintegrale verursacht wird.

Bei den betrachteten Beispielen können die Ergebnisse des Standard–Galerkin–Verfahrens

durch die Stabilisierung abhängig von dem Gittertyp und der Elementordnung verbessert werden. Dies ist für Elemente zweiter Ordnung der Fall oder für Elemente dritter Ordnung auf dem Dreiecksgitter. Im Gegensatz hierzu ist für Elemente erster Ordnung oder für Elemente dritter Ordnung auf dem Quadratgitter die Wahl $\tau = 0$ nahezu optimal, bei welcher die Stabilisierungsverfahren in das Standard–Galerkin–Verfahren übergehen.

Ferner ist untersucht worden, wie das Quadrat– im Vergleich zum Dreiecksgitter abschneidet. Bei beiden Stabilisierungsverfahren sind die Fehler auf dem Quadratgitter kleiner als auf dem Dreiecksgitter. Die Rechnungen auf dem Dreiecksgitter sind jedoch außer bei Elementen erster Ordnung bei der Streamline–Diffusion–Methode schneller als auf dem Quadratgitter. Dies ist darauf zurückzuführen, dass auf dem Quadratgitter mehr Freiheitsgrade miteinander verkoppelt werden und somit die zugehörige Systemmatrix eine größere Anzahl von Null verschiedener Einträge besitzt.
Die Simulationen für die Beispiele ohne Grenzschichten sind für die Wahl $\epsilon = 10^{-6}$ durchgeführt worden. Gemäß weiteren Tests ändern sich die Ergebnisse vernachlässigbar, wenn ϵ weiter verringert wird.

Bei Problemen mit Grenzschichten beobachtet man bei Verwendung des Standard–Galerkin–Verfahrens im gesamten Lösungsgebiet unphysikalische Oszillationen. Diese können durch die Stabilisierung weitgehend unterdrückt werden. Um jedoch auch in den Grenzschichten gute Näherungen zu erhalten, benötigt man sehr kleine Gitterweiten. Dies ist oft nicht praktikabel. Daher sollten für Probleme mit Grenzschichten andere Verfahren vorgezogen werden. Beispielsweise Stabilisierungsverfahren mit shock capturing Termen haben in der Literatur bessere Ergebnisse geliefert.

Auch für die Oseen–Gleichung sind zwei Stabilisierungsverfahren untersucht worden: die residuale Stabilisierung und eine Variante der CIP–Methode. Nach einem Überblick über die Analysis der Verfahren sind numerische Rechnungen anhand dreier Beispiele durchgeführt worden. Betrachtet wurden Elemente erster bis dritter Ordnung auf dem Quadrat- sowie dem Dreiecksgitter für verschiedene Gitterweiten. Der Einfachheit halber haben wir die Stabilisierungsverfahren zunächst nur mit einem freien Stabilisierungsparameter τ betrachtet.
Eine Frage dabei war, ob die Stabilisierungsverfahren die von der Analysis vorausgesagten Konvergenzraten erfüllen. Gemäß unseren Rechnungen werden manchmal sehr gute Konvergenzraten erreicht, obwohl τ so gewählt ist, dass die Fehler relativ groß sind. Dies kann beispielsweise für das sincos–Beispiel bei Elementen dritter Ordnung und kleinem τ beobachtet werden. Die Fehlerordnung in diesen Bereichen ist weniger interessant. Von Interesse sind vielmehr die Ordnungen der Verfahren in den τ–Bereichen, in welchen die betrachteten Fehlergrößen, das heißt der L^2–Fehler und der H^1–Fehler im Geschwindigkeits– sowie im Druckfeld, klein sind. Gemäß unseren Rechnungen werden dort die von der Fehleranalysis vorausgesagten Fehlerordnungen auf einem genügend großen Gitterlevel erreicht. Nicht getestet werden konnte dies für die CIP–Methode bei dem polynomialen Beispiel mit Elementen zweiter Ordnung auf dem Quadratgitter. Hier kommt es im Bereich, in welchem die kleinsten Fehler zu erwarten sind, zu numerischen Problemen.

Als nächstes fragt sich, wie τ gewählt werden sollte, damit die Fehlerwerte möglichst klein werden. Für die CIP–Methode sollte gemäß [BBJL07] die Wahl $\tau = 1$ optimal sein. Bei dem sincos– und dem exponentiellen Beispiel werden die von der Fehleranalysis vorausgesagten Konvergenzraten für $\tau = 1$ erreicht. Die kleinsten Fehler findet man jedoch oft bei anderer Wahl von τ. So können bei dem sincos–Beispiel die Fehlerwerte für Element dritter Ord-

nung auf dem Dreiecksgitter um mindestens eine Größenordnung verbessert werden, wenn statt $\tau = 1$ die Wahl $\tau \approx 5.6 \cdot 10^{-3}$ getroffen wird. Schlechte Resultate erhält man mit der Wahl $\tau = 1$ für Finite–Elementen erster Ordnung bei dem polynomialen Beispiel. Insbesondere fallen hier die Konvergenzraten gering aus. Gute Konvergenzraten und möglichst kleine Fehlerwerte erhält man hier hingegen für genügend kleine τ, also etwa $\tau \leq 10^{-8}$. Wird stattdessen τ auf 1 erhöht, vergrößern sich die Fehler im Geschwindigkeitsfeld um mehr als vier Größenordnungen. Folglich ist die Wahl $\tau = 1$ nach unsere Ergebnissen im Allgemeinen nicht zu empfehlen.
Ein a-priori Vorschlag für τ konnte in dieser Arbeit nicht gegeben werden. Gemäß unserer Ergebnisse hängt das optimale τ bei beiden Stabilisierungsverfahren von vielen verschiedenen Faktoren ab. Eine Rolle spielen die Elementordnung, der Gittertyp und die Form der Lösung beziehungsweise die Problemdaten. Außerdem verschiebt sich die optimale Wert von τ zuweilen mit dem Gitterlevel (siehe das sincos–Beispiel bei Verwendung von Elementen dritter Ordnung auf dem Quadratgitter).

Gemessen an den Fehlerwerten schneidet die residuale Stabilisierung insgesamt gesehen etwas besser als die CIP–Methode ab. Welches der Stabilisierungsverfahren bei jeweils optimaler Wahl von τ die kleineren Fehlerwerte liefert, hängt zwar von dem betrachteten Beispiel, der Elementordnung und dem Gittertyp ab, die residuale Stabilisierung besitzt bei optimal gewähltem τ jedoch in der Mehrzahl der betrachteten Fälle die kleineren Fehler. Zudem ist der τ–Bereich, in welchem die kleinsten Fehler angenommen werden, häufig größer als bei der CIP–Methode.

Gemessen an der Laufzeit ist die residuale Stabilisierung der CIP–Methode klar vorzuziehen. Getestet wurde dies für das sincos–Beispiel. Bei einem bezüglich der Fehler optimalem τ besitzt dort die residuale Stabilisierung für Elemente erster bis dritter Ordnung auf dem Dreiecks- sowie Quadratgitter die kürzere Laufzeit. Der prozentuale Zuwachs, der hinzunehmen ist, wenn statt der residualen Stabilisierung die CIP–Methode benutzt wird, beträgt dabei in den meisten Fällen mehr als 50 und teilweise sogar mehr als 200 Prozent. Grund hierfür sind wie bei der Konvektions–Diffusionsgleichung die Kantenintegrale der CIP–Methode, welche zu einem im Vergleich zur residualen Stabilisierung dichteren Besetzungsschema der Systemmatrix führen.
Nimmt man die Fehlerwerte und die Laufzeiten der Stabilisierungsverfahren als Maßstab, ist die residuale Stabilisierung nach unseren Ergebnissen vorzuziehen. Wie in Abschnitt 8.4 erwähnt, besitzt die CIP–Methode dafür die größere Kompatibilität mit anderen numerischen Verfahren. Demnach gibt es Fälle in denen die CIP–Methode noch benutzt werden kann, während die residuale Stabilisierung nicht mehr anwendbar ist.

Obige Untersuchungen sind im konvektions–dominanten Fall mit einer Viskosität $\nu = 10^{-6}$ durchgeführt worden. Im diffusions–dominanten Fall hat sich hingegen gezeigt, dass Taylor–Hood Elemente ohne Stabilisierung den Stabilisierungsverfahren vorzuziehen sind.

Für die CIP–Methode wurde ferner geprüft, ob die Fehlerwerte durch eine unterschiedliche Wahl der Stabilisierungsparameter verringert werden können. Bei keinem der Tests konnte hierdurch eine deutliche Verbesserung festgestellt werden. Interessant hat sich erwiesen, auf die Divergenzstabilisierung zu verzichten. Ohne Divergenzstabilisierung erzielt man in einigen Fällen gleich gute oder sogar geringfügig bessere Ergebnisse als bei der vollen Stabilisierung. Da es aber auch Beispiele gibt, bei denen die Fehlerwerte mit Divergenzstabilisierung kleiner sind, sollte im Allgemeinen nicht auf eine Stabilisierung der Divergenz verzichtet werden.

Schließlich wurde die Verwendung von Taylor–Hood–Elementen für die CIP–Methode untersucht. Gemäß den erreichten Konvergenzraten sollte es möglich sein, mindestens die gleichen Fehlerordnungen zu zeigen, wie sie für die Taylor–Hood–Elemente bei der residuale Stabilisierung bewiesen worden sind. Gemäß weiteren Tests sind die Fehler ohne eine Stabilisierung des Drucks stets kleiner als mit der Druckstabilisierung. Auf eine solche sollte demnach verzichtet werden. Durch eine Stabilisierung der Divergenz und des Gradienten der Geschwindigkeit können die Fehler hingegen vermindert werden.

Ausblick

Durch die vorliegende Arbeit werden unter anderem die folgenden Untersuchungen motiviert:

- Für die CIP–Methode bei der Oseen–Gleichung wird in [BBJL07] die Wahl $\tau = 1$ für die Stabilisierungsparameter vorgeschlagen. Grundlage für den Vorschlag sind numerische Untersuchungen, auf deren Ergebnisse in [BBJL07] nicht näher eingegangen wird. In der vorliegenden Arbeit haben wir aber Beispiele gesehen, bei welchen durch eine von $\tau = 1$ verschiedene Wahl deutlich bessere Ergebnisse erzielt werden konnten. Dies wirft die Frage auf, ob beide Programme überhaupt näherungsweise die gleichen Ergebnisse liefern und wo es gegebenenfalls zu Differenzen kommt. Daher wäre es interessant, unsere Ergebnisse mit den Resultaten zu vergleichen, welche dem Parametervorschlag in [BBJL07] zugrunde liegen.

- Die CIP–Methode besitzt eine deutlich längere Laufzeit als die Streamline–Diffusion–Methode beziehungsweise als die residuale Stabiliserung. Der Großteil der Laufzeit wird dazu verwandt, das zugrunde liegende lineare Gleichungssystem zu lösen. Die dafür zuständige Routine nimmt folglich entscheidenden Einfluss auf die Laufzeit des Programms. In der vorliegenden Arbeit ist zur Lösung stets der direkte Löser aus dem Paket UMFPACK [Dav04] verwandt worden. Möglicherweise kann ein Großteil der Laufzeit eingespart werden, wenn ein geeigneterer Löser benutzt wird.

- In dieser Arbeit sind Rechnungen mit Taylor–Hood–Elementen für die CIP–Methode durchgeführt worden. Die beobachteten Fehlerordnungen sehen viel versprechend aus. In einem nächsten Schritt könnte die zugehörige Fehleranalysis entwickelt werden.

Anhang A

Testbeispiel für die Implementierung der CIP–Methode

Beispiel A.1.
In diesem Beispiel gehen wir von der Situation auf der rechten Seite von Abbildung A.1 aus. Betrachtet wird dort das quadratische P^2-Element.

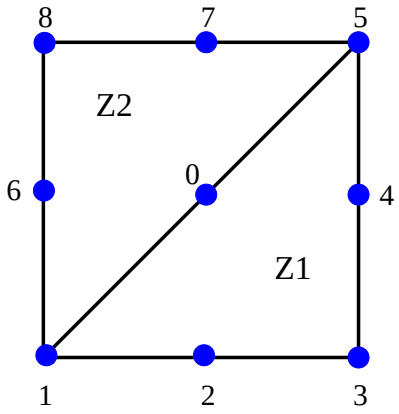

Abbildung A.1: Knotenverteilung bei Beispiel A.1. Die blauen Punkte entsprechen den Knoten. Für weitere Details siehe Text.

Die Randbedingungen nehmen wir als stark vorgegeben an. Starke Randbedingungen werden direkt im Ansatzraum des Finite–Element–Verfahrens berücksichtigt. Im zugehörigen Code wird dies wie folgt realisiert: Das lineare Gleichungssystem, welches aus dem Galerkin–Verfahren hervorgeht, sei gegeben durch $Au = f$. Zu bestimmen ist u, um die Koeffizienten für die Ansatzfunktionen zu erhalten. Liegt der 1-Knoten der n-ten Ansatzfunktion auf dem Gebietsrand, so wird ein 1-Eintrag in die $n \times n$-te Komponente von A eingetragen sowie auf der rechten Seite der vorgegebene Randwert r_n in die n-te Komponente von f. In unserem Beispiel liegen alle Knoten bis auf denjenigen mit der Nummer 0

auf dem Gebietsrand und das lineare Gleichungssystem nimmt die folgende Form an:

$$\begin{pmatrix} * & * & * & * & * & * & * & * & * & * \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u_0 \\ u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \\ u_7 \\ u_8 \\ u_9 \end{pmatrix} = \begin{pmatrix} f_0 \\ r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \\ r_6 \\ r_7 \\ r_8 \\ r_9 \end{pmatrix}, \tag{A.1}$$

wobei f_0 der Wert von f im Punkt 0 ist. Die Sterne in der Matrix kennzeichnen die Einträge, die noch assembliert werden müssen. Die Beiträge des Kantenintegrals berechnen wir im Folgenden für den Fall $b = (1, 2)$. Als Produkt zweier Polynome zweiten Grades wird der Integrand ein Polynom vierten Grades sein. Ein Polynom vierten Grades lässt sich exakt durch die Gaußquadraturformel der Ordnung 3 integrieren, welche lautet:

$$\int_a^b p(s)ds = \frac{b-a}{2}(w_0 p(s_0) + w_1 p(s_1) + w_2 p(s_2)) \tag{A.2}$$

mit den Gewichten $w_0 = 5/9$, $w_1 = 8/9$, $w_0 = 5/9$ und den Quadraturpunkten

$$s_0 = -\sqrt{\frac{3}{5}}\frac{b-a}{2} + \frac{a+b}{2}, \quad s_1 = \frac{a+b}{2}, \quad s_3 = \sqrt{\frac{3}{5}}\frac{b-a}{2} + \frac{a+b}{2}.$$

Für die Kantenintegrale ergibt sich daraus die Formel

$$I = \sum_{j=1}^{3} \frac{h_E^3}{2} \left(b_{x,j}[\partial_x u_j] + b_{y,j}[\partial_y u_j]\right) \left(b_{x,j}[\partial_x v_j] + b_{y,j}[\partial_y v_j]\right) \mid_{s=s_i} \tag{A.3}$$

mit den Quadraturpunkten

$$s_0 = (0.8873, 0.8873), \;\; s_1 = (0.5, 0.5) \quad \text{und} \quad s_2 = (0.1127, 0.1127).$$

Als nächstes bestimmen wir die Basisfunktionen und ihre Sprünge in obiger Formel. Gemäß Abschnitt 5.2 besitzt die Basisfunktion ϕ, welche in der Ecke $\mathbf{a}^i$ in Gitterzelle K ihren 1-Knoten hat, die folgende Darstellung auf K:

$$\phi|_K(\lambda) = \lambda_i(2\lambda_i - 1), \tag{A.4}$$

Wenn der 1-Knoten hingegen der Mittelpunkt $\mathbf{a}^{ij}$ der Kante mit den Eckpunkten $\mathbf{a}^i$ und $\mathbf{a}^j$ ist, lautet die lokale Darstellung von ϕ:

$$\phi|_K(\lambda) = \zeta_{ij}(\lambda) = 4\lambda_i\lambda_j. \tag{A.5}$$

Die baryzentrischen Koordinaten λ_i wählen wir wie in Beispiel 7.1 und bestimmen damit

aus obigen Formeln die Basisfunktionen zu:

$$\phi_0(x,y) = \begin{cases} 4(1-x-y)(y-x) & \text{in Z1} \\ x - xy & \text{in Z2} \end{cases},$$

$$\phi_1(x,y) = \begin{cases} (1-x-y)(1-2x-2y) & \text{in Z1} \\ 2y^2 - 3y + 1 & \text{in Z2} \end{cases},$$

$$\phi_2(x,y) = \begin{cases} 4(1-x-y)x & \text{in Z1} \\ 0 & \text{in Z2} \end{cases}, \qquad \phi_3(x,y) = \begin{cases} 2x^2 - x & \text{in Z1} \\ 0 & \text{in Z2} \end{cases},$$

$$\phi_4(x,y) = \begin{cases} 4xy & \text{in Z1} \\ 0 & \text{in Z2} \end{cases}, \qquad \phi_5(x,y) = \begin{cases} 2y^2 - y & \text{in Z1} \\ 0 & \text{in Z2} \end{cases},$$

$$\phi_6(x,y) = \begin{cases} 0 & \text{in Z1} \\ 4(1-y)(y-x) & \text{in Z2} \end{cases}, \qquad \phi_7(x,y) = \begin{cases} 0 & \text{in Z1} \\ 4(xy - x^2) & \text{in Z2} \end{cases},$$

$$\phi_8(x,y) = \begin{cases} 0 & \text{in Z1} \\ (y-x)(2y-2x-1) & \text{in Z2} \end{cases}.$$

Nun können die Sprünge der Basisfunktionen in den Quadraturpunkten berechnet werden. Im Punkt s_0 haben wir

Nr. Basisfunktion	0	1	2	3	4	5	6	7	8
$[\partial_x u]$	-4	0.5492	0.4508	-1	3.5492	-2.5492	0.4508	3.5492	-1
$[\partial_y u]$	4	-0.5492	-0.4508	1	-3.5492	2.5492	-0.4508	-3.5492	1

sowie im Punkt s_1

Nr. Basisfunktion	0	1	2	3	4	5	6	7	8
$[\partial_x u]$	-4	-1	2	-1	2	-1	2	2	-1
$[\partial_y u]$	4	1	-2	1	-2	1	-2	-2	1

und im Punkt s_2

Nr. Basisfunktion	0	1	2	3	4	5	6	7	8
$[\partial_x u]$	-4	-2.5492	3.5492	-1	0.4508	0.5492	3.5492	0.4508	-1
$[\partial_y u]$	4	2.5492	-3.5492	1	-0.4508	-0.5492	-3.5492	-0.4508	1

Setzt man die Sprünge in Formel (A.3) ein, so erhält man folgende Werte der Kantenintegrale bezüglich ϕ_0 als Ansatzfunktion.

Nr. Testfunktion	Integral
0	45.254
1	11.314
2	-22.627
3	11.314
4	-22.627
5	11.314
6	-22.627
7	-22.627
8	11.314

Anhang B

Das Tangens–Hyperbolicus–Beispiel

Wir betrachten das Tangens–Hyperbolicus–Beispiel aus [BH04] für die Konvektions–Diffusionsgleichung. Dieses Beispiel besitzt auf dem Gebiet $[0,1]^2$ die Lösung

$$u(x,y) = 0.5(1 - \tanh(5(x - 0.5))). \tag{B.1}$$

Die Vorgaben lauten wie in [BH04] $b = (1,0)^T$, $c = 1$ und $\nu = 10^{-6}$ auf Ω. Die rechte Seite f der Konvektions–Diffusionsgleichung wird so angepasst, dass u eine Lösung der Konvektions–Diffusionsgleichung ist. Die Randbedingungen sind über die Randfunktion u_b festgelegt, wobei $u_b = u$ auf $\partial\Omega$ gesetzt wird.

Für die Streamline–Diffusion–Methode wählen wir mit $\delta_0 = \tau$ dasselbe Symbol wie für den Stabilisierungsparameter der CIP–Methode. Die Abbildungen B.1 bis B.6 zeigen die Ergebnisse der Streamline–Diffusion–Methode und der CIP–Methode für das Tangens–Hyperbolicus–Beispiel. Weitere Details sind den Bildunterschriften zu entnehmen.

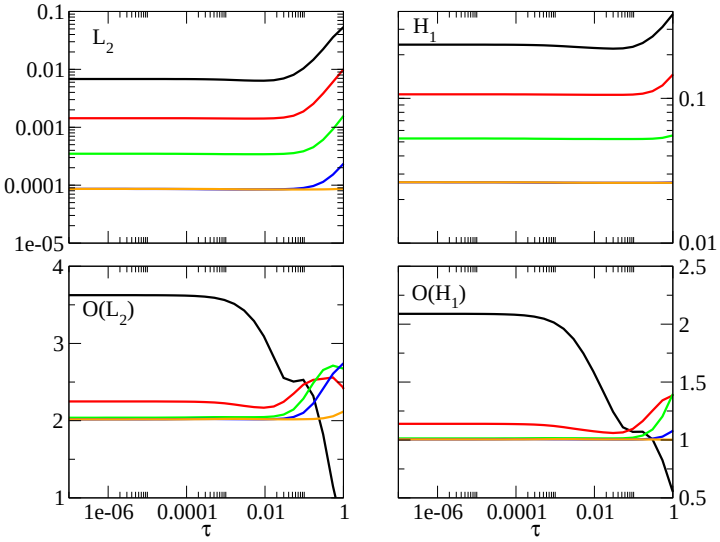

Abbildung B.1: Fehler und Fehlerordnungen Finiter–Elemente erster Ordnung auf dem Dreiecksgitter aufgetragen gegen den Stabilisierungsparameter τ. In der oberen Reihe werden der L^2– und der H^1–Fehler gezeigt, in der unteren die zugehörigen Ordnungen. Die schwarzen, roten, grünen und blauen Kurven stellen die Ergebnisse der CIP–Methode dar und beziehen sich der Reihe nach auf Gitterlevel 2,3,4 und 5. Die orangefarbenen Kurven zeigt zum Vergleich die Resultate der Streamline–Diffusion–Methode auf dem Level–5–Gitter.

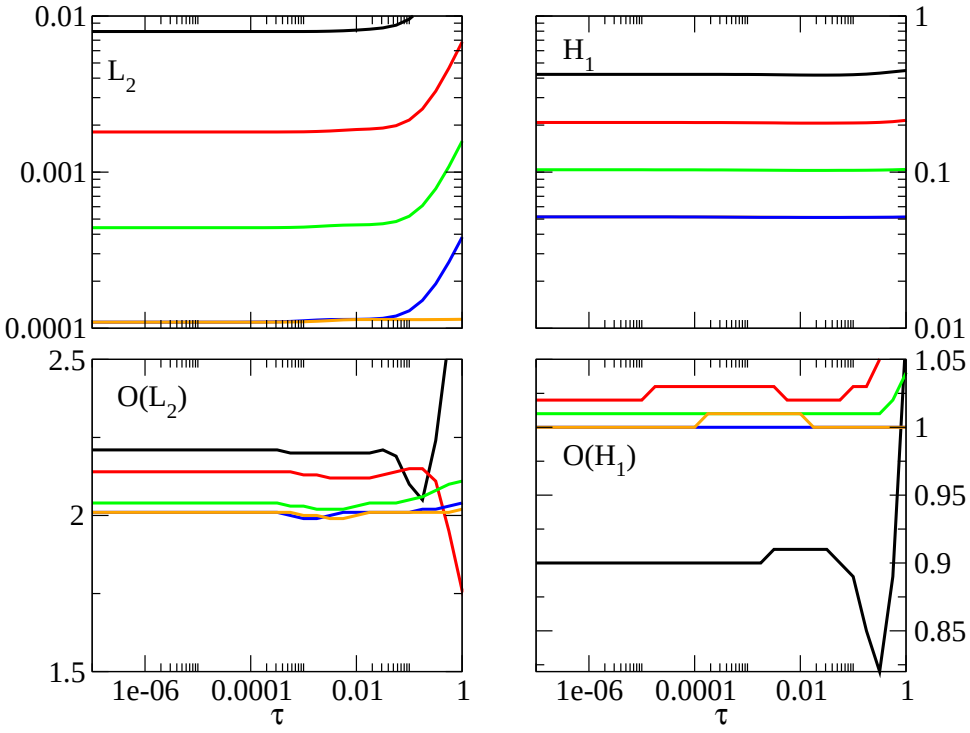

Abbildung B.2: Zu sehen ist das gleiche wie in Abbildung 7.4 nur für Finite–Elemente erster Ordnung auf dem Quadratgitter.

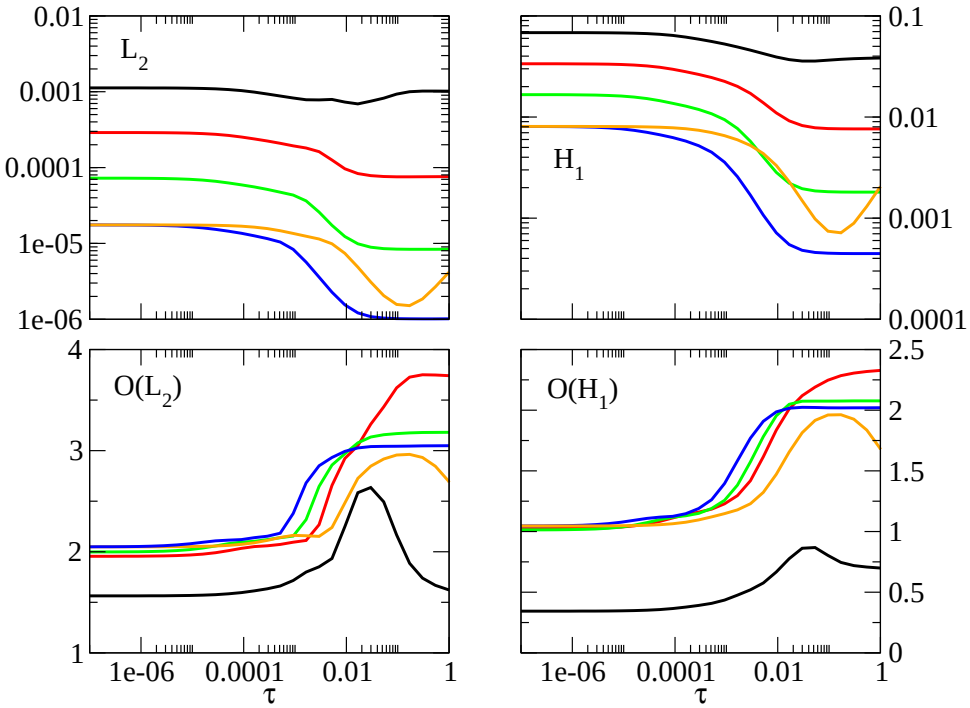

Abbildung B.3: Zu sehen ist das gleiche wie in Abbildung 7.4 nur für Finite–Elemente zweiter Ordnung auf dem Dreiecksgitter.

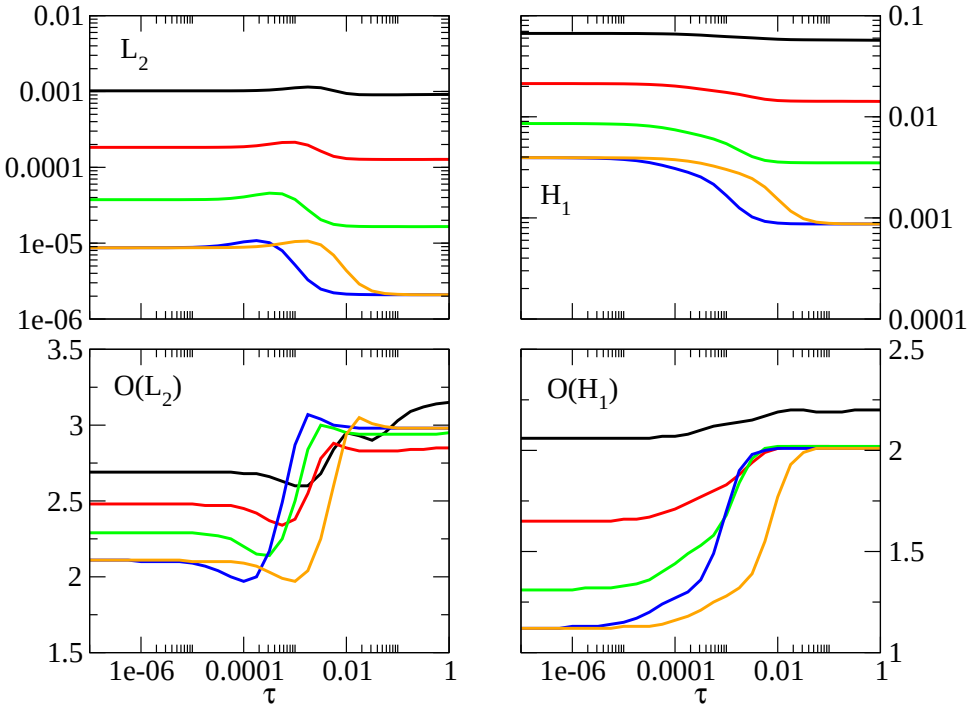

Abbildung B.4: Zu sehen ist das gleiche wie in Abbildung 7.4 nur für Finite–Elemente zweiter Ordnung auf dem Quadratgitter.

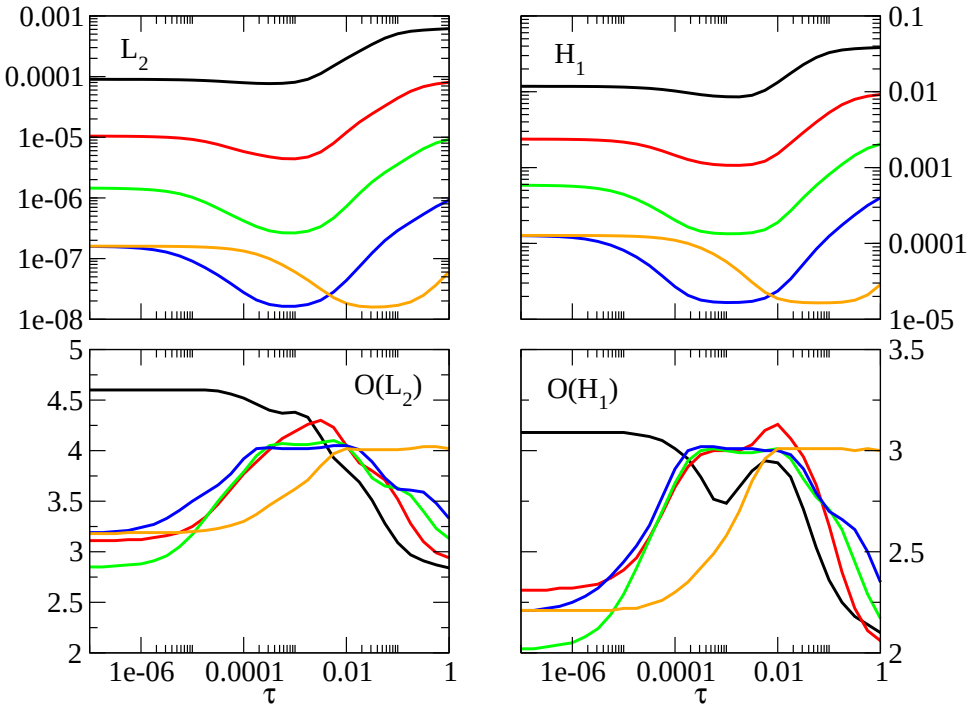

Abbildung B.5: Zu sehen ist das gleiche wie in Abbildung 7.4 nur für Finite–Elemente zweiter Ordnung auf dem Dreiecksgitter.

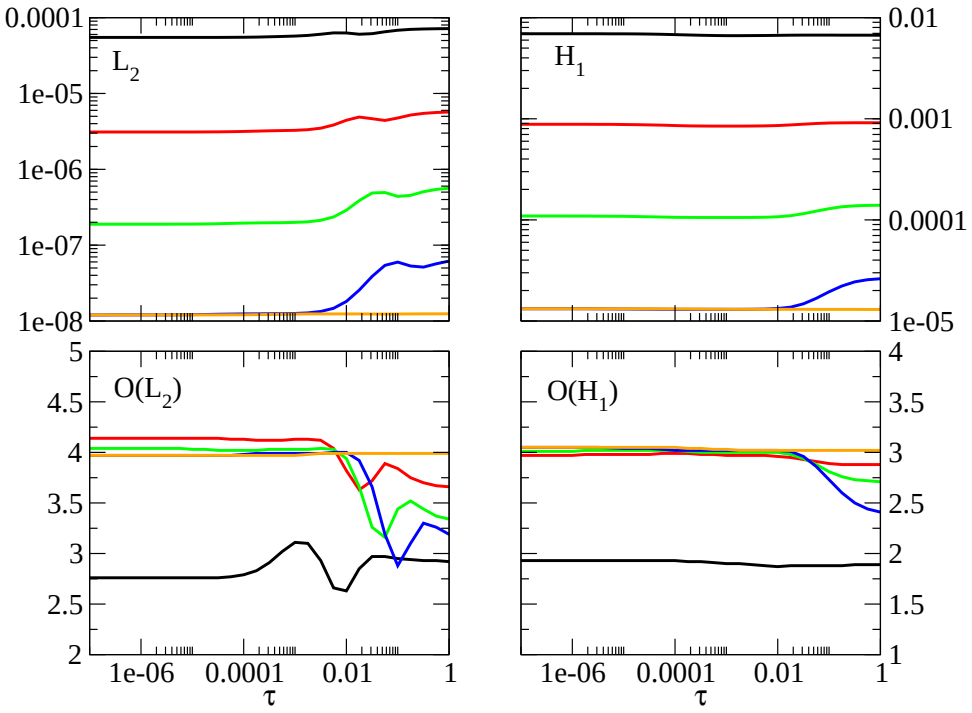

Abbildung B.6: Zu sehen ist dasselbe wie in Abbildung 7.4 nur für Finite–Elemente dritter Ordnung auf dem Quadratgitter.

Anhang C

Das exponentielle Beispiel für die Oseen–Gleichung

Für die Oseen–Gleichung betrachten wir ein exponentielles Beispiel mit der Lösung

$$\begin{aligned} u_x &= e^{x+y}, \\ u_y &= -e^{x+y}, \\ p &= 0 \end{aligned}$$

auf dem Gebiet $\Omega = [0,1]^2$ und den Vorgaben $b = (1, 0.5)$, $c = 0$ sowie $\nu = 10^{-6}$ auf Ω. Die rechte Seite f der Oseen–Gleichung erhält man durch Einsetzen in (8.1). Die Randbedingungen sind über die Randfunktion u_b festgelegt, wobei $u_b = u$ auf $\partial\Omega$ gesetzt wird.

Für die CIP–Methode wählen wir $\tau_{\text{grad}} = \tau_{\text{div}} = \tau_p = \tau$ und für die residuale Stabilisierung $\delta = \gamma = \tau$. Siehe hierzu auch Abschnitt 9.2. Die Ergebnisse der residualen Stabilisierung und der CIP–Methode für das exponentielle Beispiel sind in Abbildung C.1 bis C.3 zu sehen. Weitere Details sind den Bildunterschriften zu entnehmen.

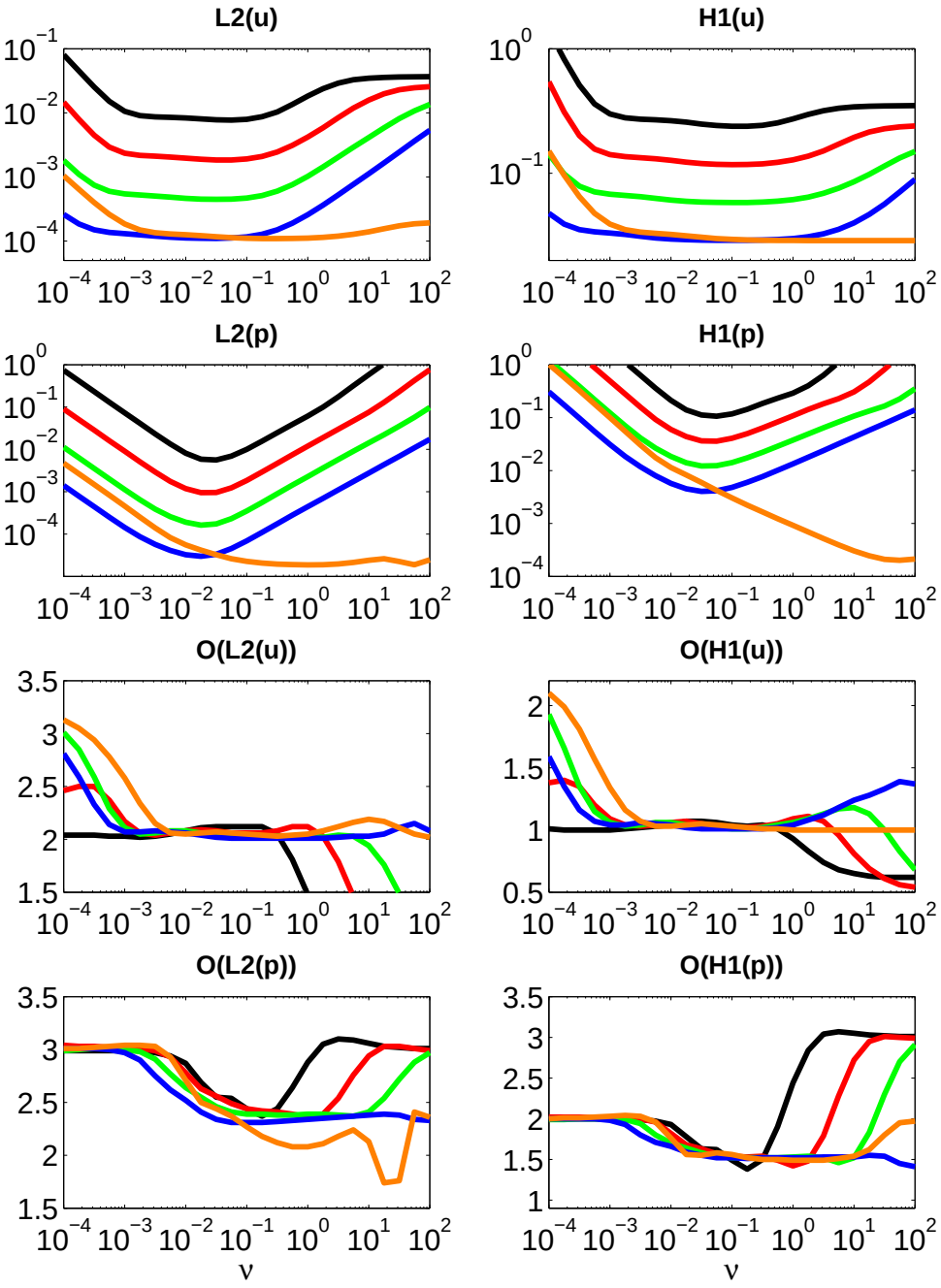

Abbildung C.1: Fehler und Fehlerraten für Finite–Elemente erster Ordnung auf dem Quadratgitter gegen den Stabilisierungsparameter τ. Die Graphen in den ersten beiden Reihen zeigen die Fehlerwerte im Geschwindigkeitsfeld u und im Druckfeld p in der L^2– beziehungsweise H^1–Norm. Die Graphen in den beiden unteren Reihen geben die zugehörigen Fehlerordnungen an. Die orangefarbene Linie bezieht sich auf die residuale Stabilisierung auf dem Level–5–Gitter. Die restlichen Kurven sind Resultate der CIP–Methode. Für die schwarze, rote, grüne, blaue Linie wurde der Reihe nach auf Level 2,3,4,5 gerechnet.

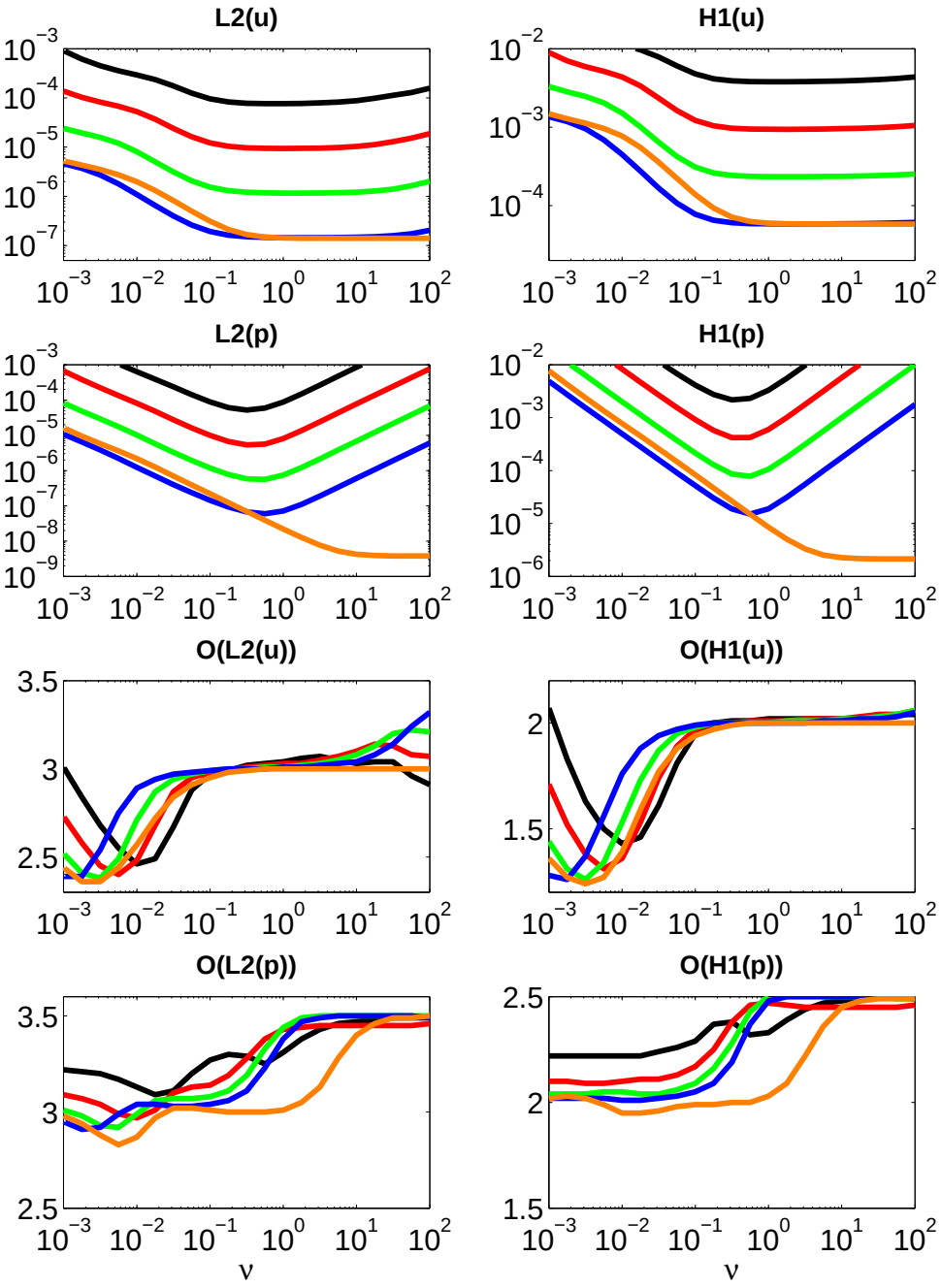

Abbildung C.2: Die Abbildung zeigt das gleiche wie Abbildung C.1 nur für Elemente zweiter Ordnung.

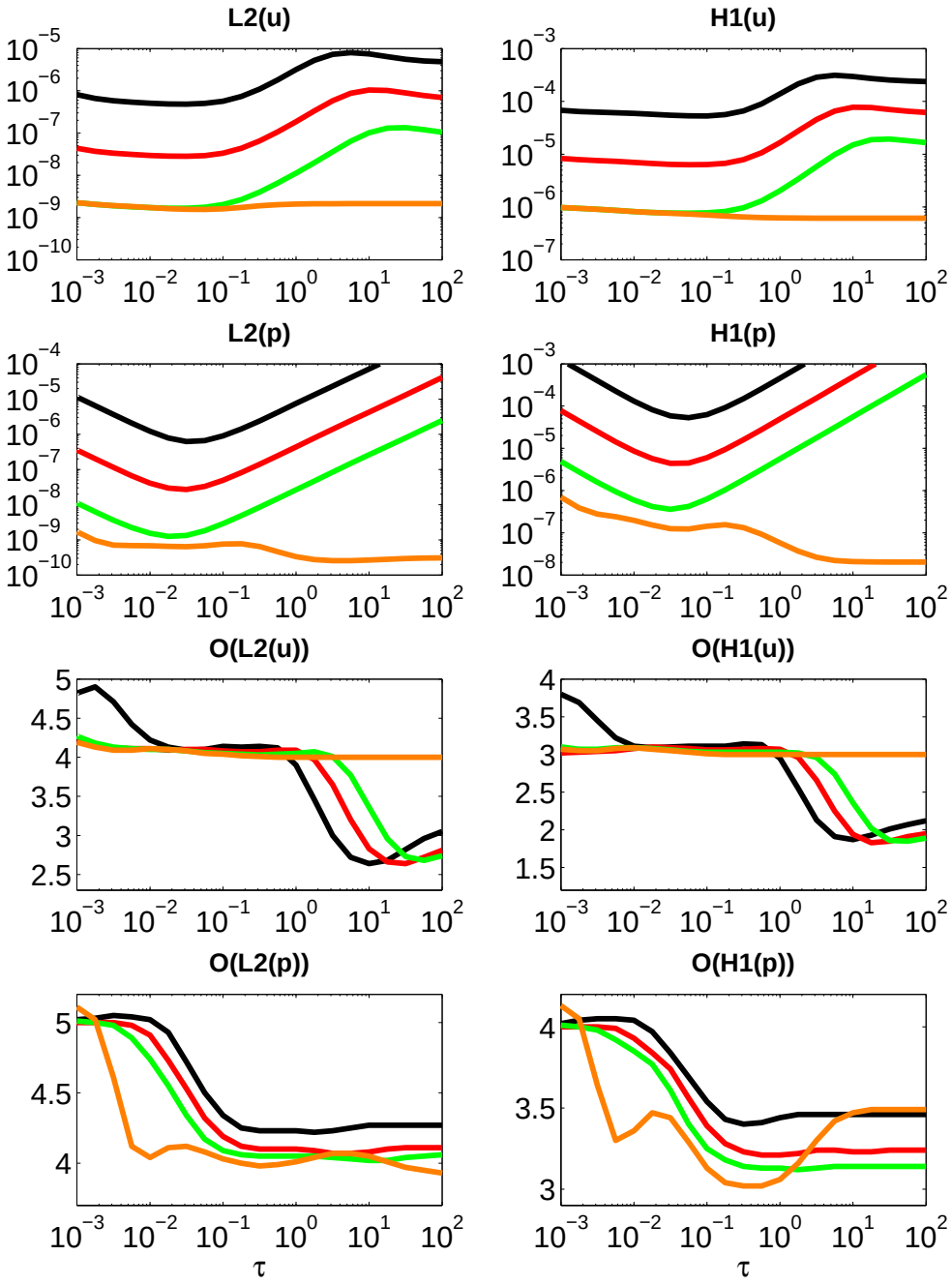

Abbildung C.3: Die Abbildung zeigt das gleiche wie Abbildung C.1 nur für Elemente dritter Ordnung.

Anhang D

Einfluss der Viskosität

Bei den Rechnungen zur Oseen–Gleichung in Kapitel 9 ist die Viskosität stets auf $\nu = 10^{-6}$ fixiert. Hier prüfen wir den Einfluss der Viskosität, wobei nur Finite–Elemente auf Quadratgittern betrachtet werden. Wie in weiteren Tests untersucht worden ist, erhält man für Finite–Elemente auf Dreiecksgittern qualitativ ähnliche Resultate.

Zunächst betrachten wir exemplarisch das sincos–Beispiel für Finite–Elemente zweiter Ordnung auf dem Level–5–Quadratgitter. Für $\nu = 10^{-6}$ wurden die aus der Analysis vorausgesagten Konvergenzraten durch beide Stabilisierungsverfahren in einem relativ großen τ–Intervall erreicht. Abbildung D.1 zeigt die Fehler und Fehlerordnungen neben $\nu = 10^{-6}$ nun auch für $\nu = 10^{-12}$ und $\nu = 1$. Verringert man die Viskosität von 10^{-6} auf 10^{-12}, so ändern sich die Ergebnisse beider Stabilisierungsverfahren kaum. Erhöht man hingegen die Viskosität auf 1 treten die folgenden Veränderungen auf:
Die Fehler im Geschwindigkeitsfeld behalten ihren Wert bei optimaler Wahl von τ bei der CIP–Methode bei, während sie bei der residualen Stabilisierung geringfügig abnehmen. Das τ–Intervall in dem die Geschwindigkeitsfehler annähernd optimal sind, wächst für beide Stabilisierungsverfahren deutlich. Auch die Konvergenzordnungen der Geschwindigkeitsfehler verbessern sich. Im Druckfeld ist der Trend hingegen umgekehrt. Hier wachsen die Fehler beider Verfahren. Die Fehler im Druckfeld sind für die CIP–Methode über das gesamte dargestellte τ–Intervall kleiner als diejenigen der residualen Stabilisierung. Zudem sind die Konvergenzordnungen der Druckfehler besser für die CIP–Methode. Letztere ist also für $\nu = 1$ der residualen Stabilisierung klar vorzuziehen.
Weiterhin beobachtet man für $\nu = 1$ eine Abnahme der Konvergenzordnungen im Druckfeld. Dieses Verhalten wollen wir nun mit den Voraussagen der Fehleranalsyis vergleichen. Da die Rechnungen auf dem Level–5–Gitter mit einer Gitterweite $h = 0.5^5$ durchgeführt werden, ergibt sich für $\nu = 1$ und die Vorgaben des sincos–Beispiels die Relation $\nu \geq h\|b\|_{L^\infty(\Omega)}$. Demnach liegt hier der diffusions–dominante Fall vor. Für den diffusions–dominanten Fall werden durch die Fehlerabschätzungen (8.30) und (8.62) für beide Stabilisierungsverfahren eine Fehlerordnung von h^r im L^2– und im H^1–Fehler der Geschwindigkeit sowie im L^2–Fehler des Druckfeldes vorausgesagt. r ist hierbei die Elementordnung. Daher sollten die Fehlerordnungen im vorliegenden Fall zumindest zweiter Ordnung in der Gitterweite h sein. Die CIP–Methode erreicht diese Ordnungen fast überall in der Abbildung. Nur die Konvergenzrate des L^2–Druckfehlers wird für kleine τ knapp verfehlt. Die residuale Stabilisierung verfehlt die Rate im Druckfehler hingegen im Großteil der Abbildung um etwa 0.3.

Rechnungen mit verschiedenen Viskositäten sind auch für das polynomiale Beispiel durchgeführt worden. Abbildung D.2 zeigt hierzu beispielhaft die Ergebnisse der CIP–Methode

für Elemente zweiter Ordnung auf dem Level–5–Quadratgitter. Das Verhalten auf dem Dreiecksgitter beziehungsweise das Verhalten der residualen Stabilisierung sind qualitativ gleich und werden daher nicht gezeigt. Für die Viskosität $\nu = 10^{-9}$ sind die Fehler im Geschwindigkeitsfeld noch größer als bei den Rechnungen aus dem letzten Abschnitt mit $\nu = 10^{-6}$. Außerdem sind die Fehlerordnungen der Geschwindigkeit weiter geschrumpft. Vergrößert man hingegen die Viskosität von $\nu = 10^{-6}$ auf 10^{-4}, so nehmen die Fehler im Geschwindigkeitsfeld ab und die aus der Theorie vorausgesagten Fehlerordnungen werden in einem deutlich größeren τ–Intervall angenommen. Für eine Viskosität von $\nu = 1$ verbessern sich die Fehlerwerte und Fehlerordnungen im Geschwindigkeitsfeld weiter. Die durch die Theorie vorausgesagten Konvergenzraten werden nun im ganzen dargestellten Bereich erreicht. Bemerkenswert ist aber, dass die Fehlerwerte im Druckfeld gestiegen sind.
Auch wenn nicht dargestellt, sind obige Rechnungen auch für das polynomiale Beispiel mit Finite-Elementen erster Ordnung durchgeführt worden. Im letzten Abschnitt hatten wir gesehen, dass die Wahl $\tau = 1$ hier im Vergleich zu einer optimalen Wahl von τ zu deutlich größeren Fehlerwerten geführt hat. Erhöht man die Viskosität, so werden die Fehler im Geschwindigkeitsfeld bei jeweils optimaler Wahl von τ zwar nicht kleiner, der Bereich beinahe optimaler τ dehnt sich jedoch aus. Spätestens bei $\nu = 1$ hat sich dieser Bereich soweit ausgedehnt, dass auch $\tau = 1$ dazu gezählt werden kann.

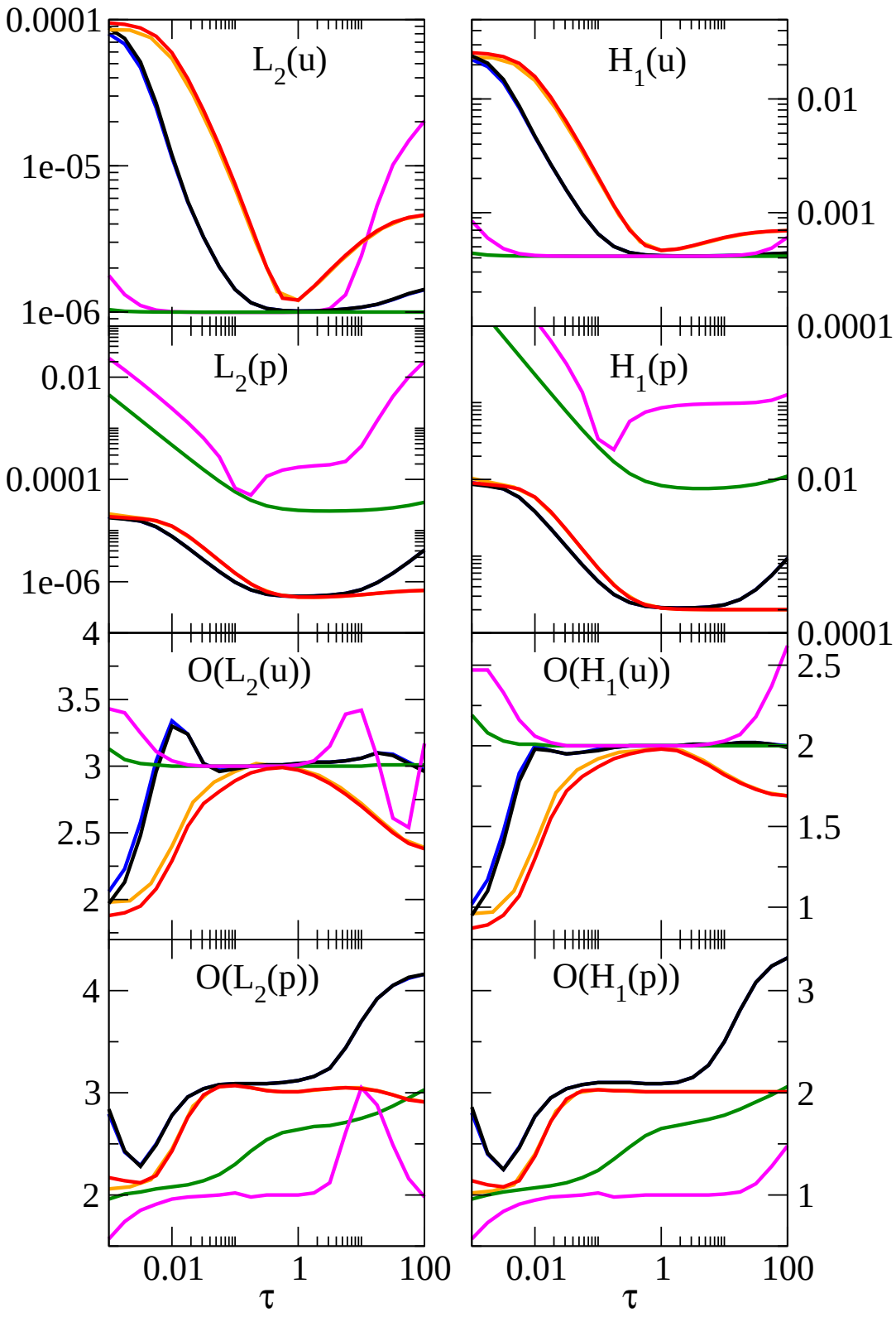

Abbildung D.1: Fehler und Fehlerordnungen der Stabilisierungsverfahren für das sincos–Beispiel (9.8) und Finite–Elemente zweiter Ordnung auf dem Level–5–Quadratgitter. Die Daten für die CIP–Methode sind die blauen Kurven mit einer Viskosität ν von 10^{-6}, die schwarzen mit $\nu = 10^{-12}$ und die dunkelgrünen mit $\nu = 1$. Die Daten der residualen Stabilisierung sind die orangefarbenen Kurven mit $\nu = 10^{-6}$, die roten mit $\nu = 10^{-12}$ und die magentafarbenen mit $\nu = 1$.

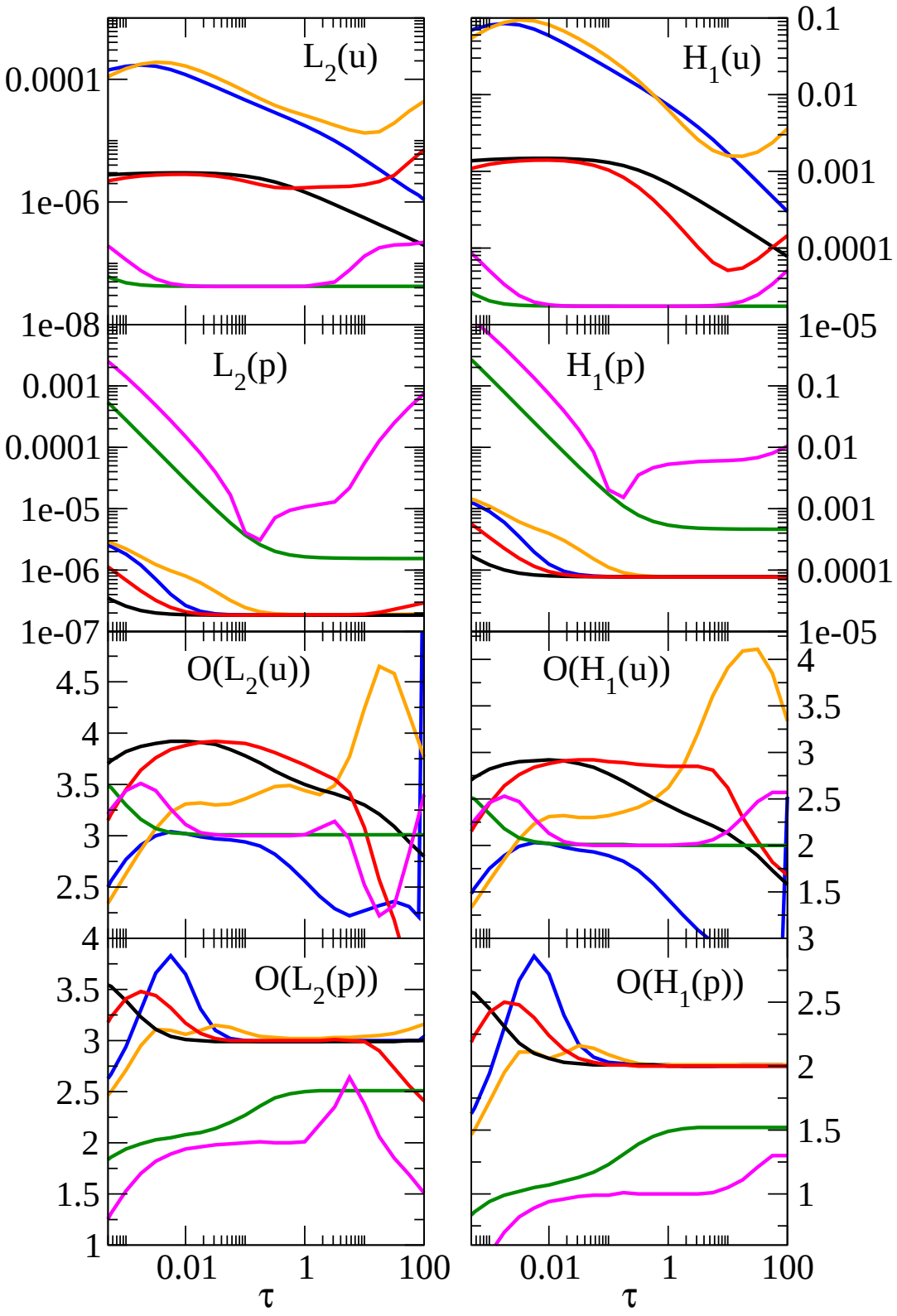

Abbildung D.2: Fehler und Fehlerordnungen der Stabilisierungsverfahren für das polynomiale Beispiel (9.9) und Finite–Elemente zweiter Ordnung auf dem Level–5–Quadratgitter. Die Daten für die CIP–Methode sind die blauen Kurven mit einer Viskosität ν von 10^{-6}, die schwarzen mit $\nu = 10^{-4}$ und die dunkelgrünen mit $\nu = 1$. Die Daten der residualen Stabilisierung sind die orangefarbenen Kurven mit $\nu = 10^{-6}$, die roten mit $\nu = 10^{-4}$ und die magentafarbenen mit $\nu = 1$.

Wie oben beschrieben, wachsen bei beiden Stabilisierungsverfahren die Fehler im Druckfeld, während die Fehler im Geschwindigkeitsfeld etwa gleich bleiben, wenn bei dem sincos–Beispiel die Viskosität von $\nu = 10^{-6}$ auf $\nu = 1$ erhöht wird. Demnach sind hier die Stabilisierungsverfahren im diffusions–dominanten Fall weniger geeignet als im konvektions–dominanten. Die Frage kommt daher auf, ob die Stabilisierungsverfahren im diffusions–dominanten Fall überhaupt noch benutzt werden sollten. Wie die Abbildung D.3 exemplarisch für das sincos–Beispiel zeigt, gibt es eine alternative Methode, welche beiden Stabilisierungsverfahren für ausreichend kleine ν überlegen zu sein scheint. Die blauen Linien beziehen sich dort auf die CIP–Methode, die orangefarbenen auf die residuale Stabilisierung. Gerechnet wir für beide Verfahren mit Elementen zweiter Ordnung auf dem Level–5–Quadratgitter. Die schwarze Linie zeigt für dasselbe Gitter zum Vergleich die Taylor–Hood–Elemente mit einem Geschwindigkeitsraum der Ordnung 2 und einem Druckraum der Ordnung 1. Auf eine Stabilisierung wird bei der Taylor–Hood Methode verzichtet. Dies sollte zumindest im diffusions–dominanten Fall sinnvoll sein, da die Taylor–Hood Methode gemäß Kapitel 8 der diskreten Babuška–Brezzi–Bedingung genügt. Aufgetragen werden in der Abbildung die Fehler und Fehlerordnungen gegen die Viskosität. Der Parameter τ der CIP–Methode wird konstant gleich 1 gehalten, derjenige der residualen Stabilisierung konstant gleich 0.1. Gemäß weiteren Tests ist diese Wahl bei beiden Verfahren für Viskositäten $\nu \geq 100$ optimal.
Trotz der optimalen Wahl für des Stabilisierungsparameters werden die Fehler der Stabilisierungsverfahren im Druckfeld oberhalb einer Viskosität von etwa 1 größer als die Fehler der Taylor–Hood Methode, während die Fehler im Geschwindigkeitsfeld für alle Verfahren oberhalb von $\nu = 0.01$ gleich sind. Ein ähnliches Verhalten hat sich auch für das polynomiale Beispiel gezeigt. So sind die Fehler der Taylor–Hood Methode hier etwa oberhalb $\nu = 10$ kleiner als diejenigen der Stabilisierungsverfahren zweiter Ordnung, während die Fehler im Geschwindigkeitsfeld aller Verfahren etwa gleich sind. Ferner ist der numerische Aufwand der Taylor–Hood Methode geringer, da weniger Terme assembliert werden müssen und die Systemmatrix eine kleinere Anzahl von Null verschiedener Einträge besitzt. Die Taylor–Hood Methode ist also für ausreichend große Viskositäten den Stabilisierungverfahren überlegen. Weitere numerische Tests sollten zeigen, dass diese Tatsache auch gültig bleibt, wenn die Stabilisierungsparameter nicht auf einen gemeinsamen Wert fixiert werden. Im Rahmen dieser Arbeit wurde dies jedoch nicht systematisch überprüft.

Abbildung D.3 zeigt ferner, dass im konvektions–dominanten Fall eine Stabilisierung benötigt wird. In allen Fehlerwerten sind die Stabilisierungsverfahren für $\nu \leq 0.01$ der Taylor–Hood Methode überlegen. Besonders deutlich erkennt man den Einfluss der Stabilisierung in den Geschwindigkeitsfehlern: Die Fehler der Stabilisierungsverfahren bleiben entweder konstant oder wachsen wenig mit kleiner werdender Viskosität, während die Fehler der Taylor–Hood Methode deutlich ansteigen.[1]
Gemäß Abbildung D.4 gilt oben beschriebenes Verhalten auch für Finite–Elemente höherer Ordnung. Gerechnet wird hier auf dem Level–4–Quadratgitter. Bei den Stabilisierungsverfahren werden Finite–Elementen dritter Ordnung benutzt, bei der Taylor–Hood Methode hingegen Finite–Elemente dritter Ordnung für die Geschwindigkeit und Elemente zweiter Ordnung für das Druckfeld. Die Stabilisierungsparameter werden jeweils wieder so gewählt, dass sie für $\nu \geq 100$ beinahe optimal sind. Bei der CIP–Methode ist die entsprechende Wahl $\tau = 1$, bei der residualen Stabilisierung $\tau = 0.056$. Für eine Viskosität größer als etwa 1 ist die Taylor–Hood Methode vorzuziehen, andernfalls verdienen die Stabilisierungsverfahren den Vorzug.

[1] Wählt man den Stabilisierungsparameter der residualen Stabilisierung bei größeren Viskositäten optimal, so bleiben die Fehler im dargestellten Bereich sogar konstant.

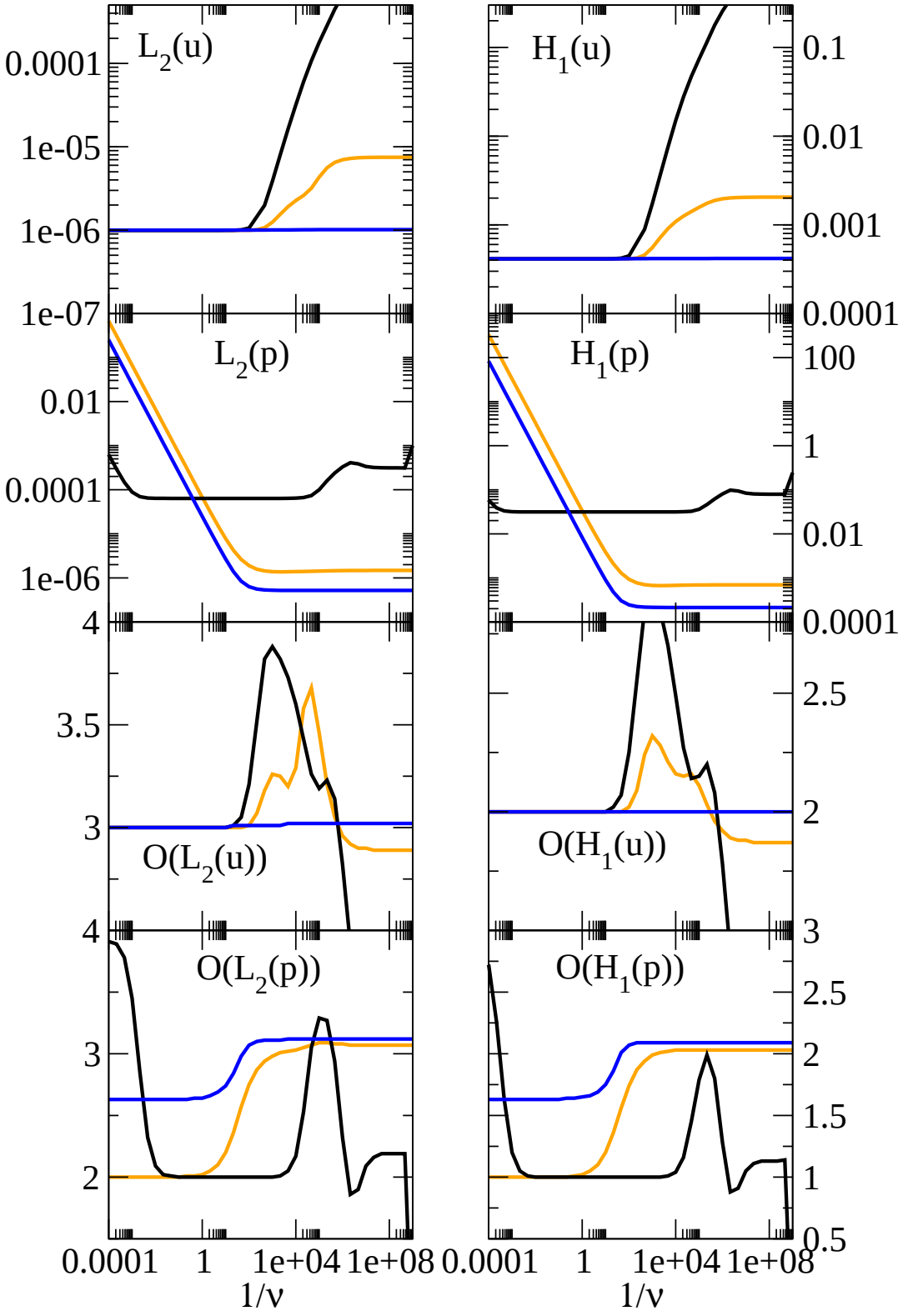

Abbildung D.3: Fehler und Fehlerordnungen für das sincos–Beispiel im Auftrag gegen $1/\nu$. Die schwarze Kurve gibt die Ergebnisse der Taylor–Hood Methode an. Gerechnet wird hier mit Elementen zweiter Ordnung im Geschwindigkeitsraum und Elementen erster Ordnung im Druckraum. Die blauen beziehungsweise die orangefarbenen Kurven beziehen sich auf CIP–Methode beziehungsweise auf die residuale Stabilisierung, wobei jeweils mit Elementen zweiter Ordnung gerechnet wird. Für weitere Details siehe Text.

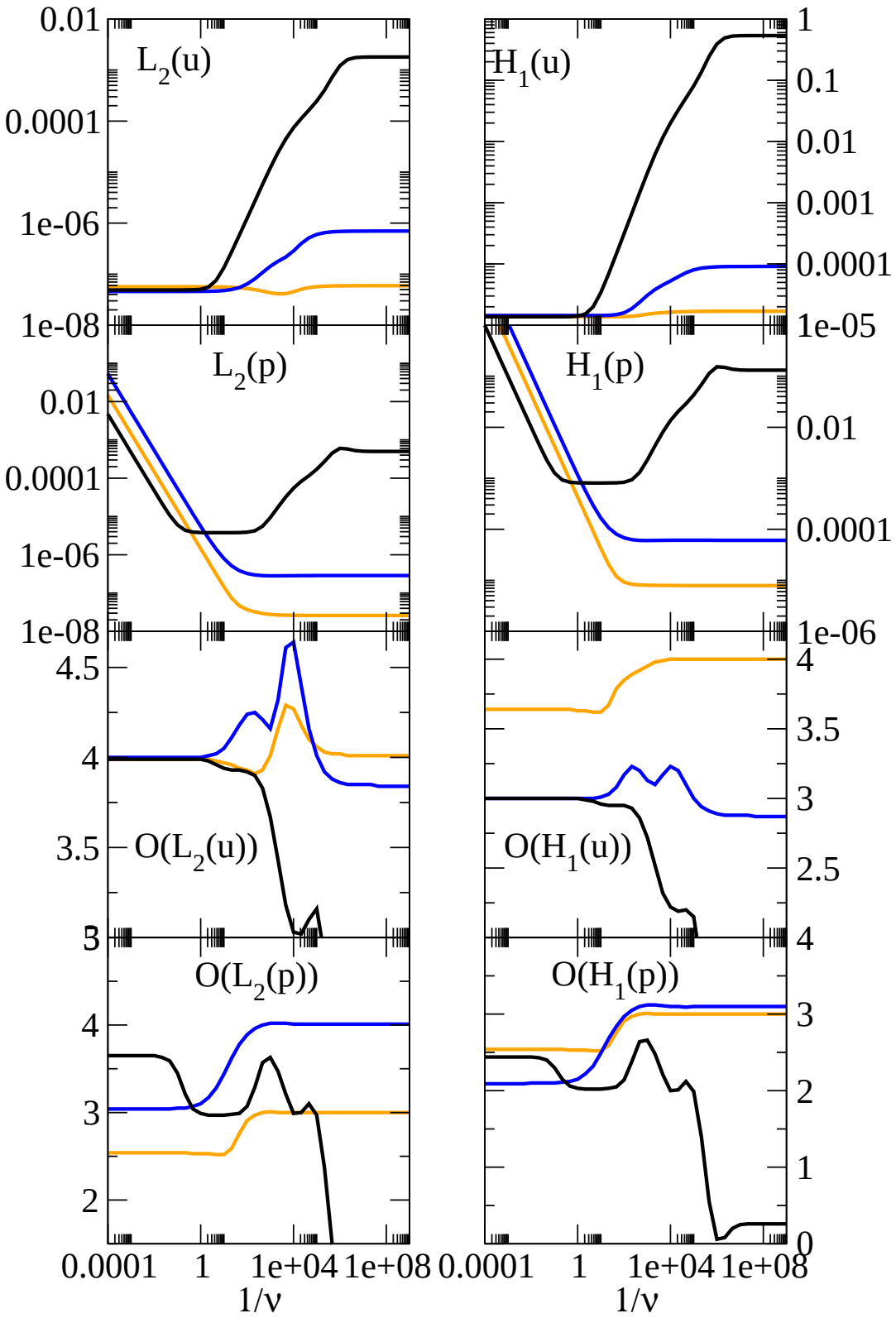

Abbildung D.4: Fehler und Fehlerordnungen für das sincos–Beispiel im Auftrag gegen $1/\nu$. Die schwarze Kurve gibt die Ergebnisse der Taylor–Hood Methode an. Gerechnet wird hier mit Elementen dritter Ordnung im Geschwindigkeitsraum und Elementen zweiter Ordnung im Druckraum. Die blauen beziehungsweise die orangefarbenen Kurven beziehen sich auf CIP–Methode beziehungsweise auf die residuale Stabilisierung, wobei jeweils mit Elementen dritter Ordnung gerechnet wird. Für weitere Details siehe Text.

Literaturverzeichnis

[Ada75] R.A. Adams. *Sobolev spaces.* Academic Press, New York, 1975.

[Alt85] H.W. Alt. *Lineare Funktionalanalysis. Eine anwendungsorientierte Einführung.* Springer Berlin, 1985.

[Azz80] A. Azzam. On differentiability properties of solutions of elliptic differential equations. *J. Math. Anal. Appl.*, 75:431–440, 1980.

[BA72] I. Babuska and A.K. Azizi. *The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations.* Survey Lectures on the Mathematical Foundations of the Finite Element Method. Academic Press, New York, 1972.

[Bat00] G.K. Batechlor. *An introduction to fluid dynamics.* Cambridge university press, 2000.

[BBJL07] M. Braack, E. Burman, V. John, and G. Lube. Stabilized finite element methods for the generalized oseen problem. *Comput. Methods Appl. Mech. Engrg.*, 196:853 – 866, 2007.

[BBO02] Ivo M. Babuska, U. Banerjee, and J.E. Osborn. On principles for the selection of shape functions for the generalized finite element method. *Computer methods in applied mechanics and engineering*, 191(49-50):5595–5629, 2002.

[BE02] E. Burman and A. Ern. Nonlinear diffusion and discrete maximum principle for stabilized Galerkin approximations of the convection–diffusion–reaction equation. *Comput. Meth. Appl. Mech. Engrg.*, 191:3833 – 3855, 2002.

[BF91] F. Brezzi and M. Fortin. *Mixed and Hybrid Finite Element Methods*, volume 15 of *Springer Series in Computational Mathematics.* Springer-Verlag, 1991.

[BFH06] E. Burman, M. A. Fernández, and P. Hansbo. Continuous interior penalty finite element method for the Oseen's equations . *SIAM Journal on Numerical Analysis*, 44(3):1621–1638, 2006.

[BG09] C. Bernardi and V. Girault. A local regularization operator for triangular and quadrilateral finite elements. *SIAM J. Numer. Anal.*, 35(5), 2009.

[BH04] E. Burman and P. Hansbo. Edge stabilization for Galerkin approximations of convection–diffusion–reaction problems. *Comput. Methods Appl. Mech. Engrg.*, 193:1437 – 1453, 2004.

[BH06] E. Burman and P. Hansbo. Edge stabilization for the generalized Stokes problem: A continuous interior penalty finite element method. *Comput. Methods Appl. Mech. Engrg.*, 195:2393 – 2410, 2006.

[BH07] Y. Bazilevs and T. J. R. Hughes. Weak imposition of dirichlet boundary conditions in fluid mechanics. *Computers and Fluids*, 36:12–26, 2007.

[Bom04] M. Boman. Estimates for the l_2-projection onto piecewise linear finite elements spaces ina weighted l_p-norm. *Preprint*, 2004.

[BP79] M. Bercovier and O. Pironneau. Error estimates for finite element method solution of the stokes problem in the primitve variables. *Numer. Math.*, 33:211–224, 1979.

[Bra07] D. Braess. *Finite Elemente: Theorie, schnelle Löser und Anwendungen in der Elastizitäststheorie*. Springer Verlag, Heidelberg, 2007.

[Bre74] F. Brezzi. On the existence, uniqueness and approximation of saddle–point problems arising from Lagrange multipliers. *R.A.I.R.O., Anal. Numer. R2*, pages 129 – 151, 1974.

[CGMZ76] I. Christie, D.F. Griffiths, A.R. Mitchell, and O.C. Zienkiewicz. Finite element methods for second order differential equations with significant first derivatives. *Int. J. Numer. Meth. Engrg.*, 10:1389 – 1396, 1976.

[Cia78] P.G. Ciarlet. *The finite element method for elliptic problems*. North-Holland Publishing Company, Amsterdam – New York – Oxford, 1978.

[Cio99] D. Cioranescu. *An Introduction to Homogenization*. Number 17 in Oxford Lecture Series in Mathematics and Its Applications. Oxford University Press, 1999.

[Clé75] Ph. Clément. Approximation by finite element functions using local regularization. *RAIRO Anal. Numer.*, 2:77 – 84, 1975.

[Cod00] R. Codina. Stabilization of incompressibility and convection through orthogonal subscales in finite element methods. *Comput. Methods Appl. Mech. Engrg.*, 190:1579–1599, 2000.

[CR72] P.G. Ciarlet and P.-A. Raviart. General lagrange and hermite interpolation in R^n with applications to finite element methods. *Arch. Rat. Mech. Anal.*, 46:177–199, 1972.

[CR73] M. Crouzeix and P.-A. Raviart. Conforming and nonconforming finite element methods for solving the stationary Stokes equations I. *R.A.I.R.O. Anal. Numér.*, 7:33–76, 1973.

[Dav04] T.A. Davis. Algorithm 832: UMFPACK V4.3–an unsymmetric–pattern multifrontal method. *ACM Trans. Math. Software*, 30:196 – 199, 2004.

[DD75] J. Douglas and T. Dupont. *Interior penalty procedures for elliptic and parabolic Galerkin methods*. 1975.

[EG04] A. Ern and J.-L. Guermond. *Theory and practice of finite elements*, volume 159 of *Applied Mathematical Sciences*. Springer Verlag, New York, 2004.

[Fai78] G. Fairweather. *Finite Element Galerkin Methods for Differential Equations*. Number 34 in Lecture Notes in Pure and Applied Mathematics. Marcel Dekker, 1978.

[FF92] L.P. Franca and S.L. Frey. Stabilized finite element methods: II. The incompressible Navier-Stokes equations. *Computer Methods in Applied Mechanics and Engineering*, 99:209 – 233, 1992.

[FS91] L.P. Franca and R. Stenberg. Error analysis of some galerkin-least-squares methods for the elasticity equations. *SIAM J. Numer. Anal.*, 28, 1991.

[FS95] J. Freund and R. Stenberg. On weakly imposed boundary conditions for second order problems. *Proceedings of the International Conference on Finite Elements in Fluids - New trends an applications*, pages 99–119, 1995.

[GFL+83] H. Goering, A. Felgenhauer, G. Lube, H. G. Roos, and L. Tobiska. *Singularly perturbed differential equations.* Akademie Verlag, Berlin, 1983.

[GR86] V. Girault and P.-A. Raviart. *Finite Element Methods for Navier-Stokes equations.* Springer-Verlag, Berlin-Heidelberg-New York, 1986.

[Gri85] P. Grisvard. *Elliptic problems in nonsmooth domains.* Pitman, Boston, 1985.

[GR05] C. Großmann and H.-G. Roos. *Numerische Behandlung partieller Differentialgleichungen.* Teubner-Verlag, 2005.
P. Grisvard. *Elliptic problems in nonsmooth domains.* Pitman, Boston, 1985.

[GS00] P.M. Gresho and R.L. Sani. *Incompressible Flow and the Finite Element Method.* Wiley, Chichester, 2000.

[HB79] T.J.R. Hughes and A.N. Brooks. A multidimensional upwind scheme with no crosswind diffusion. In T.J.R. Hughes, editor, *Finite Element Methods for Convection Dominated Flows, AMD vol.34*, pages 19 – 35. ASME, New York, 1979.

[HFB86] T.J. Hughes, L.P. Franca, and M. Balestra. A new finite element formulation for computational fluid dynamics: V. circumventing the babuska-brezzi condition: a stable petrov-galerkin formulation of the stokes problem accomodating equal-order interpolations. *Comput. Methods Appl. Mech. Engrg.*, 59:85–99, 1986.

[HHZM77] J.C. Heinrich, P.S. Huyakorn, O.C Zienkiewicz, and A.R. Mitchell. An 'upwind' finite element scheme for two-dimensional convective transport equation. *Int. J. Numer. Methods Engrg.*, 11:131 – 143, 1977.

[HKOS08] A.F. Hegarty, N. Kopteva, E. O'Riordan, and M. Stynes. *BAIL 2008: Boundary and Interior Layers*, volume 69 of *Lecture Notes in Computational Science and Engineering.* Springer Verlag, Berlin, 2008.

[HS01] P. Houston and E. Süli. Stabilised hp-finite element approximations of partial differential equations with nonnegative characteristic form. *Computing*, 66:99–119, 2001.

[JK07a] V. John and P. Knobloch. A comparison of spurious oscillations at layers diminishing (sold) methods for convection–diffusion equations: Part I – a review. *Comput. Methods Appl. Mech. Engrg.*, 196:2197 – 2215, 2007.

[JK07b] V. John and P. Knobloch. On the performance of SOLD methods for convection–diffusion problems with interior layers. *International Journal of Computing Science and Mathematics*, 1:245 – 258, 2007.

[JS86] C. Johnson and J. Saranen. Streamline diffusion methods for the incompressible euler and navier-stokes equations. *Math. Comp.*, 47:1 – 18, 1986.

[JSW87] C. Johnson, A.H. Schatz, and L.B. Wahlbin. Crosswind smear and pointwise errors in streamline diffusion finite element methods. *Math. Comput.*, 49:25 – 38, 1987.

[Li06] B. Q. Li. Discontinuous finite elements in fluid dynamics and heat transfer (computational fluid and solid mechanics). 2006.

[LR06] G. Lube and G. Rapin. Residual–based stabilized higher–order FEM for advection–dominated problems. *Comput. Methods Appl. Mech. Engrg.*, 195:4124 – 4138, 2006.

[Mic77] J. H. Michael. A general theory for linear elliptic partial differential equations. *J. diff. equations*, 23:1–29, 1977.

[MLR09] G. Matthies, G. Lube, and L. Roehe. Some remarks on streamline-diffusion methods for inf-sup stable discretisations of the generalised oseen problem. *CMAM, to appear*, 9, 2009.

[MS07] G. Matthies and F. Schieweck. Nonconforming finite elements of higher order satisfying a new compatibility condition. *International Journal for Numerical Methods in Engineering*, 16(1):23–50, 2007.

[Näv92] U. Nävert. *A finite element method for convection-diffusion problems.* PhD thesis, Chalmers university of Technology, Göteborg, 1992.

[Nit72] J. Nitsche. *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 / Heidelsberg, 1972.

[Ose10] C. W. Oseen. Über die Stoke'sche Formel und über eine verwandte Aufgabe in der Hydrodynamik. *Arkiv för matematik, astronomi och fysik vi*, 29, 1910.

[Riv08] B. Riviere. *Discontinuous Galerkin Methods For Solving Elliptic And Parabolic Equations: Theory and Implementation.* Number 35 in Frontiers in Applied Mathematics. SIAM, 2008.

[Roe07] L. Roehe. Residuale Stabilisierung für Finite Elemente Verfahren bei inkompressiblen Strömungen. Diplomarbeit, Georg–August–Universität zu Göttingen, 2007.

[RST08] H.-G. Roos, M. Stynes, and L. Tobiska. *Robust Numerical Methods for Singularly Perturbed Differential Equations*, volume 24 of *Springer Series in Computational Mathematics.* Springer, 2nd edition, 2008.

[Sch98] C. Schwab. *P- and hp- Finite Element Methods: Theory and Applications in Solid and Fluid Dynamics.* Calderon Press, Oxford, 1998.

[SE99] Y.-T. Shih and H.C. Elman. Modified streamline diffusion schemes for convection–diffusion problems. *Comput. Methods Appl. Mech. Engrg.*, 174:137 – 151, 1999.

[SFH07] R. Sevilla, S. Fernandéz, and A. Huerta. Nurbs-enhanced finite element method (nefem). *International Journal for Numerical Methods in Engineering*, 76(1):56–83, 2007.

[ST95] M. Stynes and L. Tobiska. Necessary L^2–uniform convergence conditions for difference schemes for two–dimensional convection–diffusion problems. *Comput. Math. Appl.*, 29:45 – 53, 1995.

[Sty05] M. Stynes. Steady-state convection–diffusion problems. *Acta Numerica*, 15, 2005.

[SZ90] L.R. Scott and S. Zhang. Finite element interpolation of nonsmooth functions satisfying boundary conditions. *Math. Comp.*, 54:483 – 493, 1990.

[Tab77] M. Tabata. A finite element approximation corresponding to the upwind differencing. *Memoirs of Numerical Mathematics*, 1:47 – 63, 1977.

[Tho97] V. Thomée. *Galerkin finite element methods for parabolic problems*, volume 25 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 1997.

[TL91] L. Tobiska and G. Lube. A modified streamline diffusion method for solving the stationary navier-stokes equations. *Numer. Math.*, 59:13–29, 1991.

[TP86] T.E. Tezduyar and Y.J. Park. Discontinuity–capturing finite element formulations for nonlinear convection–diffusion–reaction equations. *Comput. Methods Appl. Mech. Engrg.*, 59:307 – 325, 1986.

[Tre75] F. Treves. *Basic Linear Partial Differential Equations*. Academic Press, New York, 1975.

[TV96] L. Tobiska and R. Verfürth. Analysis of a streamline diffusion finite element method for the Stokes and Navier-Stokes equations. *SIAM J. Numer. Anal.*, 33(1):107 – 127, 1996.

[Wer95] D. Werner. *Funktional Analysis*. Springer-Verlag, 1995.

[Wig70] N.M. Wigley. Mixed boundary value problems in domains with corners. *Math. Z.*, 115:33–52, 1970.

[Wlo82] J. Wloka. *Partielle Differentialgleichungen*. B.G. Teubner Stuttgart, 1982.

[Yos65] K. Yoshida. *Functional Analysis*. Springer-Verlag, 1965.

[Zho97] G. Zhou. How accurate is the streamline diffusion finite element method? *Math. Comp.*, 66:31 – 44, 1997.

[ZR96] G. Zhou and R. Rannacher. Pointwise superconvergence of the streamline diffusion finite element method . *Numer. Methods Partial Differ. Equations*, 12(1):123–145, 1996.